jipitec

4 | 2021

Volume 12 (2021)
Issue 4 ISSN 2190-3387

Special Issue:
Impact of Technological Advances on Individuals:
Interaction of Law & Informatics

Journal of Intellectual Property, Information Technology, and Electronic Commerce Law

www.jipitec.eu

jipitec

Journal of Intellectual Property, Information Technology and Electronic Commerce Law

Volume 12 Issue 4 December 2021
www.jipitec.eu
contact@jipitec.eu

A joint publication of:

Prof. Dr. Thomas Dreier, M. C. J. (NYU)
KIT - Karlsruher Institut für Technologie, Zentrum für Angewandte Rechtswissenschaft (ZAR), Vincenz-Prießnitz-Str. 3, 76131 Karlsruhe Germany

Prof. Dr. Axel Metzger, LL. M. (Harvard)
Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin

Prof. Dr. Gerald Spindler
Dipl.-Ökonom, Georg-August-Universität Göttingen, Platz der Göttinger Sieben 6, 37073 Göttingen

Karlsruhe Institute of Technology, Humboldt-Universität zu Berlin and Georg-August-Universität Göttingen are corporations under public law, and represented by their respective presidents.

ISSN 2190-3387

Funded by

DGRI Deutsche Gesellschaft für Recht und Informatik e.V.

Table Of Contents

Editorial

Articles

Editorial

by **Golnaz A. Jafari, David Roth-Isigkeit and Ronny Thomale***

Recommended citation: Golnaz A. Jafari, David Roth-Isigkeit and Ronny Thomale, Editorial, 12 (2021) JIPITEC 241 para 1.

1 Digitalisation and automation is becoming increasingly embedded in the societal sphere and infrastructure, a process largely enabled and facilitated by technological advances in the fields of Information and Communication Technology (ICT) and informatics in general. The resulting catch-up process in the existing legal and regulatory landscape requires establishing a level playing field for different actors and stakeholders. A fair balance has to be struck between the interests of the public and private sectors in favour of innovation and digital transformation, and the need for a clear pattern of legal and regulatory standards that would safeguard the rights and interests of individuals and communities within the well-established values of economic and democratic diversity and equality.

2 Based on the flux of novel business, governance and economic models being defined and put in practice underscoring the prevalence of the data-driven economy, a collaborative discourse between the disciplines of law and informatics is (inevitably) required allowing for a better understanding of the associated implications and repercussions, affecting individuals in particular. The associated implications on individuals are predominantly the consequence of processing their personal data[1] on a large scale using algorithms, Artificial Intelligence (AI) or machine learning. Here, the effects become even more potentially detrimental when algorithms are utilised in order to detect correlations between separate datasets, which would in turn set the grounds for patterns from behaviour prediction[2] to the exercise of control over access to a service, to name a few. Algorithms as the core element of AI entailing machine learning have seen a rapid evolution, from automated sets of instructions with mathematical logic -based execution triggers to rule-based expert systems and neural networks.

3 In this context, the concept of automated decision making, the varying levels and scope of human intervention throughout these processes and the counterbalancing of associated risks and benefits have in recent years been subject of regulatory scrutiny in various jurisdictions, including the European Union (EU) legal and regulatory landscape. These decisions form and impact an integral part of daily lives of the public, yet in practice remain largely unnoticed. In principle, a decision facilitated by an automated pro-

* Golnaz A. Jafari, LL.M., doctoral researcher at Lucernaiuris, University of Lucerne, Switzerland; David Roth-Isigkeit, PhD, head of a junior interdisciplinary research group "SOCAI centre for social implications of artificial intelligence", JMU, Würzburg, Germany; Ronny Thomale, PhD, full Professor at the chair for Theoretical Physics I (TP1), JMU, Würzburg, Germany.

1 The terms 'personal data' and 'personally identifiable information (PII)' are used interchangeably, with former given preference in the EU regulatory landscape. Under the General Data Protection Regulation (EU) 2016/679 (GDPR), Art. 4(1) the term 'personal data' means "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"; Under Art. 4(2) the term 'processing' means "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction."

2 GDPR, Art. 4(4) refers to the term 'profiling' as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements"; Note: automated decision making does not always and by default involve profiling.

cessing would need to protect the individuals' rights, freedoms and legitimate interests, which ought to be achieved via implementation of safeguarding measures. Of course, in practice, this would not in itself suffice to render and strengthen individuals' stance, which would in addition require providing for legal certainty and effective judicial recourse.

4 Questions then arise, among others, as to *system transparency* on the one hand, and the levels of *intelligibility* of complex software systems on the other. In this regard, various (implied) individual rights become relevant, such as the information and explanation requirements of the GDPR[3] or the constitutional *right to informational autonomy or self-determination.*[4] The latter is primarily conceived from national constitutions as well as from Article 8 of the Charter of the Fundamental Rights of the EU, which in essence empowers individuals to decide themselves about issues of collection, disclosure and use of their personal data, albeit in the form of a non-absolute right.

5 A modern perception of privacy must take account of individuals' existence within their societal surroundings. In this context, it is rightly argued[5] that "privacy as a legal right, should be conceived essentially as an instrument for fostering the specific yet changing *autonomic capabilities* of individuals that are, in a given society at a given time, necessary for sustaining a vivid democracy." Such capabilities are increasingly threatened by technological tools that provide for vast possibilities of, among others, surveillance and monitoring both for the public and private sectors. Here, in order to strike a balance between competing interests and the right to privacy, and whether legitimate and sufficiently compelling reasons exist for allowing interferences with that right, a normative inquiry would be required on the basis of Article 8 of the European Convention on Human Right (ECHR).

6 Questions also emanate concerning fairness and bias of algorithms, and the quality of input and output datasets in terms of e.g. accuracy and balance, with direct implications for potential risks associated with discrimination in automated decision making systems against individuals and the targeted audience at large.

7 On the other hand, emerging developments in ICT have allowed for distribution in network participation, communication and governance in given contexts. Peer to peer (P2P) network infrastructures are no longer seen as exclusively embedding 'technical distribution' among network participants, while maintaining centralised governance, risking a *single point of failure.* Instead, varying levels of decentralisation in governance could in principle be enabled through the deployment of algorithmic protocols. Distributed Ledger Technology (DLT)[6] denotes a distributed record (ledger) of databases shared among computer nodes outside jurisdictional boundaries, run and maintained according to defined algorithmic consensus protocols. Depending on the form a DLT architecture would take, i.e. public, private, permission-less, permissioned or hybrid, network participation and governance rules as well as the definition of actors and stakeholders and their respective roles, next to network security and scalability would greatly vary. Therefore, legal uncertainty exists, in particular as to the attribution of liability and responsibility which would in turn have an impact on the establishment of *public trust* in these network infrastructures.

8 The present special edition has been put together as a collective effort and team work between the two academic research centres at the University of Würzburg in Germany, namely the Würzburg Centre for Legal and Social Implications of AI (SOCAI) and the CT.QMAT Cluster of Excellence which deals with topics related to complexity and topology in quantum matter. The joint effort includes a conference venue, bringing together academic scholars predominantly from the disciplines of law, computer science and business informatics, and a collaborative publication with the Journal of Intellectual Property, Information Technology and Electronic Commerce Law (JIPITEC).

9 The SOCAI centre has been founded in 2019 with the intention to foster interdisciplinary dialogue between law and technical disciplines to assess the legal framework of cutting-edge developments in

3 See the contribution by LK Kumkar & D Roth-Isigkeit in this volume.

4 See an early reference to the German Federal Constitutional Court Decision of 1983, BVerfG, judgment of the First Senate of December 15, 1983 - 1 BvR 209/83 - Rn. 1-215, for non-authoritative English summary <https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/1983/12/rs19831215_1bvr020983en.html;jsessionid=8EE69329DD3CC934B0D1321957DB249D.1_cid386>; see also A Rouvroy & Y Poullet, 'The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy' in S Gutwirth et al. (eds.), *Reinventing Data Protection?* (Springer Netherlands, 2009) ch.2, 45-76; see also C de Terwangne (on behalf of European Commission JRC), 'The Right to be Forgotten and the Informational Autonomy in the Digital Environment' (2013) 4ff; see also the contribution by F Thouvenin in this volume.

5 Ibid Rouvroy (2009) 46.

6 The terms 'blockchain' and 'distributed ledger technology (DLT)' may be utilised interchangeably throughout this draft.

hardware and software technology. Such a focus can potentially prove fruitful since – as has become clearly visible in the EU regulatory efforts on AI or data protection – law making procedures come with a considerable backlog that makes it practically difficult to orient legislation on the newest technological advancements. Yet, it is precisely this knowledge about possible legislative directions that could provide certainty to businesses and individuals and thereby accelerate and direct technological progress.

10 In this context, previous projects in cooperation with the CT.QMAT Cluster of Excellence[7] have focused on the joint development of legal norms and latest advancements in hardware. Contemporary progress towards applications of AI, deep learning and digital transformation predominantly necessitates the processing of a vast amount of data. This constitutes a major challenge, given that the steady growth of computing efficiency and higher integrated circuitry for central processing unit (CPU) power has reached its physical boundaries. As such, the prospective unfolding of the digital transformation in society will not only decisively depend on technological progress at the software, but also at the hardware level. For any useful societal support and regulation of the digital transformation, the availability and cost efficiency of future material platforms for next generation computing and data processing are thus the crucial parameters that will impact the scope of applications and the actual user group.

11 Therefore, the frame would need to be extended beyond the dimension of software. We believe that only an integrated perspective of law, hardware and software development would be fit to provide an understanding of the complex societal challenges that are embodied in technological progress. Largely speaking, social science research on digital transformation addresses the question of how we can use the technology-driven transformational uprising to create a state that is beneficial for humanity from a long-term perspective. In order to succeed, both crucial aspects of technical progress would need to be taken into account. In other words, awareness would need to be raised as to the inherent volatility of digital transformation mostly due to the fundamental uncertainties in hardware and software innovation.

12 Following this background, the idea behind our present conference has been to embed a number of central themes providing for a platform for further discourse. These include, but are not limited to, a) technical concept of 'distributed by design' and legal uncertainties as to jurisdictional boundaries, b) attribution of liability in DLT-based networks in the absence of a clear definition of roles and responsibilities among actors and stakeholders, c) digitisation of the state and potential consequences on fundamental rights of individuals, d) growing dominance of corporate entities in big data analytics and implications on the concepts of individual consent and control, e) data inaccuracy and bias in automated decision making processes and possible technical tools for detection and mitigation thereof, and f) identity management systems and individuals' control over all matters related to processing of personal data.

13 Contributions to this edition have mostly taken an interdisciplinary approach addressing, directly or indirectly, a combination of any of the above themes or more, synopses of which could be encapsulated as follows, without any particular order.

- With reference to automated decision making, in particular methods that are enabled by machine learning, a first paper acknowledges increased threats to the fundamental rights of data subjects. In doing so, the authors Kumkar and Roth-Isigkeit take the view that *explanation* requirements are merely a necessary starting point for a human review, arguing that the subjective legal asset discussed under the term *right to explanation* actually turns out to be a preparatory *right to justification*. On the one hand, this viewpoint would allow the law to reflect the general opacity of intelligent decision making systems in order to provide for a practical way of dealing with the limited explicability. On the other hand, law recognises the *autonomy* of intelligent decision making systems to the extent that the procedural and deterministic explanation of decision making is replaced by the subsequent substantive legality test. Law thus finds its mode of dealing with the non-explicability of machine decisions in converting its procedures to the model of justification adapted to human decisions.

- *Informational self-determination* is seen as the underlying rationale of the fundamental right to the protection of personal data as enshrined in Article 8 of the Charter of Fundamental Rights of the EU. The author Thouvenin adopts the stance that acknowledging informational self-determination as a fundamental right would mean that the state may not require citizens to provide information about themselves and government agencies may not use such information without a sound legal basis, leaving out any obligation on the part of the private actors. Contrary to a widespread assumption, the author stipulates that most data processing of private actors is not based on data subjects' consent but on the legitimate interests of the controller. The

7 Reference to the research group coordinated by Ronny Thomale and Giorgio Sangiovanni, Lehrstuhl für Theoretische Physik 1, JMU, Würzburg.

relation between data subjects and private actors, namely businesses that process personal data about their customers, is therefore hardly ever based on exercising informational self-determination. This factual finding is supported by a normative analysis which demonstrates that the idea of informational self-determination can hardly be reconciled with the principle of private autonomy and the resulting need to provide a justification for the granting of a right that allows one private actor to control the activity of another. Thus, while informational self-determination may be acknowledged as a fundamental right, the author concludes that the concept cannot serve as a convincing rationale for an all-encompassing regulation of the processing of personal data by private actors.

- Over the last two decades, the number of organisations, both in the public and private sector, which have automated decisional processes, has grown notably. The phenomenon has been enabled by the availability of significant amounts of personal data and the development of software systems that use those data in order to optimise decisions with respect to certain optimisation goals. Today, software systems are involved in a wide realm of decisions that are relevant for the lives of people and the exercise of their rights and freedoms. The approach taken in this paper by the author Vetrò shifts the focus away from the outcomes of automated decision making systems and instead concentrates on inputs and processes. The foundations of a risk assessment approach are then laid based on a measurable characteristic of input data, i.e. *imbalance*, which can lead to discriminating automated decisions.

- A significant opportunity to engage in greater scrutiny of the digital transformation of the state, and its impact on fundamental rights, presented itself in a landmark judgment from the Netherlands. In the said case, the automated welfare-fraud detection system called *Systeem Risico Indicatie* (SyRI) was considered, allowing for the linking and analysis of data from an array of sources in order to generate fraud-risk reports on the public. In its judgment, the Court held that the legislation underpinning SyRI violated *the right to private life*, guaranteed under Article 8 ECHR. Taking a case study approach, the authors Appelman, Ó Fathaigh and van Hoboken highlight an important principle taken into account by the Court, namely the *special responsibility* that would need to be assumed by the government when applying new technologies to strike the right balance between the benefits the use of such technologies brings, and the potential interference with the exercise of *the right to private life*.

- By definition, blockchain[8] platforms offer secure and reliable data exchange between stakeholders without a trusted third party. Private and consortium blockchains implement access restrictions, so that private data would in principle be kept from the public. However, due to its distributed structure, only by means of one node all blockchain data could risk being leaked, due to a faulty configuration. This study by authors Hofmann, Gwinner, Winkelmann and Janiesch depicts ways in which confidential information could be revealed from blockchains, which should not be exposed to the public and which would potentially include identities, contract data as well as legal data. Thereby, the legal and social implications of data leakage by this distributed and supposedly secure technology are illustrated. In summary, the paper concludes that the large attack surface of private or consortium blockchains poses a threat to the security of the networks, raising the question whether (private) blockchain networks can reach a consensus without sharing all data between nodes and what data distribution strategies defend best against weak links in the chain.

- Blockchain technology is associated with the emergence of Decentralised Autonomous Organisations (DAO) as sovereign and software-based agents. A blockchain-based peer to peer vending machine as a physical marketplace, governed by a DAO, serving as both a testing ground and a speculative artefact is posited and analysed from a *de lege lata* perspective, taking into account the foremost liability questions from both Swiss private law (tort and contractual) and public law (criminal and tax law) perspectives. For this, the authors Schuppli and A. Jafari propose a hypothetical case study upon which the legal analysis is applied. As a result of the analysis the paper highlights where the current Swiss legal framework produces unsatisfactory results. From a private law perspective, the fact that contracting parties have little to no factual recourse in case of a purchase of counterfeit goods is an undesirable state from a public policy perspective. In other words, neither consumer protection nor good faith in commercial dealings would be viably upheld in this scenario. From a public law perspective, on the other hand, it is depicted that the state faces insurmountable challenges in taxing and collecting the taxable transactions involving a blockchain-based vending machine. Also, perpetrators of criminal offences, i.e., members of a DAO or unidentifiable

8 See n 6.

associates of a DAO, could likely not be brought to justice – an outcome which directly infringes on the public good of legal protection and undermines trust in government. The authors take the position that Swiss substantive law currently does not offer a satisfactory framework to deal with such novel decentralised market infrastructures. Individuals interacting with the proposed infrastructure, be it as vendors, buyers or members of the DAO, would face uncertainty related to both private and public law enforcement. Thus, the overall functioning of the legal economy and the rule of law would be infringed upon.

- The ultimate contribution puts Facebook's Diem[9] project under scrutiny. On the one hand, many critics have recognised dangers to state currency sovereignty and the stability of the financial system; on the other hand, they fear negative developments regarding money laundering and the financing of terrorism. In addition, there are considerable concerns about an ever deeper erosion of privacy, consumer and data protection, which reaches a new dimension by linking such world currencies with already existing social networks governed and controlled by private entities. Under these circumstances, the chance of success of the Diem project clearly depends on the extent to which the aforementioned concerns can be dispelled and whether *public trust* can be established. Moreover, it is argued that the level of control by end users over their digital representations and online footprints remains untested in the context of a worldwide digital financial infrastructure as proposed by Diem. The authors A. Jafari and Gruber further elaborate and put data protection and privacy of end users under scrutiny, outlining the need for a self-sovereign identity (SSI) management system in order to address the risks associated with correlation and profiling of individuals concerning their behaviour in payment systems. The paper then concludes that for Diem to experience a realistic mass adoption and to serve as a complementary infrastructure to the established monetary systems, it must itself prove to be a constitutive part of the *lex digitalis*. Evolving into the *lex cryptographia*, it will depend on the *pouvoir constituant* of the digital world whether it succeeds in further developing a digital civil constitution in the medium of DLT. Such a constitution, not least with its respective identity management, will determine what human life will be like in a truly *vibrant ecosystem*.

14 Lastly, we would like to take the opportunity to forward our special regards to everyone who has contributed and provided us with support throughout this process, in particular special thanks to the SOCAI team members, Alina Machert, Sarah Eisert, Jason Coombe, Ruben Maass and Dominik Klaus as well as each and every one of the contributors to the SOCAI/JIPITEC special edition. Further thanks extend to Thomas Dreier and Gerald Spindler for accepting this volume as a special edition of the *Journal of Intellectual Property, Information Technology and E-Commerce Law* (JIPITEC) and Lydia Förster for her support of the editorial work at JIPITEC. The publication of this edition was supported by the Würzburg-Dresden Cluster of Excellence on Complexity and Topology in Quantum Matter — ct.qmat (EXC 2147, project ID 390858490).

9 Diem is formerly known as Libra.

Informational Self-Determination: A Convincing Rationale for Data Protection Law?

by **Florent Thouvenin***

Abstract: European data protection law rests on the assumption that individuals should have control of personal data about them. This control is often labelled "informational self-determination". The idea of informational self-determination sounds convincing and promising at first. However, a closer look reveals that this idea can hardly serve as a convincing rationale for the European approach to data protection law which aims to regulate all processing of personal data by government agencies and private actors. Rather, an important distinction must be made.

Informational self-determination may well be the underlying rationale of the fundamental right to the protection of personal data as enshrined in Art. 8 of the Charter of Fundamental Rights of the European Union and it may even be qualified as a fundamental right in itself. Acknowledging such a fundamental right, however, only means that the state may not require citizens to provide information about themselves and government agencies may not use such information without a sound legal basis. But since private actors are not bound by fundamental rights, it does not entail that the relation between private actors should be based on the idea of informational self-determination. In fact, a closer look at the most important provisions of the GDPR reveals that only some of them can be based on the idea of control or informational self-determination. Most importantly and contrary to a widespread assumption, most data processing of private actors is not based on data subjects' consent but on the legitimate interests of the controller. The relation between data subjects and private actors, namely businesses that process personal data about their customers, is therefore hardly ever based on exercising informational self-determination. This factual finding is supported by a normative analysis which demonstrates that the idea of informational self-determination can hardly be reconciled with the principle of private autonomy and the resulting need to provide a justification for the granting of a right that allows one private actor to control the activity of another. If one acknowledges that all social interaction is based on the processing of personal data, that most individuals have little interest in exercising control of personal data about them, and that data is a public good, it is hard to find a convincing reason for the granting of a right to informational self-determination which should govern the relation between private actors. Thus, while informational self-determination may be acknowledged as a fundamental right, it cannot serve as a convincing rationale for an all-encompassing regulation of the processing of personal data by private actors.

Keywords: Informational Self-Determination; Data Protection Law; Rationale; Privacy; Protection against Harms

Recommended citation: Florent Thouvenin, Informational Self-Determination: A Convincing Rationale for Data Protection Law?, 12 (2021) JIPITEC 246 para 1.

A. Introduction

1 For quite some time, data protection received little attention in law and was largely disregarded by the public. In recent years, this has fundamentally changed. The digitalisation of multiple activities and the enactment of the General Data Protection Regulation (GDPR) sparked an intense debate in academia, the media and the public. Numerous scholarly papers and newspaper articles have been published – both in law and in other disciplines. However, despite the growing interest and importance of data protection law, fundamental questions remain unanswered. Arguably the most significant one being that of the theoretical foundation of this field of law.

2 In Europe, surprisingly little time and effort has been devoted to investigate the theoretical foundation of data protection law and to identify a convincing rationale for the European approach which consists of an all-encompassing regulation of the processing of all personal data by government agencies and private actors[1]. The lack of in-depth analysis is quite striking given that the EU introduced a fundamental right to the protection of personal data[2] and enacted the GDPR which is regarded the single most important piece of regulation the EU has issued so far. As opposed to Europe, the notion and concept of privacy have been debated in the US since the publication of the seminal article of Warren and Brandeis in 1890[3]. While it is certainly true that privacy is a broader concept than data protection as it also covers issues such as bodily privacy, locational privacy, or solitude[4], the US-American concept of informational privacy is quite closely related to the European concept of data protection law. While informational privacy and data protection law are often treated as identical concepts in the media and in public and private debate, it is well understood today that the two concepts need to be distinguished[5].

3 This paper focuses on the idea of informational self-determination and questions this concept's ability to serve as a rationale for European data protection law. It thereby focusses on the all-encompassing regulation of the processing of personal data by private actors as provided for by the GDPR[6]. To this

* Prof. Dr., Professor of Information and Communications Law, Chair of the Executive Board of the Center for Information Technology, Society, and Law (ITSL) and Director of the Digital Society Initiative (DSI) of the University of Zurich. I thank Dr. Stephanie Volz, managing director of the ITSL, for research assistance.

1 While there is quite some debate in Germany, there seems to be an almost complete lack of discussion, especially in the UK, and to a lesser extent also in France. Note that most German authors focus on the public sector when discussing the theoretical foundation of data protection law; see: Alexander Rossnagel, Kein "Verbotsprinzip" und kein "Verbot mit Erlaubnisvorbehalt", Zur Dogmatik der Datenverarbeitung als Grundrechtseingriff, NJW 2019, 1–5; Wolfgang Hoffmann-Riem, 'Informationelle Selbstbestimmung in der Informationsgesellschaft' in Wolfgang Hoffmann-Riem (ed), *Offene Rechtswissenschaft* (Mohr Siebeck 2010); Gabriel Britz, 'Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts' in Wolfgang Hoffmann-Riem (ed), *Offene Rechtswissenschaft* (Mohr Siebeck 2010) 561-596; Marion Albers, 'Umgang mit personenbezogenen Informationen und Daten' in Wolfgang Hoffmann-Riem and others (eds), *Grundlagen des Verwaltungsrechts* (2nd edn, C.H. Beck 2012) 107-234; Johannes Masing, 'Herausforderungen des Datenschutzes' [2012] NJW 2305-2311; Karl-Heinz Ladeur 'Das Recht auf informationelle Selbstbestimmung: Eine juristische Fehlkonstruktion? (2009) 2 DöV 45-55.

2 Art. 8 Charter of Fundamental Rights of the European Union.

3 Samuel D Warren and Louis D Brandeis, 'The Right to Privacy' (1890) 193 Harvard Law Review 22.

4 For the different concepts of privacy see: Daniel J Solove, 'Understanding Privacy' (2008) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127888> accessed 15 November 2021, 13ff; Helen Nissenbaum, *Privacy in Context - Technology, Policy, and the Integrity of Social Life* (Stanford Law Books 2010) 67ff; Alan F Westin, *Privacy and Freedom* (Atheneum 1967) 77; Charles Fried, 'Privacy' [1968] The Yale Law Journal 475; Ruth Gavinson, 'Privacy and the Limits of Law' (1980) 89 The Yale Law Journal 421; Randall P Bezanson, 'The Right to Privacy Revisited: Privacy, News and Social Change 1890-1990' (1992) 80 California Law Review 1133; Adam Moore, 'Defining Privacy' (2008) 39 Journal of Social Philosophy 411; Bert-Jaap Koops and others, 'A Typology of Privacy' (2017) 38 (2) University of Pennsylvania Journal of International Law 483.

5 Gernot Sydow, 'Artikel 1 DSGVO' in Gernot Sydow (ed), *Europäische Datenschutzgrundverordnung* (2nd edn, Nomos 2018) para 10ff; Orla Lynskey 'Deconstructing data protection: The 'added value' of a right to data protection in the EU legal order' (2014) 63 International and Comparative Law Quarterly 567ff.; Raphaël Gellert and Serge Gutwirth, 'The legal construction of privacy and data protection' (2013) 29 Computer Law & Security Review 522, 523ff.

6 For a critical evaluation of the right to informational self-determination as a fundamental right and a governing principle for the processing of data by government agencies see: Ladeur (n 1) 45; Albers (n 1) 107; Marion Albers, 'Realizing the Complexity of Data Protection' in Serge Gutwirth and others (eds), *Reloading Data Protection* (Springer 2014) 213-235; Britz (n 1) 561-596; Paul De Hert and Serge Gutwirth, 'Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of the

end, the paper first outlines the idea and concept of informational self-determination (B.); second, analyses the fractional implementation of this concept in the GDPR (C.); and third, demonstrates that informational self-determination cannot be considered a feasible rationale for data protection law (D.). The paper concludes with a call for the development of alternatives, both with regard to the need for a convincing rationale and alternative regulatory approaches that can build upon and properly implement such rationale (E.).

B. Idea and Concept

4 The idea and concept of informational self-determination refers to every individual's right and opportunity to determine which information about him- or herself is disclosed to others and for what purposes such information may be used[7]. In Europe, the notion of an individual's right to informational self-determination was first articulated by the Federal Constitutional Court of Germany in its landmark decision on the Federal Census Act of 1983[8]. Herein the Court suspended the carrying out of a population census and ruled that the Federal Census Act must be amended before census may resume. The Court based its ruling on the argument that the rights to human dignity and integrity as enshrined in the Basic Law of Germany provides for a more specific fundamental right of every individual to decide on the disclosure and use of his or her personal information[9].

5 Since 1983, the term and idea of informational self-determination have had a successful career in legal thinking and in public debate, at least in Europe where the right to informational self-determination has become one of the conceptual foundations for the right to the protection of personal data as enshrined in Art. 8 of the Charter of Fundamental Rights of the European Union[10]. Following the decision of the German Constitutional Court in 1983, many even argue that the right to informational self-determination is a fundamental right in itself[11]. This

Power' in Erik Claes and others (eds) *Privacy and the Criminal Law* (Intersentia 2006) 61-104; Gellert and Gutwirth (n 5) 522-530; Gloria Gonzáles Fuster *The Emergence of Personal Data Protection as a Fundamental Right of the EU* (Springer 2014); Gloria Gonzáles Fuster and Serge Gutwirth, 'Opening up personal data protection: A conceptual controversy' (2013) 29 Computer Law & Security Review 531-539; Nikolaus Marsch, *Das europäische Datenschutzgrundrecht* (Mohr Siebeck 2018) 98ff.; Nikolaus Marsch, 'Artificial Intelligence and the Fundamental Right to Data Protection' in Thomas Wischmeyer and Timo Rademacher (eds) *Regulating Artificial Intelligence* (Springer 2020) 33-52; Ralf Poscher, 'The Right to Data Protection: A No-Right Thesis' in Russel A. Miller (ed) *Privacy and Power: A Transatlantic Dialogue in the Shadow of the NSA-Affair* (Cambridge University Press 2017) 129-142; Ralf Poscher, 'Artificial Intelligence and the Right to Data Protection' (2021) Max Plank Institute for the Study of Crime, Security and Law Working Paper No. 2021/03 <https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3806531_code386115.pdf?abstractid=3769159&mirid=1> accessed 15 November 2021.

7 Schwartz, 'Regulating Governmental Data Mining in the United States and Germany: Constitutional Courts, the State, and New Technology' [2011] William and Mary Law Review 351, 368; Kenneth A Bamberger and Deirdre K Mulligan, 'Privacy in Europe: Initial Data on Governance Choices and Corporate Practices' [2013] The George Washington Law Review 1529, 1539.

8 Decision of the Federal Constitutional Court of Germany of 15 December 1983, Az 1 BvR 209/83, 1 BvR 484/83, 1 BvR 420/83, 1 BvR 362/83, 1 vR 269/83, 1 BvR 440/83, BVerfGE 65, 1 – Volkszählung.

9 BVerfGE 65 (n 8) 43 - Volkszählung.

10 Peter Gola, 'Einleitung' in Peter Gola (ed), *Datenschutz-Grundverordnung: DS-GVO* (2nd edn, C.H. Beck 2018) para 6; Bernd Schmid, 'Art. 1 DSGVO' in Jürgen Taeger and Detlev Gabel (eds), *DSGVO BDSG* (3rd edn, Deutscher Fachverlag GmbH, Fachmedien Recht und Wirtschaft 2019) para 25; Jürgen Kühling and Johannes Raab, 'Einführung' in Jürgen Kühling and Benedikt Buchner (eds), *Datenschutz-Grundverordnung BDSG Kommentar* (3rd edn, C.H. Beck 2020) para 26, see also Antoinette Rouvroy and Yves Poullet 'The Right to Informational Self-Determination and the Value of Self-Development' in Serge Gutwirth and others (eds), *Reinventing Data Protection* (Springer 2009) 51, 68.

11 Schwartz (n 7) 364, 367ff; Brendan Van Alsenoy and Eleni Kosta and Jos Dumortier, 'Privacy notices versus informational self-determination: Minding the gap' [2014] International Review of Law, Computers & Technology 185, 188; Markus Thiel, *Die „Entgrenzung" der Gefahrenabwehr* (Mohr Siebeck 2011) 221; Claudio Franzius, 'Das Recht auf informationelle Selbstbestimmung' [2015] Zeitschrift für das juristische Studium 259; René Rhinow and Markus Schefer and Peter Übersax (eds), *Schweizerisches Verfassungsrecht* (3rd edn, Helbing & Lichtenhahn 2016) para 1376ff; Eva Maria Belser, ‹Zur rechtlichen Tragweite des Grundrechts auf Datenschutz: Missbrauchsschutz oder Schutz der informationellen Selbstbestimmung?› in Astrid Epiney and others (eds), *Instrumente zur Umsetzung des Rechts auf informationelle Selbstbestimmung/Instruments de mise en oeuvre du droit à l'autodétermination informationnelle* (Schulthess 2013) 25; critical of the characterisation as a fundamental right: Hans Peter Bull, *Informationelle Selbstbestimmung – Vision oder Illusion?* (2nd edn, Mohr Siebeck 2011) 45ff; Alexandre Flückiger, 'L'autodétermi-

approach has also been adopted by the Swiss Federal Supreme Court[12] even though the Swiss Federal Constitution solely provides for a right of every person to be protected against the misuse of his or her personal data (Art. 13 (2) Swiss Federal Constitution). The right to informational self-determination has also evolved regarding its content. For some authors, this right does not only allow individuals to decide on the disclosure and use of information about them but grants them full control of the use of "their" personal data[13].

6 In contrast, the German Federal Constitutional Court has significantly attenuated its understanding of the right to informational self-determination in a relatively recent decision by stating that this right does not confer a general or even comprehensive right to self-determination with regard to the use of one's own personal data; instead, it shall only provide individuals a right to have a substantial say in the making available and the use of their personal data[14].

7 Regardless of this remarkable confinement, the aforementioned view according to which the right to informational self-determination grants every individual a right to decide on the disclosure and use of his or her personal information is still the predominant understanding of the idea and concept of informational self-determination in Europe. Most prominently, this "classical" understanding of informational self-determination has been adopted by the French legislator who explicitly states in its law on electronic data processing, files and freedoms that every individual has a right to decide on and control the use of their personal data and that this right must be exercised within the framework of the GDPR and the aforementioned national law[15].

8 Even if one agrees that the right to informational self-determination is a fundamental right, this right may only serve as a rationale for regulating the processing of personal data by government agencies. Such regulation(s) would have to define what personal data government agencies may collect about their citizens and under what conditions and for which purposes the data may be processed. But as private actors are not (directly) bound by fundamental rights[16], informational self-determination cannot readily serve as a rationale for regulating the processing of personal data by private actors[17]. Instead, a more in-depth analysis is needed.

C. Actual Implementation

9 The GDPR hardly provides any guidance as to its rationale. The wording of its objective is very broad and general. According to Art. 1 (2) GDPR the regulation aims at protecting "fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data". Even though this objective may serve as a (quite unspecific) guidance for the processing of personal

nation en matière de données personnelles: un droit (plus si) fondamental à l'ère digitale ou un nouveau droit de propriété?' [2013] Aktuelle Juristische Praxis, 837 passim; Thomas Gächter and Philipp Egli, 'Informationsaustausch im Umfeld der Sozialhilfe – Rechtsgutachten' (Jusletter, 6 September 2010) <https://jusletter.weblaw.ch/juslissues/2010/583/_8587.html> accessed 15 November 2021, para 21ff.

12 Swiss Federal Supreme Court (BGE 146 I 11) [2019] at 3; Swiss Federal Supreme Court (BGE 145 IV 42) [2018] at 4.2; Swiss Federal Supreme Court (BGE 143 I 253) [2017] at 4.8; Swiss Federal Supreme Court (BGE 142 II 340) [2016] at 4.2; Swiss Federal Supreme Court (BGE 140 I 2) [2014] at 9, all with further references.

13 Rouvroy and Poullet (n 10) 45.

14 BVerfGE – 1 BvR 16/13, 87.

15 Art. 1 al. 2 de la loi n. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés: «Les droits des personnes de décider et de contrôler les usages qui sont faits des données à caractère personnel les concernant et les obligations incombant aux personnes qui traitent ces données s'exercent dans le cadre du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, de la directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 et de la présente loi.

16 Art. 51 Charter of Fundamental Rights of the European Union; Art. 16 para 2 Treaty of the Functioning of the European Union; Art. 1 para 3 Basic Law of the Federal Republic of Germany; Art. 35 para 3 Swiss Federal Constitution *e contrario*; see also Stefanie-Daniela Waldmeier, *Informationelle Selbstbestimmung – ein Grundrecht im Wandel* (Dissertation, 2015), 104 <https://www.zora.uzh.ch/id/eprint/122636/> accessed 15 November 2021 and Masing (n 1) 2305.

17 This is disregarded by the Federal Constitutional Court of Germany and the Swiss Federal Supreme Court. In BVerfG, 1 BvR 16/13, 85, the German Constitutional Court has stated that there is no reason for not applying the fundamental right to informational self-determination in the relation between private actors. The Swiss Federal Supreme Court has repeatedly stated that the fundamental right to informational self-determination implies that every individual has a right to decide about the processing of personal data about them by government agencies and private actors; see Swiss Federal Supreme Court (BGE 146 I 11) [2019] at 3; Swiss Federal Supreme Court (BGE 144 II 91) [2017] at 4.4; Swiss Federal Supreme Court (BGE 140 I 2) [2014] at 9, all with further references.

data by government agencies, it can hardly serve as a rationale for the all-encompassing regulation of the processing of personal data by private actors given that they are not directly bound by fundamental rights.

10 While the provision on the objective of the GDPR does not give clear guidance as to the regulation's rationale, recital 7 provides some by stating that "Natural persons should have control of their own personal data". Although the GDPR does not mention informational self-determination, the idea of natural persons controlling their own personal data is to be considered an identical concept labelled less eloquently. Accordingly, at least in the German speaking part of Europe, many scholars agree that the idea of informational self-determination is the underlying rationale of the GDPR[18].

11 The search for a convincing rationale is not merely a theoretical problem since the often very broad notions of the GDPR require an interpretation of the legal text which must be carried out (amongst others) with regard to the purpose of the law[19]. By applying these notions in one way or another, scholars, practitioners and – most importantly – supervisory authorities and courts, make implicit assumptions about the rationale of data protection law. Given their impact on the interpretation and application of the GDPR, these assumptions should be made explicit to allow for a critical assessment of the assumed rationale and the resulting decisions.

12 If the GDPR aims to put the idea of control or informational self-determination into action, this raises the question if this concept is duly implemented and able to provide a sound theoretical basis for the most important rules and procedures established in the GDPR. The key provisions that must be analysed for this assessment are the principles relating to the processing of personal data (Art. 5 GDPR), the rules on the lawfulness of processing, including the specific provisions on consent (Art. 6 et seqq. GDPR), the rights of the data subjects (Art. 12 et seqq. GDPR), and the rules on the enforcement of the provisions, namely the ones on the competence, tasks and powers of the supervisory authorities (Art. 55 et seqq. GDPR) and the ones on remedies, liability and penalties (Art. 77 et seqq. GDPR).

13 The *principles relating to the processing of personal data* (Art. 5 GDPR) can only be explained to a very limited extent by the idea of informational self-determination. Transparency (Art. 5 (1) (a) GDPR), purpose limitation (Art. 5 (1) (b) GDPR) and security of data processing (Art. 5 (1) (f) GDPR) are key prerequisites for informational self-determination as exercising control requires that data subjects are informed about the processing of personal data about them, that this data is not processed for purposes which are incompatible with the ones the data subjects have been informed about, and appropriate security measures are implemented to prevent unauthorised processing and accidental loss or destruction of the data. But the other principles, namely the principles of lawfulness and fairness (Art. 5 (1) (a) GDPR), data minimisation (Art. 5 (1) (c) GDPR), accuracy (Art. 5 (1) (d) GDPR) and storage limitation (Art. 5 (1) (e) GDPR) do not aim at establishing control of data subjects. While these principles may serve legitimate goals, they cannot be based on the concept of informational self-determination.

14 Together with the principles of data protection, the rules on the *lawfulness of processing* (Art. 6 GDPR) form the normative core of the GDPR. The most prominently regulated and most intensively discussed reason for the lawfulness of processing of personal data is the data subject's consent (Art. 6 (1) (a); Art. 7 et seq. GDPR). The requirement of consent is evidently a straightforward implementation of informational self-determination. Though many data subjects believe that the processing of personal data about them is usually based on their consent, consent is far from being the prevailing basis for the lawfulness of processing. Unfortunately, there is no empirical evidence available on the relative importance of the various legal grounds for the processing of personal data. But for all practitioners – data protection officers, data protection lawyers and supervisory authorities – it is clear that in the vast majority of cases the lawfulness of processing is not based on consent but on the legitimate interests pursued by the controller (Art. 6 (1) (f) GDPR). In informal exchanges, prominent data protection commissioners have assumed that this is true for more than 90% of data processing activities. Regardless of how accurate this number may be, the relative importance of data subjects' consent and the legitimate interests of controllers as a legal basis for the processing of personal data is very clear. Evidently, the most important legal ground for the processing of personal data is not based on the idea of informational self-determination but on the need of controllers to process personal data in a wide range of situations. The fact that the data subjects' interests are considered when assessing the legitimate interests of the controller does not make any difference as the data subjects have no means to influence the balancing of interests, e.g. by

18 Kühling and Raab (n 10) para 26; Masing (n 1) 2305; Jan Philip Albrecht 'Die EU-Datenschutzgrundverordnung rettet die informationelle Selbstbestimmung!' [2013] Zeitschrift für Datenschutz 587; critical Winfried Veil, 'Die Datenschutz-Grundverordnung: des Kaisers neue Kleider' [2018] Neue Zeitschrift für Verwaltungsrecht 686, 691.

19 Schmid (n 10) para 19; Pötters, 'Art. 1' in Gola (n 10) para 20.

providing their own point of view on the processing subject to the assessment. The concept of informational self-determination cannot serve as a basis for the other reasons for the lawfulness of processing either, namely the processing for the performance of a contract (Art. 6 (1) (b) GDPR), for compliance with a legal obligation of the controller (Art. 6 (1) (c) GDPR), and for the performance of a task carried out in the public interest (Art. 6 (1) (e) GDPR). The only exception is the processing of personal data for protecting the vital interests of the data subject (Art. 6 (1) (d) GDPR), which is based on the assumption of the data subject's consent[20]. Given the very limited importance of consent for the lawfulness of processing, it proves impossible to ground the assessment of the legal basis for the processing of personal data on the concept of informational self-determination.

15 As opposed to the lawfulness of processing, the *rights of data subjects* (Art. 12 et seqq. GDPR) can clearly be based on the concept of informational self-determination. This holds true for the obligation of controllers to provide data subjects with a wide range of information (Art. 13 et seq. GDPR) and for the specific rights of data subjects, namely the right of access (Art. 15 GDPR), the right to rectification (Art. 16 GDPR), the right to erasure (Art. 17 GDPR), the right to restriction of processing (Art. 18 GDPR), and the right to object (Art. 20 GDPR). But even here, the control of data subjects is limited as some rights come with important restrictions. Namely, the right to erasure is merely granted if one out of a limited set of situations is given, e.g. if personal data is no longer necessary in relation to the purpose for which it was collected (Art. 17 (1) (a) GDPR) or if the data subject withdraws consent and there is no other legal ground for the lawfulness of processing (Art. 17 (1) (b) GDPR). The same is true for the right to restriction of processing even though the situations in which such a right takes effect are different (Art. 18 GDPR). Most importantly, data subjects have no general right to object to the processing of their personal data. Instead, this right is only granted if the processing of personal data is based on the legitimate interest of the data controller (Art. 6 (f) GDPR) or if it is necessary for the performance of a task carried out in the public interest (Art. 6 (e) GDPR). In addition, the right to object must always be exercised on grounds relating to the particular situation of the data subject (Art. 21 (1) first sentence GDPR), e.g. for reasons relating to their family life or for the protection of trade secrets. Even if such reasons are given, the right to object is subject to another very general restriction since the controller may continue to process the data if it is able to demonstrate compelling legitimate grounds for the processing of the data which override the interests, rights and freedoms of the data subject (Art. 21 (1) second sentence GDPR). Even though compelling grounds may be given in many instances as they have to be assessed in a pondering of interests of the controller on the one hand and the data subject on the other[21], personal data can be processed in many cases against the data subject's express objection.

16 The *enforcement* of data protection law is primarily ensured by supervisory authorities; they are responsible for monitoring and enforcing the application of the GDPR (Art. 57 (1) (a) GDPR). They are vested with far-reaching powers including (amongst many others) the power to carry out investigations in the form of data protection audits (Art. 58 (1) (b) GDPR), to order the controller or processor to bring processing operations into compliance with the provisions of the GDPR (Art. 58 (2) (d) GDPR), to impose a temporary or definitive limitation or a ban on the processing of personal data (Art. 58 (2) (f) GDPR), and, to impose an administrative fine (Art. 58 (2) (i) GDPR). While Supervisory authorities may act in response to complaints of data subjects (Art. 57 (1) (f) GDPR) or initiate investigations themselves as they see fit (Art. 57 (1) (h) GDPR). Although the GDPR grants every data subject a right to an effective judicial remedy against a controller or processor (Art. 79 GDPR) and the right to receive compensation for a damage suffered (Art. 82 GDPR), these rights are hardly used. Instead, the enforcement of the provisions of the GDPR almost entirely banks on the supervisory authorities. While these authorities are certainly convinced to act in the best interest and on behalf of data subjects, the concept of enforcement by an independent supervisory authority can hardly be reconciled with the idea of informational self-determination, i.e. the idea that data subjects decide by themselves about the processing of their personal data.

17 The brief analysis of the most important rules and procedures established in the GDPR has revealed that only a limited number of its key provisions can be based on the idea of control or informational self-determination. Most importantly, in about nine out of ten cases the processing of personal data by

20 See also Jürgen Taeger 'Art. 6 DSGVO' in Taeger and Gabel (n 10) para 46; Philipp Reimer, 'Art. 6 DSGVO' in Sydow (n 5) para 3; Benedikt Buchner and Thomas Petri, 'Art. 6 DSGVO', in Kühling and Buchner (n 10) para 109f; Peter Schantz, 'Art. 6 DSGVO' in Simitis and others (eds), *Datenschutzrecht: DSGVO mit BDSG* (Nomos 2019) para 61.

21 Sebastian Schulz, 'Art. 21 DSGVO' in Gola (n 10) para 12; Mario Martini 'Art. 21DSGVO' in Boris Paal and Daniel Pauly (eds), *DS-GVO BDSG* (3rd edn, C.H. Beck 2021) para 29; Tobias Herbst, 'Art. 21 DSGVO' in Kühling and Buchner (n 10) para 19ff; Johannes Caspar, 'Art. 21 DSGVO' in Simitis and others (n 20) para 19; Martin Braun and Hans-Georg Kamann 'Art. 21 DSGVO' in Eugen Ehmann and Martin Selmayr (eds), *DS-GVO: Kommentar* (2nd edn, C.H. Beck 2018) para 22ff.

private actors is based on the legitimate interests of the controller and not on data subjects' consent. Given the key importance of the legal basis for the processing of personal data under the GDPR, the finding alone that most processing of personal data is based on the legitimate interests of the controller and not on data subjects' consent clearly demonstrates that the GPPR does not implement the idea of informational self-determination. This finding is amplified by the fact that the restrictions to the right to object even allow for the processing of personal data against the explicit will of data subjects. The lack of implementation of informational self-determination in the GDPR endorses the finding that this idea and concept cannot be perceived as the underlying rationale for European data protection law.

D. Normative Analysis

18 The finding that informational self-determination is not truly implemented in the GDPR despite the intention of the European legislator to grant individuals control of "their" personal data raises serious doubts as to the feasibility of this concept. But this factual finding does not preclude that informational self-determination should be the rationale of data protection law and that the GDPR should be revised to ensure its proper implementation. However, there are also important doubts on a normative level as to whether informational self-determination is a feasible rationale.

19 At first, the idea of informational self-determination sounds very convincing. After all, liberty, dignity, autonomy or personal freedom, i.e. the right of every individual to decide about their own life within the limits of the law, are core values shared by most western societies and fundamental rights guaranteed implicitly or explicitly by most constitutions in Europe[22]. In the relationship between private actors, these core values are reflected in the principle of private autonomy. From this perspective, informational self-determination appears to be a logical, almost inevitable consequence or even part of the general right to self-determination[23]. Accordingly, scholars and courts referring to the idea of informational self-determination hardly ever provide an explanation as to why such a right should exist[24]. This is especially true for recital 7 of the GDPR which fails to provide any explanation as to why natural persons should have control of personal data about them. Given the importance and impact of the idea and concept of informational self-determination, this is astonishing. It seems that lawmakers, courts, and most scholars have been carried away by the persuasive power of an eloquent terminology. Surely, a closer analysis is needed.

20 This analysis must distinguish between the relation between individuals and the state and the relation between individuals and other private actors, namely businesses. The relation between individuals and the state is primarily determined by a set of fundamental rights and a set of laws that define and delimit the activities of government agencies. Acknowledging a fundamental right to informational self-determination thus only means that the state may not require citizens to provide information about themselves and government agencies may not use that information without a sound legal basis[25]. The situation presents differently, however,

22 For example: Art. 5 para 1 sentence 1 European Convention on Human Rights; Art. 1 and Art. 6 Charter of Fundamental Rights of the European Union; Art. 1 para 1 and Art. 2 para 2 Basic Law of the Federal Republic of Germany; Art. 7 and Art. 10 Swiss Federal Constitution.

23 In this sense also BVerfGE 65 (n 8) 42ff.

24 For Germany: Dietrich Murswiek and Stephan Rixen 'Art. 2 GG' in Michael Sachs (ed), *Grundgesetz* (9th edn, C.H. Beck 2021) para 72ff.; Udo Di Fabio 'Art. 2 Abs. 1 GG' in Theodor Maunz and Günter Dürig (eds), *Grundgesetz-Kommentar* (94th edn, C.H. Beck 2021) para 174f.. For Switzerland: Swiss Federal Supreme Court (BGE 120 II 118) [1994] at 3a; Swiss Federal Supreme Court (BGE 122 I 153) [1996] at 6b; Swiss Federal Supreme Court (BGE 138 II 346) [2012] at 8.2; Rainer Schweizer, 'Art. 13 Abs. 2 BV' in Stephan Breitenmoser and Rainer Schweizer (eds), *Die Schweizerische Bundesverfassung* (3nd edn, Dike 2014) para 72; Regina Kiener and Walter Kälin and Judith Wyttenbach, *Grundrechte – Stämpflis juristische Lehrbücher* (3rd edn, Stämpfli 2018) 178; Jörg Paul Müller and Markus Schefer (eds), *Grundrechte in der Schweiz* (4th edn, Stämpfli 2008) 164f; Waldmeier (n 16) 105.

25 For Germany: Murswiek and Rixen (n 24) para 13ff, 73; Di Fabio (n 24) para 178. For Switzerland: Schweizer (n 24) para 79; Giovanni Biaggini, 'Art. 13 BV' in Giovanni Biaggini (ed), *Bundesverfassung der Schweizerischen Eidgenossenschaft: Kommentar* (Orell Füssli 2017) para 11. Art. 52 of the Charter of Fundamental Rights of the European Union states that "Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others". With regard to data protection law see Benedikt Buchner 'Art. 1 DSGVO' in Kühling and Buchner (n 10) para 16; Heinrich Amadeus Wolff, 'AEUV Art. 16' in Matthias Pechstein and Carsten Nowak and Ulrich Häde (eds), *Frankfurter Kommentar EUV/GRC/AEUV* (Mohr Sie-

for private actors. According to the principle of private autonomy, private actors are free to pursue all activities they see fit[26] and the introduction of limitations calls for justification[27]. This also applies to the collection and use of personal data. This fundamental problem is mostly disregarded when promoting the idea of informational self-determination. Yet it is obvious that granting individuals a right to control the use of personal data about them inevitably leads to a limitation of all private actors to collect and use such data. Interestingly, such a limitation can hardly be integrated into the broad types of rights the law has developed to govern the relationship between private actors. Private law knows three basic types of rights that allow private actors to restrict the freedom of other private actors: property rights, tort law, and contracts. Of course, this categorisation is a gross simplification, and a much more detailed analysis would be needed to make the necessary distinctions. But looking at these very broad categories nevertheless reveals that a right of private actors to control the processing of personal data by other private actors is hard to integrate into our legal system. In any case, and contrary to what the notion of informational self-determination implies, a right to informational self-determination does not exist per se in the relationship between private actors and such right cannot be justified by simply stating that the fundamental right to informational self-determination should apply mutatis mutandis to the relation between private actors[28]. Instead, calling for a right to informational self-determination in the relationship between private actors requires a convincing justification. When looking for such a justification, three aspects should be considered.

21 First, all human interaction is based on the processing of personal data. We are constantly processing important amounts of data about others in our brains. But no one would consider that we should have a right to determine what others think about us[29]. This also applies to business relations, e.g. to a consumer shopping at a local grocery store. The shopkeeper will gather quite some information about the habits, preferences, moods, and financial resources of its customers and no one would call for the introduction of a right that would allow the consumer to control the processing of that data in the shopkeeper's brain. Why should this be fundamentally different if the data was stored on paper or an electronic device? In fact, it is not, as demonstrated by the key importance of the legitimate interests of controllers as a legal basis for the processing of personal data[30]. It is precisely because all human interaction is based on the processing of personal data that legislators and supervisory authorities cannot help but recognize scores of various instances in which personal data can be processed without data subjects' consent, thereby depriving them of their alleged right to informational self-determination.

beck 2017) para 11, 12; Philip Kunig and Jörn Axel Kämmerer 'Art. 2 GG' in Ingo von Münch and Philip Kunig (eds), *Grundgesetz Kommentar: GG* (7th edn, C.H. Beck 2021) para 78.

26 In Switzerland, private autonomy is the basis for economic freedom according to Art. 27 Swiss Federal Constitution; Kurt Vallender 'Art. 27 BV' in Bernhard Ehrenzeller and others (eds), *Die Schweizerische Bundesverfassung: St. Galler Kommentar* (4th edn, Dike 2014) para 51; Bernhard Waldmann, 'Art. 35 BV' in Bernhard Waldmann and Eva Maria Belser and Astrid Epiney (eds), *Basler Kommentar Bundesverfassung* (Helbing Lichtenhan 2015) para 71. In Germany, the concept of private autonomy is covered by Art. 2 para 1 Basic Law of the Federal Republic of Germany, see Udo Di Fabio 'Art. 2 Abs. 1 GG' in Maunz and Dürig (n 24) para 101; Christian Starck 'Art. 2 GG' in Hermann von Mangoldt and Friedrich Klein and Christian Starck (eds), *Grundgesetz* (7th edn, C.H. Beck 2018) para 145; Horst Dreier, 'Art. 2 Abs. 1 GG' in Horst Dreier (ed), *Grundgesetz-Kommentar* (3rd edn, Mohr Siebeck 2013) para 35, 62; Hans Jarass, 'Art. 2 GG' in Hans Jarass and Bodo Pieroth (eds), *Grundgesetz für die Bundesrepublik Deutschland – Kommentar* (16th ed, C.H. Beck 2020) para 22; Kunig and Kämmerer (n 25) para 78.

27 According to art 36 para 2 Swiss Federal Constitution restrictions on fundamental rights such as economic freedom must be justified by a public interest or by the protection of the fundamental rights of third parties; see also Biaggini (n 25) para 29; Vallender (n 26) para 57. The same applies in German law, see Di Fabio (n 24) para 104; Starck (n 26) para 19ff; Horst Dreier, 'Art. 2 Abs. 2 GG' in Dreier (n 26) para 47.

28 In this sense, for Germany: Masing (n 1) 2307f. For Switzerland: Swiss Federal Supreme Court (BGE 146 I 11) [2019] at 3.1.1; Swiss Federal Supreme Court (BGE 144 II 91) [2017] at 4.4; Swiss Federal Supreme Court (BGE 142 II 340) [2016] at 4.2, all with further references. In a similar way, but without referring to the idea of informational self-determination: De Hert and Gutwirth (n 6), stating that similar rationales apply with regard to the regulation of the processing of personal data in the public and the private sector.

29 Likewise: Masing (n 1) 2307. This problem has already been addressed in the seminal decision of the Federal Constitutional Court of Germany. The court has rightly pointed out that personal information is a reflection of social reality that cannot be exclusively assigned to a specific individual, which is why all individuals must accept restrictions on their right to informational self-determination; BVerfGE 65 (n 8) 1, 44 – Volkszählung.

30 See above, C.

22 Second, the concept of informational self-determination only makes sense if individuals care about the collection and usage of personal data about them. In other words, granting individuals control over the use of personal data is only meaningful if they actually exercise this control. Yet this is hardly the case. Only few data subjects use the rights granted by the GDPR and many studies show that privacy policies are hardly read[31]. Instead of exercising our supposedly important right to informational self-determination, most of us just click *accept* whenever we are asked if we agree to the processing of data about us. The lack of exercise also raises the question of whether the (limited) amount of control which is granted today is of any benefit to individuals. Even if one assumes that the mere possibility to exercise (some) control has a certain value for data subjects, the benefits created must be weighed against the costs incurred for granting that control. While reliable numbers are not available, one may infer from anecdotal evidence that the costs for establishing compliance with the GDPR are in the three-digit million range for the big tech companies and in the two-digit million range for many other large companies that serve customers in the EU[32]. And this solely includes the direct costs for compliance while disregarding the much greater costs of lost opportunities. Namely the costs for research and development and innovative business models which are not possible at all or are not carried out because of the limitations for the use of personal data and the liability risks caused by the GDPR. From this perspective, it can hardly be assumed that a regulation which is built on the concept of informational self-determination will create greater benefits than costs for society at large.

23 Third, data is a public good[33]. Such goods are characterised by two features: they can be used simultaneously by an unlimited number of persons without the use by one person affecting the use by another (non-rivalrous use) and no one can exclude others from the use of these goods (non-excludable use). Given the non-rivalrous use, the benefit of a public good for society is greatest, if it can be used by everyone. Accordingly, legal instruments that allow an individual to restrict the use of such goods should be granted only if such restrictions are needed to achieve other important policy goals. With regard to private actors, two aspects are key. First, a legal intervention is necessary if needed to protect individuals from harms caused by others; second, an intervention is needed in case of market failure, e.g. if a good valuable to society would not be produced if the producer were unable to reap the benefits it created[34]. The latter need for intervention has been debated in connection with the demand for the cre-

31 Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change, Recommendations for Businesses and Policy Makers, FTC Report, March 2012 2, 61; Daniel J Solove, 'Introduction: Privacy Self-Management and the Consent Dilemma' (2013) 126 Harvard Law Review 1880, 1884ff; Aleecia M McDonald and Lorrie F Cranor, 'The Cost of Reading Privacy Policies' (2008) 4 I/S: A Journal of Law and Policy for the Information Society 543, 565, estimate that it would take 201 hours annually for an American Internet user to read the privacy policies of all the services they use.

32 Concrete and reliable figures are not yet available and most companies will be reluctant to publish them. However, some indications can be gained from few publicly available statements. For example, according to an estimate by Forbes, compliance with the requirements of the GDPR costs Fortune 500 companies around $16 million; see Oliver Smith, 'The GDPR Racket: Who's Making Money From This $9bn Business Shakedown' (Forbes, 2 May 2018) <https://www.forbes.com/sites/oliversmith/2018/05/02/the-gdpr-racket-whos-making-money-from-this-9bn-business-shakedown/#4727a86434a2> accessed 15 November 2021; similarly, Jeremy Kahn and Stephanie Bodoni and Stefan Nicola, 'It'll Cost Billions for Companies to Comply With Europe's New Data Law' (*Bloomberg Businessweek*, 22 March 2018) <https://www.bloomberg.com/news/articles/2018-03-22/it-ll-cost-billions-for-companies-to-comply-with-europe-s-new-data-law> accessed 15 November 2021; Rita Heimes and Sam Pfeifle, 'Study: GDPR's global reach to require at least 75,000 DPOs worldwide' (*iapp*, 9 November 2016) <https://iapp.org/news/a/study-gdprs-global-reach-to-require-at-least-75000-dpos-worldwide/> accessed 15 November 2021.

33 For the notion of a public good: Richard A Posner, *Economic Analysis of Law* (9th edn, Aspen Publ 2014) 402; Robert Cooter and Thomas Ulen, *Law and Economics* (6th edn, Berkeley Law Books 2016) 40; Hans-Bernd Schäfer and Claus Ott, *Lehrbuch der* ökonomischen *Analyse des Zivilrechts* (6th edn, Springer 2020) 86f. With regard to data: Herbert Zech, *Information als Schutzgegenstand* (Mohr Siebeck 2012) 107ff; Thomas Heymann, 'Rechte an Daten, Warum Daten keiner eigentumsrechtlichen Logik folgen' [2016] Computer und Recht 650, 652ff; Wolfgang Kerber, 'A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis' [2016] Gewerblicher Rechtsschutz und Urheberrecht Internationaler Teil 989, 992ff; Lothar Determann, 'No One Owns Data' (2018) 70 Hastings Law Review 1, 41; Florent Thouvenin and Rolf H Weber and Alfred Früh, *Elemente einer Datenpolitik* (Schulthess 2019) 9ff, with further references.

34 Josef Drexl and others, 'Data Ownership and Access to Data' (2016) Max Planck Institute for Innovation and Competition Research Paper No. 16-10, 2ff; Wolfgang Kerber, 'Governance of Data: Exclusive Property vs. Access' (2016) 47(7) International Review of Intellectual Property and Competition Law 759, 760; Thouvenin and Weber and Früh (n 33) 36ff.

ation of some kind of "data ownership"[35]. Today, it is widely accepted, however, that there is no market failure with regard to the production of personal data,[36] and that legislators should not grant any property rights in data, neither to businesses for the data they have collected nor to individuals with respect to data about them[37].

24 If one lets go of the idea that personal data somehow "belongs" to the data subject, there is no convincing reason why an individual should be able to control the use of data about it by a private actor as long as the processing of such data does not cause the individual any harm. While the latter rationale for legal intervention can hardly be doubted and cases of harm such as discrimination or manipulation based on the processing of personal data actually occur, it is also obvious that the need to avoid and remedy harm is unable to support the idea of informational self-determination and to justify the granting of a right that allows individuals to control the processing of personal data about them by other private actors.

E. Conclusion

25 The above analysis demonstrated that the idea and concept of informational self-determination cannot serve as a convincing rationale for the all-encompassing regulation of the processing of personal data by private actors. With regard to private actors, informational self-determination is not properly implemented in the GDPR and there are no convincing reasons why this should be the case. As a consequence, the idea and concept of informational self-determination should be abandoned.

26 This raises the question as to potential alternatives both regarding the rationale of data protection law and the implementation of such rationale in an alternative regulatory framework. While such alternatives cannot be developed in this paper, it seems possible to identify the most important goals of an alternative approach. First, the law should protect the informational privacy of all individuals and second, it should ensure that no one is harmed by the processing of personal data about them. In addition, some sector-specific rules may be necessary to contain the market power of the big tech companies, namely platform providers. As opposed to hopes and promises voiced when enacting the GDPR, data protection law is not a suitable instrument to achieve this goal.

27 The importance of informational privacy and the need to protect it against unwanted interference is hardly contested. While there is some overlap between the idea of informational self-determination and the idea of informational privacy, the latter concept becomes much clearer if the former is abandoned. The protection of informational privacy would ground on every individual's right to decide what information about them is made available to others, but it would not allow for individuals to control the further use of such information once it has been made available to others. This rationale would allow to abandon some of the most important and most questionable approaches of the GDPR and other data protection laws, namely the need to provide a legal basis for every processing of personal data and the obligation to process such data according to some very general principles such as purpose limitation, data minimisation and storage limitation. Other concepts of data protection law would still be key, namely the principle of transparency, which allows individuals to know what personal data is being collected, and the principle of security, which requires controllers and processors to ensure a sufficient level of data security.

28 As with the need to protect informational privacy, the need to ensure that no one is harmed by the processing of personal data about them is widely recognised. The GDPR tries to achieve this goal through its comprehensive regulation which seeks to mitigate the risks that may be caused by the processing of personal data (risk-based approach[38]).

35 Thouvenin and Weber and Früh (n 33) 36ff, with further references; Michael Dorner, 'Big Data und «Dateneigentum», Grundfragen des modernen Daten- und Informationshandels [2014] Computer and Recht 617, 625 with further references.

36 Drexl and others (n 34) 2ff; Josef Drexl, 'Designing Competitive Markets for Industrial Data – Between Propertisation and Access' (2016) Max Planck Institute for Innovation and Competition Reserach Paper No. 2016/13 30ff; Florian Faust, 'Ausschliesslichkeitsrecht an Daten?' in Stiftung Datenschutz (ed), *Dateneigentum und Datenhandel* (Erich Schmidt Verlag 2019) 85, 99; Kerber (n 33) 992ff; Thouvenin and Weber and Früh (n 33) 56ff.

37 Thouvenin and Weber and Früh (n 33) 89ff. For an overview of the scholarly papers and the opinion of the Swiss legislator see Thouvenin and Weber and Früh (n 33) 21ff.

38 Horst Heberlein, 'Art. 5 DSGVO' in Ehmann and Selmayr (n 21) para 30; Markus Schröder, 'Der risikobasierte Ansatz in der DSGVO' [2019] Zeitschrift für Datenschutz 503; the risk-based approach is also reflected in Art. 35 GDPR on the data protection impact assessment, see Moritz Karg, 'Art. 35 DSGVO' in Simitis and others (n 20) para 2; Mario Martini, 'Art. 35 DSGVO' in Paal and Pauly (n 21) para 2; Ulrich Baumgartner 'Art. 35 DSGVO' in Ehmann and Selmayr (n 21) para 12.

However, by focussing on mitigating largely unknown and unspecific risks, data protection law often fails to protect individuals against the realisation of these risks, i.e. from the actual harms that may be caused by the processing of personal data such as discrimination and manipulation. By providing specific legal remedies, an alternative approach could not only grant individuals appropriate means to remedy such harms but also provide powerful incentives for businesses to avoid the occurance of such harms in the first place.

Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands

by **Naomi Appelman, Ronan Ó Fathaigh and Joris van Hoboken***

Abstract: This article discusses the use of automated decisioning-making (ADM) systems by public administrative bodies, particularly systems designed to combat social-welfare fraud, from a European fundamental rights law perspective. The article begins by outlining the emerging fundamental rights issues in relation to ADM systems used by public administrative bodies. Building upon this, the article critically analyses a recent landmark judgment from the Netherlands and uses this as a case study for discussion of the application of fundamental rights law to ADM systems by public authorities more generally. In the so-called SyRI judgment, the District Court of The Hague held that a controversial automated welfare-fraud detection system (SyRI), which allows the linking and analysing of data from an array of government agencies to generate fraud-risk reports on people, violated the right to private life, guaranteed under Article 8 of the European Convention on Human Rights (ECHR). The Court held that SyRI was insufficiently transparent, and contained insufficient safeguards, to protect the right to privacy, in violation of Article 8 ECHR. This was one of the first times an ADM system being used by welfare authorities has been halted on the basis of Article 8 ECHR. The article critically analyses the SyRI judgment from a fundamental rights perspective, including by examining how the Court brought principles contained in the General Data Protection Regulation within the rubric of Article 8 ECHR as well as the importance the Court attaches to the principle of transparency under Article 8 ECHR. Finally, the article discusses how the Dutch government responded to the judgment. and discusses proposed new legislation, which is arguably more invasive, with the article concluding with some lessons that can be drawn for the broader policy and legal debate on ADM systems used by public authorities.

Keywords: Automated decision-making; Fundamental rights; Social welfare ; Risk profiling; Digital administrative state

Recommended citation: Naomi Appelman, Ronan Ó Fathaigh, and Joris van Hoboken, Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands, 12 (2021) JIPITEC 257 para 1.

A. Introduction

1 In October 2019, the UN Special Rapporteur on extreme poverty and human rights warned about the dangers of the digital transformation of the State, where digital technologies are being used to "automate, predict, identify, surveil, detect, target and punish" individuals.[1] Indeed, the UN Special Rapporteur on the right to privacy has recommended that because more and more decisions affecting the daily lives of individuals are being automated, "their impact on human rights needs to be carefully and continuously evaluated".[2] For the public and pri-

* Naomi Appelman, PhD Researcher, Institute for Information Law (IViR), Faculty of Law, University of Amsterdam; Dr. Ronan Ó Fathaigh, Senior Researcher, Institute for Information Law (IViR), Faculty of Law, University of Amsterdam; and Prof. Dr. Joris van Hoboken, Professor of Law, Chair "Fundamental Rights and Digital Transformation," Vrije Universiteit Brussel (VUB), and Associate Professor, Institute for Information Law (IViR), Faculty of Law, University of Amsterdam (the Chair at VUB is established at the Interdisciplinary Research Group on Law Science Technology & Society, with the support of Microsoft).

1 Report of the Special Rapporteur on extreme poverty and human rights, UN Doc A/74/493 (11 October 2019), para 3.

2 Report of the Special Rapporteur on the right to privacy, UN Doc A/73/438 (17 October 2018), para 41.

vate sector, the digital transformation involves the processing of "vast quantities" of data from numerous sources, and using "predictive analytics to foresee risk, automate decision-making and remove discretion from human decision makers".[3] This digital transformation has only accelerated during the Covid-19 pandemic. Indeed, in the summer of 2020, five UN Special Rapporteurs, including the UN Special Rapporteur on the right to privacy, expressed their deep concern over "patterns of abuse" that had emerged through States leveraging digital technologies during the pandemic, and called for greater scrutiny of the gap between State commitments to fundamental rights and "actual practices".[4]

2 A recent landmark judgment from the Netherlands creates an opportunity to scrutinise in detail the use of ADM systems by administrative authorities, and its impact on fundamental rights. In the SyRI case,[5] the District Court of The Hague considered a controversial automated welfare-fraud detection system called *Systeem Risico Indicatie* (SyRI), which allows the linking and analysing of data from an array of government agencies to generate fraud-risk reports on people. These risk reports result in individuals being subject to investigation by authorities for possible fraud.[6] The system was criticised for its lack of transparency, the fact it was "used exclusively in areas with a high proportion of low-income residents, migrants and ethnic minorities", had "hugely negative impact on the rights of poor individuals without according them due process", and as such, was labelled as an implementation of a "surveillance state for the poor".[7] In its judgment, The Hague Court held that the legislation underpinning SyRI violated the right to private life, guaranteed under Article 8 of the European Convention on Human Rights (ECHR).[8]

3 The purpose of this article is to analyse the use of machine-learning algorithms and ADM systems by public administrative bodies, particularly systems to combat social-welfare fraud. We analyse such use from a fundamental rights perspective, using the landmark SyRI judgment in the Netherlands as a case study. First, (Section B) the article outlines the emerging fundamental rights issues in relation to the use of ADM systems by the administrative state and discusses the legal and standard-setting instruments at European level in relation to ADM systems and fundamental rights, under both the Council of Europe (COE) and European Union (EU) legal frameworks. Next, (Section C) the article discusses the SyRI judgment and focuses in particular on (I.) how The Hague Court brought principles contained in the EU's General Data Protection Regulation within the rubric of Article 8 ECHR; (II.) the importance the Court attaches to the principle of transparency; and (III.) the finding that the legislation lacked sufficient safeguards, in violation of Article 8(2) ECHR. Finally (IV.), the article critically analyses how the Dutch government responded to the judgment, with further legislation which is arguably more draconian than the SyRI legislation. We conclude with some lessons that can be drawn for the broader policy and legal debates on the digital transformation in Europe.

B. The Digital Transformation and Fundamental Rights

4 This article is focused on the digital transformation of the administrative state, involving the use of machine-learning algorithms and ADM systems by public administrative bodies, for decisions by a range of authorities, such as in the area of welfare, health, education and taxation.[9] As UN Special Rapporteur on extreme poverty and human rights Philip Alston describes, the digital transformation involves "processing of vast quantities of digital data" from many sources, and use "predictive analytics to foresee risk, automate decision-making".[10] In addition to this technological dimension, Alston notes how it tends to "remove discretion from human decision makers".[11] Notably, Coglianese and Lehr highlighted in 2017 that the use of machine-learning algorithms and ADM systems by public administrative bodies

3 UN Doc A/74/493 (n 1) para 3.

4 UN Office of the High Commissioner, 'UN experts warn of closing digital space amid COVID-19 pandemic' (30 July 2020) <www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26139&LangID=E>.

5 Rb Den Haag 5 February 2020, ECLI:NL:RBDHA:2020:1878 (hereinafter: *SyRI*).

6 ibid para 3.2

7 Special Rapporteur on extreme poverty and human rights, 'The Netherlands is building a surveillance state for the poor, says UN rights expert' (16 October 2019) <www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=25152&LangID=E>.

8 Convention for the Protection of Human Rights and Fundamental Freedoms, 4 November 1950, ETS No 5.

9 For an overview of ADM systems being used in public administration in Europe, see Fabio Chiusi, Sarah Fischer, Nicolas Kayser-Bril, and Matthias Spielkamp (eds), *Automating Society Report 2020* (AlgorithmWatch and Bertelsmann Stiftung 2020) <https://automatingsociety.algorithmwatch.org/>.

10 UN Doc A/74/493 (n 1) para 3.

11 ibid.

has "escaped sustained analysis".[12] Similarly, Alston has stated that the use of ADM systems in public administrative contexts, for example in relation to the welfare state, has "garnered remarkably little attention".[13] However, scholars have been recently examining the use of ADM systems by public administrative bodies from a number of important perspectives, such as the influential work by Eubanks, who has examined the impact of ADM systems by public authorities on those living in poverty.[14] In Europe, Choroszewicz & Mäihäniemi have approached the use of ADM systems by public authorities from a sociolegal perspective, and examined specific national legislation in EU member states on ADM in the public sector.[15] In this regard, Ranchordás has argued how digitisation of the administrate state can lead to digital exclusion in Europe.[16]

Further, Ranchordás and Schuurmans have highlighted the influential role of private actors in automated welfare-fraud systems.[17]

5 We build upon this work and approach the question of the ADM systems operated by public administration specifically from a European fundamental rights perspective, in order to understand the fundamental rights frameworks that exist at European level for ensuring that ADM systems operated by national governments do not violate fundamental rights. This is because ADM systems can impact upon an array of rights and freedoms guaranteed under European fundamental rights law, including the right to a fair trial and due process, the rights to private life, freedom of expression, freedom of assembly, the right to an effective remedy, and the prohibition of discrimination. Indeed, we focus on the SyRI judgment as a case study in order to demonstrate the distinct issues and difficulties that national courts may encounter in applying European fundamental rights law to ADM systems operated by administrative bodies.

6 We also build on the law and technology scholarship that has focused on the discriminatory impact of algorithms, the surveillance state, the use of algorithms by large platforms and the emerging regime of surveillance capitalism.[18] Finally, we take into account recent research by civil society organisations, such as the Berlin-based AlgorithmWatch, has started to shine a light on the widespread use of ADM systems by governments in Europe.[19] In its 2020 report on ADM systems in Europe, AlgorithmWatch warned that the "vast majority of uses tend to put people at risk rather than help them", including risks of discrimination and disproportionate interferences with privacy.[20]

C. The applicable European fundamental rights framework

7 In order to begin our analysis, the first question that must be posed is what legal frameworks exist at European level for ensuring that ADM systems operated by national governments do not violate fundamental rights? In this regard, national gov-

12 Cary Coglianese and David Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 *Georgetown Law Journal* 1147, 1152. See also Lorna McGregor, Daragh Murray and Vivian Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68(2) *International & Comparative Law Quarterly* 309.

13 UN Doc A/74/493 (n 1) para 3. For scholarship on the impact of ADM systems on individuals in poverty, see, for example, Virginia Eubanks, *Automating Inequality: How high-tech tools profile, police, and punish the poor* (St Martin's Press 2018); and Virginia Eubanks, 'Algorithms Designed to Fight Poverty Can Actually Make It Worse' (2018) 319 *Scientific American* 68.

14 Virginia Eubanks, *Automating Inequality: How high-tech tools profile, police, and punish the poor* (St Martin's Press 2018); and Virginia Eubanks, 'Algorithms Designed to Fight Poverty Can Actually Make It Worse' (2018) 319 *Scientific American* 68.

15 Marta Choroszewicz and Beata Mäihäniemi, 'Developing a Digital Welfare State: Data Protection and the Use of Automated Decision-Making in the Public Sector across Six EU' (2020) 1(1) *Global Perspectives* 12910.

16 Sofia Ranchordás, 'The Digitalization of Government and Digital Exclusion: Setting the Scene' forthcoming in G Ferreira Mendes & C Blanco de Morais (eds.) *Direito Publico e Internet: Democracia, Redes Sociais e Regulação do Ciberespaço* (FGV /IDP/ Univ. Lisboa, 2020) <http://dx.doi.org/10.2139/ssrn.3663051>. See also Sofia Ranchordás, 'Automation of Public Services and Digital Exclusion' (*I-CONnect: Blog of the International Journal of Constitutional Law*, 11 March 2020) <www.iconnectblog.com/2020/03/automation-of-public-services-and-digital-exclusion/>; Sofia Ranchordás, 'Public Law and Technology: Automating Welfare, Outsourcing the State' (*I-CONnect: Blog of the International Journal of Constitutional Law*, 15 January 2020).

17 Sofia Ranchordás and Ymre Schuurmans, 'Outsourcing the Welfare State: The Role of Private Actors in Welfare Fraud Investigations' (2020) 7(1) *European Journal of Comparative Law and Governance* 5.

18 UN Doc A/74/493 (n 1) para 3.

19 See Chiusi, Fischer, Kayser-Bril and Spielkamp (n 9).

20 ibid 7.

ernments have binding legal obligations pursuant to both membership of the COE and the EU. Beginning with the COE, its Committee of Ministers has been quite explicit in emphasising the basic principle that its member states have a legal obligation under the ECHR to ensure that the use of algorithmic systems by public authorities does not violate the ECHR rights of individuals within their jurisdiction, such as the right to private life under Article 8 and right to a fair trial and due process under Article 6.[21] As Wagner et al. have examined, ADM systems can impact upon an array of rights and freedoms guaranteed under the ECHR, including the right to a fair trial and due process, the right to private life, freedom of expression, freedom of assembly, the right to an effective remedy, and the prohibition of discrimination.[22] Thus, any national legislation relating to the use of ADM systems, national court judgments interpreting such legislation, and decisions of administrative authorities, must be consistent with the rights guaranteed under the ECHR.

8 The European Court of Human Rights (ECtHR) is tasked with interpreting the ECHR, and while the ECtHR has not yet considered an ADM system operated by a public authority, it has delivered numerous judgments on the use of automated systems and data collection systems used for government surveillance. For example, the ECtHR has held that an electronic-surveillance system in operation in Hungary violated the right to respect for private life under Article 8 ECHR. Crucially, the ECtHR emphasised that surveillance systems using "automated and systemic data collection" had "reached a level of sophistication which is hardly conceivable for the average citizen".[23] Indeed, the Court warned about the capacity of governments to acquire "detailed profile[s] of the most intimate aspects of citizens' lives", which may result in "particularly invasive" interferences with the right to private life.[24] Similarly, the ECtHR has found a violation of Article 8 over a system in the United Kingdom allowing storing of a person's photograph in a police database, where the police could apply facial recognition and facial mapping techniques to the image.[25] The Court emphasised the essential importance of Article 8 to guard against the "risk of arbitrariness" which flows from vesting "obscure" powers with the State, and "especially where the technology available is continually becoming more sophisticated".[26]

9 Article 8 (1) ECHR guarantees the right to respect for private life, and Article 8 (2) ECHR allows interferences with the right to private life only under certain conditions. For an interference with private life to be consistent with Article 8 ECHR, it must be "in accordance with law", "pursue a legitimate aim", and "necessary in a democratic society".[27] Crucially, for an interference to be in accordance with law, it is simply not enough, for example, for a system of surveillance to be set out in legislation. This test also encompasses whether there are sufficient safeguards to protect against "arbitrary interference by public authorities."[28] Indeed, the Court has found national legislation in specific cases to be deficient in this regard, such as legislation on surveillance failing to have appropriate safeguards to protect specific groups of individuals, such as journalists, from government surveillance.[29]

10 Notably, the COE's Committee of Ministers adopted an important Recommendation in 2020 on the human rights impacts of algorithmic systems, given the current "digital transformation" European societies are undergoing.[30] This is important, as the ECtHR can rely upon recommendations from the Committee of Ministers to provide "guidance as to the approach

21 Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (8 April 2020), preamble.

22 Ben Wagner et al, *Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications* (Council of Europe 2017) 10.

23 *Szabó and Vissy v Hungary* App no 37138/14 (ECtHR, 12 January 2016) para 68.

24 ibid para 70.

25 *Gaughran v UK* App no 45245/15 (ECtHR, 13 February 2020) para 70.

26 ibid para 86.

27 See, for example, *Telegraaf Media Nederland Landelijke Media BV and Others v the Netherlands* App no 39315/06 (ECtHR, 22 November 2012) para 89.

28 ibid para 90.

29 ibid para 102. Van der Sloot has even argued that the ECtHR has transformed into a "European Constitutional Court" with its recent case law on government surveillance, by "formally assesses the quality of Member States' laws and even advises Member States' legislative branch on how to amend its legal system in order to be Convention-compliance" (see Bart van der Sloot, 'The Quality of Law: How the European Court of Human Rights gradually became a European Constitutional Court for privacy cases' (2020) JIPITEC 160, 177. See also, Eleni Kosta, 'Algorithmic state surveillance: Challenging the notion of agency in human rights' (2020) *Regulation & Governance* <https://doi.org/10.1111/rego.12331>).

30 Recommendation CM/Rec(2020)1 (n 15).

which should be taken to interpreting" ECHR rights, and has applied these recommendations in its case law.[31] Notably, the Recommendation singles out the use of algorithmic systems by States for their public services, warning that such algorithmic systems can prompt a "particular, higher risk to human rights", because an individual may "not have a possibility to opt out," where its use is prescribed by law, or when she/he "suffers negative consequences as a result of the decision to opt out".[32]

11 The Recommendation defines "high risk" as including the use of algorithmic systems in situations where the lack of alternatives "prompts a particularly high probability of infringement of human rights, including by introducing or amplifying distributive injustice".[33] This is the case where the ADM system produces "serious consequences for individuals", such as legal consequences, or for predictive or individual risk assessment by public authorities.[34] Thus, the Committee of Ministers is acutely aware of the possibility of violations of ECHR rights through ADM systems used in public services, and how such systems can perpetuate existing inequalities. This view echoes the observation from the UN Special Rapporteur on extreme poverty that the use of algorithmic systems for risk calculation and need classification by welfare authorities can "reinforce or exacerbate existing inequalities and discrimination".[35] This is because such ADM systems may be used to target poor and marginalised individuals already subject to discrimination and most likely to be in need of state aid. Indeed, as discussed below, the SyRI system deployed in the Netherlands exclusively targeted so-called "problem" neighbourhoods, with the Court recognising that the system could "inadvertently" be based on bias, such as a lower socio-economic status or an immigration background.[36]

12 The final COE instrument to be mentioned is the COE's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.[37] In 2018, new Protocol was adopted amending the Convention, which inserts a new Article 9(1)(a), and guarantees a right for every individual not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration.[38] However, there is an exception under Article 9(2), that the right shall not apply if the decision is authorised by a law to which the controller is subject, and which also lays down suitable measures to safeguard the data subject's rights, freedoms and legitimate interests. Thus, any national legislation permitting the use of ADM systems by public administrative authorities which does not allow an individual to exercise their right under Article 9(1)(a) would need to include measure to safeguard an individual's data rights.

13 In addition to the COE framework, the EU legal framework is particularly important.[39] The EU Charter of Fundamental Rights guarantees many of the rights contained in the ECHR, including the right to a fair trial, respect for private life, freedom of expression and freedom of assembly; in addition to rights not specifically enumerated in the ECHR, such as the right to the protection of personal data.[40] Further, the most significant secondary EU legislation on ADM systems is the GDPR,[41] which applies to the

31 See, for example, *Manole and Others v Moldova* App no 13936/02 (ECtHR, 17 September 2009) para 101 and 102.

32 Recommendation CM/Rec(2020)1 (n 15) s A(11) (Appendix). It should also be recongised that it can be similarly difficult to opt out of ADM-type systems operated by the private sector, and even where there are mechanisms to opt out, these mechanisms may not operate fully as stipulated (see, e.g., Paresh Dave, 'Google faces lawsuit over tracking in apps even when users opted out' *Reuters* (14 July 2020) <www.reuters.com/article/us-alphabet-google-privacy-lawsuit-idUSKCN24F2N4>.

33 ibid.

34 ibid.

35 UN Doc A/74/493 (n 1) para 28.

36 *SyRi* (n 5) para 6.93.

37 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 28 January 1981, ETS No 108. See Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, *Guidelines on Artificial Intelligence and Data Protection*, T-PD(2019)01 (25 January 2019).

38 Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 10 October 2018, CETS No 223, art 9(1).

39 See also High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (European Commission 2019).

40 Charter of Fundamental Rights of the European Union [2012] OJ C326/391. See also Consolidated Version of the Treaty on the Functioning of the European Union [2012] OJ C326/1.

41 Regulation (EU) 2016/679 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) [2016] OJ L119/1.

processing of personal data wholly or partly by automated means.[42]

14 Crucially, Article 22(1) GDPR provides (subject to exceptions in Article 22(2) GDPR) that data subjects have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning them or similarly significantly affects them.[43] Profiling is defined as any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.[44] However, Article 22(2) GDPR contains important exceptions to the prohibition on ADM and profiling, including when it is authorised by national law which "lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests".[45] These measures include the right to obtain human intervention, and to express one's point of view and to contest the decision.[46] In relation to ADM systems used by public authorities, it is important to note that Recital 71 of the GDPR expressly recognises that ADM and profiling "should be allowed", where it is authorised by national law, including for "fraud and tax-evasion monitoring and prevention purposes".[47] Thus, the drafters of the GDPR clearly envisaged that ADM, and specifically, profiling, by public authorities should not be in principle interfered with, especially where it is deployed in the fight against fraud.

15 In 2021, the European Parliament adopted a Resolution on artificial intelligence, including AI systems in the decision-making process of public authorities.[48] The Resolution warns of many risks associated with ADM systems specially used by public authorities. Significantly, the Parliament called on the European Commission, and the European Data Protection Board, to issue guidelines and recommendations on the criteria and conditions applicable to decisions based on profiling and the use of AI by public authorities.[49] First, the Resolution stressed that AI systems in the decision-making process of public authorities can result in "biased decisions that negatively affect citizens".[50] As such, the Parliament recommended that such ADM systems should be subject to "strict" control criteria in terms of security, transparency, accountability, non-discrimination, and social responsibility.[51] Indeed, EU member states were urged to assess the risks related to AI-driven decisions by public authorities "before" automating activities connected with the exercise of state authority.[52] Further, the Resolution recommends that there should be safeguards, including meaningful human supervision, transparency and the possibility to contest a decision.[53] Finally, the Parliament called for the explainability of algorithms, transparency and regulatory oversight when AI is used by public authorities, and for impact assessments to be conducted before tools using AI technologies are deployed by state authorities.[54]

16 In terms of the risk of ADM systems used by public authorities, the UN Special Rapporteur on poverty points out that seemingly neutral terms such as the "digital transformation" should not conceal the "politically driven character" of ADM systems.[55] These systems are promoted as improving "efficiency" and "rooting out fraud".[56] However, the Rapporteur argues that digital technologies are presented as neutral and scientific, but may in fact facilitate, justify and shield "values and assumptions that are far removed from, and may be antithetical to, the prin-

42 ibid art 2(1).

43 ibid art 22(1).

44 ibid art 4(4).

45 ibid art 22(2)(b).

46 ibid art 22(3). There has been considerable debate over these provisions: see, for example, Andrew Selbst and Julia Powles, 'Meaningful information and the right to explanation' (2017) 7 *International Data Privacy Law* 233; and Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" is Probably Not the Remedy You Are Looking For' (2017) 16 *Duke Law & Technology Review* 18.

47 ibid recital 71.

48 European Parliament resolution of 20 January 2021 on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice, P9_TA-PROV(2021)0009.

49 ibid para 62.

50 ibid para 52.

51 ibid.

52 ibid para 71.

53 ibid para 52.

54 ibid para 62.

55 UN Doc A/74/493 (n 1) para 6.

56 ibid.

ciples of human rights".[57] As such, diverging views around the risk and benefits of ADM systems used by public authorities may go some way to explaining how national courts struggle to apply COE and EU legal frameworks when considering the compatibility of these systems with European fundamental rights law. As the following section demonstrates, The Hague Court did indeed struggle on how to apply such frameworks.

D. The SyRI Judgment

17 Before delving into different aspects of the SyRI Judgment relevant for understanding the impact of fundamental rights law on the use of ADM systems by public authorities, this section will outline the facts of the case, focussing on how the SyRI system operates. The case was initiated by a coalition of civil society organisations who brought legal proceeding against the Dutch Government in The Hague District Court in March 2018, over the operation of the SyRI system, claiming a violation of Article 8 ECHR. Crucially, the Court ruled in favour of the coalition, and declared the SyRI legislation is in violation of Article 8 ECHR due to, which will be discussed in depth in the following sections, a lack of transparency and appropriate safeguards in the connection with the linking of personal data across government agencies.[58]

18 First, as to what SyRI actually is, the Court defined SyRI as a "legal instrument",[59] which the Dutch government created with the purpose of preventing and combating illegal use of government funds and government schemes in the area of social security and income-dependent schemes, and in order to prevent and combat "taxes and social security fraud and non-compliance with labour laws."[60] The Court went on to explain how the actual SyRI-projects work to achieve these aims. Concretely, when, based on the SyRI legislation, a SyRI-project is started, data from different government agencies is linked and analysed in order to produce a risk report of people. When a risk report is filed on an individual, this means they are "deemed worthy of investigating with regard to possible fraud".[61] The aim is that this automated analysis would help in tracking down social welfare fraud.

19 Importantly, in its history SyRI has only been used to analyse people in specific neighbourhoods, referred to as "problem" neighbourhoods (i.e. with lower socio-economic inhabitants), which was confirmed by the government in its submissions to the Court.[62] As to the government agencies involved, these range from municipal governments, the Netherlands Tax and Customs Administration, the Social Insurance Bank, the Immigration and Naturalisation Service, the Employee Insurance Agency, as well as supervisory authorities such as the Social Affairs and Employment Inspectorate.[63] The data shared by these government agencies covers an enormous range, totalling on 17 general types of data, including data on health, finance, education, fiscal payments, employment and "integration".[64]

20 The different steps involved in a SyRI project are as follows. Importantly, a SyRI project starts when a number of the government agencies involved organise in a "collaborative alliance" and create a proposal to use SyRI in a specific neighbourhood.[65] This proposal is submitted to the Minister who, after hearing the advice from the steering group consisting of all the government agencies involved,[66] then officially decides to apply SyRI.[67] The relevant data of the different agencies is then collected, pseudonymised and analysed according to the risk indicators and model as outlined in the proposal.[68] The cases of people flagged by the risk model are then analysed by the Ministry before a definite risk report is submitted, and the relevant government agency conduct further research into possible fraud.[69] The people whose data is involved in the project are only informed when an official investigation follows upon a risk report.[70] Importantly, the risk model and indi-

57 ibid.

58 *SyRI* (n 5) para 5.1.

59 ibid para 3.1.

60 ibid para 4.4.

61 ibid para 3.2.

62 ibid para 3.9-10, 4.24, 6.93. Notably, the Court did *not* find that the use of SyRI in "problem" neighbourhoods in and of itself was disproportionate or in violation of Article 8. However, it did find that *there is a risk* that SyRI "inadvertently creates links based on bias, such as a lower socio-economic status or an immigration background". (see para 6.93).

63 ibid para 3.3; art 64 lid 1 Wet SUWI.

64 Art 5a.1(2) Besluit SUWI.

65 *SyRI* (n 5) para 3.3, 4.20-22; art 64 lid 2 Wet SUWI.

66 ibid para 3.6.

67 ibid para 3.3.

68 ibid para 4.22, 4.28 - 4.29.

69 ibid para 4.29 - 30.

70 ibid para 6.54. Notably, Dutch media has reported that SyRI has not led to the discovery of a single case of fraud.

cators, threshold values, types of data and people involved are unknown to both the Court, the citizens involved and wider society.[71]

21 Having thus set out the operation of SyRI, the Court then turned to compatibility of the system with the right to private life under Article 8 ECHR, to which we now turn. Concretely, the following sections will focus on four aspects related to the SyRI judgment. First, the way in which the Court involved the general principles of the GDPR in its application of Article 8 ECHR will be considered. Then, the different ways in which the Court ran into issues related to a lack of transparency on how the SyRI systems operate concretely is analysed, followed by a discussion of the possible safeguards for the protection of the right to private life that could be employed. Finally, proposed legislation in the Netherlands that is following in the footsteps of the now void SyRI legislation is discussed.

I. The relationship between the EU Charter, GDPR and Article 8 ECHR

22 One of the most striking aspects of the SyRI judgment is the way in which The Hague Court related the GDPR to Article 8 ECHR. The Court used the general principles of data protection from Article 5 GDPR to substantiate the requirements of Article 8 ECHR, more specifically, the criterion that any interference should be "necessary in a democratic society".[72] Using (secondary) EU legislation to interpret ECHR provisions is not uncontroversial as, despite many connections, the EU and the COE remain distinct legal orders.[73] It would seem more appropriate to interpret Article 8 ECHR based on the case law of the ECtHR, and the principles established in the case law, rather than relying on a piece of EU secondary legislation.

This section will trace how the Court came to this line of reasoning and what the possible consequences can be.

23 As the claimants based their main claim on a violation of Article 8 ECHR, the Court, subsequently centred the judgment around the question whether the SyRI legislation constituted a violation of the fundamental right to a private life as protected by Article 8 ECHR.[74] Basing claims directly on international human rights obligations and, especially, the ECHR, instead of the Dutch Constitution, is common legal practice in the Netherlands. This is due to constitutional provisions that prohibit Dutch Courts from constitutional review of Dutch legal provisions, but does allow for direct application of international human rights treaties.[75] In the judgment, the Court extensively discussed the applicable legal framework, differentiating between, on the one hand, the COE with Article 8 ECHR, and on the other, the EU with Article 7 and 8 of the EU Charter, and the GDPR as relevant secondary legislation.[76] The Court recognised the nature of the ECHR as providing "for a minimum level of protection of the fundamental right to respect for private life",[77] and that within the EU Charter, there is "at least the same minimum level of protection as the ECHR", although the Charter and the GDPR do provide protection that is "specified in more detail and in some instances extends beyond the protection under the ECHR".[78] The Court, more specifically, considered the general principles of data protection in GDPR to be an extension of the fundamental rights protection of the Charter.[79]

24 As stated, the Court took the striking step to take into account the general principles of data protection from the EU Charter and the GDPR in its review of whether the SyRI was compatibility with Article 8 ECHR. Thus, applying Article 8 ECHR entailed that the SyRI legislation "must meet the aforementioned general principles of data protection, as laid down in Union law in the Charter and the GDPR, such as the principle of transparency, the principle of purpose limitation and the principle of data minimisation."[80] The Court used this conclusion to employ the general principles of data protection from the GDPR to substantiate the "necessary in a democratic society" criterion as part of the Article 8 ECHR test.[81] More specifically, the principles of transparency, data minimisation and purpose limitation were used

See Charlotte Huisman, 'Fraudesysteem Overheid Faalt' *de Volkskrant* (Amsterdam, 27 June 2019) 6-7.

71 ibid para 6.100.

72 *SyRI* (n 5) para 6.7.

73 See also *Nederlandse Jurisprudentie* 2020/386, Note by E.J. Dommering (Case Comment).

74 *SyRI* (n 5) para 5.1, 6.38.

75 Art 93, 94 and 120 Grondwet.

76 *SyRI* (n 5) para 6.19 – 6.41.

77 ibid para 6.37

78 ibid, referencing EU Charter (n 34) art 52(3).

79 *SyRI* (n 5) para 6.27- 36.

80 ibid para 6.40.

81 ibid para 6.80.

to assess whether the requirements of necessity, proportionality and subsidiarity were met as part of this criterion.[82]

25 The Court seemed to assume a reciprocity between the ECHR and the EU Charter, including EU secondary legislation such as the GDPR, based on the notion that the Charter explicitly provides that the meaning and scope of its rights also guaranteed in the ECHR must be, at a minimum, the same as those in the ECHR.[83] However, this does not mean that the level of protection offered by the ECHR should be supplemented by the additional protection offered within the EU framework when applying ECHR provisions. As such, the assumed reciprocal relation between the ECHR and the EU Charter, including secondary legislation, is not sufficiently substantiated in the judgment. This begs the question to what extent straining to establish this interdependent relationship between the two different legal orders was necessary when the Court could also have opted to apply the GDPR directly, in parallel to its Article 8 ECHR assessment. An explanation for this notable step by the Court might be found in a combination of the ECHR tradition in the Netherlands in combination with the greater flexibility offered by the ECHR as opposed to the GDPR. As stated, basing claims directly on ECHR provisions is common legal practice in the Netherlands due to the direct effect of these international treaties in the Dutch legal system. This would have made the step of further substantiating with principles from the GDPR shorter. Additionally, including the principles of data protection from the GDPR gave The Hague Court more solid ground in assessing the SyRI ADM-system and allowed for a detailed analysis without having to go through the technical analysis and possible prejudicial questions as when the GDPR would have been directly applied. This allowed the Court to include data protection principles while still sticking solidly to the fundamental rights perspective. However, it remains to be seen whether this step will be followed by other Courts. At this point, we can continue to another striking element: the way in which the concept of transparency functioned throughout the judgment.

II. The principle of transparency and Article 8 ECHR

26 Transparency forms an essential element of the SyRI judgment, due mainly to the fact that the system itself is inaccessible and its workings are kept secret from The Hague Court, the citizens involved and wider society.[84] This section will analyse the different problems this posed to the Court on several steps of its legal analysis and how these were dealt with. The lack of transparency consisted of the fact that the risk model and indicators, threshold values, types of data and people involved were and, to this day, are unknown and that the citizens involved are not informed of their involvement.[85] Although a rich body of ECtHR case law exists on how to apply the test of Article 8 ECHR (whether the interference of SyRI amounts to a violation of private life) to secretive government measures,[86] testing this ADM system used in the context of government welfare gave rise to apparent difficulties for The Hague Court in several steps of its Article 8 ECtHR analysis: the extent and seriousness of the interference, whether it was in accordance with law, and whether the interference was necessary in a democratic society. The lack of transparency in how the system operated (models, indicators and data used) and in communications to citizens proved fatal as it was one of the main arguments for the Court's conclusion that the automated social welfare fraud system violated Article 8 ECHR.[87] The judgment reveals both the differentiated and pivotal role transparency plays in adjudicating such a government ADM system, but it also leaves many questions unanswered on the scope of protection Article 8 ECHR affords to the government's use of ADM systems. This section will analyse at which points transparency, or the lack thereof, played an important role in the judgment in order to draw out lessons on the fundamental rights dimension of the use of ADM systems by the administrative state.

27 Immediately, at the first substantive step the lack of transparency on how the ADM system functions led to difficulties for the Court in assessing the extent and seriousness of the interference. The lack of transparency on how the SyRI ADM-systems actually operate meant that the Court, at several points, was

82 ibid para 6.80 -6.107.

83 EU Charter (n 34) art 52(3).

84 *SyRI* (n 5) para 6.65.

85 ibid para 6.100.

86 *S and Marper v UK* App nos 30562/04 and 30566/04 (ECtHR 4 December 2008). See Van der Sloot (n 29); and also, for example, *Szabó and Vissy v Hungary* (n 18), and *Big Brother Watch and Others v UK* App nos 58170/13, 62322/14 and 24960/15 (ECtHR 13 September 2018) (referred to ECtHR Grand Chamber). See Bart van der Sloot and Eleni Kosta, 'Big brother watch and others v UK: Lessons from the latest Strasbourg ruling on bulk surveillance' (2019) 5 *European Data Protection Law Review* 252.

87 *SyRI* (n 5) para 6.7, 6.83, 6.95.

unable to verify the opposing parties' positions.[88] This difficulty in assessing the extent of the interference poses an interesting contrast to established ECtHR case law on mass surveillance where the unknown factors were *when* and against *whom* the interference occurred, but the operation of the mass surveillance system's interference itself was clearly established.[89] In this judgment, The Hague Court was confronted with a complex discussion on what (speculative) elements of the SyRI systems are legally relevant as the parties differed widely on not only the nature, but also the legal definition of SyRI.[90] For example, does SyRI make use of big data, profiling, automated decision-making, machine learning, data mining, unstructured data collection and, if so, which of these elements are relevant for the legal assessment of the system?[91] The remaining question is to what extent this debate, in future cases, would be solved with more technical transparency as, in the end, the legally relevant question is what is the impact of these automated risk assessments on an individual citizen's private life and fundamental rights more generally. Putting most of the focus on the legal characterisation of the technology risks decentring the actual effect of their involvement in the projects, the eventual risk report and possible subsequent fraud investigation on citizens' private lives, and the responsibility of the government. As mentioned above, the UN Special Rapporteur on extreme poverty similarly warns that digital technologies in welfare systems are often presented as "scientific" and neutral, although "they can reflect values and assumption that are far removed from, and may be antithetical to, the principles of human rights."[92]

28 For now, the Court was able to circumvent most of these discussions by either declaring that it was unable to verify the claims due to the government's secrecy, or stating that the claim was irrelevant for the legal question at hand.[93] The Court concluded that SyRI consists of "structured data processing based on existing, available files" and a risk model which "consist of predetermined risk indicators and which gives an indication of whether there is an increased risk" of social welfare fraud.[94] Further, the Court included the government's secrecy towards the Court, and towards the people involved who are at no point informed, as part of the extent and seriousness of the interference with private life.[95]

29 Subsequently, the Court proceeded to assess whether SyRI is in accordance with law and, again, the secrecy surrounding the actual functioning of the system inhibited the Court applying the Article 8 ECHR criteria straightforwardly. Following the claimants' arguments, the Court based its assessment on mass surveillance case law from the ECtHR, specifically the case of *S and Marper v UK*.[96] Even though the Court emphasised that the context of mass surveillance is substantially different from the SyRI case, it stated that the *S and Marper* judgment contains "considerations of the ECtHR on data protection of a more general nature".[97] The case shows, according to The Hague Court, that the assessment of "whether the interference is in accordance with the law may be closely connected to the assessment whether the interference is necessary in a democratic society".[98] This led The Hague Court to the conclusion that in this particular instance it did not need to make this assessment, as further analysis would show that the legislation was not "necessary in a democratic society".[99]

30 The reasoning applied by The Hague Court meant the substantive analysis of the "in accordance with law" criterion was sidestepped, or rather skipped over, in favour of the "necessary in a democratic society" criterion. As recognised by the Court, the substantive requirements contained in the "in accordance with law" criterion are to a large extent dependent on the "content of the instrument in question, the field it is designed to cover and the number and status of those to whom it is addressed".[100] It is not clear that the criteria developed for mass surveillance can be applied to the context of social welfare fraud detection and, consequently, that passing over the

88 ibid para 6.49, 6.53-54.

89 See n 64 above.

90 *SyRI* (n 5) para 6.44.

91 ibid para 6.42-6.65.

92 UN Doc A/74/493 (n 1) para 6.

93 *SyRI* (n 5) para 6.56, 60, 63.

94 ibid para 6.62.

95 ibid para 6.60.

96 ibid para 6.67, citing *S and Marper v UK* (n 86).

97 ibid para 6.67. The plaintiffs had argued that processing personal data for the use of SyRI violated various provision of the GDPR, including Article 6 (Lawfulness of processing), 13 (Information to be provided where personal data are collected from the data subject), and Article 22 (Automated individual decision-making, including profiling). However, the Court held it would 'not assess whether the SyRI legislation is contrary to one or more specific provisions of the GDPR on which' the plaintiffs relied.

98 ibid para 6.71.

99 ibid para 6.72.

100 ibid para 6.69, citing *S and Marper v UK* (n 86) para 96.

"in accordance with law" criterion was warranted in this case. Especially as this line of reasoning offers a form of legitimacy to the government's lack of transparency, framing it as a defect that can be amended with sufficient safeguards.

31 Finally, as elaborated in the previous section, the Court took the remarkable step of letting transparency play a crucial role in the last step of the analysis: assessing whether the interference was necessary in a democratic society. As elaborated on in the previous section, this criterion was substantiated with the principles of data protection, and transparency in particular, as found in Article 5 GDPR.[101] The Court made clear that, at a minimum, insight must be given into "the risk indicators and the risk model, or at least ... further legal safeguards to compensate for this lack of insight".[102] Additionally, insight needed to be given into "which objective factual data can justifiably lead to the conclusion that there is an increased risk".[103] Absent this information, the Court concluded it was unable to "verify how the simple decision tree [in the risk model], to which the State refers, is generated and of which steps it is comprised".[104] This opacity and lack of information also greatly inhibits the ability of the people involved to exercise their rights or defend themselves, especially since they are at no point informed of their (passive) involvement.[105] As such, the judgment requires governments to provide all necessary information, such as their involvement in an ADM-system to detect social welfare fraud and what their risks scores are, to people in order to enable them to exercise their rights and contest unwanted data processing, which is a core aim of the more specific GDPR transparency provisions.[106]

32 Further, the Court connected both the potential discriminatory biases (e.g. lower socio-economic status or an immigration background) in the system itself, and the discriminatory and stigmatising effect of the system's implementation, as pointed out by the UN Special Rapporteur, to the apparent lack of transparency.[107] Especially considering SyRI's sole implementation in "problem districts" of the Netherlands, and the large amount of (sensitive) data used, the Court explicitly recognised the "risk that SyRI inadvertently creates links based on bias, such as a lower socio-economic status or an immigration background".[108] The Court concluded that "due to the absence of verifiable insights into the risk indicators and the risk model as well as the function of the risk model" it was unable to ascertain whether the risk of stigmatisation and discrimination was "sufficiently neutralised"[109] (i.e., the risk would be neutralized if the specific risk indicators used by SyRI were made public, so it could be analysed properly whether the system is not discriminatory against individuals based on race, national or social origin, or association with a national minority). Via this procedural line of reasoning, the Court, remarkably, brought issues pertaining to racism, discrimination and classism into the fold of Article 8 ECHR, instead of Article 14 ECHR (which prohibits discrimination). Explicitly taking into account the potential harmful social and political effects of these types of ADM systems in the context of social welfare is of crucial importance. However, putting them into a procedural context of "transparency" and "sufficient safeguards" runs the risk of falling into the frame that discrimination or social stigmatisation can be technically solved.[110] This focus on the technological aspect ignores and neutralises the deeply political and social aspects that are especially relevant in the context of social welfare systems that serve the most vulnerable in society (especially the discriminatory element of these systems).

33 Transparency plays a pivotal role in the SyRI judgment, taking on many guises and appearing at every step of the Court's analysis. What is beyond dispute is the conclusion that the extent to which the Dutch government withheld information and insight into its ADM system does not pass the test under Article 8 ECHR of whether the interference constituted a violation. A minimum of insights into the workings of the system (i.e., the risk indicators and model) is necessary for courts to perform their supervisory role over the executive branch of government, and for the people involved to defend themselves against government overreach. However important this transparency is, it does to a large extent only form a precondition for the truly substantive assessment of whether the impact on people's private life of suspecting them of social welfare fraud, their (passive) involvement in the ADM-system, their risk-report and possible investigation is justified. A possible risk is that the discussion on what transparency should concretely mean, or on how to legally characterise the ADM

101 *SyRI* (n 5) para 6.30.

102 ibid para 6.95.

103 ibid para 6.87.

104 ibid para 6.90.

105 ibid para 6.90.

106 GDPR (n 35) art 13-14.

107 *SyRI* (n 5) para 6.92-94.

108 ibid para 6.93.

109 ibid para 6.94.

110 See UN Doc A/74/493 (n 1).

systems themselves, once they are more transparent, deflects attention from this substantive assessment. For example, the assessment of how increased transparency towards the people involved will actually translate into contestable systems, or what the relation between the government and its citizens should be in the context of social welfare, and what privacy and treatment people can expect.[111]

III. Lack of safeguards

34 In addition to transparency, a crucial aspect of the SyRI judgment was the Court's finding that the SyRI legislation contained "insufficient safeguards" to protect the right to private life, in violation of Article 8(2) ECHR.[112] This was because, as the Court held, the SyRI legislation paid "insufficient attention to the principle of purpose limitation and the principle of data minimisation", and thus, violated Article 8(2) ECHR.[113]

35 First, while the legislation contained an "exhaustive enumeration" of the data categories that qualify for processing,[114] the Court pointedly held it was "hard to imagine any type of personal data that is not eligible for processing in SyRI".[115] Importantly, the Court criticised the SyRI legislation for not providing for a "comprehensive review," or a review by an independent third party, prior to the data processing by the Minister, in order for an assessment of whether or not the interference with private life was "necessary, proportionate and subsidiary in light of all the files that are linked in a project considering the specific purpose of that project".[116] The Court noted that a body called the National Intervention Teams Steering Group (LSI) advises the Minister about the application of SyRI in a specific SyRI project. However, the Court stated that the LSI is "merely an advisory organ", and its advice was "non-binding and lacks an explicit legal basis".[117] Thus, the Court held that the lack of independent assessment prior to the approval by the Minister violated Article 8(2) ECHR, which requires such a safeguard. Crucially, the Court held the LSI was comprised of "representatives of organs which also have an interest in combating and preventing abuse and fraud," including the Social Affairs and Employment Inspectorate, the Tax and Customs Administration, and the police.[118] Moreover, in relation to data protection impact assessments (DPIA), the Court harshly criticised the State's approach. The Court held that the State had "failed to elucidate why, considering the extent and seriousness of the invasion of private life, occasioned by the processing of data in SyRI," a data protection impact assessment was not carried out for each individual project.[119] However, the Court stopped short of finding a violation of Article 8(2) ECHR on the basis of the lack of individual data protection assessments.

36 The Court concluded that in "view of the large amount of data that qualify for processing in SyRI," no comprehensive and no independent assessment prior to the approval by the Minister, the SyRI legislation therefore contained "insufficient safeguards", in light of the principles of purpose limitation and data minimisation under Article 8 ECHR.[120] This focus on insufficient safeguards was entirely justified, as the SyRI legislation lacked any independent oversight to assess whether it was proportionate to link such a vast amount of personal data from different government agencies for the purpose of a specific SyRI project to develop individual risk profiles of social welfare fraud. Especially important are safeguards that allow for not just a discussion focussed on the workings of the technology used (e.g., the technical properties of the SyRI system) but allows for a substantive

111 Notably, the Court nowhere referred to case law on how the ECtHR conceptualises and protects rights in relation to social welfare, other than through the frame of transparency and privacy under Article 8 ECHR. This case law includes, for example, in relation to Article 6 ECHR, *Zednik v the Czech Republic* App no 74328/01 (ECtHR, 28 June 2005); in relation to Article 1 of Protocol No. 1 ECHR, *Azinas v Cyprus* App no 59498/00 (ECtHR, 20 June 2002); and in relation to Article 14 ECHR, *Van Raalte v the Netherlands* App no 20060/92 (ECtHR, 21 February 1997). For analysis of this case law, see Ingrid Leijten, 'The right to minimum subsistence and property protection under the ECHR: Never the twain shall meet?' (2019) 21 *European Journal of Social Security* 307; Ingrid Leijten, *Core Socio-Economic Rights and the European Court of Human Rights* (CUP 2018); Antonia Baraggia and Maria Elena Gennusa, 'Social Rights Protection in Europe in Times of Crisis: "A Tale of Two Cities"' (2017) 11 *Vienna Journal on International Constitutional Law* 479; and Ana Gómez Heredero, *Social security as a human right: the protection afforded by the European Convention on Human Rights* (Council of Europe Publishing 2007).

112 *SyRI* (n 5) ibid para 6.106.

113 ibid para. 6.96.

114 ibid para 6.98.

115 ibid.

116 ibid para 6.99.

117 ibid para 6.101.

118 ibid para 6.101.

119 ibid para 6.105.

120 ibid para 6.106.

discussion on whether the use of such systems is warranted. Thus, the Court concluded that the SyRI legislation violated Article 8(2) ECHR because (a) it was insufficiently transparent; and (b) contained insufficient safeguards to protect the right to private life, as required under Article 8(2) ECHR. However, the Court's analysis of (and supposed concern for) sufficient safeguards was somewhat undermined, as mentioned above, by its refusal to examine whether the SyRI legislation was "in accordance with law" under the first limb of Article 8(2) review. The Court decided that it would leave "undiscussed in its review whether the SyRI legislation is sufficiently accessible and foreseeable and as such affords an adequate legal basis".[121] Finally, the Court stated, it would not assess whether the SyRI legislation was in violation of specific provisions of the GDPR[122].

IV. SyRI 2.0

37 Not long after the of SyRI judgment was delivered, the Dutch government proposed legislation to the Dutch Parliament which critics have dubbed "Super SyRI".[123] The law - *wet gegevensverwerking door samenwerkingsverbanden* (WGS) - is intended to function as a framework for data sharing and the use of ADM systems.[124] The government considers the WGS is needed as it creates a legal basis for the data processing which is currently lacking,[125] and the data sharing and analysis across government agencies is deemed necessary for a more integrated approach to societal problems.[126] This proposed legislation clearly shows the impact of the SyRI judgment, and the fast pace of the digital transformation of the administrative state.

38 Due to several waves of severe criticism,[127] the proposed WGS has been amended twice, with its most recent version currently being discussed in the Dutch Senate.[128] The latest WGS proposal addresses several of these criticisms and, in essence, functions similarly to SyRI: creating a legal framework basis for data sharing and the use of ADM systems across government agencies. A notable difference is that the WGS is not specifically geared towards social welfare fraud, but is currently aimed at government partnerships in the domain of financial fraud, money laundering, organised crime and complex health and safety cases.[129] However, the law does contain the explicit possibility of adding other partnerships in a broad range of domains, including social welfare, by means of government decree.[130] Despite the substantial reforms to the proposed WGS, the current version is persistently receiving considerable criticism from NGOs, wider society, and the Dutch Parliament itself.[131] The criticism focusses on, still, the reliance on delegated competencies (a framework-law structure) and the vast scope of different domains or goals included in the framework. The combination of both these qualities means the possible scopes of partnerships, types of data and ADM systems are nearly unlimited.

39 Viewing the proposed WGS in light of the SyRI judgment brings up many questions with regards to the concrete functioning of several of the proposed safeguards, and the de facto extent of the transparency of possible ADM systems. However, the most interesting connection to make will be

121 ibid para 6.72.

122 ibid para 6.107.

123 Peter te Lintel Hekkert, 'Zet Super SyRI op de Lijst met Controversiële Wetsvoorstellen' (*FNV*, 1 Febuary 2021) <https://www.fnv.nl/nieuwsbericht/sectornieuws/uitkeringsgerechtigden/2021/02/verklaar-super-syri-controversieel>; 'Super SyRI: Bestuurd door Black Boxes' (*Bij voorbaat verdacht*, 12 November 2020) <https://bijvoorbaatverdacht.nl/super-syri-bestuurd-door-black-boxes/>.

124 TK 2019-2020, 35 447, nr. 2.

125 Werkgroep verkenning kaderwet gegevensuitwisseling, 'Kennis delen geeft kracht' (2014), bijlage bij TK 2014-2015, 32 761, nr. 79, p. 5; M. P. Beijer, 'Het voorstel voor een nieuw regelgevend kader voor de gegevensverwerking door samenwerkingsverbanden' (2020), TvBSH 6; TK, 2019-2020, 35 447 nr 3, p. 2.

126 TK 2019-2020, 35447, nr. 3, p. 2.

127 See bijlagen bij TK 2019-2020, 35 447, nr. 3; TK 2020-2021, 35 447, nr. 20; TK 2019-2020, 35 447, nr. 4; 'SyRI-coalitie maant kabinet: stop overhaaste invoering 'Super SyRI'' (*Bij voorbaat verdacht*, 25 May 2020) <https://bijvoorbaatverdacht.nl/syri-coalitie-maant-kabinet-stop-overhaaste-invoering-super-syri/>; and Harriet Duurvoort, 'Hoe de Overheid Inbreuk maakt op Privacy is Dubieuzer dan Facebook en Google' *de Volkskrant* (Amsterdam, 27 May 2020). For a summary of the earlier criticism see: M. P. Beijer (2020) (n 116) p. 311.

128 EK 2020-2021, 35 447, nr. Al; TK 2019-2020, 35 447, nr. 1; TW 2019-2020, 35 447, nr. 4.

129 Hoofdstuk 2 WGS.

130 Art 3.1 WGS.

131 'SyRi-coalitie aan Eerste Kamer: 'Super SyRI' Blauwdruk voor meer Toeslagenaffaires' (*Platform Bescherming Burgerrechten*, 11 January 2021) <https://platformburgerrechten.nl/2021/01/11/syri-coalitie-aan-eerste-kamer-super-syri-blauwdruk-voor-meer-toeslagenaffaires/?s=SyRI>; Tommy Wieringa, 'De Wet is een Slang die Alleen Mensen Zonder Schoenen Bijt' *NRC Handelsblad* (Amsterdam, 23 January 2021) 2. TK 2020-2021, 35510, nr. 27.

with a not previously discussed element of the SyRI judgment. As the Court emphasises at several points in the judgment, that it considers, based on ECHR case law, the government to have "a special responsibility when applying new technologies to strike the right balance between the benefits the use of such technologies brings as regards preventing and combating fraud on the one hand, and the potential interference with the exercise of the right to respect for private life through such use on the other hand."[132] This "special responsibility" plays an important role in the Court's weighing of whether SyRI's interference in people's private lives is to be considered necessary in a democratic society.[133] The Court substantiates this responsibility further by emphasising the speed of developments in data-linking and automated analysis, which increases the risk for people's private lives, whilst simultaneously making it more difficult to understand what effect these systems have on people's lives.[134] This is why, according to the Court, the government has a "special responsibility" with the implementation of such technologies, which can be interpreted as raising the bar for a government in those circumstances.

40 Considering this special responsibility, especially the instrument of a framework law which leaves most particulars to delegated government decrees can be seen as problematic. Any government system geared at fraud detection needs to balance this aim with the fundamental right to private life. This special responsibility seems to imply that the exercise of ensuring this "fair balance" must be conducted with more care or more extensively when implementing ADM systems. The structure of a framework law precludes the possibility of an extensive parliamentary and societal debate, and detailed context-specific deliberations on the implementation of an ADM system by a given (private) partnership. As such, a framework law allowing for the use of ADM systems by the government seems to not take sufficient heed of this special responsibility to substantiate how the "fair balance" between a specific aim and the right a private life is achieved. Interpreted in this way, this idea of a special responsibility as developed in the SyRI judgment is fully in line with the original advice of the Council of State in 2019, where it advised against a framework law, favouring specific sectoral legislation.[135]

132 *SyRI* (n 5) para 6.84, citing *S and Marper v UK* (n 86) para 112.

133 ibid para 6.84 - 85.

134 ibid para 6.85.

135 TK 2019-2020, 35 447, nr. 4.

E. Conclusion

41 This article has critiqued the SyRI system in the Netherlands and used The Hague Court's landmark judgment as a lens through which to examine the broader issues arising from the digitisation of the State through the use of ADM systems by public authorities. This discussion raises three concluding points. First, the conceptualisation of the problems and issues with ADM systems seems to be over-focused on the inner workings of the technology used (e.g., the technical properties of the SyRI system), an over-focus on attempting to fit technological questions into specific legal classification regimes (primarily under the GDPR), and with the technology itself being unquestionably connected to progress and efficiency i.e., technological-solutionism. However, this approach risks law becoming merely an overly technologically-centred analysis. Instead, we argue that when looking at the use of ADM systems by public authorities, we should treat the technology as a mere starting point, with the role of law (and human rights law in particular) being to bring in other perspectives, including the role of the technology in its social context and people's actually experience with these systems. This occurred in the SyRI case for instance when the Court was able to take into account the SyRI system was only being used in so-called problem neighbourhoods, and that such uses meant the system could create links based on bias, including lower socio-economic status or an immigration background. Of course, there are limitations to human rights law analysis of ADM systems, as this form of legal review does not allow for a questioning of the underlying policy choices for introducing these systems (beyond the cursory examination of whether an ADM system pursues a legitimate aim). A second connected point concerns the "special responsibility" governments have to safeguard the private life of their citizens when implementing ADM systems. This increased responsibility concretely translates to the need for the government to take extra care in establishing there is a fair balance between the aim the ADM system seeks to fulfil, and any interference with citizens' private lives. General framework laws, such as those implemented in the Netherlands, that leave many of the concrete weighing of these interests and rights to delegated ministerial competencies, do not easily seem to be compatible with this special responsibility, and are a model that should not be followed in other EU member states.

42 Finally, the analysis demonstrates the difficulty of the application of data protection frameworks (especially Article 22 GDPR on automated individual decision-making) to ADM systems deployed by the administrative state, and to the digital transformation of the State more broadly. This was epitomised by The Hague Court's convoluted approach to the GDPR

and Article 8 ECHR, and choosing the latter as the most appropriate framework for its examination of the SyRI legislation. However, as discussed above, the suitability of current data protection frameworks for protecting individuals from disproportionate interferences with their private life must be questioned. Instead, an assessment of these technologies should recognize that their use "prompts a particularly high probability of infringement of human rights, including by introducing or amplifying distributive injustice", especially where the ADM system produces serious consequences for people, such as legal consequences, losing social welfare, or people forced by law to be subjected to risk profiling by public authorities.[136] As such, we must move beyond treating these technologies are simply "scientific" and "neutral",[137] and question more structural aspects, including the underlying policy choices involved in their deployment. This approach could hopefully obviate the need for courts, such as The Hague Court, to step in to protect individual citizens from the excesses of the use of ADM systems by the State.

136 ibid.

137 UN Doc A/74/493 (n 1) para 6.

Imbalanced data as risk factor of discriminating automated decisions

A measurement-based approach

by **Antonio Vetrò***

Abstract: Over the last two decades, the number of organizations -both in the public and private sector- which have automated decisional processes has grown notably. The phenomenon has been enabled by the availability of massive amounts of personal data and the development of software systems that use those data to optimize decisions with respect to certain optimization goals. Today, software systems are involved in a wide realm of decisions that are relevant for the lives of people and the exercise of their rights and freedoms. Illustrative examples are systems that score individuals for their possibility to pay back a debt, recommenders of the best candidates for a job or a house rent advertisement, or tools for automatic moderation of online debates. While advantages for using algorithmic decision making concern mainly scalability and economic affordability, on the other hand, several critical aspects have emerged, including systematic adverse impact for individuals belonging to minorities and disadvantaged groups. In this context, the terms data and algorithm bias have become familiar to researchers, industry leaders and policy makers, and much ink has been spelled on the concept of algorithm fairness, in order to produce more equitable results and to avoid discrimination. Our approach is different from the main corpus of research on algorithm fairness because we shift the focus from the outcomes of automated decision making systems to its inputs and processes. Instead, we lay the foundations of a risk assessment approach based on a measurable characteristic of input data, i.e. imbalance, which can lead to discriminating automated decisions. We then relate the imbalance to existing standards and risk assessment procedures. We believe that the proposed approach can be useful to a variety of stakeholders, e.g. producers and adopters of automated decision making software, policy makers, certification or audit authorities. This would allow for the assessment of the risk level of discriminations when using imbalanced data in decision making software. This assessment should prompt all the involved stakeholders to take appropriate actions to prevent adverse effects. Such discriminations, in fact, pose a significant obstacle to human rights and freedoms, as our societies increasingly rely on automated decision making. This work is intended to help mitigate this problem, and to contribute to the development of software systems that are socially sustainable and are in line with the shared values of our democratic societies.

Keywords: discrimination risk; data bias; algorithm fairness; digital policy; data ethics; data governance

Recommended citation: Antonio Vetrò, Imbalanced data as risk factor of discriminating automated decisions: A measurement-based approach 12 (2021) JIPITEC 272 para 1.

A. Background and Motivations

1 A large number of decisional processes -both in the public and private sector- are based on software elaborated recommendations, or they are completely automated, and it is likely that the phenomenon will further increase in the future [1] [2] [3]. This phenomenon has been enabled by the large availability of data and of the technical means in order to analyze them for building the predictive, classification and ranking models that are at the core of automated decision making (ADM) systems[1]. The decisions delegated or supported by these systems range from predicting debt repayment capability [4] to identifying the best candidates for a job position [5], from detecting social welfare frauds [6] to suggesting which university to attend [7], to name a few. While advantages for using ADM systems are evident and they concern mainly scalability of the operations and consequential economic efficiency, on the other hand, several critical aspects have emerged, including for instance transparency and accountability [8]. Yet another major controversy concerns discriminatory behavior, in terms of "unjustified distinction of individuals based on their membership, or perceived membership, in a certain group or category" [9]. This issue emerged from a large amount of evidence both in scientific literature [10] and journalistic investigations [11], which showed how ADM systems may systematically discriminate the weakest segments of society and exacerbate existing inequalities. Such problem would often occur as a result of imbalanced input datasets [12], which is the focus of this paper. Data imbalance is an unequal distribution of data between classes [13], which occurs when the number of data points available is very different among different classes. Causes of imbalance can be errors or limitations of the data collection design and operation, alternatively no other reason than disparities in the current reality that the data itself reproduce. Imbalance is between-class when only two classes are taken into consideration and one class is over-represented with respect to the other or multiclass when imbalances exist between multiple classes. In this paper, we focus on the more general case, i.e. multiclass imbalance.

2 Imbalanced data is known to be problematic in the machine learning domain since long [14], and is still relevant [15], especially because it can corrupt the performances of supervised learning algorithms in terms of heterogeneous accuracies across the classes of data. For example, consider an algorithm for predictive maintenance that labels a certain product component either as close to breakage or not close to breakage, and is trained with historical data from three different suppliers. A is a well-known company which sells several million pieces of that component per year. B is a company with a few thousand sales, and C is a company with less than a thousand sold components of that product. It is reasonable to expect that the algorithm trained with the historical data from the three companies could perform with higher prediction accuracy for components of supplier A and lower accuracy for products of suppliers B and C. In this fictitious example, imbalance in the input data could be the major cause for the disparate performance of the predictive algorithm, due to the fact that the model has been trained with significantly more data from Company A[2].

3 Now imagine a context where the objects of the prediction are not products but individuals, and an organization uses historical data on employees to predict which candidates' CVs most likely correspond to future successful software engineers. It comes as no surprise that the large majority of predicted candidates will be male, due to the disproportionate gender ratio in the sector. Indeed, this is not a fictitious example but rather a very blatant case of discrimination caused by data imbalance. Namely, the development of a software system by Amazon to evaluate the CVs of potential employees retrieved from the web [16]. The goal of the system was to find successful future employees, whereby the predictors were word patterns extracted from CVs of the past 10 years. According to the news agency report [16], the project started in 2014 and was stopped in 2017 because female profiles were systematically downgraded, regardless of a certain number of attempts to make technical adjustments. Here, the problem was that training data came mostly from men, since the majority of employees in the technology sector is male.

4 A similar unequal treatment due to gender imbalance in the input data has been found in a scientific

* Antonio Vetrò is a Senior Research Fellow at Nexa Center for Internet & Society and Assistant Professor at the Department of Control and Computer Engineering of Politecnico di Torino, Italy. ORCID: 0000-0003-2027-3308.

1 We follow the definition of Automated Decision Making provided by Algorithm Watch[1]: "Systems of automated decision-making (ADM) are always a combination of the following social and technological parts: i)a decision-making model ; ii) algorithms that make this model applicable in the form of software code ; iii) data sets that are entered into this software, be it for the purpose of training via Machine learning or for analysis by the software; iv) the whole of the political and economic ecosystems that ADM systems are embedded in (elements of these ecosystems include: the development of ADM systems by public authorities or commercial actors, the procurement of ADM systems, and their specific use)."

2 Due to the large difference of available data from the three companies, concurrent causes as incomplete data or different defectiveness ratios might play a minor role in explaining the divergence of performance measures.

experiment on the search engine Common Crawl [17]. The authors compared three techniques of machine learning for occupational classification with almost 400.000 collected biographies. In all cases, even without explicitly using gender indicators, the rate of correct classifications followed the existing gender imbalances of the occupational groups. In another study [18] it was reported that Facebook advertisements for employment opportunities were significantly skewed among ethnic and gender groups, leading to persistent discriminatory treatment and unequal job opportunities along the lifetime of the advertisements. This study was partially replicated by Algorithm Watch, with similar results [19]. For example, an advertisement for truck driver jobs was shown about ten times more to men than to women (4,864 times vs 386), which confirms that Facebook optimizes its target audience with past users' reactions to similar announcements, thus replicating imbalances in the data. The consequence of such a conservative mechanism is that people are deprived of opportunities based on gender, ethnic origin or other personal traits, in practice infringing Article 21 of the EU Charter of Human Rights [20]. In the United States (US), the discriminatory effect of the Facebook advertisement platform has been scrutinized by the Department of Housing and Urban Development. It sued Facebook in March 2019 for violating the Fair Housing Act, whereby the allegations were based on the evidence that housing advertisements were disproportionally targeted with respect to race, gender and other personal characteristics [21].

5 Amplifications of input data imbalance in software outputs have also been reported in general purpose search engines. A study by Kay et al. [22] on Google search results showed that in the occupational groups typically dominated by men, women were significantly under-represented, in comparison to the real gender ratio retrieved from the official employment statistics. The authors showed that such disproportion influences the perceptions of actual gender relations in occupations, with possible amplification effects on inequalities in jobs. Discrimination issues in the Google search engine are not a novel fact, as demonstrated in an empirical study from 2013 [23], which showed that advertisements for commercial products of arrest records were displayed with relevant different rates for names usually referred to non-Caucasian people than for names usually referred to Caucasian people. The opacity of the search algorithm did not allow the authors to isolate and validate the causes. However, they had confidence in reporting that the past clicks behavior of Google search users (used by Google AdSense service) might have played a major role and propagated a societal bias in the search algorithm results.

6 The negative effects illustrated in these cases could become worse or even life-altering in fields like justice or medicine, where the combined use of ADMs and historical data is rapidly increasing. The most famous case in the justice field is the investigation on COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), conducted by the non-profit organization Pro-Publica [24]. COMPAS is an algorithm used by judges to assess the probability of recidivism of defendants. The COMPAS algorithm was distorted in favor of white defendants, whereby those who were rearrested were nearly twice as likely to be misclassified as low risk than black defendants. Furthermore, the black defendants who did not get rearrested were nearly twice as likely to be misclassified as higher risk (false positive) than white defendants. The major cause was that the number of records in the dataset related to black defendants was much higher than the number of records of white defendants.

7 Regarding the medical field, a recent study [25] found evidence of ethnicity-based discrimination in a widely-used commercial system for deciding which patients should get into an intensive care program. Medical doctors applied risk scores generated by an algorithm trained on historical data about medical expenditure and the use of health services. In cases of an equivalent health status, white patients were significantly more likely than black patients to be assigned to the intensive care program. In fact, the risk score reflected more the expected cost of treatment than health conditions, with former being highly correlated to the economic wealth of the patients. In another empirical study in the medical field [26], the amount of data used for training classification algorithms in six different clinical disciplines showed that most of it came from only three geographic areas in the US, with no representation for the majority of states. Hence, automating a diagnosis on patients from states not included in the training data, would lead to wrong results and to missing health issues that are more common in the excluded geographic areas.

8 The cases summarized above, although exemplificative and not exhaustive, clearly show how imbalance in data can propagate and be reflected in the output ADM systems. When this occurs, it ceases to be a mere problem of data engineering and it becomes a socio-technical issue, particularly important when systems automate high stake decisions that can produce serious consequences for individuals. As our societies increasingly rely on ADMs, this phenomenon poses a significant challenge for the values on which our societies are based and for fundamental human rights and freedoms. The deployment of ADM systems embeds the risk to create an adverse impact for individuals belonging to minorities and marginalized groups, and to introduce or amplify distribu-

tive injustice [27] [28]. In this paper we face this important issue by focusing on the specific problem of data imbalance. We propose a measurement-based risk assessment approach, by measuring imbalance in input data, whereby we highlight the potential risk of discriminating automated decisions. We describe the theoretical foundations of the risk assessment approach, which resides in existing standards on software system quality and risk assessment. We identify three measures of imbalance and we apply them with an illustrative example.

9 The measures can be applied both before the deployment (i.e., during development) and after the deployment of ADM systems: for this reason, we believe that the proposed approach can be useful to a variety of stakeholders for assessing the risk of discriminations, including the creators or commissioners of the system, researchers, policymakers, regulators, certification or audit authorities. Assessments should prompt taking appropriate action to prevent adverse effects.

10 The paper is organized as follows: in Section B we lay the theoretical foundations of our proposal, followed in Section C by the explanation of three imbalance measures and an example of their application. In Section D we explain how this research contributes to the literature of algorithm bias and fairness, while in Section E we briefly report on the relations to the most recent policy efforts in Europe for regulating ADM systems. We conclude in Section F with a discussion of the limitations and share our roadmap for future work.

B. Data imbalance as risk factor of discriminations by automated decision making systems

11 The ADM systems described in the previous section systematically discriminate against certain groups of individuals because of imbalances in the input data. For this reason, we consider data imbalance as a risk factor and we propose measures to address it. This proposal has its foundations in software quality and risk management standards.

12 The cornerstone of the conceptual model is the series of standards ISO/IEC 25000:2014 Software Engineering — Software Product Quality Requirements and Evaluation (SQuaRE) [29]. SQuaRE includes quality modeling and measurements of software products,[3] data and software services. According to the philosophy and organization of this family of standards, quality is categorized into one or more quantifiable characteristics and sub-characteristics. For example, the standard ISO/IEC 25010:2011 formalizes the product quality model as composed of eight characteristics, which are further subdivided into sub-characteristics. Each (sub) characteristic relates to static properties of software and dynamic properties of the computer system[4]. An example of product quality characteristics is reliability, and one of its sub-characteristics is maturity[5]. Characteristics and sub-characteristics can be quantified by measurable properties of the software. For example, "failure" is a dynamic property of the software, and the number of failures is a quality measure element, which is used to measure maturity in terms of mean time between failures[6]. Reliability is quantified through the measures of its sub-characteristics.

13 Similar to product quality, data quality in ISO/IEC 25012:2008 is categorized into 15 characteristics, such as completeness, efficiency, recoverability. Each of these characteristics is quantifiable through measures of quality-related properties, defined in ISO/IEC 25024:2015. The characteristics can belong either to the "Inherent" point of view if dependent only on the data themselves, such as completeness. Alternatively, they can belong to the "System-dependent" point of view, such as recoverability. They can also belong to both, such as efficiency. Data imbalance is not a characteristic of data quality in ISO/IEC 25012:2008, however the SQuaRE standards have a structure which fits our purpose, and it defines a principle that is relevant in our context, which is the propagation principle. This principle entails that the quality of the software product, service and data would affect the quality in use and would thus have consequences for the users of a software system[7].

3 A software product is a "set of computer programs, procedures, and possibly associated documentation and data" as defined in ISO/IEC 12207:1998. In SQuaRE standards, software quality stands for software product quality.

4 A system is the "combination of interacting elements organized to achieve one or more stated purposes" (ISO/IEC 15288:2008), for example the aircraft system. It follows that a *computer system* is "a system containing one or more components and elements such as computers (hardware), associated software, and data", for example a conference registration system. An ADM system that determines eligibility for economic aid for paying drinking water bills is a software system.

5 Reliability is defined in ISO/IEC 25010:2011 as the degree to which a system, product or component performs specified functions under specified conditions for a specified period of time"; Maturity is defined in ISO/IEC 25010:2011 as "degree to which a system, product or component meets needs for reliability under normal operation".

6 Number of failures/average min-max duration.

7 In practice evaluating and improving product/service/

Figure 1 represents how this chain of effects is formalized in SQuaRE. In the realm of data quality, a simplification of this concept is the GIGO principle, which is the "garbage in, garbage out" principle. In other words, data that is outdated, inaccurate and incomplete make the output of the software unreliable.

14 We apply this principle to data imbalance because it can cause biased software outputs that negatively affect the final users, in the same way bad data quality affects the quality in use and thus has an impact on the final users. In fact, imbalanced datasets may lead to imbalanced results, which in the context of ADM means differentiation of products, information and services based on personal characteristics. In specific applications such as wages, insurance, education, working positions, tariffs, etc. such differentiations can lead to unjustified unequal treatment or discrimination. For this reason data imbalance shall be considered as a risk factor in all those ADM systems that rely on historical data and operate in relevant aspects of the lives of individuals.

15 The second conceptual pillar of the proposal is the ISO 31000:2018 standard [31] which identifies guiding principles for risk management. The proposal consists of a framework for integrating risk management into organizational contexts, and a process for managing risks at "strategic, operational, program or project levels". In the context of this discussion, data imbalance shall be explicitly taken into account within the risk management process, which we reproduce from the standard in Figure 2. Risk assessment is therefore at the center of our proposal. The process consists of risk identification, analysis and evaluation. Here, we briefly describe them and specify the relation with our approach.

- Risk identification refers to finding, recognizing and describing risks within a certain context and scope, and with respect to specific criteria defined prior to risk assessment. In our case, it is the risk associated with discriminating individuals or groups of individuals by operating ADM systems in contexts in which the impact on the lives of people would be relevant. Section A contains examples of these situations.

- Risk analysis aims to understand the characteristics of the risk and, when possible, its levels. This is the phase where measures of data imbalance are used as indicators for the risks of discrimination, due to the bias propagation effect previously described. In Section C we will introduce three measures and we will show them in action on a real dataset.

- Risk evaluation, as the last step, is a process in which the results of the analysis are taken into consideration in order to decide whether additional action is required. If affirmative, this process would then outline available risk treatment options and the need for conducting additional analyses. In addition, the process would define other types of required actions and the actors who would undertake those actions. In our case, the indicators of data imbalance should be analyzed in the context of the specific prediction/classification algorithms used, the social context, the legal requirements of the domain, etc.[8]

16 Figure 3 summarizes the approach and the connections with the international ISO/IEC standards used as reference frameworks. In the upper layer, we represent the elements of the SQuaRE series (2500n) which are most relevant for our scope. In the bottom layer, we report the main elements of the risk management process of ISO 31000. The constitutive elements of our approach - in the middle of Figure 3- are mapped to the concepts of SQuaRE and the phases of ISO 31000:

- the ADM systems constitute the context of use in terms of SQuaRE terminology, and they are specified in the context definition phase of the ISO 31000;

- the discrimination operated by ADM systems is the specific object of the risk identification process in ISO 31000 (given the context), and it decreases the quality in use of the software;

- data imbalance extends the SQuaRE data quality model because it is an inherent data characteristic: as such, i) it preserves the propagation principle and ii) it is measurable; the identified measures can be used as risk indicators in the risk analysis phase;

data quality is one mean of improving the system quality in use. It shall be clarified that in this text we refer only to the effects related to quality characteristics of the SQuaRE standards. However, the same principle can be applied to other aspects of software development that are treated in other standards, for instance the improvement of any of the lifecycle processes defined in ISO/IEC 12207:2008 and ISO/IEC 15288:2015 will determine an improvement of product quality, which in turn contributes to improving system quality in use and has a positive effect on final users (users can be direct and indirect). Although this aspect is out of our scope here, it could be relevant for techniques/procedures applied in software development processes to identify negative societal effects of software since its early development phase (for instance, in requirements definition [30]).

8 This part is not in the scope of this paper; however, we will provide some details for future work needed in this direction in Section F.

- the criteria for activating mitigation actions (e.g., thresholds for the indexes) and the mitigation actions are mapped respectively to the risk evaluation and risk treatment phases.

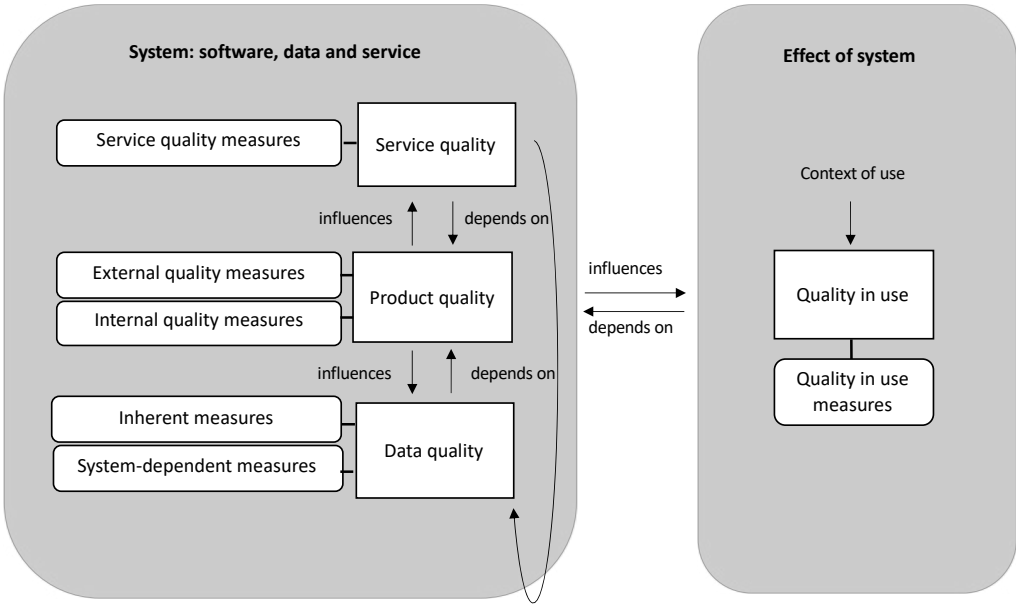


Figure 1. Quality effects in SQuaRE

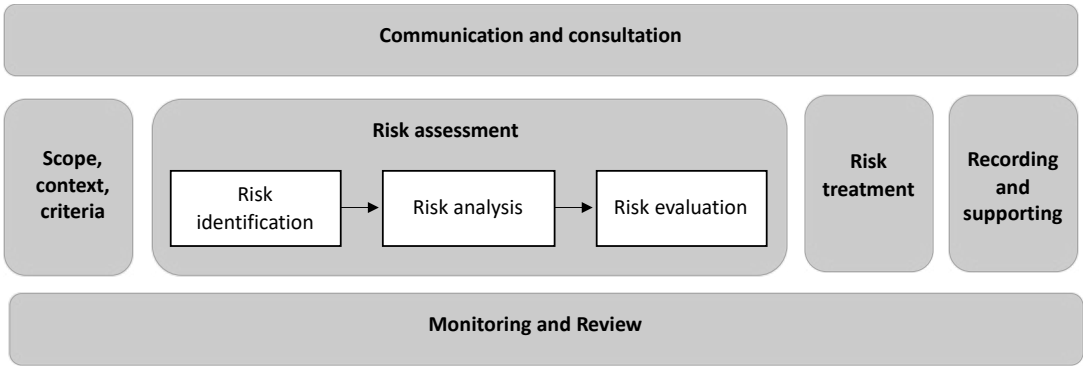


Figure 2. Risk management process in ISO 31000:2018

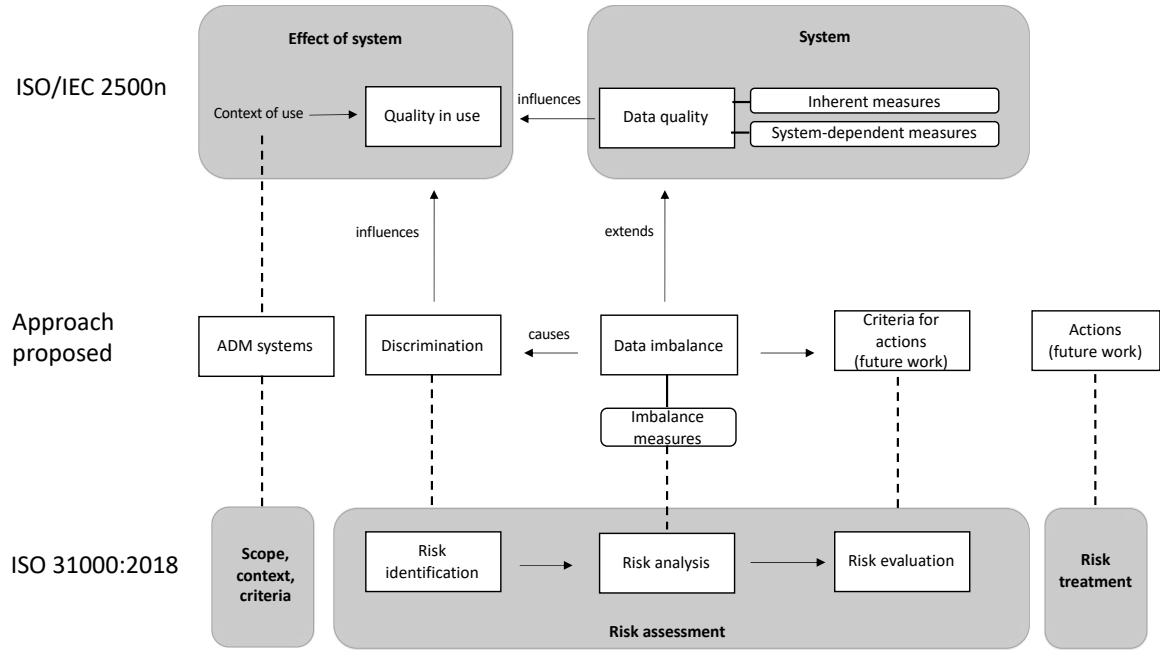


Figure 3. Approach proposed in relation to ISO standards of reference

C. Measures of imbalance for categorical data

17 According to our line of reasoning, imbalance in input data can propagate downstream to software output. As a consequence, measures of imbalance are interpreted as risk indicators.

18 Since imbalance is defined as an unequal distribution between classes [13], we focus on categorical data. In fact, most of the sensitive attributes are considered categorical data, such as gender, home town, marital status, and job. Alternatively, if they are numeric, they are either discrete and within a short range, such as family size, or they are continuous but often re-conducted to distinct categories, such as information on "age" which is often discretized into ranges such as "< 25", "25-40", "41-60". We identified three measures from the literature of social and natural sciences, where imbalance is known in terms of (lack of) heterogeneity and diversity: the identified measures are the Gini, Shannon and Simpson indexes. We provide details in Table 1, whereby we specify their formula and normalized versions, i.e. in the range 0-1, respectively in the second and third columns. In the fourth column, we provide notes for value interpretations.

19 We briefly comment on the measures here:

- Gini index measures how many different types are represented in a dataset. It has been conceived as a measure of heterogeneity, whereby it is used for different purposes in several disciplines, for example, to measure political polarization, market share in competition, ecological diversity, and racial discrimination. It increases if probabilities/frequencies become as equal as possible e.g. when different attributes would have similar representations.

- Shannon index has been proposed as a measure of diversity, and it provides information about community composition, taking the relative abundances of different classes into account. It is a concept widely employed in biology, phylogenetics, and ecology.

- Simpson index is another measure of diversity in ecology, which measures the probability that two individuals randomly selected from a sample belong to the same species or the same class/category. It has been used in ecology for measuring the diversity of living beings in a given place, as well as in social and economic sciences for measuring wealth, uniformity, and equity.

20 In order to show the three measures at work, we make an example with the widely used data from COMPAS as they are provided by the US based non-profit organization ProPublica [32]. The data contain variables used by the COMPAS algorithm in scoring criminal defendants in Broward County (Florida), along with their outcomes within two years of the decision. The original dataset includes 28 variables, eight of which are considered as protected attributes[9], such as last name, race, or marital status.

9 Protected attributes are qualities, traits or characteristics of individuals that, by law, cannot be discriminated against.

Table 1 Indexes of imbalance.

Index	Formula	Normalized formula	Notes
Gini	$G = 1 - \sum_{i=1}^{m} f_i^2$	$G_n = \frac{m}{m-1} \cdot \left(1 - \sum_{i=1}^{m} f_i^2\right)$	m is the number of classes f is the relative frequency of each class $f_i = \frac{n_i}{\sum_{i=1}^{m} n_i} f_i = \frac{n_i}{\sum_{i=1}^{m} n_i}$ n_i= absolute frequency The higher G and G_n, the higher is the heterogeneity: it means that categories have similar frequencies The lower the index, the lower is the heterogeneity: a few classes account for majority of instances
Shannon	$S = -\sum_{i=1}^{m} f_i \ln f_i$	$S_n = -\frac{1}{\ln m} \sum_{i=1}^{m} f_i \ln f_i$	For m, f, f_i and n_i check Gini Higher values of S and S_n indicate higher diversity in terms of similar abundances in classes The lower the index, the lower is the diversity, because a few classes account for most of the data
Simpson	$D = \frac{1}{\sum_{i=1}^{m} f_i^2}$	$D_n = \frac{1}{m-1}\left(\frac{1}{\sum_{i=1}^{m} f_i^2} - 1\right)$	For m, f, f_i and n_i check Gini Higher values of D and D_n indicate higher diversity in terms of probability of belonging to different classes The lower the index, the lower is the diversity, because frequencies are concentrated in a few classes

The identification of protected attributes can be related to the characteristics listed in Article 21 - Non- discrimination of the EU Charter of Human Rights [20].

We chose the COMPAS dataset because it is probably the most known source in the scientific communities that study bias and fairness of algorithms. As we summarized previously, Pro Publica showed that the COMPAS algorithm classified black people with a much higher risk of recidivism than white people. Here, the probability of being predicted high risk was 47% for black people and 24% for white people, and a similar difference was observed in the false positives rate, i.e. 31% black people vs 14% white people. This occurred mainly because input data is highly imbalanced. In other words, not only black defendants in the dataset are many more than white defendants, with a 51% vs 34% ratio, but the ratio of black recidivist in the whole dataset was double the ratio of white recidivist with 27% against 13%. Similar, although less striking, considerations can be made for the gender attributes whereby women labeled high-risk got a much lower risk of recidivating than men classified as high-risk. The age attribute, on the other hand, was the stronger predictor of high score for violent recidivism (details are available in [32]).

21 Taking into considerations these problematic aspects, we make the following computations:

- We summarize the frequencies of ethnicity, gender, and age categories in Table 2 , both in terms of the overall percentage and as to the ratio of recidivists;
- We compute the imbalance measures on ethnicity and gender categories, both in the whole dataset and on recidivists only, and we report results in Table 3, embedding histograms.

22 We first look at the measurements in the whole dataset. There, all indexes are able to detect the imbalance in the three classes. However, each index has a different sensibility.

- Simpson values are much lower than Gini and Shannon and they point to ethnicity as the most imbalanced data, followed by sex and age categories. This index is more sensitive to the number of possible instances, i.e. six for ethnicity, two for gender and three for the age category;
- Shannon provides the same order of risk provided by Simpson, however, with higher values and shorter distances between rank positions, namely 0.08 between 1st-2nd and 0.19 between 2nd-3rd positions vs respectively 0.14 and 0.24 in Simpson, which results in more distinct values;
- Gini is different because it highlights a higher risk for the sex column, thus reflecting its strongest influence as a predictor, followed by ethnicity and age categories.

23 Looking at the column "percentage recidivists" in Table 2, we observe that measures are lower than the previous column, reflecting an even higher imbalance in the values of the three classes:

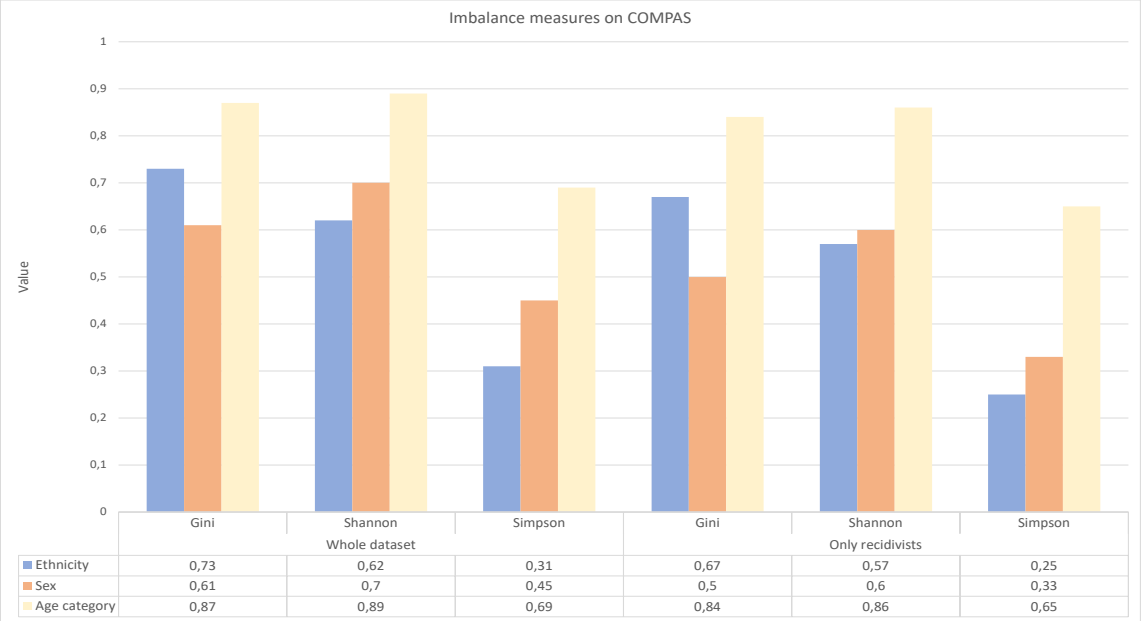
- Simpson preserves the previous rank, but the distance between ethnicity and sex is closer, while the age category has only a slight decrease;
- Shannon keeps being very similar to Simpson, however with higher values;
- Gini also preserves its rank of values, but the difference between the first and second positions is larger now.

24 The question is which index to use. Given that in COMPAS the most severe problem occurred with ethnicity, the answer for this specific dataset would be the Simpson index, due to the fact that it identifies the highest imbalance in a more distinct way. However, this is a consideration made *a posteriori*, on a well-known case with a well-established problem. In view of the future real cases, especially in the design and production phase of an ADM system where there is no information on how the system behaves in operation, a certain number of further considerations should be made, with the most relevant being how to handle a divergence of index values, how to choose meaningful severity thresholds for each index, and which actions to take after the risk is recognized as a relevant concern. To resolve these issues and to make the measures trustable as risk indicators, their reliability shall be extensively investigated, taking also into consideration different types of data and classification/prediction algorithms, the application domain and the groups of stakeholders who are potentially impacted. We will make a further mention to this future work in the last section of the manuscript.

Table 2 Frequency of occurrences for attributes in ethnicity, sex, age categories in COMPAS

ATTRIBUTE	ATTRIBUTE VALUE	OVERALL PERCENTAGE	PERCENTAGE RECIDIVISTS
ETHNICITY	African-American	51.4%	26.9%
	Caucasian	34.1%	13.3%
	Hispanic	8.2%	3.1%
	Asian	0.5%	0.1%
	Native American	0.2%	0.1%
	Other	5.6%	2.0%
SEX	Male	81.0%	38.8%
	Female	19.0%	6.7%
AGE CATEGORY	Less than 25	21.8%	12.2%
	Between 25 and 45	57.2%	26.6%
	Greater than 45	20.9%	6.7%

Table 3 Application of the indexes to COMPAS database



	Whole dataset			Only recidivists		
	Gini	Shannon	Simpson	Gini	Shannon	Simpson
Ethnicity	0,73	0,62	0,31	0,67	0,57	0,25
Sex	0,61	0,7	0,45	0,5	0,6	0,33
Age category	0,87	0,89	0,69	0,84	0,86	0,65

D. Relations to research in algorithmic bias and fairness

25 In recent years, much ink has been spilled on bias and fairness in algorithms. An impressive amount of scientific research has been carried out, especially in the machine learning communities, in order to elaborate strategies that would lead to more equitable results of ADM systems. Efforts mainly focus on techniques to detect systematic discriminations and mitigating them according to different definitions of fairness. Excellent references for getting an overall picture are the survey on bias and fairness in machine learning by Mehrabi et al. [33] , the comprehensive, and still ongoing, work on fairness in machine learning by Barocas et al. [34] and the review of discrimination measures for algorithm decision making by Žliobaitė [35]. A common limitation of these approaches is that mathematical formalizations of fairness cannot be simultaneously satisfied [36][37]. In other words, no universally accepted notion of fairness exists, since defining "fair impact" implicitly embodies political, economic or cultural visions [38]. The ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT[10]) has recognized this issue and has been designed and promoted not only for computer scientists working in the area, but also for scholars and practitioners from "law, social sciences and humanities to investigate and tackle issues in this emerging area". Our approach can be located in this space of inter-disciplinary discussion. It contributes to the main corpus of researches on algorithmic bias and fairness by moving the focus from the outcomes of ADM systems to their inputs, and by making a first step to fill a well-recognized existing gap in the literature, as reported in recent studies, such as in [39] ("There is a need to consider social-minded measures along the whole data pipeline") and in [40] ("returning to the idea of unfairness suggests several new areas of inquiry [...] a shift in focus from outcomes to inputs and processes").

26 In addition, we aim at reaching a higher generalizability of what we currently observe in the field of research. Namely, i) our approach can be applied to any ADM system which is data-based, and not only in machine learning; ii) we build our theoretical framework upon a series of international standards, which incorporate *by design* a multi-stakeholder perspective; iii) we look at data imbalance as risk factor and not as a technical fix, despite the fact that there are well-established techniques for reducing data imbalance in the field of data engineering, especially for machine learning, where the problem has been spotted since the beginning of the 2000's [13]). In this context, we think it is preferable to keep the ultimate responsibility in the realm of human agency. We believe that a risk approach is more suitable for the scope, as it creates space for active human considerations and interventions, rather than delegating the mitigation of the problem to yet another algorithm.

10 See <https://facctconference.org/>.

27 An approach similar to ours and with a wider scope is the work of Takashi Matsumoto and Arisa Ema [41], who proposed a risk chain model for risk reduction in Artificial Intelligence (AI) services, named RCM. By applying RCM in a given risk scenario, it can be proven that a propagation occurs from the technical components of AI systems (data and model) up to the user's understanding, behavior, and usage environment, passing through the service operation management and aspects related to the code of conduct of the service provider as well as the communication with users. The authors consider both data quality and data imbalance as risk factors, whereby they stress the importance of visualizing the relations between risk factors for the purpose of a better planned risk control. While our work is smaller in scope, we think that it can be easily plugged into the RCM framework, due to the fact that we offer a quantitative way to measure imbalance, backed by a structural relation to the ISO/IEC standards on software quality requirements and risk management. Furthermore, it shall be clarified that we did not address data quality as a risk factor given that data quality metrics are well-established in SQuaRE. Nevertheless, we recognize that specific studies would be necessary for selecting the types of measures for data quality that are suitable in the management of ADM system risks.

28 Other approaches which can be related to ours are in the direction of labeling datasets. Two of our previously published studies suggest i) the "Ethically and socially-aware labeling" (EASAL) [42] which aims at developing datasets metadata in order to raise the awareness of the risks of discriminative operations by ADM systems. And secondly, ii) an exploratory analysis of imbalance metrics on two datasets [43], on the basis of which we better specified the theoretical foundations of our approach, and extended the analysis to cover COMPAS. In the context of dataset labeling, the "The Dataset Nutrition Label Project"[11] has been an inspiring work for us. Similar to nutrition labels on food, this initiative aims to identify the "key ingredients" in a dataset such as provenance, populations, and missing data. The label takes the form of an interactive visualization that allows for exploring the previously mentioned aspects. Here, the ultimate goal is to avoid the fact that flawed, incomplete, skewed or problematic data would have a negative impact on automated decision systems, and to drive to the creation of more inclusive algorithms. Notably, our measures could be integrated in this project. Yet another labeling approach is "Datasheets for Datasets" [44]. With respect to other initiatives, this proposal consists of more discursive technical sheets for the purpose of encouraging an increasingly clear and comprehensive communication between users of a dataset and its creators. Eventually, it is worth mentioning the project called "DataTags - Share Sensitive Data with Confidence". [12]The aim of this project is to support researchers who are not legal or technical experts in investigating considerations about proper handling of human subjects' data, and to make informed decisions when collecting, storing, and sharing sensitive data.

E. Relations to European Union policy

29 We extensively reported on how and why bias (imbalance) in data used by ADM systems challenge a founding element of the rule of law of our democratic societies: the principle of non-discrimination [20]. The "Recommendation of the Committee of Ministers to member states on the human rights impacts of algorithmic systems" [45], published by the Council of Europe (CoE) on 8 April 2020, emphasizes the impact of algorithmic systems on human rights and the need for additional normative protections. Although the CoE cannot issue binding laws, it is the main organization for safeguarding human rights in the Europe, and for this reason the recommendation is of particular interest for our purposes. The document defines "high risk" in correspondence with "the use of algorithmic systems in processes or decisions that can produce serious consequences for individuals or in situations where the lack of alternatives prompts a particularly high probability of infringement of human rights, including by introducing or amplifying distributive injustice" (p.5). In these situations, "risk-management processes should detect and prevent the detrimental use of algorithmic systems and their negative impacts" (p.6). The recommended obligations for the states include a continuous review of algorithmic systems throughout their entire lifecycle. In terms of data management, bias in the data as risk factor for systematic discrimination is explicitly cited: "States should carefully assess what human rights and non-discrimination rules may be affected as a result of the quality of data that are being put into and extracted from an algorithmic system, as these often contain bias and may stand in as a proxy for classifiers such as gender, race, religion, political opinion or social origin" (p.7). The document adds that bias and discriminatory outputs should be properly tested since the analysis and modeling phase and even "discontinued if testing or deployment involves the externalization of risks or costs to specific individuals, groups, populations and their environments" (p.8). Precautionary measures should include risk assessment procedures to evaluate potential risks

11 It is a joint initiative of MIT Media Lab and Berkman Klein Center at Harvard University <https://datanutrition.org/>.

12 See <https://techscience.org/a/2015101601/>.

and minimize adverse effects, in cooperation with all relevant stakeholders. Similar obligations are recommended to the private sector.

30 Looking at the Institutions of the European Union (EU), the problem of biased ADM systems is widely recognized, as acknowledged by the words of Margrethe Vestager[13] : "If they're trained on biased data then they can learn to repeat those same biases. Sadly, our societies have such a history of prejudice that you need to work very hard to get that bias out" [46]. The words of M. Vestager should be considered in the context of the ongoing efforts of the EU to redefine the markets rules in response to the rapid technological advancements related to the emergence of automated decision making processes. As a matter of fact, we report the "Resolution on automated decision making processes and consumer protection" [47] which was approved by the EU Parliament on 6 February 2020. The document is relevant because it comes from the highest legislative Institution in the EU and because therein, we find explicit references to the two foundational elements of our proposals. More precisely, the Parliament stresses:

- "the need for a risk-based approach to regulation, in light of the varied nature and complexity of the challenges created by different types and applications of AI and automated decision-making systems" (p. 4);
- "the importance of using only high-quality and unbiased data sets in order to improve the output of algorithmic systems and boost consumer trust and acceptance" (p.11-12).

31 Although the general context of the Resolution is market surveillance, it is still within the ambit of the EU Charter of Fundamental Rights, and in particular Article 38 on consumer protection [48]. It is worth reminding that the European Commission acknowledged the problem of biased ADM since the publication of its communication "Artificial Intelligence for Europe" [49] on 25 April 2018 by stipulating "Whilst AI clearly generates new opportunities, it also poses challenges and risks, for example [...] bias and discrimination" (p.15). Notwithstanding the non-binding value of the document, this communication paved the way to several other policy documents[14]. In the given policy document examples, the term "risk management" recurred often and hitherto it is indicated as the more suitable approach for regulating algorithmic systems[15].

32 This short overview of the most recent efforts on regulating algorithmic systems in Europe, although not exhaustive, defines a further perspective from which our proposal should be derived. In fact, we showed that the risk-based approach is a cornerstone element of the European approach to regulating algorithmic systems, which is currently under redefinition. As a consequence, our proposal can potentially cross this path, whereby balance measures can be suitable risk indicators of propagation (or even amplification) of bias in the input data of ADM systems. In addition, they can be used for certification and labeling purposes, as our notes in the preceding section highlighted.

F. Conclusions: limitations and future work

33 This study faces a problem of wide impact, but it has a well limited boundary of applicability. We take action concerning the problem of systematic discriminations caused by the use of ADM systems, and we focus on a very specific cause, i.e., the imbalance in the data used as input. We propose a metric-based approach in order to evaluate imbalance in a given dataset as a risk factor of discriminatory output of ADM systems. This approach has its foundations on the ISO standards on software quality and risk management. We identify three measures for categorical data, and we run an illustrative example on three columns of the COMPAS dataset, a well-recognized and widely debated case, where imbalance was the main cause of discriminative software output. The example shows that all the indexes detect imbalance, however with different severity and with little variation in the rank of risks. The example, and the study in general, falls short in defining how to effectively manage the risk after the identification. This is a structural, albeit temporary, limitation of the proposal. In fact, in order to derive criteria for action, a systematic investigation is necessary to assess the reliability of the indexes, to identify how their sensibility to imbalance changes in correspondence with different types of data and algorithms used, and to find meaningful thresholds of risks in relation to the context of use and the severity of the impact on individuals. We are working in this direction and we

13 Margrethe Vestager is the Executive Vice President of the European Commission for A Europe Fit for the Digital Age since December 2019 and European Commissioner for Competition since 2014.

14 Including the Coordinated Plan on Artificial Intelligence, the Strategy for Artificial Intelligence, and the very recent (15 December 2020) Digital Services Act draft proposal of the European Commission.

15 Risk management is also a cornerstone element of the AI regulation proposal by the European Commission, which was intentionally left out of the scope of the policy overview because subject to numerous future negotiations.

will be able to elaborate the first guidelines in the following months, thus increasing the internal validity of the present study. Extensive analyses on real systems and replications from third parties will be necessary in order to improve the external validity. We will therefore try to engage researchers in a community effort for testing the measures and to build an open benchmark.

34 We conclude remarking that a much wider number of technical and societal risk factors connected to the deployment of ADM systems exist. For the reader who would like to get an overarching vision, we recommend policy and research reports which investigate the impact of ADM systems, including AI systems, on human rights[16]. For all other readers who stumbled upon this manuscript, we hope that the proposal, despite its current limitations, provided useful insights as a valuable contribution in the common effort of building and regulating algorithmic decision making in a socially sustainable way. More importantly, such is aimed in the direction of protecting individual and collective rights, as well as the promotion of freedoms and the flourishing of our democratic societies.

Acknowledgments

We would like to thank Prof. Marco Ricolfi for his careful review and precious suggestions to the text. We are also grateful to Eleonora Bassi and Giovanni Battista Gallus for inspiring the risk management approach, to Prof. Juan Carlos De Martin for his recommendations on data labeling, and to Prof. Marco Torchiano for his contributions to the data quality aspect of the approach. A special mention also to Elena Beretta and Mariachiara Mecati, who are working on their PhD on translating the concepts described here into practice.

16 Some illustrative but not exhaustive references: [50] [51] [52] [53].

Bibliography

[1] F. Chiusi, S. Fischer, N. Kayser-Bril, and M. Spielkamp, "Automating Society Report 2020," Berlin, Oct. 2020. Accessed: Nov. 10, 2020. [Online]. Available: https://automatingsociety.algorithmwatch.org.

[2] B. Reese, *The fourth age: Smart robots, conscious computers, and the future of humanity*. Simon and Schuster, 2018.

[3] E. Brynjolfsson and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, Reprint edition. New York London: W. W. Norton & Company, 2016.

[4] I. I. Makrygianni and A. P. Markopoulos, "Loan Evaluation Applying Artificial Neural Networks," in *Proceedings of the SouthEast European Design Automation, Computer Engineering, Computer Networks and Social Media Conference*, New York, NY, USA, Sep. 2016, pp. 124–128, doi: 10.1145/2984393.2984407.

[5] Z. Siting, H. Wenxing, Z. Ning, and Y. Fan, "Job recommender systems: A survey," in *2012 7th International Conference on Computer Science Education (ICCSE)*, Jul. 2012, pp. 920–924, doi: 10.1109/ICCSE.2012.6295216.

[6] D. Abu Elyounes, "'Computer Says No!': The Impact of Automation on the Discretionary Power of Public Officers," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3692792, Sep. 2020. [Online]. Available: https://papers.ssrn.com/abstract=3692792.

[7] S. Kanoje, D. Mukhopadhyay, and S. Girase, "User Profiling for University Recommender System Using Automatic Information Retrieval," *Procedia Comput. Sci.*, vol. 78, pp. 5–12, Jan. 2016, doi: 10.1016/j.procs.2016.02.002.

[8] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.

[9] B. Friedman and H. Nissenbaum, "Bias in Computer Systems," *ACM Trans Inf Syst*, vol. 14, no. 3, pp. 330–347, Jul. 1996, doi: 10.1145/230538.230561.

[10] S. Barocas and A. D. Selbst, "Big Data's Disparate Impact," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2477899, 2016. Accessed: Jul. 16, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2477899.

[11] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Reprint edition. New York: Broadway Books, 2017.

[12] G. Ristanoski, W. Liu, and J. Bailey, "Discrimination aware classification for imbalanced datasets," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, New York, NY, USA, Oct. 2013, pp. 1529–1532, doi: 10.1145/2505515.2507836.

[13] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[14] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, Oct. 2002.

[15] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, Nov. 2016, doi: 10.1007/s13748-016-0094-0.

[16] J. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters*, Oct. 10, 2018. https://reut.rs/2Od9fPr (accessed Nov. 10, 2020).

[17] M. De-Arteaga *et al.*, "Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, Jan. 2019, pp. 120–128, doi: 10.1145/3287560.3287572.

[18] M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke, "Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. CSCW, p. 199:1-199:30, Nov. 2019, doi: 10.1145/3359301.

[19] N. Kayser-Bril, "Automated discrimination: Facebook uses gross stereotypes to optimize ad delivery," *AlgorithmWatch*. https://algorithmwatch.org/en/story/automated-discrimination-facebook-google/ (accessed Nov. 10, 2020).

[20] European Union, "EU Charter of Fundamental Rights - Article 21 - Non-discrimination," *European Union Agency for Fundamental Rights*, 2007. https://fra.europa.eu/en/eu-charter/article/21-non-discrimination (accessed Nov. 10, 2020).

[21] T. Jan and E. Dwoskin, "Facebook is sued by HUD for housing discrimination," *The Washington Post*. https://www.washingtonpost.com/business/2019/03/28/hud-charges-facebook-with-housing-discrimination (accessed Nov. 10, 2020).

[22] M. Kay, C. Matuszek, and S. A. Munson, "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2015, pp. 3819–3828, doi: 10.1145/2702123.2702520.

[23] L. Sweeney, "Discrimination in online ad delivery," *Commun. ACM*, vol. 56, no. 5, pp. 44–54, May 2013, doi: 10.1145/2447976.2447990.

[24] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine Bias," *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed Nov. 10, 2020).

[25] Z. Obermeyer and S. Mullainathan, "Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 Million People," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, Jan. 2019, p. 89, doi: 10.1145/3287560.3287593.

[26] A. Kaushal, R. Altman, and C. Langlotz, "Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms," *JAMA*, vol. 324, no. 12, p. 1212, Sep. 2020, doi: 10.1001/jama.2020.12067.

[27] "World stumbling zombie-like into a digital welfare dystopia, warns UN human rights expert," *OHCHR | United Nations Human Rights - Office of the High Commissioner*, Oct. 17, 2019. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25156&LangID=E (accessed Nov. 10, 2020).

[28] V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press, 2018.

[29] International Organization for Standardization, "ISO/IEC 25000:2014 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guide to SQuaRE," *ISO-International Organization for Standardization*, 2014. https://www.iso.org/standard/64764.html (accessed Nov. 10, 2020).

[30] L. Duboc, C. McCord, C. Becker, and S. I. Ahmed, "Critical Requirements Engineering in Practice," *IEEE Softw.*, vol. 37, no. 1, pp. 17–24, Jan. 2020, doi: 10.1109/MS.2019.2944784.

[31] International Organization for Standardization, "ISO 31000:2018 Risk management — Guidelines," *ISO - International Organization for Standardization*, 2018. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/56/65694.html (accessed Nov. 10, 2020).

[32] J. Larson, S. Mattu, L. Kirchner, and J. Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, May 23, 2016. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed Nov. 10, 2020).

[33] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ArXiv190809635 Cs*, Sep. 2019, [Online]. Available: http://arxiv.org/abs/1908.09635.

[34] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*. fairmlbook.org, 2019.

[35] I. Žliobaitė, "Measuring discrimination in algorithmic decision making," *Data Min. Knowl. Discov.*, vol. 31, no. 4, pp. 1060–1089, Jul. 2017, doi: 10.1007/s10618-017-0506-1.

[36] S. A. Friedler, C. Scheidegger, and S. Venkatasubramanian, "On the (im)possibility of fairness," *ArXiv160907236 Cs Stat*, Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.07236.

[37] J. Kleinberg, "Inherent Trade-Offs in Algorithmic Fairness," in *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, New York, NY, USA, Jun. 2018, p. 40, doi: 10.1145/3219617.3219634.

[38] E. Beretta, A. Santangelo, B. Lepri, A. Vetrò, and J. C. De Martin, "The Invisible Power of Fairness. How Machine Learning Shapes Democracy," in *Advances in Artificial Intelligence*, Cham, 2019, pp. 238–250.

[39] E. Pitoura, "Social-minded Measures of Data Quality: Fairness, Diversity, and Lack of Bias," *J. Data Inf. Qual.*, vol. 12, no. 3, p. 12:1-12:8, Jul. 2020, doi: 10.1145/3404193.

[40] B. Hutchinson and M. Mitchell, "50 Years of Test (Un)fairness: Lessons for Machine Learning," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, Jan. 2019, pp. 49–58, doi: 10.1145/3287560.3287600.

[41] T. Matsumoto and A. Ema, "RCModel, a Risk Chain Model for Risk Reduction in AI Services," *ArXiv200703215 Cs*, Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.03215.

[42] E. Beretta, A. Vetrò, B. Lepri, and J. C. De Martin, "Ethical and Socially-Aware Data Labels," in *Information Management and Big Data*, Cham, 2019, pp. 320–327.

[43] M. Mecati, F. E. Cannavò, A. Vetrò, and M. Torchiano, "Identifying Risks in Datasets for Automated Decision–Making," in *Electronic Government*, Cham, 2020, pp. 332–344, doi: 10.1007/978-3-030-57599-1_25.

[44] T. Gebru *et al.*, "Datasheets for Datasets," *ArXiv180309010 Cs*, Mar. 2020, [Online]. Available: http://arxiv.org/abs/1803.09010.

[45] Council of Europe, "Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems." Apr. 08, 2020, [Online]. Available: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016809e1154.

[46] M. Vestager, "Algorithms and democracy - AlgorithmWatch Online Policy Dialogue." Oct. 30, 2020, [Online]. Available: https://ec.europa.eu/commission/commissioners/2019-2024/vestager/announcements/algorithms-and-democracy-algorithmwatch-online-policy-dialogue-30-october-2020_en.

[47] P. De Sutter, "Motion for a Resolution - on automated decision-making processes: ensuring consumer protection and free movement of goods and services." Feb. 06, 2020, [Online]. Available: https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_EN.pdf.

[48] European Union, "EU Charter of Fundamental Rights - Article 38 - Consumer protection," *European Union Agency for Fundamental Rights*, 2007. https://fra.europa.eu/en/eu-charter/article/38-consumer-protection.

[49] European Commission, "Artificial Intelligence for Europe." Apr. 25, 2018, [Online]. Available: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625.

[50] G. Noto La Diega, "Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3188080, May 2018. [Online]. Available: https://papers.ssrn.com/abstract=3188080.

[51] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3518482, Jan. 2020. doi: 10.2139/ssrn.3518482.

[52] A. Mantelero, "AI and Big Data: A blueprint for a human rights, social and ethical impact assessment," *Comput. Law Secur. Rev.*, vol. 34, no. 4, pp. 754–772, Aug. 2018, doi: 10.1016/j.clsr.2018.05.017.

[53] D. Allison-Hope, "Artificial Intelligence: A Rights-Based Blueprint for Business," *BSR*, Aug. 28, 2018.

https://www.bsr.org/en/our-insights/report-view/artificial-intelligence-a-rights-based-blueprint-for-business.

A criterion-based approach to GDPR's explanation requirements for automated individual decision-making

by **Lea Katharina Kumkar and David Roth-Isigkeit***

Abstract: Automation of decision-making processes represents an essential element of the digital transformation. However, automated data processing based on machine learning methods poses increased threats to the fundamental rights of data subjects. One main reason for this is the fact that tracing and explaining the solution path responsible for a certain machine output requires high technical effort. The new European data protection law provides a framework for explanation requirements that apply to users of the new – automated – technologies. This article outlines the current state of discussion on explanation requirements for automated decisions and advocates a restrictive interpretation of the corresponding provisions in the GDPR.

Keywords: GDPR; artificial intelligence; automated individual decision-making; right to explanation

Recommended citation: Lea Katharina Kumkar and David Roth-Isigkeit, A criterion-based approach to GDPR's explanation requirements for automated individual decision-making, 12 (2021) JIPITEC 289 para 1

A. GDPR and the „right to explanation"

1 Methods of automated data processing are on the rise in public and private spheres. In particular, the interest in machine learning applications has exponentially grown in the last years. Main drivers for this are increased availability of large amounts of data and better computing power. Yet, since the European data protection legislator did not consider the developments towards (full) automation in detail, the relationship between data protection law and new possibilities of automated data processing applications remains largely unclear. The resulting conflicts on existence and possible scope of concrete rights and duties are unfortunate not only for the data controller, but also for data subjects, as legal uncertainty could deter data subjects from asserting their rights.

2 One crucial aspect for the protection of data subjects that the General Data Protection Regulation (GDPR) brought up, yet does not resolve, is the question in how far the controller of automated individual decision-making applications has to fulfil certain explanation requirements.[1] Here, the buzzword of the "right to

* The author Kumkar is an assistant professor of civil law, business law and legal aspects of digitalization at Trier University and an affiliated member of the Institute for Digital Law Trier (IRDT). The author Roth-Isigkeit leads a junior research group and the interdisciplinary SOCAI centre for social implications of artificial intelligence, both at Würzburg University. This piece further develops the argument of a previous article of the authors, published in Juristenzeitung 6/2020, pp. 277-286.

1 Under the umbrella term of "explanation requirements", we include here both the duties to inform under Art. 13(2)(f), 14(2)(g) GDPR and the right to information under Art. 15(1)(h) GDPR as well as a possible right to explanation under Art. 22(3) in conjunction with Recital 71 (4) GDPR. For extensive references to the German commentary literature, see L. Kumkar and D. Roth-Isigkeit, 'Erklärungspflichten bei automatisierten Datenverarbeitungen nach der DSGVO'

explanation" has received particular attention in the literature.[2] Underlying this discussion is the so far unanswered question to what extent users of automated decision-making and recommendation systems must be able to disclose the functioning of the system and, if necessary, also the specific decision-making path for individual cases. This question is key especially for advanced automation applications (such as artificial neural networks or complex decision trees). Here, the outcome may lead to so-called black box constellations.[3] In these, the process that leads the machine to dispense a certain output is – if at all – only traceable with a high level of technical effort.

3 The GDPR provides for special explanation requirements of the data controller for certain cases of automated data processing. According to Art. 13(2)(f), 14(2)(g),[4] when collecting personal data, the data controller shall provide the data subject with information on "the existence of automated decision-making pursuant to Art. 22(1) and (4) and – at least in these cases – meaningful information about the logic involved and the scope and intended effects of such processing for the data subject". Art. 15(1)(h) provides for a corresponding right of access. Furthermore, Recital 71(4) mentions the "explanation of the decision reached" as part of the "suitable safeguards" for automated processing. This leads parts of the literature to accepting the existence of a case-by-case requirement to provide detailed and specific explanations for processing operations.[5]

4 Such an obligation would entail quite dramatic consequences for the way in which data controllers will be able to use automated processing in the future. It would imply considerable financial risks for data controllers, in particular due to the strict penalty provisions in Art. 83(5)(b). In addition, disclosure of the decision path may be not possible due to technical difficulties or (legitimate) interests of the processor in preserving business and trade secrets.[6]

5 Referring to the current discussion on explanation requirements, this paper advocates a restrictive interpretation of the relevant provisions of the GDPR. In contrast to the procedural approaches suggested by large parts of the literature, we propose a criterion-based approach, which requires the *ex ante* disclosure of possible decision criteria, but not the *ex post* disclosure of the detailed process of decision-making and weighing in the individual case. For a better understanding of the relevant disputes, we first outline the general principles on automated individual decision-making pursuant to Art. 22 (B.). This is followed by a description of potential points of reference to derive a "right to explanation" for automated decisions (C.). Building on considerations on function and technical limits (D.), we discuss potential implications of the existence of a "right to explanation" (E.).

(2020) 75 (6) Juristenzeitung 277-286.

2 See, for an introduction to this debate with further references B. Casey et al., 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Tech LJ 143, 189.

3 For the social problem of black box constellations, see e.g. F. Pasquale, 'The Black Box Society – The Secret Algorithms that Control Money and Information' (2015), Harvard University Press, 1-18.

4 The following article denominations refer to the General Data Protection Regulation (GDPR) if not otherwise stated.

5 Notably B. Casey et al., 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Tech LJ 143; M. Brkan, 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond' (2019) 27 Int J Law Info Tech 91; M. Brkan and G. Bonnet, Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas (2020) 11 European Journal of Risk Regulation 18; M. Kaminski, The right to explanation, explained (2019) 34 Berkeley Tech. L.J. 189; T. Kim and B. Routledge, Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach (2021) Business Ethics Quarterly, Fist View 1. The term presumably derives from an initially unpublished conference paper by Goodman/Flaxman, EU Regulations on Algorithmic Decision Making and "a Right to an Explanation," available at https://arxiv.org/pdf/1606.08813.pdf (last accessed June 29, 2021). The paper focused mainly on technical issues and was primarily intended to draw attention to the difficulties of explaining complex algorithmic processes.

6 In this sense, the "qualified transparency" called for by e.g. F. Pasquale, 'The Black Box Society – The Secret Algorithms that Control Money and Information' (2015), Harvard University Press, 140 ff. should also be understood as a balancing between different interest groups. See further on the challenges of trade secret protection in the data-driven economy, A. Wiebe and N. Schur, 'Protection of trade secrets in a data-driven, networked environment – Is the update already out-dated?' (2019) 14 (10) Journal of Intellectual Property Law & Practice 814–821.

B. The legal framework for automated individual decision-making pursuant to Art. 22 GDPR

6 Art. 22 provides a framework for automated individual decision-making and profiling measures. A similar provision was already provided for in Art. 15 Data Protection Directive (DPD).[7] Pursuant to Art. 22(1), the data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. Paras. 2 and 3 provide for exceptions to this principle, in particular in situations where the data subject has given consent. Para. 4 sets out specific requirements for particularly sensitive data (cf. Art. 9(1)).

I. Content and Meaning

7 Art. 22 does not establish a separate basis of permission for the processing of personal data, but establishes an additional prerequisite that must be observed during processing. Despite its systematic localization among the data subjects' rights, the provision – at least indirectly – has the character of a prohibition. Processing that does not comply with the requirements of Art. 22 is prohibited even without explicit statement by the data subject.

8 An even more comprehensive prohibition is provided for in Art. 22(4) for special categories of personal data, which according to Art. 9(1) include in particular data revealing racial or ethnic origin, political opinions, religious beliefs, trade union membership, genetic predispositions or health status and sexual orientation. Deviating from other requirements of Art. 22, the inclusion of data pursuant to Art. 9(1) in automated decisions is generally only permissible if the data subject has expressly consented (Art. 22(4) in conjunction with Art. 9(2)(a)), or if the processing is both necessary due to substantial public interest and is proportionate (Art. 22(4) in conjunction with Art. 9(2)(g)).

II. Profiling and automated decision-making

9 The wording of the official title of Art. 22 ("automated individual decision-making, including profiling") is misleading. "Profiling" is not a specific application of automated decision-making, but a special data processing operation.[8] According to Art. 4 (4) profiling includes "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements".

10 "Automated decisions" within the meaning of Art. 22 build on a data processing operation by linking automated processing with consequences for the data subject. This understanding is also supported by the fact that Art. 22(1) mentions "profiling" as an example following the term "automated processing" – and not the subsequent "decision". Profiling is therefore only one possible manifestation of the processing covered by Art. 22.[9] Regarding the concrete scope of the prohibition contained in Art. 22, there is agreement that neither the profiling process nor the automated processing as such is covered, but only the decision based *solely* on this automated processing – at least if and to the extent that this has legal or similarly significant adverse effects.

11 The GDPR does not define the term "automated individual decision-making." Yet, it can be assumed that it makes a distinction between "decisions" in a narrow sense as opposed to "processing". Not every type of data processing automatically qualifies as a decision within the meaning of Art. 22(1). Rather, a minimum degree of complexity must be inherent, since otherwise even simple if-then connections such as dispensing money at an ATM would fall under the regulation.[10] That such is not intended, also becomes clear from the comparison with profiling, as mentioned in Art. 22(1). Profiling aims

7 Directive 95/46/EC. For the development of the requirements of Art. 22 GDPR from Art. 15 DPD, see also I. Mendoza and L. Bygrave, in: T.E. Sydodinou et al. (eds.), *EU Internet Law* (2017), Springer International Publishing, 77 ff.

8 Expressing its favor for the independence of the two terms: Art. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 Rev.01 (6 February 2018) 8: "Automated decisions can be made with or without profiling; profiling can take place without making automated decisions."

9 Differently I. Mendoza and L. Bygrave, in: T.E. Sydodinou et al. (eds.), *EU Internet Law* (2017), Springer International Publishing, 77 ff., 90 f., identifying a drafting error and stating the provision should be understood as referring only to profiling.

10 S. Schulz in P. Gola (ed.), *DSGVO* (2018), C.H. Beck, Art. 22 para 21; B. Buchner in J. Kühling and B. Buchner (eds.), *DSGVO* (2020), C.H. Beck, Art. 22 para 18.

at the evaluation of personality traits and implies a certain materiality of the processing.[11]

12 Consequently, a decision only exists in the case of an act that selects from (at least) two variants and has a final impact on the external world, which can be attributed to a (natural or legal) person. According to *Bygrave*, a decision means that a "particular attitude or stance is taken towards a person and this attitude/stance has a degree of binding effect in the sense that it must— or, at the very least, is likely to— be acted upon."[12]

13 According to the wording of Art. 22(1), the decision must be based on "solely automated" processing, which means that the decision is taken "without any human intervention", as also clarified in Recital 71.[13] This means that Art. 22 does not apply if the knowledge gained in the course of automated processing is only used as basis or for the preparation of a decision to be taken by a natural person. Here, the natural person involved has to apply a margin of discretion.

14 If an actual review of the content takes place by a human employee with the corresponding decision-making powers to change the processing result, the automated data processing only becomes the (working) basis for the decision of a natural person, and is thus no longer "solely" automated.[14] The situation is different if the result found by the machine is merely accepted by a human administrator without any examination of the content. Also, random checks or interventions in neural networks to improve decisions, such as in supervised learning, do not constitute sufficient human intervention. Since the content of the decision remains unchanged, the situation merely resembles a "maintenance" of the system.[15]

III. Legal effect

15 Art. 22 covers only automated decisions with legal effects or the ones that "similarly significantly affect" the data subject. While the GDPR does not explicitly specify when a decision is to be considered as having "legal effects," it can be assumed that this implies that the legal status of the data subject is altered in any way.[16] Assessing this in more detail, however, much remains unclear. For example, the question arises as to whether this includes only adverse decisions so that (purely) favorable legal consequences remain outside the scope of the provision.[17] A general definition of "similarly significantly affected" has neither yet emerged. However, there is wide agreement that the threshold of mere nuisance must be exceeded.[18]

11 An interim definition of profiling is contained in Recital 24 (2) GDPR where it reads "In order to determine whether a processing activity can be considered to monitor the behaviour of data subjects, it should be ascertained whether natural persons are tracked on the internet including potential subsequent use of personal data processing techniques which consist of profiling a natural person, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes."

12 For a detailed discussion, see L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 532.

13 However, the occasionally expressed demand that this implies the absence of human intervention from the collection of the data to the issuing of the decision must be rejected. Here, only the decision-making process in the narrower sense is decisive. The dangers emanating from automated decisions are not less serious for the person concerned if the preceding data collection is (still) manual or only partially automated. Only this kind of understanding does satisfy the comprehensive protective purpose of Art. 22 GDPR.

14 See L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 532-533.

15 T. Hoeren and M. Niehoff, 'KI und Datenschutz – Begründungserfordernisse automatisierter Entscheidungen' (2018) 9 *Rechtswissenschaft* 47, 53.

16 L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 532.

17 Against this, it could be argued that the wording of Art. 22(1) GDPR does not contain a corresponding restriction. On the other hand, the protective purpose of Art. 22 GDPR contradicts the inclusion of favorable legal consequences, as the data subject does not need to be protected from (purely) favorable decisions.

18 Art. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 Rev.01 (6 February 2018) 21: "For data processing to significantly affect someone the effects of the processing must be sufficiently great or important to be worthy of attention." Another subject of discussion is the extent a legal effect can be assumed for automated decisions in contractual relationships. While the legal effects in the case of termination and acceptance of contractual offers are predominantly affirmed, opinions are divided in the case of refusal to conclude a contract (under certain conditions). It is convincingly argued against the existence of a legal effect within the meaning of Art. 22(1) GDPR in the cases of a refusal to conclude a contract or a refusal to accept certain conditions that legal effects do not "unfold" as intended but, on the contrary, do not occur at all.

IV. Exceptions and suitable measures to safeguard

16 Art. 22(2) provides for three exceptions to the prohibition of Art. 22(1). The norm does not apply if a decision is necessary for entering into, or performance of, contract between the data subject and the data controller (lit. a), if a decision is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests (lit. b) or if the decision is based on the data subject's explicit consent (lit. c).[19]

17 If the data controller intends to base the data processing on one of the exceptions under Art. 22(2)(a) or 22(2)(c), the controller shall, according to Art. 22(3), implement suitable measures to ensure that the data subject is provided with adequate safeguards. This includes at least that the data subject has (1) the right to obtain personal intervention on the part of the controller, (2) the opportunity to put forward his or her own point of view, and (3) the right to contest the decision (so-called minimum safeguards). As can be seen from the wording of Art. 22(3), the aforementioned list is not exhaustive, i.e. the "suitable measures to safeguard" to be taken by the controller may also require further measures.

C. Possible starting points for explanation requirements

18 Some argue that Art. 22(3), in conjunction with Recital 71(4), provides a "right to explanation" for the data subject, which is intended to apply comprehensively and retrospectively to the entire individual decision-making process (I.) Other authors assume that the data controller's duty to explain can be derived solely from the general information rights of Arts. 13 to 15 (II.). While the two approaches differ significantly in terms of scope and timing of the explanation requirement, they both largely leave open the required content and depth of the explanation (III.).

19 According to Art. 4 No. 11 GDPR consent means "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."

I. A „Right to explanation" pursuant to Art. 22 (3) in combination with Recital 71 of the GDPR

19 In the context of the rights of data subjects pursuant to Art. 22(3), it is being discussed whether a "right to explanation" against the data controller in the form of a case-by-case requirement to justify is being established.[20] Indications for the existence of such a right or – correspondingly – an equivalent requirement to explain on the part of the controller are not to be found directly in Art. 22. The provision only mentions the right to intervention of a person, explanation of one's own position and contestation of the decision. Yet, an indication could be found in Recital 71, which states:

> *"However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, [...], or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision."*

20 In this context, the wording of Recital 71(4) suggests that the controller is obliged to justify the specific decision *ex post* and on a case-by-case basis, when it refers to "obtaining an explanation of the decision reached after such assessment". The explanation should therefore not only include the abstract functionality of the device used,[21] but also a justification of the concrete decision in the individual case.[22]

21 However, the question arises in which cases the requirement could be applied at all. This appears problematic because the explanation requirement is *only* contained in the recitals and does not find a counterpart in the wording of Art. 22(3). In particular, referring to the lack of binding effect of the recitals, *Wachter et al.* took the view that a right to explanation is currently not legally imposed by

20 See references in note 5.

21 This is the case with the explanation requirements pursuant to Art. 12 to 15 GDPR, cf. below. C. II.

22 S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 81.

Art. 22.[23] They see this supported by the fact that the requirement to explain specific individual decisions was omitted from the normative text of Art. 22 during the deliberations on the drafting of the GDPR. This would suggest that the legislator did not intend to make such a right binding. Even though *Wachter et al.* consider it possible that case law will establish a right to explanation in the future as part of the interpretation of the "adequate safeguards"[24] they do not see it currently imposed by the GDPR.[25]

1. A „Right to explanation" as minimum guarantee?

22 This view is convincing – at least against the backdrop of the current practice of the European Court of Justice (ECJ) on the status of the recitals in the interpretation of substantive guarantees. In European legal acts, recitals are placed before the determining norms as an anticipated statement of reasons (cf. the usual introductory wording "considering the following reasons"). The ECJ has consistently held that "[...] the preamble to a Community act has no binding legal force and cannot be relied on either as a ground for derogating from the actual provisions of the act in question or for interpreting those provisions in a manner clearly contrary to their wording".[26] Starting point and limitation of a teleological interpretation based on the recitals is thus always the wording of the norm in question. In a decision from 1989, the ECJ clarified – albeit with regard to the recitals of a regulation – that a recital "may cast light on the interpretation to be given to a legal rule, it cannot in itself constitute such a rule."[27]

23 For a "right to explanation", the wording of Art. 22(3) is the decisive limit. The enumeration of the "minimum safeguards" to be guaranteed by the controller in Art. 22(3) is exhaustive. Here, if the European legislator had wanted to make the minimum safeguards open-ended, it would have expressed this – following good custom – by adding words such as "for instance", "for example" or "in particular". A non-exclusive understanding would ultimately also undermine the goal of creating binding minimum standards for data controllers and data subjects (cf. Recital 10(1)). There is thus no room for a broadening teleological interpretation. According to this understanding, Art. 22(3), in conjunction with Recital 71(4), does not generally provide for a "right to explanation" in the form of a minimum guarantee.

2. A „Right to explanation" as suitable measure?

24 However, the fact that a "right to explanation" is not mentioned in the enacting terms of Art. 22(3) does not necessarily suggest that a requirement to explain could not exist in any conceivable case.[28] This is because the rights of the data subject are not exhausted by the (minimum) rights explicitly mentioned in Art. 22(3) – as the statutory use of the word "at least" suggests. Rather, according to Art. 22, the data controller must take all reasonable measures necessary to safeguard the rights and freedoms as well as the legitimate interests of the data subject. This does not exclude, at least not systematically, that the "reasonable measures" could in some cases also include an *ex post* and case-by-case explanation of the decision.

25 Art. 22 suggests that the legislator did not intend to include the explanation requirements among the measures to be taken in *every case* to protect the data subject. Rather, they belong to the group of "suitable measures" that go beyond the minimum guarantees. This means whether the explanation requirements under Art. 22(3) apply in an individual case depends on the broader question in which cases the explanation is considered necessary to protect the rights and freedoms as well as the legitimate interests of the data subject. This depends very significantly, on which function can be attributed to the "right to explanation" in the overall structure of the legal protection of data subjects.[29]

23 Ibid., 79 ff.

24 Ibid., 81.

25 Ibid., 80.

26 Case C-345/13 *Karen Millen Fashions* [2014] ECLI:EU:C:2014:2013 para 31; see also Case C-136/04, *Deutsches Milch-Kontor* [2005] EU:C:2005:716 para 32.

27 Case C-215/88, *Casa Fleischhandels-GmbH v Bundesanstalt für landwirtschaftliche Marktordnung* [1989] ECLI:EU:C:1989:331 para 31; see also T. Klimas and J. Vaiciukaite, 'The Law of Recitals In European Community Legislation' (2008) 15 ILSA Journal of International & Comparative Law 92 f. and H. Rösler in J. Basedow, K. Hopt and R. Zimmermann (eds.) *Max Planck Encyclopedia of European Private Law* (2012), Oxford University Press, 979 ff.

28 Likewise M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 204.

29 *Infra*, D. I.

II. Information rights according to Arts. 13 to 15 GDPR

26 While the "right to explanation" according to Art. 22(3) in conjunction with Recital 71(4) is in dogmatically uncertain territory, the mandatory nature of the information rights in Arts. 13 to 15 is (at least) *ipso iure* beyond question. According to Art. 13(2)(f) and 14(2)(g), the data controller shall provide the data subject with information on "the existence of automated decision-making, including profiling, referred to in Art. 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.". Art. 15(1)(h) grants the data subject a corresponding right of access against the data controller. On closer examination, however, several details remain unclear.

1. Scope of the Provisions

27 Since the material preconditions of Art. 13(2)(f), 14(2)(g) and 15(1)(h) explicitly refer to Art. 22(1) and (4), the question arises whether the requirements apply solely in the (narrow) cases of automated individual decision-making which are also covered by the preconditions of Art. 22(1); i.e. whether a decision with legal effects or a similarly significant impairment is always required. The wording "at least in these cases" in Art. 13(2)(f), 14(2)(g) and 15(1)(h) could suggest the provisions cover (automated) processing below the threshold of "decision". However, it remains completely undefined according to which criteria such further cases are to be determined. Against the backdrop of the strict penalty for the information duties (Art. 83(5) (b)), it seems unconvincing to extend the information duties to other processing operations.

28 Rather, it must be assumed that the wording was simply copied from the preceding provision in Art. 12 lit. a 2nd Alt. DPD. Yet, while the wording in the DPD was intended to give the Member States room for manoeuvre in implementation, it cannot fulfil this function in the (directly applicable) GDPR. This means with regard to the rights and obligations under Art. 12 to 22, Member States only keep the competence to restrict, but not the right to extend. [30] The fact that only Art. 22 (1) and (4) are explicitly referred to (but not Art. 22(3)) also indicates that the requirement to provide information does not extend to the minimum guarantees contained in Art. 22(3).[31]

2. Relevant timing

29 The relevant timing of the information differs between the various provisions. Pursuant to Art. 13(2)(f), the information must be provided "at the time when personal data are obtained." In light of Art. 14(2)(g), the information must be provided pursuant to Art. 14(3), namely "within a reasonable period after obtaining the personal data, but at the latest within one month" (lit. a). However, if use of the data for communication with the data subject (lit. b) or disclosure to another recipient (lit. c) is intended beforehand, this triggers an immediate obligation to provide information from the time of first communication or disclosure. The right to information pursuant to Art. 15(1)(h), on the other hand, is not limited to the moment of data collection, but can also be exercised after the conclusion of the data processing or the automated decision resulting therefrom.

3. Implications for the content of information requirements

30 In the case of Art. 13(2)(f), relevant inference on the content of the information to be provided can be drawn from the time at which the information is provided. Since the information must be provided at the time of the data collection, i.e. before the actual processing operation, the obligation to provide information in Art. 13(2)(f) cannot be directed at a (subsequent) explanation of the processing operation, but is exhausted in the mere announcement of the forthcoming automated decision. [32] It can be further concluded from this that the declaration of the data controller in the sense of the logic of the norm must also only take into account the general functioning of the decision-making and not the (not yet

30 M. Martini, 'Blackbox Algorithmus' (2019), Springer, 182 f. In cases that cannot be subsumed under Art. 22 (1) GDPR, information can nevertheless be provided on a voluntary basis.

31 Ibid., 187, arguing in favor of an addition in this regard. However, some of the German commentaries derive a corresponding obligation to provide an explanation directly from Art. 22(3) GDPR, see S. Schulz, in: P. Gola (ed.), *DSGVO* (2018), C.H. Beck, Art. 22 para 41 f.

32 S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 82. Cf. M. Martini, 'Blackbox Algorithmus' (2019), Springer, 191.

determined) specific circumstances of the (still imminent) individual decision.[33]

31 For Art. 14(2)(g) and 15(1)(h) the relevant points in time do not allow for such a conclusion on the content of the information requirement. The information can also be provided after processing with a concrete processing result already available. It could be argued that although the wording of the provisions of Art. 14(2)(g) and Art. 15(1)(h) is the same as in the case of Art. 13(2)(f), the "meaningful information" covered varies depending on the point in time at which the information is provided. Thus, especially the right of access in Art. 15(1)(h) could potentially cover information on the specific circumstances of the individual decision.[34]

32 However, upon closer examination this is not convincing. The information requirements pursuant to Art. 14(2)(g) and 15(1)(h) cannot be attributed to a broader content than the obligations in Art. 13(2)(f). This is not only supported by the fact that the wording of the norms is identical, but also by the circumstance that according to Art. 14(2)(g) and 15(1)(h), only information on the "intended" effects must be provided, which suggests a future orientation. [35] This interpretation corresponds to the assumptions made in the guidelines of the Art. 29 Data Protection Working Party.[36] In summary, it can be stated that Art. 13 to 15 – unlike the "right to explanation" derived from Art. 22(3) – require a prior declaration by the data controller, which is directed at the abstract functionality of the data processing.

III. The search for a common ground in explanation requirements

33 In light of the above, there are differences between the two approaches on explanation requirements for automated decisions. Art. 22(3) in combination with Recital 71(4) intends a *subsequent* explanation of the *specific* decision. Art. 13(2)(f), 14(2)(f), 15(1) (h), on the other hand, require a *prior* explanation of the functionality of the data processing and thus provide for *abstract* information rights. From this perspective, there is no connection between the different explanation requirements.

34 For both, the data subject and controller, such a conclusion seems unrealistic from the perspective of practical data protection. Irrespective of their temporal validity and scope, both requirements concern a common basic question: What level of explanation must the controller of automated data processing (be able to) provide? What information about the data processing must (be able to) be shared with the data subjects? The GDPR is silent on the concrete content of these requirements – and yet endows them with the threat of a hefty fine (see Art. 83(5)). It is therefore crucial to develop a pragmatic standard that both users and data subjects can use as a guideline when providing or requesting explanations for automated data processing and that at the same time fulfils the legal demands of both, Art. 22(3) in conjunction with Recital 71(4) as well as Arts. 13(2)f, 14(2) (f) and 15(1)(h).

D. A joint answer to the required explanation depth for automated decision-making

35 The proposal presented here attempts to combine these requirements in order to develop a joint answer to the question of the necessary depth of explanation based on the previous considerations. With respect to the basic functions of the explanation requirements (I.) and the technical limitations of the traceability of automated decisions (II.), it seems reasonable to limit explanation requirements to outlining the decision criteria that form the basis of the (planned) automated processing (III.).

I. A functional view on explanation requirements

36 Looking at the explanation in connection with automated decisions from a functional perspective, similar purposes can be identified in the cases

33 S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 78 ff., who distinguish between *system functionality* (ex ante and ex post) and *specific decision* (ex post).

34 This is argued in particular in the German commentary literature, cf. M. Bäcker in Kühling/Buchner DS-GVO BDSG (2020), Art. 15 Rn. 27 with further references.

35 Cf. S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76, 83. M. Martini, 'Blackbox Algorithmus' (2019), Springer, 192.

36 Art. 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 Rev.01 (6 February 2018) 26.

of Art. 22(3) in conjunction with Recital 71(4) as well as Art. 13(2)(f), 14(2)(f), and 15(1)(h).

1. „Legibility" of the decision

37 First of all, the explanation enables the data subjects to understand the basis of the (automated) decision. This can be derived from the requirement in Art. 12(1) on how the information should be provided, i.e. "in a concise, transparent, intelligible and easily accessible form, using clear and plain language". Even automated decisions within the limits of Art. 22, which are permissible in principle, entail an increased risk of non-transparency for the data subject. Since the data subject usually has no knowledge on how the upcoming decision will be taken, it is difficult for him or her to assess in advance what risks to his or her data will be associated with the planned processing.[37] Without knowledge of the decision-making process, it will be impossible to control whether a decision may be linked to inadmissible criteria, such as a feature of Art. 9 or the non-discrimination criteria of Art. 21 of the European Charter of Fundamental Rights.

38 The wording of Art. 13(2)(f), 14(2)(f), 15(1)(h) ("meaningful information about the logic involved") is indicative. "Meaning" takes the perspective of the understanding data subject, who should be enabled to draw conclusions about the essential decisional factors from the transmitted information.[38] These contexts of meaning must – which does not seem obvious from the formulation – be available in a form that is comprehensible to humans.[39]

39 Following this line of argumentation, *Martini* adopts a narrow understanding of the explanation requirement:[40] According to him, "explanation" means describing the content of the decision in more detail, but not disclosing the reasons for the decision to its full extent. The phrase "an explanation of the decision reached" refers grammatically to the "individual presentation of the case" of the person concerned. This means in consequence that the right to an explanation only exists to the extent that it is necessary in order to explain to an individual how his or her own point of view has been taken into account in the decision and why the result of the assessment has turned in that specific way.

40 Such an understanding of "explanation" requires outlining the essential basis for the decision in a form that is comprehensible to humans, and thus a kind of "legibility".[41] In this way, the information contributes to the data subject's autonomy that had been endangered through the opacity of processing. With *Bygrave*, one could understand this as a requirement of a concept of cognitive sovereignty pervasive in data protection law, "a human being's ability and entitlement to comprehend with a reasonable degree of accuracy their environs and their place therein."[42]

41 Neither disclosure of the raw data nor the technical aspects of the decision-making mechanism would meet this requirement, because the person concerned usually does not have the technical means to put it into a comprehensible form. Examining the explanation requirement from the perspective of the data subject, it soon becomes clear that the literature opinion that asks for a complete breakdown of the decision program or disclosure of the algorithm to fulfill this requirement misses this aspect.[43] From a functional perspective, only those considerations can be covered by the explanation requirement that contribute to a (human) "legibility" of the automated decision.

2. Due process

42 Further, explanation requirements stand in connection with the right to challenge the (automated) decision, which is highlighted as a "minimum guarantee"[44] in Art. 22(3). In this context, the scope of the explanation requirement can be clarified in a similar manner based on the required information

37 See M. Martini, 'Blackbox Algorithmus' (2019), Springer, 176. Likewise M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 211: "They need to be given enough information to be able to understand what they are agreeing to [...]."

38 A. Selbst and J. Powles, 'Meaningful information and the right to explanation' (2017) 7 International Data Privacy Law 233, 239.

39 Ibid., 240.

40 M. Martini, 'Blackbox Algorithmus' (2019), Springer, 191.

41 See also G. Malgieri and G. Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243 ff.

42 L. Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions' in Ienca et al. (eds.), Cambridge Handbook of Life Sciences, Information Technology and Human Rights (forthcoming).

43 See e.g. M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 189 ff.

44 Supra, C. I. 1.

for the data subject to effectively make use of this right to challenge.[45] The data subject "must be able to recognize on the basis of this information whether incorrect data has found its way into the procedure or whether the individual particularities of his or her situation have not been sufficiently taken into account."[46] The key aspect here is that the information may be used to raise substantiated objections and to trigger a human review in a second step.[47]

43 The enumeration of rights of the data subject in Art. 22(3) provides further indication on the required depth of explanation. The wording implies a need for suitable measures, "to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."

44 It is then key whether one understands these various aspects as a unit or as separate rights.[48] While the presentation as a list suggests that they are separate, this interpretation is not very plausible, as it would lead to a kind of circle of decisions and challenges. If the data subject's rights under Art. 22(3) could not be advanced uniformly, the data subject would be confronted with a renewed automated decision on the same factual basis after the challenge, against which the challenge would again be admissible.[49] However, it is precisely here that automated decision-making systems are not (yet) capable of automatic self-correction. If the factual basis remains unchanged, the decision will remain unaltered after repeated runs of the system. The "right to challenge" in the common reading of the rights from Art. 22(3) thus only becomes plausible if it demands a human decision *replacing* the automated decision.[50]

45 This argument in turn allows drawing conclusions on the required depth of explanation. In the context of a human re-decision, a subsequent explanation of the original decision path would be superfluous, as the new decision would be taken uninfluenced by the machine output result.[51] For the effective legal protection of the respective person, an explanation of the algorithmic decision-making mechanism is neither necessary nor expedient.[52]

II. Technical limitations regarding the ability to explain automated decision-making

46 Further indications of limited explanation requirements are the technical limitations regarding the ability to explain automated data processing. Particularly in advanced applications of machine learning, the complexity of the system means that it is only possible with the greatest technical difficulty to find a form of explanation that is understandable for humans.

47 Solutions for this problem are discussed under the umbrella topic of "explainable AI".[53] Contemporary advances allow, for example in image recognition by machine intelligence, revealing certain patterns of decision-making, such as determining which pixel patterns were observed for the recognition of

45 S. Schulz, in: P. Gola (ed.), *DSGVO* (2018), C.H. Beck, Art. 22 para 42. For a recent application highlighting the goal of public accountability, Talia B. Gillis and Josh Simons, 'Explanation < Justification: GDPR and the Perils of Privacy' (2019) 2 J.L & Innovation 72 (80).

46 Author's translation: P. Scholz, in: Simitis/Hornung/Spiecker gen. Döhmann (eds.), *Datenschutzrecht* (2019), Nomos, Art. 22 para 57.

47 P. Scholz, in: Simitis/Hornung/Spiecker gen. Döhmann (eds.), *Datenschutzrecht* (2019), Nomos, Art. 22 para 57.

48 S. Wachter, B. Mittelstadt and C. Russell, 'Counterfactual Explanations without opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 842, 873.

49 S. Wachter, B. Mittelstadt and C. Russell, 'Counterfactual Explanations without opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 842, 873.

50 L. Bygrave in Kuner et al (eds.), *GDPR* (2019), Oxford University Press, Art. 22, 538.

51 S. Wachter, B. Mittelstadt and C. Russell, 'Counterfactual Explanations without opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law and Technology 842, 874.

52 See also L. Edwards and M. Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law and Technology Review 18, 81, who argue that, with respect to the subjective legal protection of data subjects, the traceability of decisions is not the decisive criterion.

53 On the current state of the legal discussion B. Waltl and R. Vogl, 'Explainable Artificial Intelligence–the New Frontier in Legal Informatics' (2018) Jusletter IT (22 February 2018); P. Hackerl et al, 'Explainable AI under contract and tort law: legal incentives and technical challenges' (2020) 28 Artificial Intelligence and Law 415–439; see further A. Deeks 'The judicial Demand for explainable Artificial Intelligence' (2019) 119 Columbia Law Review 1829–1850.

certain shapes.[54] In the case of complex deliberation processes, on the other hand, as would be required in the applications discussed here, it is largely unclear to what extent the output result of the work process made visible would be comprehensible to humans. In general, it can be said that we are currently in a state where greater performance of a program corresponds with a reduced comprehensibility of its internal processes. It is therefore not necessarily to be assumed that technical progress will produce explainable data processing, but the opposite of complete opacity is also conceivable, if not likely.

48 The legal value of this technical limitation of the actual comprehensibility of automated data processing is admittedly rather low. It is only suitable to a very limited extent to determine explanation requirements, otherwise one would also fall into a naturalistic fallacy, deriving norms from facts. This principle is also reflected in the GDPR. For example, Recital 58(3) sets particularly high requirements on transparency for situations of high complexity.

49 Nevertheless, technical feasibility can allow conclusions on what the European legislator intended in the context of the explanation requirements. Here it is unlikely – though not impossible – that the GDPR establishes a legal standard that is not technically feasible. In this respect, the above explanations are helpful supplementary information for the interpretation of the standard, which, just like the functional analysis, point to a limited explanation requirement.

III. Consequences for the depth and direction of the explanation

50 Based on these considerations we propose a standard for the depth and direction of the explanation that fulfils two criteria. On the one hand, it ensures the "legibility" of the decision for the data subject and the ability to challenge it. On the other hand, it is technically feasible for the controller. Both conditions indicate that the explanation requirements should be understood in such a limited way that they require an outline of the decision criteria in a form accessible to humans.[55]

51 Decision criteria can help render the mode of operation of a program transparent and traceable. Admittedly, in such a model not all discrimination risks associated with automated data processing may be avoided. Although all decisions can be traced back to the direct or indirect interaction of decision criteria, an all-encompassing control of the decision program in a way that it could be traced how exactly the interaction of individual criteria led to a certain output result is neither technically nor legally feasible.

52 Furthermore, the imposition of a comprehensive requirement to explain the functionality of the data processing and the concrete outcome of the decision would also be questionable from a legal policy perspective. It would create an appearance of controllability of the internal mechanisms of automated data processing and shift burdens of justification onto data subjects.

53 In practical terms, the criterion-based approach advocated here means that the data controller must disclose the (real-world) criteria that the decision program takes into account for its calculations. On the one hand, this imposes a transformational task on the controller to translate the criteria from the digitized form into a linguistic representation. On the other hand, disclosure of the program's concrete mode of operation is not required. Regarding the question of how specific the disclosure of these decision criteria must be, the sanction practice of the data protection supervisory authorities is likely to become a decisive factor for the further development of the law.

E. Implications

54 The view adopted here understands the explanation requirements as a necessary starting point for a human review. If one considers the requirement to present a catalogue of criteria as the basis for this, the term "explanation" (derived from Recital 71 (4)) is misleading, since this represents only the starting point for the intervention directed at a human decision. It is therefore reasonable to assume that the subjective legal asset discussed under the term "right to explanation" actually turns out to be a preparatory *right to justification*.[56]

55 This understanding entails both opportunities and risks.[57] On the one hand, it allows the law to reflect

54 W. Samek, T. Wiegand and K.-R. Müller, 'Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models' (2018) 1 ITU Journal: ICT Discoveries 39 ff.

55 Differently M. Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkeley Tech LJ 189, 209 ff., referring to the high value placed on transparency in the GDPR.

56 See, for the conceptual background, R. Forst, 'The Right to Justification' (2007), transl., Columbia University Press.

57 Under certain circumstances, the considerations made here could also gain significance beyond the scope of the GDPR through the so-called Brussels effect. On this point, see B. Casey et al., 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic

the general opacity of intelligent decision-making systems in order to provide for a practical way of dealing with the limited explicability. It thus offers a possibility for the social integration of technical progress. On the other hand, law thus recognizes the "autonomy" of intelligent decision-making systems to the extent that the procedural and deterministic explanation of decision-making is replaced by the – comparable to legal protection against human decisions – subsequent substantive legality test. Law thus finds its mode of dealing with the non-explicability of machine decisions in converting its procedures to the model of justification adapted to human decisions. Time will show whether this approach will also prove sustainable in practical terms.

Audits in Enterprise' (2019) 34 Berkeley Tech LJ 143, 185.

The Case of Diem
A Distributed Ledger Technology-based Alternative Financial Infrastructure Built by a Centralised Multisided Platform

by **Golnaz A. Jafari and Malte-C. Gruber***

Abstract: In pursuing its declared mission "to enable a simple global currency and financial infrastructure with a safe, secure and compliant payment system that empowers billions of people," Diem has encountered apparent resistance from various social fields and politics. On the one hand, many critics recognise dangers to state currency sovereignty and the stability of the financial system; on the other hand, they fear negative developments regarding money laundering and the financing of terrorism. In addition, there are considerable concerns about an ever deeper erosion of privacy, consumer and data protection, which reaches a new dimension by linking such world currencies with already existing social networks governed and controlled by private entities. Under these circumstances, the chance of success of the Diem project clearly depends on the extent to which the aforementioned concerns can be dispelled and whether public trust can be established. Together with an overview of the developments of the Diem project since the inception of the underlying idea, the authors highlight the actors and their respective roles in an infrastructure primarily run and operated on distributed ledger technology (DLT), with computer nodes distributed across different jurisdictions. Moreover, it is argued that the level of control by end users over their digital representations and online footprints remains untested in the context of a worldwide digital financial infrastructure as proposed by Diem. The paper further elaborates and puts data protection and privacy of end users under scrutiny, outlining the need for a self-sovereign identity (SSI) management system in order to address the risks associated with correlation and profiling of individuals concerning their behaviour in payment systems.

Keywords: Libra; Diem; Facebook; Distributed Ledger Technology; DLT; blockchain; network governance; trust; digital identity; self-sovereign identity; SSI; central bank digital currency; CBDCs; crypto-assets

Recommended citation: Golnaz A. Jafari and Malte-C. Gruber, The Case of Diem: A Distributed Ledger Technology-based Alternative Financial Infrastructure Built by a Centralised Multisided Platform, 12 (2021) JIPITEC 301 para 1

A. Introduction

1 In order to achieve its set objective to design a location-independent alternative worldwide system for digital finance, Diem is set to be built on distributed ledger technology (DLT), and was initially structured to be governed by a Swiss based member association, the Diem[1] Association, and its subsidiary and primary operating entity Diem Networks. Diem Net-

* Golnaz A. Jafari, LL.M., is a doctoral researcher at Lucernaiuris, University of Lucerne, Switzerland, formerly a research associate at SOCAI, University of Würzburg, Germany & a research fellow at NRCCL, University of Oslo, Norway; Malte-C Gruber, Dr. iur., is a Professor of Legal Philosophy and Commercial Law with a focus on Intellectual Property Law and the Law of New Technologies at Lucernaiuris, University of Lucerne, Switzerland. The views and opinions expressed in this article remain those of the authors. Authors specifically thank Josephine Heinzelmann, and in particular, Dr. Steven Howe of Lucernaiuris as well as Pieter v Ysseldijk for their insightful comments.

works was designed to take the role of a regulated payment systems operator, activation of which required a payment systems licence from the Swiss Financial Market Supervisory Authority (FINMA). As of May 2021, the FINMA application for a payment systems licence has reportedly[2] been withdrawn in an attempt to limit the jurisdictional scope of the project to the United States of America (USA), at least during the initial phase. Nevertheless, both entities seem to currently hold active status on the Swiss commercial registries.

2 Against this background, one would also need to consider the possibility of central banks introducing central bank digital currencies (CBDCs). Once operational, Diem could presumably have an impact on the worldwide financial sector and further popularise the inception of CBDCs, most probably in a form consisting of public-private partnerships.

3 The paper will first provide an overview of the organisational structure of the Diem project as well as its technical typology in an attempt to define main actors and stakeholders, respectively to distinguish between the two phases of the project, namely Libra 1.0 and Libra 2.0, now known as Diem. In the subsequent section, attention is given to the definition of the fundamental legal nature of the Diem design, which took a two-fold form. The two-fold design model consisted of single fiat currency stablecoins and an intra-network Diem crypto token acting as a "digital composite" for some of the network's stablecoins. The digital composite would then be backed by a basket of fiat currencies and other assets. As for the project's initial phase, and with a primary focus to comply with the US regulatory landscape, Diem is set to take off in the form of a single US dollar -backed token.

4 Trust as the elementary fact of social life and, more specifically, as the central factor in the context of money creation as well as financial services, is addressed in the following section. Trust in the functioning of a given system would bear a direct link with the transparency of the system's governance. Here, the authors argue that Diem's chances for mass adoption would depend in particular on its prospects of gaining trust as a new and alternative digital form of *private* currency alongside the established monetary systems. This would require a number of constituents, such as comprehensive accessibility and trustworthiness based on legal certainty, clear attributions of responsibility, appropriate models of liability, and effective legal mechanisms of enforceability.

5 Trust and transparency are closely linked with consumer protection as well as compliance with end users' privacy and personal data protection. Identity management is therefore pivotal. In the final section of this paper, the authors argue that as it stands, despite minimal information being publicly available, Diem's identity management system would fail to give end users an effective control over their digital representation on its network. The paper concludes by highlighting the significance of recent technological developments in the digital identity sphere, whereby a standardised and interoperable self-sovereign identity (SSI) management could be a way forward in such network infrastructure.

B. The Case of Diem

I. Organisational Structure in a Nutshell

6 Initially branded as Libra 1.0, the project was incepted in June 2019[3] by the social network platform Facebook in an attempt to provide cross-border financial services enabled through the means of technologies such as distributed ledger technology (DLT).[4]

7 Headquartered[5] in Geneva, Switzerland, the Diem Association, previously known as the Libra Association, was formed in July 2019 as a non-profit (independent) membership association to be responsible for the development and governance

1 Announced by Libra Association, the name of the project has been changed from 'Libra' to 'Diem' in an attempt "to reinforce organisational independence"(1 December 2020) <https://www.diem.com/en-us/updates/diem-association/>; notably the name change has triggered the possibility for a legal action by a London based fintech company which operates finance application software also named Diem < https://cointelegraph.com/news/carpe-diem-lawsuit-threatened-over-facebook-s-libra-rebrand-plan>.

2 FINMA, 'Diem withdraws licence application in Switzerland' (12 May 2021) <https://www.finma.ch/en/news/2021/05/20210512-mm-diem/>.

3 Libra Engineering Team, 'Libra: The path forward' (18 June 2019) <https://www.diem.com/en-us/blog/the-path-forward/>.

4 The terms 'blockchain' and 'DLT' are often used interchangeably. The authors take the view that blockchain could be considered a subcategory of DLT, whereby entries to ledger (or chain) are primarily bundled in the form of blocks.

5 See entity registration in the commercial registry of the canton of Geneva (31 July 2019) <https://www.shab.ch/api/v1/publications/5626ee28-a9a2-4193-b05b-e5dc0679f155/pdf>.

of the project. Members of the association,[6] mostly businesses and enterprises, were set to be represented by one representative per entity with a right to one vote. These representatives would participate in the governance and key decision making areas of the project, develop its long-term strategy, and respectively validate all the transactions on the Diem network. The day to day management of the Diem Association would be carried out by its designated board of directors, with the association's operational leadership remaining in the hands of an appointed executive team.

8 As a subsidiary to Diem Association, Diem Networks,[7] previously known as Libra Networks, was registered in the form of a limited liability company (LLC)[8] by Facebook in Geneva in May 2019.[9] Initially a stakeholder of Diem Networks, Facebook Global Holdings II LLC later transferred its shares to the Diem Association in October 2019.[10] Diem Networks was founded to become a financial technology (fintech) entity, for the pursuit of a number of objectives. These include the development and production of software and infrastructure, particularly in line with investment activities, payment operations, financing, identity management, data analytics, big data, blockchain and other technologies. With the recent withdrawal from the FINMA application for a payment systems licence[11] which was submitted on the legal basis of the Swiss Financial Market Infrastructure Act (FMIA), the sister subsidiary Diem Networks US, Inc.[12] has recently been registered as money services business (MSB) administered by the Financial Crimes Enforcement Network (FinCEN) under the Bank Secrecy Act (BSA). Diem Networks US, Inc., wholly owned by Diem Association, has recently partnered with Silvergate Capital Corporation[13] whereby latter is set to be the exclusive issuer of Diem's single US dollar-backed token and the manager of the associated reserve.

9 Given the significant user base, global reach and network effect of Facebook and its associated group of entities, such as Instagram, WhatsApp and Messenger, with over a quarter of the world's population at its disposal,[14] the Diem project has, since its inception, been subject to extensive regulatory scrutiny from various jurisdictions including, but not limited to, the USA,[15] the European Union (EU)[16] and Switzerland. The project's potential effect on

6 See <https://www.diem.com/en-us/association/>; the membership profile of the association has changed since its inception in June 2019 with companies such as Visa, Mastercard, Paypal and eBay, among others, eventually opting out of the project.

7 See name change update in the commercial registry of the canton of Geneva (8 December 2020) <https://www.shab.ch/shabforms/servlet/Search?EID=7&DOCID=1005042618>.

8 LLC equals Société à responsabilité limitée (SARL) and Gesellschaft mit beschränkter Haftung (GmbH) in Swiss company law.

9 See entity registration in the commercial registry of the canton of Geneva (2 May 2019) <https://www.shab.ch/shabforms/servlet/Search?EID=7&DOCID=1004624965>.

10 See stakeholder update in the commercial registry of the canton of Geneva (21 October 2019) <https://www.shab.ch/shabforms/servlet/Search?EID=7&DOCID=1004742062>. Note: The shares have later been transferred from Diem Association to Diem GmbH, a new entity in Lucerne (01.02.2021) <https://www.zefix.admin.ch/de/search/entity/list/firm/1391882>.

11 FINMA, 'Libra Association: FINMA licensing process initiated' (16 April 2020) <https://www.finma.ch/en/news/2020/04/20200416-mm-libra/>. See n 2.

12 Diem Networks US, Inc. is incorporated in Washington (16 September 2020) <https://opencorporates.com/companies/us_va/11109939>.

13 On 12 May 2021 Diem announced its withdrawal from the ongoing FINMA application for a payment systems licence, bringing the project during its initial phase into the US regulatory perimeter. Diem Association's subsidiary and primary operating entity Diem Networks US has now partnered with Silvergate, a California based state-chartered bank, in a plan to first issue its US dollar-backed tokens. Diem's US dollar-backed tokens and the associated reserve is set to be exclusively issued and managed by Silvergate. <https://www.diem.com/en-us/updates/diem-silvergate-partnership/>. Note: no Diem tokens have been issued as of October 2021.

14 See Statista, 'Cumulative number of monthly Facebook product users as of 3rd quarter 2020' (4 November 2020) <https://www.statista.com/statistics/947869/facebook-product-mau/>; during the last reported quarter, Facebook stated that 3.21 billion people were using at least one of the company's core products.

15 'Facebook's Zuckerburg grilled by Congress on Libra – as it happened' (*Financial Times,* 23 October 2019) <https://www.ft.com/content/bf16f8ec-6897-38be-9ff4-0f40cc4c779d>.

16 European Council, 'Joint statement by the Council and the Commission on "stablecoins"' (5 December 2019) <https://www.consilium.europa.eu/en/press/press-releases/2019/12/05/joint-statement-by-the-council-and-the-commission-on-stablecoins/#>; "when an initiative has the potential to reach a global scale, the concerns are likely to be amplified and new potential risks to monetary sovereignty, monetary policy, the safety and efficiency of payment systems and financial stability can rise."

worldwide financial stability and state sovereignty in the control of money creation are placed at the centre of public discourse.[17]

10 Money[18] is seen as a *public good* that is built on *public trust* in order to carry out its socioeconomic functions. In this context, a distinction would need to be made between *account-based* and *token-based*[19] forms,[20] on the basis of their respective identification and verification requirements. The account-based form relies on the identification of the payer's identity, i.e. bank deposits. The token-based form depends on the verification of authenticity of the object that is being exchanged, i.e. physical cash, respectively cryptographically generated payment token models.

11 In this regard, the European Central Bank (ECB) made a reference[21] to Diem as an infrastructure for the creation of stateless money by conglomerates of corporate entities, with potential conflict of interests, whereby the entities would only be accountable to their respective stakeholders and members. Here, control over the distribution network would arguably be maintained by these entities "acting as quasi-sovereign issuers of currency,"[22] which would then exercise privileged access to user private data for, among others, monetisation purposes.[23]

12 As has been pointed out,[24] the technical protocol of Diem has made extensive references to the term 'account'[25], mostly leaving out the term 'token'. On the Diem network, the object to be exchanged would then take the form of a payment instruction that would need to be authenticated and processed in accordance with the applicable rules of the network. In this case, identity verification may not be necessary in order to process transactions. Nevertheless, as will be discussed in subsequent sections, the Diem network is set to put in place a strict identity verification procedure through its in-house digital wallet application. As a result, classification of the *private* currency designed by Diem as either account-based, or token-based, seems not to be a straightforward task.

13 Moreover, the concerns raised are arguably not immaterial, given, not least because of Facebook's business model as a for-profit multisided platform. This model enables the company to effectively govern the interactions among participants in its network, namely users, developers and marketers.[26] Here, users utilise the platform free of charge in return for their metadata which can contain behavioural information. This metadata is then used by Facebook for the generation of analytics, on the basis of which marketers place advertisements. On the other hand, developers are charged by Facebook in order to integrate and monetise their application software on its platform. Facebook therefore largely depends on its user volume and advertising revenue[27] for maintaining its operations.

17 ECB, 'Money and private currencies: reflections on Libra' speech by Yves Mersch (2 September 2019) <https://www.ecb.europa.eu/press/key/date/2019/html/ecb.sp190902~aedded9219.en.html>.

18 In economic theory, a functional definition of money would consist of three elements of a) a unit of account, b) a means of payment (exchange), and c) a store of value. Money can either be physical (cash) or non-physical (scriptural or electronic).

19 See section C.1 for further details.

20 MK Brunnermeier et al., 'The Digitalisation of Money' (2019) NBER Working Paper Series nr. 26300, 4f.; CM Kahn et al., 'Should the central bank issue e-money?' (October 2018), 8-11 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3271654>.

21 ECB, 'Money and private currencies' (n 17).

22 Ibid.

23 V Khan & G Goodell, 'Libra: Is it really about money?' (August 2019) <https://arxiv.org/ftp/arxiv/papers/1908/1908.07474.pdf>.

24 D Jackson, 'Global 'stablecoin' Challenges: Response to FSB Consultation Document' (12 July 2020), 3f. <https://www.fsb.org/wp-content/uploads/Dr.-Douglas-Jackson.pdf>.

25 'The Libra Blockchain' (23 July 2019), 15f < The Libra Blockchain>; "at the logical level, an account is a collection of resources and modules stored under the account address. At the physical level, an account is treated as an ordered map of access paths to byte array values. An access path is a delimited string similar to a path in a file system."

26 A Hagiu & J Wright, 'Multisided Platforms' (2015) Working Paper, Harvard Business School; notably, multisided platforms are distinguished from vertically integrated platforms in that the former do not exercise control over interactions but rather govern them; in 5, two features seen inherent in multisided platforms are "a) they enable direct interaction between two or more distinct sides & b) each side is affiliated with the platform."

27 S Ghosh, 'Understanding Multi-sided Platforms: Social Networks and more' (12 October 2015) <https://samghoshblog.wordpress.com/2015/10/12/understanding-multi-sided-platforms-social-networks-and-more/>.

14 By venturing into financial services, Facebook's potential expansion of access to users' financial and behavioural data in payment services would arguably aggregate the existing risks associated with the correlation of users' profiles to a wide range of their activities spanning from social networks to spending patterns and monetary transaction records.[28]

15 In light of Facebook's demonstrated pattern of failing to keep consumer data private,[29] risks associated with user data privacy in the context of the proposed Diem project have been subject to numerous Congress hearings[30] in the USA, primarily by the House Financial Services Committee.

16 In order to tone down the ongoing discussions, Facebook shifted away from Libra 1.0 and rolled out Libra 2.0,[31] now named Diem, with an updated whitepaper[32] on technical and organisational matters published in April 2020. The downgraded version of the project was initially expected to launch by January 2021,[33] pending an affirmative outcome of its licence application with FINMA. With the recent shift from its FINMA application for a payment systems licence in Switzerland to the MSB licence registry with FinCEN in the USA, the project's initial phase is yet to materialise as of the date of this writing.

17 Nevertheless, during the latest G7[34] (virtual) roundtable among finance ministers of participating states, central bank governors, the European Commission and the Eurogroup, hosted by the Treasury Secretary of the USA, the need for an effective regulatory landscape prior to inception of any such project was reiterated. Notably, the German representative[35] took the view that merely rebranding Libra would certainly not render sufficient the project's admissibility within the German as well as the EU markets.

18 In this context, Facebook's intention to expand its scope of activities to financial services worldwide is not restricted to its Diem project. It has recently launched[36] an electronic payments system for fiat currency[37] transfers via one of its core products, WhatsApp. The initiative is set to initially target the Brazilian market and is enabled through Facebook's in-house software application, Facebook Pay. The distinction between Diem and WhatsApp's electronic payments system rests on the nature of the particular currency in circulation. The former is built based on a cryptographically generated *private* currency model (otherwise known as cryptocurrency), whereas the latter integrates payments based on digital representation of underlying fiat currency that is both *public* and *private* money.

19 More importantly, Facebook's subsidiary Calibra was founded in June 2019[38] with the aim of providing in-house financial services, including digital wallet services, to the Diem network. Later rebranded as

28 See also DA Zetzsche, RP Buckley & DW Arner, 'Regulating Libra: the Transformative Potential of Facebook's Cryptocurrency and Possible Regulatory Responses' (2019), UNSW Law, 15f.

29 US House Committee on Financial Services, 'Waters Statement on Facebook's Cryptocurrency Announcement' Press Release (18 June 2019) <https://financialservices.house.gov/news/documentsingle.aspx?DocumentID=403943>.

30 See S Hrg. 116-71, 'Examining Facebook's proposed digital currency and data privacy considerations' (16 July 2019) <https://www.congress.gov/event/116th-congress/senate-event/LC64460/text?s=1&r=2>.

31 Libra Association, 'Libra developers: The path forward' (16 April 2020) <https://www.diem.com/en-us/blog/libra-developers-the-path-forward/>.

32 Libra whitepaper v2.0, 'Cover Letter' (April 2020) <https://wp.diem.com/en-US/wp-content/uploads/sites/23/2020/04/Libra_WhitePaperV2_April2020.pdf>.

33 'Facebook's Libra currency to launch next year in limited format' (*Financial Times,* 27 November 2020) <https://www.ft.com/content/cfe4ca11-139a-4d4e-8a65-b3be3a0166be>.

34 US Department of the Treasury, 'Readout from a Treasury Spokesperson on Secretary Mnuchin's Discussion with G7 Finance Ministers and Central Bank Governors' Press Release (7 December 2020) <https://home.treasury.gov/news/press-releases/sm1203>.

35 'Facebook's renamed cryptocurrency is still 'wolf in sheep's clothing': German Finance Minister' (*Reuters,* 7 December 2020) <https://www.reuters.com/article/g7-digitial-facebook/facebooks-renamed-cryptocurrency-is-still-wolf-in-sheeps-clothing-german-finance-minister-idUSKBN-28H20B>.

36 WhatsApp Blog, 'Bringing payments to WhatsApp for people and small businesses in Brazil' (15 June 2020) <https://blog.whatsapp.com/bringing-payments-to-whatsapp-for-people-and-small-businesses-in-brazil>.

37 See the definition of fiat money (currency): one that is declared legal tender and issued by a central bank. Fiat money derives its value from public trust in central banks in order to maintain price stability.

38 Facebook, 'Coming in 2020: Calibra' (18 June 2019) <https://about.fb.com/news/2019/06/coming-in-2020-calibra/>.

Novi Financial[39] and headquartered in the state of California, USA, the entity has also been registered as MSB[40] by FinCEN which is passport-able among all states.

20 The first product of the company, the Novi digital custodial wallet, is set to be rolled out as a stand-alone software application, yet duly integrate-able in Facebook's core products of Messenger and WhatsApp.[41] In other words, operational interoperability would in principle be ensured between Diem, Messenger and WhatsApp infrastructures. The Novi wallet is a crucial element in the operability of the project given that it will serve as the main user interface upon which services would be built based on smart contract codes.

21 Furthermore, Facebook Financial (F2)[42] has been established as an internal group mandated with streamlining and managing Facebook's payments projects including Facebook Pay. The group will be led by the head of Novi Financial, who will also be involved in WhatsApp's electronic payment system initiative. Notably, Novi Financial is one of the members of the Diem Association.

22 As an interim remark, it has become increasingly apparent from the organisational breakdown of Diem, as it stands to date, that Facebook is arguably set to maintain a certain degree of governance and control, albeit indirectly, over the project. Through the bundling of its in-house software applications with the company's core products, the dynamics of user dependency seem to emerge, despite Facebook's absence from the membership of the Diem Association, respectively considering the fact that the company seems to no longer own stakes in Diem Networks in Switzerland and the USA. Once users would be enabled to engage in spending behaviour across Facebook's core products free of charge in return for their metadata, the company's advertising revenue would be set to experience an exponential growth.

II. Technical Typology & Taxonomy

1. Main Characteristics

23 The proposed Diem alternative financial infrastructure is set to be designed and built based on distributed ledger technology (DLT).

24 DLT could be defined as a shared database (or ledger) of records, distributed among computer nodes outside jurisdictional boundaries. A subset of involved interactions between these nodes is defined by consensus protocols. Every entry, update or transaction to the ledger would be time stamped, cryptographically hashed,[43] cryptographically signed[44] and authorised prior to its addition to the ledger.

25 An algorithmic consensus would represent the agreed-upon true state among all participants and stakeholders, which could either be reached on a system level or on an individual deal level, depending on the type of DLT deployed.

26 DLT can take various forms depending on the deployed participation and governance protocols, among which is a typical public and permission-less model. Here, the ledger would essentially operate on the basis of 'data broadcasting', where data is in principle broadcast to every single computer node on the ledger, irrespective of any associated interest or stake. DLT could also be designed in a hybrid public and permissioned format, or, in a rather stricter sense, in a private and permissioned format, or organised as a consortium.[45]

39 Facebook, 'Welcome to Novi' (26 May 2020) <https://about.fb.com/news/2020/05/welcome-to-novi/>.

40 See <https://www.docdroid.net/544Gxxg/calibra-msb-registration-pdf>; reference to FinCEN's definition of the term 'money services business (MSB)' as "a person wherever located doing business, whether or not on a regular basis or as an organised or licensed business concern, wholly or in substantial part within the United States, operating directly or through an agent, agency branch, or office, who functions as, among other things, a money transmitter." in FinCEN Guidance (FIN-2019-G001, 9 May 2019), 3.

41 Facebook, 'Welcome to Novi' (n 39).

42 'Facebook Financial Formed to Pursue Company's Payments Plans' (*Bloomberg,* 10 August 2020) <https://www.bloomberg.com/news/articles/2020-08-10/facebook-financial-formed-to-pursue-company-s-commerce-ambitions>.

43 A cryptographic hash function is a mathematical function used in cryptography. Typical hash functions take inputs of variable lengths to return outputs of a fixed length. A cryptographic hash function combines the message-passing capabilities of hash functions with security properties. Cryptographic hash adds security features to typical hash functions with stronger mathematical guarantees for collisions etc.

44 This process is made possible through the creation of encryption schemes such as asymmetric encryption or public key infrastructure (PKI) with public/private key pairs. Digital signatures are generated by private keys. Digital signatures are defined as "mathematical schemes for demonstrating the authenticity of a digital message."

45 For more on this see N Kannengisser et al., 'Trade-offs between Distributed Ledger Technology Characteristics'

27 In addition, from a legal and regulatory compliance perspective, in 'data propagation', as opposed to data broadcasting, transaction records would only be shared with nodes on a *need-to-know* basis depending on their stake, which would then enhance privacy and data protection thresholds. The form a DLT takes defines the scope of 'read' and 'write' privileges and restrictions granted to the network participants. The internal governance of a given DLT network would therefore be closely interlinked with the factual dynamics surrounding its nodes. Disintermediation associated with DLT effectively lays the ground for poly-directional relationships among nodes that are connected through software programmes.[46]

28 Notably, the choice of the architectural form of an underlying DLT would also have potential implications, directly or indirectly, in the way a smart contract code is defined and operated. A smart contract code is essentially a decentralised application running on a DLT network.

29 A smart contract code could be defined as a computer programme written based on a number of predefined terms and conditions as well as oracles. These programmes can facilitate, verify and enforce the negotiation and execution of legal contracts.[47] They can have interfaces to handle input from parties to contracts.[48] An oracle is an agent or an interface designed to verify external data and real-life occurrences. Upon satisfaction of the pre-defined terms and conditions, and the update of external data through the means of oracles, these programmes would change their state of information and autonomously self-execute[49] the predetermined outcome. Automation is, as a result, seen as an inherent and key feature of a smart contract code. Here, pre-defined terms and conditions as well as outcomes between *trust-less*[50] network participants, i.e. in the context of parties to a given transaction, would in principle be executed without reliance upon intermediation.

30 Furthermore, DLT enables the creation of native value[51] from scratch, which is intrinsically accrued from the rules of the system as well as network participation therein. Alternatively, a real world value can be collateralised and digitally represented in the form of a token appendable on DLT, otherwise known as asset tokenisation, with the end product often referred to as a crypto-asset.

31 Here, a token[52] is defined as a piece of information recorded on DLT, and takes the form of digital representation of value or asset, respectively a claim, ownership or access right. The terms token and crypto-asset are often referenced interchangeably in different jurisdictions. In the EU, preference is given to crypto-asset,[53] whereas in Switzerland the term token[54] is mostly utilised.

(2020) ACM Computing Surveys (CSUR), 53(2), 1-37.

46 P Paech, 'The Governance of Blockchain Financial Networks' (2017) Modern Law Review, 80(6) MLR.

47 M Wöhrer & U Zdun, 'Smart Contracts: Security Patterns in the Ethereum Ecosystem and Solidity' (2018) IWBOSE, IEEE, 2.

48 Ibid.

49 In this context, 'technical enforcement' is not synonymous to 'legal enforcement'.

50 Here the term 'trust-less' strictly refers to the absence of a concentrated single intermediary. In other words, in the DLT context there are mechanisms put in place that facilitate distribution of 'trust', whereby participants in the system, without necessarily trusting one another, are able to reach consensus as to a 'true state'.

51 Known examples are Bitcoin and Ethereum blockchain networks.

52 See Liechtenstein Tokens & Trusted Technology Service Provider Act "TVTG" (January 2020), Article 2 <https://www.gesetze.li/konso/2019301000>; for 2020 unofficial translation <https://www.regierung.li/media/medienarchiv/950_6_08_01_2020.pdf?t=2>; in a general technical sense: "tokens are classified as ordinary or delimiter tokens. An ordinary token is a numeric constant, an ordinary identifier, a host identifier, or a keyword. A delimiter token is a string constant, a delimited identifier, an operator symbol, or any of the special characters shown in the syntax diagrams", see <https://www.ibm.com/support/knowledgecenter/ssw_ibm_i_73/db2/rbafzch2tok.htm>; in a DLT context: "token is a digital representation of value on a shared distributed ledger that is owned and secured using cryptography to ensure its authenticity and prevent modification or tampering without the owner's consent", see <https://www.itu.int/en/ITU-T/focusgroups/dlt/Documents/d11.pdf>.

53 European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on Markets in Crypto-assets, and amending Directive (EU) 2019/1937 (MiCA)', COM (2020) 593 final.

54 FINMA, 'Guidelines for enquiries regarding the regulatory framework for ICOs' (February 2018) <https://www.finma.ch/en/~/media/finma/dokumente/dokumentencenter/myfinma/1bewilligung/fintech/wegleitung-ico.pdf?la=en>.

32 In this respect, a token or a crypto-asset[55] whose value is derived from an underlying asset that is considered stable, in order to limit price volatility, is also referred to as stablecoin.[56] The underlying asset can take various forms of one or more fiat currencies, one or more commodities, real estate as well as securities.[57]

33 Nevertheless, an algorithmic stablecoin[58] is denoted as one which aims at maintaining a stable value via protocols, whereby the supply of the crypto-assets would increase or decrease in response to changes in demand, and one which does not reference one or more other assets. Additionally, global stablecoin[59] refers to one that has a worldwide reach, is adoptable across jurisdictions and bears the potential to achieve significant volume. Such a tailor-made definition seems on face value to tie in with Diem. Nevertheless, as noted earlier and to be elaborated further in section C., it becomes increasingly apparent that such a classification may not be entirely accurate.

2. Version 1.0

34 Diem first issued its whitepaper version 1.0[60] in June 2019 with the objective to deliver on the promise of *the internet of money*. The initial approach of the project was a DLT-based financial system backed by a reserve of assets and governed by the Libra Association, now Diem Association. The token previously called Libra "LBR" was set to be backed by a basket of bank deposits and short-term government securities held in the Libra Reserve, which would be administered by both the association and its subsidiary Diem Networks, for every LBR created.[61] Both Facebook and Calibra, now Novi Financial, were among the founding members of the association. Also among the member entities was Breakthrough Initiatives, co-founded by Facebook's founder Mark Zuckerberg.[62] The final decision making, as is currently the case, was given to the association, while Facebook was to maintain leadership of the project during the project's inception year, 2019. Once launched, Facebook and its affiliates' role in governance were to be equal to other members.[63]

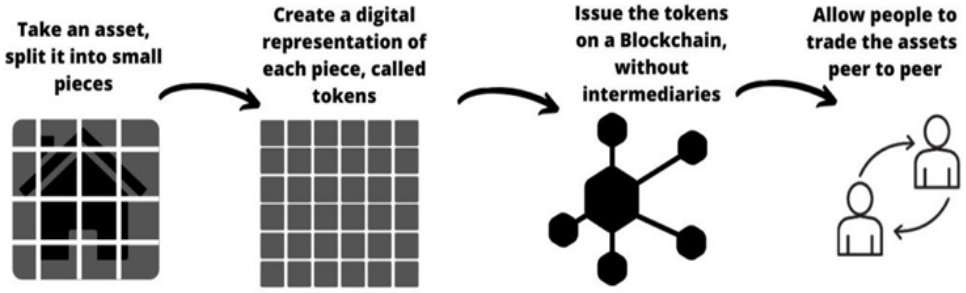


Figure i: simple example of asset tokenisation; credit: www.assetsonblockchain.com

35 The version 1.0 was set to be based on a permissioned DLT[64] with an aim to move towards a permission-less governance model. In both scenarios, the network's participation protocol was to be open access. Smart contract codes would be written based on the Move virtual machine programming language,[65] and the

55 A distinction can be made between fungible and non-fungible crypto-assets. A fungible crypto-asset can be replaced by an equivalent asset with similar market value. A non-fungible crypto-asset or token (NFT) is in principle uniquely identified to ensure its traceability and is generally irreplaceable.

56 FINMA, 'Supplement to the guidelines' (11 September 2019), 1-4, <https://www.finma.ch/en/~/media/finma/dokumente/dokumentencenter/myfinma/1bewilligung/fintech/wegleitung-stable-coins.pdf?la=en>.

57 Ibid.

58 COM (2020) 593final (n 53) Recital 26; Financial Stability Board (FSB), 'Regulation, Supervision and Oversight of "Global Stablecoin" Arrangements' (13 October 2020), 5, <https://www.fsb.org/wp-content/uploads/P131020-3.pdf>.

59 Ibid.

60 Libra whitepaper v1.0, 'An Introduction to Libra' (June 2019).

61 Ibid 3 &7; it is emphasised that LBR is not pegged to any single currency, and "...will not always be able to convert into the same amount of a given local currency. Rather, as the value of the underlying assets moves, the value of one Libra in any local currency may fluctuate"; Furthermore LBR was set to be interest bearing.

62 See <https://breakthroughinitiatives.org/board>.

63 Libra whitepaper v1.0 (n 60), 4.

64 Ibid.

65 Ibid 5 "...by making the development of critical transaction

consensus mechanism would be based on byzantine fault tolerant (BFT),[66] a variation of voting-based mechanisms, carried out by selected validator nodes, i.e. the members of the association who are publicly identified on the network. For this, validator nodes would process transactions and interact with each other in order to reach consensus on the state of the database (or ledger).[67] Notably, smart contract code risk control and management would be carried out by the Diem Association,[68] whereby only the association approved smart contracts were to be published and interact directly with the Diem payment system.

36 As an additional objective to develop and promote an open identity standard,[69] the network would enable pseudonymisation,[70] in principle allowing users to hold multiple addresses (accounts)[71] without risking correlation of these accounts with the holders' real world identities. This is made possible through generating multiple key pairs. On the other hand, a reference is made to the underlying DLT which is set to take the form of a single data structure which would record the history of transactions and states over time,[72] whereby through a unified framework applications could read any data on-ledger at any point in time for proof of integrity.

code easier, Move enables the secure implementation of the Libra ecosystem's governance policies, such as the management of the Libra currency and the network of validator nodes."; "It enables 'resource types' that constrain digital assets to the same properties as physical assets: a resource has a single owner, it can only be spent once, and the creation of new resources is restricted."

66 Ibid.

67 The Libra Blockchain (n 25), 1.

68 Libra whitepaper v2.0 (n 32), 8.

69 Libra whitepaper, v1.0 (n 60), 9.

70 Ibid 6; Regulation (EU) 2016/679 of the European Parliament and of the Council on the general 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 (GDPR)) (2016) OJ L119/1, Art. 4(5) on the definition of pseudonymisation: "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

71 The Libra Blockchain (n 25), 4.

72 Libra whitepaper, v1.0 (n 60), 6.

37 It consequently remains rather unclear as to how data visibility would be maintained on the network, which would then have direct implications regarding privacy.

3. Version 2.0

38 Shifting away from the version 1.0, version 2.0 was introduced with an update on the whitepaper in April 2020.[73] Diem would introduce multiple tokens, each backed by a single fiat currency, in the form of stablecoins. Each single currency stablecoin would then be fully backed[74] by the Diem Reserve, the administration of which is seemingly set to be transparent to the public.[75]

39 The project also seemed to perpetuate its original initiative of creating an intra-network token LBR, now Diem, in principle to be backed by a basket of multiple fiat currencies and other assets, acting as a "digital composite" of the stablecoins created on the network. This token would also be utilised as a means of settlement for cross-border transactions, in particular for jurisdictions where single currency stablecoins have not been introduced, and would be convertible to respective local currencies through third party service providers. The two-fold token model would collectively be referred to as Diem coins.[76]

40 The planned combination of a permissioned DLT platform with integrated smart contract code applications, and an intra-network Diem token together with a variety of single currency stablecoins, governed and supervised by a central authority, was intended to evolve into an ecosystem in the financial services sector, which a large part of the world's population could access via ordinary smartphones and edge devices.

41 Diem took a step further in its version 2.0 aiming at integrating central bank digital currency (CBDCs)[77]

73 See n31f.

74 Libra whitepaper v2.0 (n 32), 12; "full backing means that the Reserve will hold, in cash or cash equivalents and very short-term government securities, an amount at least equal to the face value of each Diem coin in circulation."

75 Ibid 13.

76 For taxonomic breakdown, see Jackson, 'Global 'stablecoin' Challenges' (n24), 6f; for the latest developments as to design of the Diem tokens during the initial phase of the project see n 13.

77 Libra whitepaper v2.0 (n 32), 2.

models once these begin to materialise. The initial plan of moving towards a permission-less DLT network has seemingly been omitted from the agenda of the latest version, governance of which now seems to take place collaboratively between the Diem Association and its subsidiary Diem Networks US, Inc. Facebook is no longer seen as a member of the association, without any special rights.[78] Novi Financial remains as a member together with Breakthrough Initiatives.

42 Diem Networks was mandated with the definition of policies and procedures for reconfiguring the Diem DLT network in case of critical errors, respectively in case of a need for upgrades.[79] The company would, based on contractual arrangements, mint and burn Diem tokens for the purpose of distribution to the market via designated entities called dealers, which would be regulated as financial institutions.[80] Diem Networks would therefore not enter into any contractual relationship with exchange platforms or end users, save for emergency operations.[81] Diem Association would exercise control over the process of minting and burning of Diem tokens, the mandate for which was given to Diem Networks. The association and its subsidiary, Diem Networks, would also operate a compliance infrastructure integrated in the form of a financial intelligence unit (FIU)[82] function, in order to monitor the network regarding any suspicious activity.

43 User interaction with the Diem DLT network was set to take place via regulated or certified virtual asset service providers (VASPs).[83] Alternatively, direct user access would also be made possible, albeit with limited transaction volume and account address balance, through Unhosted Wallets.[84] At protocol level, VASPs would be required to comply with the "travel rule" when transacting.[85] The travel rule[86] ensures that VASPs collect and exchange beneficiary and originator information with VASP counterparties for any transmittal exceeding USD 1,000. Under the travel rule, the required personally identifiable information (PII)[87] would include names, account numbers, physical addresses as well as unique identification numbers. In the course of facilitating transactions on behalf of users, VASPs would be given the possibility to record transactions off-ledger and internally in their respective books.[88]

44 In light of this requirement, it is arguable as to the manner in which Diem would effectively maintain user pseudonymity and respect correlation resistance between users' activities on the system and their real identities.

45 As mentioned, the Novi digital custodial wallet, which would most probably function as a hosted wallet, would act as the main user interface of the

78 Ibid 6.

79 Ibid 8; at the time of this writing Diem Networks referred to the subsidiary based in Switzerland which was the candidate for a payment systems licence application pending a decision by FINMA. At present, with the withdrawal of the FINMA licence application, Diem Networks US, Inc., a sister subsidiary wholly owned by Diem Association, has instead been registered as a money services business (MSB) licensee by FinCEN in the US.

80 Ibid 17.

81 Ibid 13f "In the context of a recovery and resolution plan, the association is considering whether to provide for two key components that could be implemented in severe stress scenarios in the unlikely case that the network is unable to convert the very short-term government securities in the Reserve into cash fast enough to satisfy all requests to burn Diem coins without incurring fire-sale losses: a) redemption stays which would delay Diem coin redemptions and allow for additional time to liquidate the Reserve's assets during a window of time without incurring large fire-sale losses, b) early redemption haircuts which would impose a fee for instant redemptions and require coin holders to internalise their negative externality (i.e., fire-sale losses) in a run."

82 The term financial intelligence unit (FIU) is defined as a "... central, national unit that is responsible for receiving and analysing information from private entities on financial transactions which are considered to be linked to money laundering and terrorist financing," see Directive (EU) 2015/849 of the European Parliament and of the Council on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing (4AMLD) (2015) OJ L141/73.

83 Financial Action Task Force (FATF), 'Guidance for a risk based approach: virtual assets and virtual asset service providers' (June 2019), 13 <https://www.fatf-gafi.org/media/fatf/documents/recommendations/RBA-VA-VASPs.pdf>.

84 Libra whitepaper v2.0 (n 32), 18f; Distinction between hosted and unhosted wallets lies in the exercise of control over private keys. In the case of hosted wallets, private keys are stored by third parties, whereas in unhosted wallets private keys remain in the control of users.

85 Ibid 20.

86 FATF, 'Guidance for a risk based approach' (n 83), Recommendation 16, 28 – 31.

87 The term 'personally identifiable information (PII)' is used interchangeably with the term 'personal data', which is used in the EU General Data Protection Regulation (GDPR).

88 Libra whitepaper v2.0 (n 32), 18.

network. In terms of user data privacy, account information and financial data would not be shared with Facebook and its core products, i.e. for the purpose of improving advertisement targeting, except in particular cases.[89] These include the prevention of crime, compliance with law, payment processing and service providers as well as in the case of data aggregation related to service performance and related products, with an in-built correlation resistance technique, details of which remain unclear.

46 The wallet would integrate an identity system named 'visual identity', with an obligatory identification of all users through government issued identities. Arguably,[90] the Novi wallet design would be implemented in the form of an off-ledger payment mechanism,[91] with Novi Financial acting as a VASP, for the provision of both exchange and custodial wallet services. Novi Financial would then "hold all Diem coin backing for Novi balances in its own accounts on the underlying Diem DLT network."

C. Legal & Regulatory Framework

I. Diem's Fundamental Nature

47 The legal and regulatory implications concerning the technical design of Diem are directly dependent upon the *substance* underlying such design. As mentioned earlier, there seems to be no clear-cut distinction to be made as to whether the intended design would fall under an *account-based*, or *token-based* private currency issuance. This distinction would be essential in understanding the applicable identification and verification requirements thereof.

48 In addition to the general civil law status of cryptographically generated tokens (crypto tokens) on DLT, it is essential to determine whether Diem could be considered *money* and in what way it will be comparable to currencies such as fiat money.[92] Not least, the question arises as to what extent an underlying *intrinsic value* could make a decisive difference in this equation.

49 In economic theory, a functional definition of money[93] generally consists of three elements of a) a unit of account, b) a means of payment (exchange), and c) a store of value. Money can either be physical (cash) or non-physical (scriptural or electronic).

50 More specifically, *public* money is distinguished from *private* money[94], whereby *public* money is defined as fiat money or fiat currency that is legal tender[95] and which is issued by central banks. *Private* money takes the form of fiat currency credit issued by licensed credit institutions such as retail and commercial banks. Fiat money derives its value from *public trust* in central banks which are primarily mandated to maintain price stability.

51 At first glance, the two-fold Diem design, in the form of single fiat currency stablecoins and an intra-network Diem token backed by a basket of multiple fiat currencies and other assets as the 'digital composite', seems to satisfy the three inherent constituents of money, albeit issued in the form of *private* currency. Here, it is pivotal to take account of the fact that the nature of Diem tokens as a *store of value* may be questionable with reference to the trust associated with the survivability of the Diem Reserve. Given the significant user base, global reach and network effect of Facebook and its associated group of entities, it would not be far-fetched to argue that the Diem design would satisfy the *means of payment (exchange)* element.

52 When defining the legal significance of crypto tokens and their classification, *substance* matters over *form*. As referred to in the preceding sections, DLT can enable the creation of native value from scratch through token representation. Such value would intrinsically be accrued from the rules of the system, network participation as well as the market response to those set rules. Here, the token is seen as an empty container.[96] Alternatively, real world

89 Facebook, 'Novi: Customer Commitment', 1f. < https://bit.ly/3826M16>.

90 Jackson, 'Global 'stablecoin' Challenges' (n 24), 22.

91 The Libra Blockchain (n 25), 22; it is anticipated that many payment transactions on Diem will occur off-ledger, for example, within a custodial wallet or by using payment channels.

92 On the regularly repeated functions of money, see eg K Langenbucher, 'Digitales Finanzwesen. Vom Bargeld zu virtuellen Währungen' (2018) 218 AcP 385, 388f.; L Müller & M Ong, 'Aktuelles zum Recht der Kryptowährungen' (2020) 29 AJP/PJA 198, 206ff.

93 See n 18.

94 See section D.III.

95 See the definition of the term 'legal tender' under Commission Recommendation on the scope and effects of legal tender of euro banknotes and coins (2010/191/EU) OJ L 83/70, para.1 regarding euro banknotes and coins "...where a payment obligation exists, the status of legal tender should imply three things: first, mandatory acceptance; second, acceptance at full face value; and third, the power to discharge from payment obligations."

96 As reflected in Liechtenstein TVTG Act (n 52).

value can be collateralised and digitally represented in the form of a token appendable on DLT. This is known as asset tokenisation, with the end product often referred to as a crypto-asset. Therefore, it is feasible to consider, and before any economic, sociological or legal classification, crypto tokens as semantic artefacts of network communication of digital platforms.

53 In this respect, initial, tentative approaches describe the likes of first category crypto tokens such as bitcoin, where value is created on DLT from scratch and maintained through the rules of the system, as "value-embodying data".[97] Here, it would certainly have to be asked whether the term *value embodiment*[98] is an oxymoron, which obviously presupposes an immaterial entity in which values could materialise. Above all, it has been argued that the monetary data value of such crypto token units should be considered distinct from the immaterial effects of data protection based on personal rights.[99]

54 There is widespread agreement in the so far sparse[100] civil law[101] literature only on what crypto tokens are not. Crypto tokens are not considered as things, since they are not separable, physical objects,[102] nor are they to be qualified as claims.[103] The latter would only be possible if a central institution owed the holders of crypto tokens a payment in the form of a contractual redemption right.[104] However, this prerequisite can not only be assumed for e-money, but also for e-money tokens, which will be discussed in the subsequent section. At best, a claim could then be derived from a relationship under corporate law between all users of a closed network.[105]

55 On the other hand, crypto-assets have been classified as property *sui generis* in a number of jurisdictions, in particular in common law systems such as the UK.[106]

56 Arguably, the lack of identifiability of a claim may lead to the assumption of an unplanned regulatory gap, which would have to be filled by analogy according to the rules of traditional legal methodology.[107] However, even the precondition of a regulatory gap can be doubted. This is because of the result of the underlying analogy, for example, an alleged equivalence[108] of crypto tokens in terms of property law, which is almost circularly based on the assumption of similarity with movable property or money.[109] Therefore, functional equivalences with traditional property ownership according to the standards of national property law provisions hardly lead any further.[110] Rather, they threaten to obscure the view of

97 Langenbucher, (2018) (n 92), 409.

98 Cf BV Enz, 'Die zivilrechtliche Einordnung von Zahlungs-Token wie dem Bitcoin als "Registerwertdaten" und deren Aussonderbarkeit im Konkurs de lege lata und de lege ferenda' (2020) 116 SJZ, 291, 294.

99 Cf S Omlor, 'Kryptowährungen im Geldrecht' (2019) 183 ZHR, 294, 311: "Such a personal rights component is missing from the sober transaction data on a payment blockchain".

100 Cf A Walter,'Bitcoin, Libra und sonstige Kryptowährungen aus zivilrechtlicher Sicht' (2019) 72 NJW, 3609 („shadowy existence").

101 See German Civil Code "BGB", Section 90; Swiss Civil Code "ZGB", Art. 641ff.

102 See n 99; see also B Beck & D König,'Bitcoin: Vertragstypologische Einordnung von kryptographischem Geld' (2015) 70 JZ, 130ff.; Langenbucher, (2018) (n 92), 405. In this respect, the German legal definition of the term "Sache" also corresponds to the Swiss private law doctrine on the interpretation of the Art. 641ff ZGB; see for instance Enz, (2020) (n 98), 293f; ibid, *Kryptowährungen im Lichte von Geldrecht und Konkursaussonderung* (Zürich 2019), paras.334ff.

103 Cf Langenbucher, (2018) (n 92) 385, 405ff; on the difficulties of classification from the perspective of U.S. law, see CS Goforth, 'U.S. Law: Crypto is Money, Property, a Commodity, and a Security, all at the Same Time' (October 25, 2018), Journal of Financial Transformation (forthcoming) <https://ssrn.com/abstract=3272975>.

104 See German Civil Code "BGB", Section 241(1) sentence 1.

105 Cf DA Zetzsche, RP Buckley & DW Arner, 'The Distributed Liability of Distributed Ledgers: Legal Risks of Blockchain' (2017) 52 UNSWLRS, 26ff <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3018214>; HC von der Crone, FJ Kessler & L Angstmann, 'Token in der Blockchain – privatrechtliche Aspekte der Distributed Ledger Technology' (2018) 114 SJZ, 340 f ("System agreement among all participants").

106 UK Jurisdictional Task Force (UKJT), 'Legal Statement on Crypto-assets and Smart Contracts' (November 2019); the UKJT statement has been endorsed by the English case law namely AA v Persons Unknown [2019] EWHC 3556, paras.57-59.

107 Cf Walter, (2019) (n 100), 3611ff.

108 Cf Walter, (2019) (n 100), 3613 ("Crypto tokens are supposed to correspond to cash in terms of their structure and thus to a movable thing").

109 For a critical view on such analogy conclusions from a Swiss perspective, see Enz, (2020) (n 98) 291, 293f; ibid, *Kryptowährungen im Lichte von Geldrecht und Konkursaussonderung* (Zürich, 2019), paras.345ff.

110 On Swiss property law, cf. B Graham-Siegenthaler & A Furrer, 'The Position of Blockchain Technology and Bitcoin in

the specific differences between the objects of two completely different media worlds.

57 In general, depending on the contingencies of the legal dogmatics of national legal systems, crypto tokens could be regarded as "other objects".[111] The conceptual commonality of "other objects" and things consists only in the fact that both types of legal objects are goods that differ precisely with regard to the characteristic of corporeality. As "incorporeal goods", "other objects" thus include all goods not covered by property law. These include, among others, intangible goods regulated by special law, but also unprotected inventions, technical know-how, as well as digital data and virtual goods.[112] In this respect, it appears obvious to also consider crypto tokens as "incorporeal goods",[113] which thus would find their place beyond objects and rights as legal objects *sui generis*.

58 Of course, this construction cannot hide the fact that the classification of property found in this way has its origin in legal relationships based on the law of obligations.[114] It would therefore be only logical that it is sometimes addressed with the rather vague term of "other property rights".[115] Thus, on the basis of "proven dogmatics", an attempt is made[116] to treat crypto tokens *de lege lata* in the context of claims under the law of condemnation as suitable objects of unjust enrichment[117]or as "other rights" protected in tort.[118]

59 Here, however, a clear distinction must be made between property law and personality law justifications of tort protection.[119] While the "guarantee of confidentiality and integrity of information technology systems" developed by the German Federal Constitutional Court[120] under personality law also forms a corporeal object of protection for tort law,[121] the more extensive classification of property-like rights of control over data or data files as "other rights" encounters some concerns.[122]

60 At the least, a corresponding protection of property-like data without attribution to personal rights requires an increased argumentative effort. Apart from the controversial discussion about a supposed new *right to one's own data*, it must be borne in mind that such *data ownership*[123] denotes something fundamentally different from data protection derived from personal rights. Even more far-reaching attempts to extend tort protection to individual units of crypto tokens may therefore seem rather far-fetched.[124] Such approaches, as well as the many other inadequate attempts at analogies, functional equations or equivalences, make clear that the current "private law system is completely undeveloped with regards to blockchain technology".[125] The widespread idea that civil law could also "keep up with the developments of modern technology"[126] in this respect would only appear as a continuation of the mantra constantly repeated in civil law that a mature jurisprudence will no longer be embarrassed by history.[127]

Swiss Law' (2017) Jusletter 8.5.2017, paras.42ff.

111 See German Civil Code "BGB", Section 453(1).

112 See, with further examples, A Peukert,'"Sonstige Gegenstände" im Rechtsverkehr' in S Leible, M Lehmann & H Zech (eds.), *Unkörperliche Güter im Zivilrecht* (Tübingen 2011), 95ff; referring to Beck & König, (2015) (n 102), 132f.

113 Cf Beck & König, (2015) (n 102).

114 In this respect, the Roman legal concept of *res corporales* (Inst. 2.2.; Dig. 1.8.1.) appears as a possible equivalent precisely because of its open inclusion of rights; differently Peukert, (2015) (n 112).

115 See Langenbucher, (2018) (n 92), 407; G Spindler & M Bille,'Rechtsprobleme von Bitcoins als virtuelle Währung' (2014) 68 WM, 1357, 1360.

116 Langenbucher, (2018) (n 92), 407ff.; Spindler & Bille, (2014) (n 107), 1363.

117 See German Civil Code "BGB", Section 812; Swiss Code of Obligations "OR", Art. 62ff.

118 See German Civil Code "BGB", Section 823(1); Swiss Code of Obligations "OR", Art. 41(1).

119 Equally unclear in this regard Langenbucher (n 92); Spindler & Bille, (2014) (n 115); cf. also G Spindler, 'Digitale Wirtschaft – analoges Recht: Braucht das BGB ein Update?' (2016) 71 JZ, 805, 813f.

120 German Federal Constitutional Court, BVerfGE 120, 274.

121 Cf M-C Gruber, *Bioinformationsrecht. Zur Persönlichkeitsentfaltung des Menschen in technisierter Verfassung* (Tübingen 2015), 158ff.

122 See eg A Spickhoff, 'Der Schutz von Daten durch das Deliktsrecht' in S Leible, M Lehmann & H Zech (eds.), *Unkörperliche Güter im Zivilrecht* (Tübingen 2011), 233ff, 243ff.

123 For a profound consideration based on legal theory, see M Amstutz, 'Dateneigentum. Funktion und Form' (2018) 218 AcP, 439ff.

124 In this sense Omlor, (2019) (n 99), 310.

125 See, with regard to German civil law, ibid; cf also Spindler, (2016) (n 119), 816.

126 Walter, (2019) (n 100), 3609.

127 Cf M-C Gruber, 'Futurities of Law. Versuche über die Zukunft

61 After the legal perspective of traditional legal doctrine has to give up its claim to develop the digital law of the future as the only authoritative perspective of observation, it will no longer be able to unilaterally determine the legal quality of crypto tokens. It would then no longer be necessary to ask primarily whether they correspond to (personal) property, whether they are similar to "coinage" or "cash", or even comparable to "forces of nature" or "energy", whether they are more like claims or securities, or to what extent they come closest to new types of intangible property.

62 Answers to the legal questions must therefore rather be sought where the new legal phenomena unfold, namely in the *second, digitalised* legal world itself. From such a perspective, primarily the specific characteristics of crypto tokens as to their substance ought to be worked out in order to draft the appropriate, independent standardisations on this basis.

63 In this context, a number of recent bespoke regulatory developments aim in principle at bringing clarity to crypto token legal classifications. Here, the new legal objects *sui generis* could be anchored, for instance, in separate civil law provisions as rival, exclusively assigned "register value data",[128] as "register value rights"[129] or as "property-like legal assets"[130], among others.

64 For the purposes of this paper, and in the context of Diem and alongside potential implications thereof, the recent regulatory developments in the jurisdictions of Switzerland, the EU and the USA have primarily been put under scrutiny.

II. Developments as to Classifications

1. Switzerland

65 The Swiss FINMA has categorised[131] crypto tokens in four groups of a) utility tokens, b) payment tokens, c) asset tokens and d) hybrid tokens. Payment tokens are means of payment, lacking any further function or link to any other development project, whereas utility tokens are those intended to provide digital access to an application or service. Asset tokens refer to underlying physical assets, company equity and rights such as dividends and interest payments, while hybrid tokens are a combination of any of the above. FINMA has also recognised[132] the emergence of stablecoin models, and in their classification, the authority has reiterated the view that *substance* matters over *form.*

66 At first glance, Diem would take a hybrid format seemingly catching features from at least two of the above categories, namely payment and asset tokens. The single fiat currency stablecoins would each be backed by a fiat currency, whereas the intra-network Diem token would act as a "digital composite" of some of those stablecoins, and would be backed by a basket of multiple fiat currencies and other assets. This would be in line with the definition of an asset token. Aimed as a complementary payments system, Diem's underlying purpose has been to provide for an alternative means of payment. Respectively, in order to be granted access to the Diem infrastructure and utilise its applications, end users would in principle need to acquire Diem tokens. Notably, purely utility tokens do not in general embody any financial purpose.

67 As a result, it seems that Diem's two-fold design incorporates characteristics from asset tokens, payment tokens and, partially, from utility tokens. In addition, Diem's two-fold design can be seen to derive its value from the underlying referenced fiat currencies and other assets. Consequently, Diem could be considered as a form of security under Swiss law, when defined[133] as a derivative or a financial contract, the price of which is set particularly according to a) assets such as shares, bonds, commodities etc., and b) reference values such as currencies, interest rates etc.

68 It can then be argued that the liability to comply with potential conversion claims by the token holders remains with the Diem Association and its subsidiaries. In this context, irrespective of contractual exonerations, it would be erroneous to consider intermediaries such as the third party service providers as independent actors, rather than agents.

69 Notably, Diem would only assume functionality by means of the underlying (implied) right to claim fiat currency or other assets. This in itself would then represent Diem as the effective embodiment of an

des Rechts' (2021) 107 ARSP (forthcoming).

128 Cf Enz, (2020) (n 98), 295.

129 See section C.II.1.

130 See, with regard to German civil law, Omlor, (2019) (n 99), 341, considering the insertion of a new Section 90b in German Civil Code "BGB".

131 FINMA, (2018) (n 54).

132 FINMA, (2019) (n 56).

133 Financial Market Infrastructure Ordinance (FMIO) (25 November 2015), Art. 2.2 (a)(b).

uncertificated security, issued by and subjected exclusively to the rules of the Diem network, thus rendering and mimicking a transfer system used for payment claims against debtors.

70 Recently, Swiss law has undergone a legislative reform process[134] that permits the exchange of asset tokens as uncertificated securities. This specific category of tokenised rights,[135] defined as uncertificated register securities,[136] and their legal transfer thereof, would therefore serve relevance in the context of Diem. The new laws will impose an obligation against the crypto token issuer, whereby holders would be given legal certainty in terms of the effect of disposal of the rights embodied in such tokens. Also, certain security standards by way of appropriate technical and organisational measures would need to be met by an underlying DLT system upon which the entries will be appended. The system would need to show resistance to manipulation, and be designed in such a way that no unauthorised intervention would be possible, in particular by the system operators and the third party service providers.

71 On the other hand, under Swiss law the validity of the underlying transaction is required in order for the disposal of a right or asset to have any legal effect. Under the principle of causality, therefore, Diem token holders would need to be able to demonstrate their legal status as holders, independently from any third party such as the third party service providers. This would imply the holder exercising a certain control over digital identifiers associated with the corresponding Diem tokens. Here, verification of such control would be available to any potential new beneficiary, without the need for the register on the Diem DLT network to be publicly accessible. This requires a particular identity management mechanism as addressed in the subsequent section.

72 The new laws in Switzerland also introduce a legal framework[137] for segregation of crypto-assets from third party service providers who provide custodial services. On the Diem network, dealers, VASPs and Novi digital custodial wallet must therefore undertake to keep the assets of third party clients available for those particular clients *at all times*.

73 Nevertheless, as mentioned, Diem Reserve seems to be fully backed by cash or cash equivalents and very short-term government securities, which can be assumed not to consist of segregated accounts specifically referring to identifiable token holders.

2. European Union

74 In the EU, the European Commission[138] has recently put forward a proposal for a Regulation on Markets in Crypto-assets (MiCA). Distinction is made between three sub categories of crypto-assets. These include a) utility tokens which provide digital access to a good or service, accepted only by the issuer of that token without a financial purpose and related to the operation of a digital platform; b) asset-referenced tokens which aim at maintaining a stable value by referencing several currencies that are legal tender,[139] one or several commodities, one or several crypto-assets, or a basket of such assets, often for the purpose of a means of payment to buy goods and services and to transfer value; and c) e-money tokens which are intended primarily as a means of exchange by referencing only one fiat currency that is legal tender, with a function arguably similar to that of electronic money (e-money).[140]

75 Notably, e-money tokens bear close similarities to e-money on the grounds that the holders of both

134 Swiss DLT Framework, parliamentary approval of 25 September 2020 <https://www.admin.ch/opc/fr/federal-gazette/2020/7559.pdf>; Note: The amendments to the Swiss Code of Obligations, the Federal Intermediated Securities Act and the Federal Act on International Private Law that are envisaged in the DLT bill have now enter into force from 1 February 2021. These provisions enable the introduction of ledger-based securities that are represented in a DLT. The remaining provisions of the DLT bill have entered into force as of 1 August 2021.

135 See Art. 973d – 973i, Swiss Code of Obligations (CO).

136 See reference to the term "Registerwertrechte"; CMS Law –Now, 'The new Swiss blockchain/DLT laws have been finalised and presumably entre into force early 2021' (15 October 2020) with reference to "uncertificated register securities have features largely analogous to traditional certificated securities. Any right that can be securitised also qualifies as an underlying right for uncertificated register securities, including asset tokens and utility tokens." < https://cms.law/en/che/blogs/law-now-blog/the-new-swiss-blockchain-dlt-laws-have-been-finalised-and-presumably-enter-into-force-early-2021>.

137 See the amended Art. 242a., Swiss Debt Enforcement and Bankruptcy Law (DEBL).

138 COM (2020) 593 final (n 53), Art.3; Recital 9.

139 See n 95.

140 In the EU, e-money is regulated under the Directive 2009/110/EC of the European Parliament and of the Council on the taking up, pursuit and prudential supervision of the business of electronic money institutions amending Directives 2005/60/EC and 2006/48/EC and repealing Directive 2000/46/EC (e-money Directive) (2009) OJ L 267/7.

would by default be entitled to a claim[141] against the issuing institution. Specifically, subject to a contractual right, e-money and e-money tokens are redeemable at any given moment against fiat currency as legal tender at par value.

76 In this respect, MiCA further makes a specific classification of *significant* asset-referenced tokens and *significant* e-money tokens. For a crypto-asset to be considered significant,[142] a number of variables such as the underlying network's customer size, value or market capitalisation, number or value of transactions, size of reserve assets, significance of cross-border activities, as well as the interconnection with the financial system, would be decisive.

77 Furthermore, the ECB has taken a rather exclusive approach[143] by denoting a crypto-asset as "any asset recorded in digital form that is not and does not represent either a financial claim on, or a financial liability of, any natural or legal person, and which does not embody a proprietary right against an entity." With this take, the ECB seems to associate the risk profile of crypto-assets with the lack of an underlying claim or liability.

78 Whether the two-fold design of Diem would by default confer a redemption right at par value or a claim against the operating subsidiary, or the entity mandated for minting and burning Diem tokens, in favour of respective holders would play a decisive role as to the potential implications under MiCA, respectively the e-money Directive.[144]

79 Variables such as the potential customer size of the Diem network, its value and market capitalisation as well as the significance of its cross-border activities, among others, could render the project *significant* under the MiCA definition. Both Diem token models would by definition be caught under MiCA's two categories of asset-referenced tokens, respectively the e-money tokens.

80 Within the possible scope of applicability of the e-money Directive in the context of the Diem's single fiat currency stablecoins, the rules laid down in the Payment Services Directive (PSDII)[145] may also become increasingly relevant, in particular from the perspective of consumer protection as to, among others, the obligation to safeguard[146] end users' funds.[147] This obligation would be effective immediately on receipt of funds by payments institutions as well as e-money institutions.

81 In addition, it would be feasible to consider Diem under the PSDII definition of payment instrument[148] denoting a personalised set of procedures agreed between the payment service user and the payment service provider, used in order to initiate a payment order. This argument can be substantiated by the fact that the two-fold Diem design will be considered as a combination of *significant* asset-referenced and e-money tokens. As explained, the *significance* relates to the worldwide reach Diem will have based on the existing network built by its founding members. One of the consequences of this would be that with the identity management system deployed by Diem, there could be a competitive advantage in its favour in consideration of the account data portability[149] facilitated under PSDII.

82 On the other hand, any crypto-asset that would fall within the remit of the definition of a financial instrument would be subject to the EU Markets in Financial Instruments Directive (MiFID II).[150] A financial instrument[151] can take the form of, among others, a transferable security or a unit in a collective investment undertaking.

141 Ibid Art. 2(2); Art. 11; COM (2020) 593 (n 53), Art. 44.2, Art. 44.4.

142 Ibid Art. 39.1; Recital 41f; Art. 50.1; Recital 49.

143 ECB Crypto-Assets Task Force, 'Crypto-Assets: Implications for financial stability, monetary policy, and payments and market infrastructures' (May 2019), 7f.

144 See n 140; for the latest developments see also n 13, 79.

145 Directive (EU) 2015/2366 of the European Parliament and of the Council on payment services in the internal market (PSDII) (2015) OJ L 337/35; Recitals 24-25; see also European Consumer Organisation (BEUC), 'Crypto-assets: BEUC response to the Commission's consultation' (13 May 2020), 7f.

146 PSDII (n 145), Art. 10.

147 See PSDII definition of the term 'funds' as "banknotes, coins, scriptural money or e-money within the meaning of the e-money Directive", Art. 4(25).

148 PSDII (n 145), Art. 4(14).

149 Ibid Art. 66 & 67.

150 Directive 2014/65/EU of the European Parliament and of the Council on Markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (MIFiD II) (2014) OJ L 173/349; European Commission has recently proposed a reform of the definitions of this directive in order to include those financial instruments that are issued utilising DLT; COM (2020) 596 final.

151 Ibid Art. 4.1(15).

83 Furthermore, from the perspective of the applicability of the EU anti-money laundering (AML) regime,[152] the definition of the term virtual currency becomes essential. Here, virtual currency is described[153] as a digital representation of value that is not issued or guaranteed by a central bank or a public authority, one that is not necessarily attached to a legally established currency and does not possess a legal status of currency or money, but is accepted by natural or legal persons as means of exchange that can be transferred, stored and traded electronically.

84 Diem is likely to fall under the scope of 5AMLD, due to the fact that the form of attachment to "a legally established currency" would be dependent upon the governance of a private entity. Thus, any actor that provides custodial wallet services, safeguards private keys and engages in exchange services between Diem tokens and fiat currencies would fall under the category of obliged entities and be subject to due diligence, disclosure and supervisory requirements. These actors are Novi Financial and the associated designated entities, i.e. dealers and VASPs.

3. United States of America

85 In the USA, a draft federal bill was introduced in Congress, known as Stablecoin Tethering and Bank Licensing Enforcement or the Stable Act.[154] The term stablecoin[155] was defined as any cryptocurrency or other privately-issued digital financial instrument that a) is directly or indirectly distributed to investors, financial institutions, or the general public; b) is denominated in or pegged to the US Dollar (USD), or to any other national or state currency; and c) is issued with a fixed nominal redemption value, with the intention[156] of establishing a reasonable expectation or belief among the general public that the instrument will retain a nominal redemption value that is so stable as to render the nominal redemption value effectively fixed.

86 Given that the initial phase of the Diem project, once launched, would take the form of a single USD dollar-backed stablecoin, the extensive licensing regime set to be introduced by the Stable Act, if and once enacted, would therefore be relevant.

87 FinCEN has also proposed[157] implementation of stricter AML requirements for certain transactions that involve convertible virtual currency (CVC)[158] or digital assets with legal tender status (LTDA). Under the proposal, banks and MSB licensees would be required to verify the identity of their customers and keep record of transactions and counterparties in relation to transactions above certain thresholds that involve either a) unhosted wallets; or b) hosted wallets where a given transaction would be greater than USD 3,000.

88 As mentioned, Novi Financial is a US-registered MSB and would be the main user interface on the Diem network acting as a digital custodial (hosted) wallet. The Diem network would also support the integration of unhosted wallets,[159] albeit with limited threshold as to transaction volume and account address balance. FinCEN rules, if and once passed, would certainly have implications on Diem, in particular from the perspective of the network's identity management.

89 Moreover, regarding potential risks associated with DLT-based transactions involving digital asset securities, as well as custodial services in digital asset securities provided by dealers and brokers, the US Securities and Exchange Commission (SEC) has issued

152 In the EU, the anti-money laundering regime is regulated under the Directive (EU) 2018/843 of the European Parliament and of the Council amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing (5AMLD) (2018) OJ L 156/43.

153 Ibid Art. 3(18).

154 116th Congress, 'Discussion Draft of 'Stablecoin Classification and Regulation Act of 2020" (19 November 2020) <https://tlaib.house.gov/sites/tlaib.house.gov/files/STABLEAct.pdf>.

155 Ibid Sec. 3(a)(aa)1.

156 Ibid "... or in such a manner that, regardless of intent, has the effect of creating a reasonable expectation or belief among the general public that the instrument will retain a nominal redemption value that is so stable as to render the nominal redemption value effectively fixed."

157 FinCEN, 'FinCEN Proposes Rule Aimed at Closing Anti Money Laundering Regulatory Gaps for Certain Convertible Virtual Currency and Digital Asset Transactions', Press Release (18 December 2020); Federal Register, 'Requirements for Certain Transactions Involving Convertible Virtual Currency or Digital Assets', Proposed Rule (23 December 2020).

158 See FinCEN's definition of the term 'virtual currency' as "a medium of exchange that can operate like currency but does not have all the attributes of "real" currency, including legal tender status; CVC is a type of virtual currency that either has an equivalent value as currency, or acts as a substitute for currency, and is therefore a type of "value that substitutes for currency.""; FinCEN Guidance (FIN 2019 –G001, 9 May 2019), 7.

159 See n 84.

a statement[160] relating to disclosure requirements in favour of customers, among others. In doing so, SEC has referred to the Rule 15c3-3[161] whereby segregation of customer securities (and related funds) would need to be ensured by dealers and brokers by maintaining physical possession or control over customer's fully paid and excess margin securities.[162] Establishing a control mechanism is particularly important in the context of digital asset securities that are issued and transferred via DLT. Such would take the form of effective maintenance of private keys and ensuring the authenticity of the recipient address prior to a digital asset transfer transaction via smart contract codes.

90 In the context of the Diem network, this control mechanism threshold would have significant implications for dealers, VASPs, and the Novi digital custodial wallet.

4. Public-Private Partnership

91 In the context of central bank[163] issued digital currencies (CBDCs)[164], these are kept specifically outside the scope of both legislative proposals, namely MiCA in the EU, respectively the Stable Act in the USA. Relevantly, the European Parliament[165] has reflected upon a need for public-private cooperation in the context of the future of money creation.

92 With reference to the concept of synthetic central bank digital currencies (sCBDCs),[166] the European Parliament takes the view that sCBDCs would have a number of advantages over CBDCs. These include[167] a) lower initial and maintenance costs, b) regulation of private stablecoin issuers by central banks, and c) lower reputational risk for central banks, given that central banks would continue focusing on their primary mandate, namely maintenance of price stability.

93 The European Parliament's stance on favouring public-private cooperation seems to tie in well with Diem's intention to eventually integrate CBDCs into its infrastructure.

D. Diem's Prospects of Trust

I. Digital Livelihoods in a "Vibrant Ecosystem"

94 If Diem's vision of *the internet of money* as a vibrant ecosystem is taken seriously, the requirements associated with it will also take on considerable significance. In this sense, digital living spaces are to be understood not only economically, but above all ecologically. What is required then, for one thing, is free and equal access to the global monetary and financial infrastructure, consequently conceived as a *public good*. And, secondly, it is of central importance to ensure the necessary trust in the functioning of the systems involved, which has a direct link to transparency in governance. From a legal perspective, therefore, what is needed is essentially the guarantee of legal certainty, the clear attribution of responsibilities, the determination of liability

160 SEC, 'Custody of Digital Asset Securities by Special Purpose Broker-Dealers' (23 December 2020) <https://www.sec.gov/rules/policy/2020/34-90788.pdf>; reference to the definition of the term 'digital asset' "...an asset that is issued and/or transferred using distributed ledger or blockchain technology ("distributed ledger technology"), including, but not limited to, so-called "virtual currencies," "coins," and "tokens."

161 Securities Exchange Act 1934, Rule 15c3-3 (Customer Protection Rule).

162 Ibid 17 CFR 240.15c3-3(b)(1).

163 Or any public authority acting in the capacity of monetary authority.

164 European Parliament, 'Public or Private? The Future of Money' Monetary Dialogue Papers (December 2019), 17, <https://www.europarl.europa.eu/cmsdata/207653/13.%20PE%20642.356%20DIW%20final%20publication-original.pdf>; "just like paper currency and coins, CBDC would be fixed in nominal terms, universally accessible, and valid as a legal tender for all public and private transactions. As with any public currency, the objective of the central bank would be that CBDC fulfil its efficiency as a medium of exchange, its security as a store of value, and its stability as the unit of account for economic and financial transactions."

165 Ibid 17f "...the main difference between CBDC and sCBDC is who maintains the end relationship with the customer: for CBDC, this is the central bank, while private entities maintain the end relationship with customers with sCBDCs."

166 T Adrian, T Mancini-Griffoli, 'The rise of digital currency' (9 September 2019) <https://voxeu.org/article/rise-digital-currency>; with reference to the proposed definition of the term 'sCBDC': "In the sCBDC model – which is a public–private partnership – central banks would go back to focusing on their core function: providing trust and efficiency by means of state-of-the-art settlement systems. The private sector – stablecoin providers – would be left to satisfy the remaining steps under appropriate supervision and oversight, and focus on their own competitive advantage – innovating and interacting with customers."

167 European Parliament (2019) (n 164),18f.

rules and, last but not least, the establishment of legal enforcement mechanisms *by design*.

95 The Diem Association justified its project with the noble goal of opening up access to financial services, especially for people in developing and emerging countries. However, this access will by no means be *free* in every respect. On the contrary, it will have its price.

96 As part of the digital, data-driven platform economy, as mentioned, Diem would contribute to an even further expansion of the Facebook empire. In Diem's single fiat currency stablecoin model, the value of which would be linked to single fiat currencies that are legal tender, identification of end users will be mandatory in numerous jurisdictions under the internationally established Know Your Customer (KYC) rules and national AML laws. In this way, mandatory legal standards would presumably help Facebook et al. to identify its now approximately 2.45 billion monthly active users and to track their business and social behaviour almost seamlessly.[168]

97 Finally, the effects of recent court decisions authorising Facebook to prohibit the use of pseudonyms[169] and thus to impose a real name requirement on its users are dramatic.[170]

98 The involuntary complicity of legislators and courts with Facebook does not only bypass privacy and data protection that is apparently considered obsolete. The consequences go much deeper, whereby the complete identification of all users and transaction information would create a comprehensive database to equip adaptive algorithms and artificial intelligence (AI) with the necessary training data and enable them to analyse, imitate and predict human behaviour. Therefore, it can be assumed that the Diem project is not primarily about building an efficient alternative financial system, but primarily about economic profit and further monopolisation of the data-driven platform economy.[171] Moreover, the suspicion is raised that Diem, as part of the digital platform economy, could be just another "colonisation project from Silicon Valley".[172]

99 Consequently, Diem's chances will depend in particular on its prospects of gaining trust as a new, alternative digital form of *private* currency alongside the established monetary systems. This would require a number of constituents, such as comprehensive accessibility and trustworthiness based on legal certainty, clear attributions of responsibility, appropriate models of liability, and effective legal mechanisms of enforceability.

II. Accessibility & Trustworthiness

100 The right to equal access to the global monetary and financial infrastructure presupposed by Diem serves not only as an individual fundamental right, but also, in an institutional sense,[173] as a necessary functional condition of social life in the digital medium. Comparable to other public goods and natural livelihoods, it presupposes the guarantee of a digital living space.

101 In order to enable action and decision making in this *bio-digital ecosystem*, it is generally necessary, as it corresponds to the *nature* of the world according to Niklas Luhmann, to stabilise behavioural expectations, i.e. to establish certainty of expectations and trust.[174] As "an elementary fact of social life",[175] trust creates the basis for "living and acting with greater complexity in relation to events":[176] "Where there is trust, there are more opportunities for experience and action".[177] In this respect, money is one of what Luhmann calls social mechanisms "that allow us to postpone decisions and yet already ensure them, that is, to live with a future of high, indeterminate event complexity".[178] Therefore, the stabilisation of such mechanisms depends on trust.[179]

168 Cf M Langer, 'Libra und des Pudels Kern' (2019) 14 IRZ 509f.

169 See n 70.

170 Higher Regional Court OLG München, Urteil vom 8.12.2020 (Az. 18 U 2822/19 Pre und 18 U 5493/19 Pre).

171 Cf Langer, (2019) (n 168), 510.

172 O Leistert, 'Hearing in the Digital Agenda Committee of the German Bundestag' (26.9.2019, hib 1052/2019).

173 Cf N Luhmann, *Grundrechte als Institution. Ein Beitrag zur politischen Soziologie* (6th ed. Berlin 2019).

174 See N Luhmann, *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität* (4th ed. Stuttgart 2000), 1ff & 61ff; N Luhmann, *Soziale Systeme. Grundriß einer allgemeinen Theorie* (Frankfurt a. M. 1984), 179ff.

175 Ibid *Vertrauen*, 1.

176 Ibid *Vertrauen*, 18.

177 Ibid *Vertrauen*, 8.

178 Ibid *Vertrauen*, 19.

179 On the fundamental importance of trust for the stabilisation of monetary transactions, see G Simmel, *Philosophie des Geldes* (5th ed. Berlin 1930), ch. 2 III, 151ff, 164ff.

102 However, this is no longer primarily a matter of guaranteeing moral-individual trust in human persons, but rather of *system trust* in the depersonalised functionality and the regular course of communicative and technical processes. In this regard, the law has the special task of generating the necessary social trust in the functioning of information technology and, what is more, of digital institutions. Such a task can only be fulfilled by seeing the possibility of mistrust, likewise dispelling it with adequate model designs.

103 Consequently, the creation of such *socio-technical trust* is one of the most prominent objectives of the so-called DLT laws. For example, the Liechtenstein TVTG Act[180] aims "to ensure trust in digital legal communication, in particular in the financial and economic sector and the protection of users in TT Systems." The trustworthiness of DLT infrastructures does not primarily result from a central or superordinate authority, but from the reliability of the communicative operations in decentralised infrastructures themselves.

104 Hybrid tokens such as Diem's two-fold design would only have a chance of success if it is possible to guarantee a stable, uninfluenceable *source of truth* (beyond central state authorisation in accordance with financial market law)[181] using the means of distributed records. The actors involved in the network must all be able to rely on the fact that the traded crypto tokens are recognised as values in accordance with general expectations. Furthermore, they must be able to trust that all value transfer transactions are factually and legally executed.

105 In this respect, the same prerequisites basically apply to the functioning of hybrid tokens as to the *decentralising mechanism* of money in general, which Luhmann characterises in a corresponding way. According to Luhmann "the mechanism, however, presupposes for its functioning that money itself enjoys trust. The individual must be able to assume that with the money symbol he really holds in his hand the possibilities it promises, so that he can confidently postpone his decision on the final use of the money and enjoy or exploit the complexity of the possibilities represented in it as such in abstract form."[182]

106 In order to put this complexity of the economic and financial system literally into the hands of the participants, corresponding legal modelling is required in addition to technical designs. What is needed are legal models that support the trust of the participants in the sense of "trusting one's own expectations"[183] by mapping their underlying assumptions reliably and consistently, i.e. by making them legally secure.

III. Responsibility, Liability & Enforceability

107 The allocation of liability responsibilities is one of the remaining legal means by which access to the independent digital self-regulation processes can succeed in the form of an *exogenous* influence from outside.[184] As became apparent in the years following the subprime mortgage crisis, attempts to control the digital and financial sector through law have often proved ineffective, at least insofar as only single causal factors of undesirable developments have been made the object of control. Advanced legal concepts must therefore take into account the multiple dynamics, not least the diverse circumvention strategies of the actors involved in these sectors, to the extent that they focus on their "internal constitution".[185]

108 From the perspective of the new *lex cryptographia*, analyses and regulatory approaches of the financial sector come into consideration on the one hand, and, on the other, the specific inherent normativities of the relevant decentralised DLT networks are now to be included.

109 Especially against the background of the discussion in the preceding section as to the functional definition of money in economic theory, it seems apparent that the Diem hybrid tokens satisfy the three inherent constituents of money, albeit issued in the form of *private* currency. As pointed out, with regards to Diem tokens' nature as a *store of value,* such would be dependent upon the trust associated with the survivability of the Diem Reserve.

180 See n 52; the term 'TT Systems' refers to Trustworthy Technologies Systems within the meaning of the TVTG Act.

181 Cf C Zellweger-Gutknecht & RH Weber, 'Private Zahlungsmittel und Zahlungssysteme. Auf dem Weg zu neuen digitalen Geldordnungen' (2020) Jusletter 11.1.2020, paras.30ff.

182 Luhmann, *Vertrauen* (2000) (n 174), 63.

183 Ibid 1.

184 On the need for an externally compelled self-limitation of the "capillary constitution", see in particular G Teubner, *Constitutional Fragments: Societal Constitutionalism and Globalization* (Oxford 2012), 73ff.

185 See, especially with regard to financial crises, G Teubner, 'A Constitutional Moment? The Logics of "Hitting the Bottom"' in P Kjaer, G Teubner & A Febbrajo (eds.), *The Financial Crisis in Constitutional Perspective: The Dark Side of Functional Differentiation* (Oxford 2011), 3ff, 5ff.

110 Therefore, this form of *private* currency is clearly distinguished from the only forms of money creation in the traditional financial system, namely *public* and *private* money.[186] In all these forms of money in use today, there is already a lack of a separate purpose that could still convey a monetary value. Here, the value of money is derived solely from *socially constituted trust*. At the latest since the end of the binding of public money to the international gold standard,[187] i.e. the covering of banknotes issued by central banks by an adequate stock of gold reserves, it has become apparent in all clarity "how little money is bound in its innermost essence to the physicality of its substrate."[188] This early observation by Georg Simmel is also confirmed today in that money is "entirely a sociological phenomenon, a form of interaction among people". Therefore, "the more condensed, the more reliable, the more easily appealing the social connections, the purer its nature emerges."[189]

111 According to Simmel, it is "the solidity and reliability of social interactions, the consistency, as it were, of the economic circle, which prepares the dissolution of the money substance."[190] Only on this basis can the necessary confidence of economic actors emerge that in daily cash transactions they are not only dealing with pieces of mostly low-value metal alloys or paper, but can pay with them at a "nominal value" - "*non aes sed fides*".[191]

112 Today, the conditions of this confidence have long since ceased to be guaranteed unilaterally by individual state institutions such as central banks. The latter no longer obtain the cover required for the money in circulation only through corresponding gold holdings, but also, for example, through the acquisition of currency reserves, government bonds, securities or refinancing credits.[192] These means of monetary policy have proven to be precarious, especially in the recent past, which has been marked by financial crises.

113 The state as a money and value creator obviously lives on preconditions that it cannot guarantee itself. On the one hand, the central banks' large-scale open market operations are intended to serve them as indirect instruments of control in the sense of "interest rate policy", for example by providing the commercial banks with corresponding credits in the expectation of increasing the overall economic money supply.[193] On the other hand, they lead to the fact that today it is primarily the active commercial banks worldwide that are *de facto* engaged in money creation. In other words, "the widespread circulation of non-cash money in current accounts, the circulation of moneyless payment transactions, the new communication technologies, and - of particular importance - the globalisation of money and capital transactions, have prised the money-creating monopoly from the hands of the national central banks."[194] Here it becomes clear that an *intrinsic value* has long since ceased to be a prerequisite for the concept of money. On the contrary, privatised money creation has virtually developed into a "creatio ex nihilo".[195]

114 It is not possible to go into further detail here on how it comes about that these money creation mechanisms sometimes lead to fatal, crisis-like growth spirals, which are determined by harmful growth pressures, e.g. excessive growth pressures in the real economy on the one hand, and excessive speculative money creation in the financial economy on the other.[196] It should be noted, however, that in order to avoid such self-destructive growth excesses, it is important "to identify the dynamics that accelerate the growth spiral of a social sector to the point where it tips over into destructiveness by colliding with other social dynamics."[197]

115 As mentioned, the underlying value of the two-fold design of Diem hybrid tokens will be derived from the Diem Reserve, which is a reserve of fiat currencies and short-term government securities. But even with such securities, which certainly have complex risks,[198] the necessary mechanisms of currency supply control and guaranteed availability of a counter value backed by liquid assets are in principle safeguarded. The remaining risks of loss are to be reduced by means of a decentralised distribution of

186 See section C.I.

187 For a historical overview, see S Omlor, *Geldprivatrecht. Entmaterialisierung, Europäisierung, Entwertung* (Tübingen 2014), 22ff.

188 G Simmel, (1930) (n 179), ch. 2 III, 156.

189 Ibid.

190 Ibid, 155.

191 Ibid 164, with reference to such an inscription on Maltese coins.

192 Cf Langenbucher, (2018) (n 92), 385, 391.

193 Ibid.

194 See Teubner, (2011) (n 185), 6, with further references.

195 Ibid.

196 Ibid 6ff.

197 Ibid 10.

198 Cf Zellweger-Gutknecht & Weber, (2020) (n 181), para.28.

the assets to a geographically distributed network of custodian banks, which not least also spreads the associated responsibilities for risks accordingly.

116 In comparison, the establishment of the "consistency of the economic circle" in the sense of "solidity and reliability of social interactions"[199] in the crypto network appears to be conceptually more demanding. Here, beyond the *system trust* to be established technically, a fundamental *social trust* is still needed, which could not be replaced by the simple mechanics of a "crypto-proof".[200] However, the "trusted technology"[201] nature of DLT would be fundamentally misunderstood if it were to be reduced to its mathematical operations and *trust-less* characteristics as a "technology of mistrust".[202] Understood correctly, trust in DLT means guaranteeing the socio-technical conditions by means of adequate regulations in order to not only secure value and assets, but also to stabilise behavioural expectations among the acting actors. However, this cannot be achieved by means of state legislation alone.[203]

117 Legal norms should be used here to ensure that crypto tokens such as Diem establish internal self-restrictions in their technical medium that are oriented towards the aforementioned "internal constitution".[204] What this means is impressively summed up by Gunther Teubner. He stipulates that "just as in political constitutions power is used to limit power, so the system-specific medium must turn against itself. Fight fire by fire; fight power by power; fight law by law; fight money by money. Such a medial self-limitation would be the real criterion differentiating the transformation of the 'inner constitution' of the economy from external political regulation."[205]

118 Could these insights be transferred to the creation of a new crypto-constitution? How could the corresponding reflexive self-limiting mechanisms - for instance as limiting constitutional functions of a fight *crypto by crypto* - be set up? Certainly, it has to be kept in mind that money creation must not remain exposed to the unbridled addiction of the global banking market to non-cash money.[206] In this context, crypto tokens can make a productive contribution to the withdrawal of the addictive drug non-cash money by offering a better secured alternative to the *creatio ex nihilo* of current account credit.

119 However, the required security is not only guaranteed by the reserve of assets, which is always emphasised in the Diem project. A complete constitutional *crypto order* also requires the guarantee of autonomy, at least in three respects. These are a) self-regulation of the crypto sector without direct attempts at control on the part of state-institutionalised politics, b) avoidance of one-sided ties to individual forms of value or money of other monetary systems, and c) independence from individual technical operators as well as the infrastructure of social networks set up behind DLT.

120 This would by no means signify leaving the crypto sector to its own devices and placing it in a normatively unregulated state of total anonymity. No one needs to fear being afflicted by the "spectre of crypto anarchy"[207] as long as cryptographically generated value also lives up to its function as a *public good* and justifies the trust that must be presupposed. In addition to the stabilisation of value through an appropriately distributed reserve, this includes a further stabilisation of expectations through reliable allocation of responsibility and liability as well as corresponding enforcement possibilities vis-à-vis the various participating entities of the network.

121 It should be noted that this does not render the abolition of user anonymity. As mentioned, it is not clear as to whether Diem will be set up in the form of an *account-based* private currency issuance, respectively *token-based*. The original advantages of the token-based model could be maintained with the help of identity management that only ever reveals the partial identity of the *data person*, and only within the limited scope of a given transaction. Furthermore, responsibilities can be specifically linked to the corresponding roles of the (non-anonymous) responsible parties and collectives involved in the DLT network.

122 In the Diem network, the Diem Association has

199 Cf Simmel, (1930) (n 179), 155.

200 Cf Langenbucher, (2018) (n 92), 395, with particular reference to N Dodd, *The Social Life of Money* (Princeton & Oxford 2014), 362ff.

201 See n 52 & 180.

202 Cf Dodd, (2014) (n 200), 362; see also n 50.

203 A different view is held by Langenbucher, (2018) (n 92), 395, who sees the success of virtual currencies as "dependent on the societal-state underpinning of trust".

204 See also Teubner, (2011) (n 185), 15: "The task would, with a bit of luck, be to combine external political, legal and social impulses with changes to the internal constitution."

205 Ibid17.

206 See Teubner, (2011) (n 185), 16 ff, with resolute demands for a restoration of the money creation monopoly of the central banks.

207 Cf TC May, *The Crypto Anarchist Manifesto* (1988) <https://www.activism.net/cypherpunk/crypto-anarchy.html>.

been mandated with the general governance of the project. Its subsidiary Diem Networks US, Inc. could be seen as a "collegial institution" as it has been licensed and registered as a MSB by FinCEN, taking the primary operating role of the project at least during the project's initial phase.[208] As such, it could act as a "reflection centre", comparable to central banks in fiat money, in order to advance self-regulation in the sense of the self-rationality and self-normativity of the Diem network and to make it compatible with society.[209] However, this still requires a clear commitment to the function of the new monetary value as a *public good*, i.e. the decisive recognition of the users and other actors involved in Diem as the *public*. This does not mean, of course, that monetary value creation has to be a state matter at the same time. Rather, it belongs in the "public infrastructure of the economic sector" and is a "genuine component of the constitution of the economy because it takes part in determining the public function of the economy".[210]

123 The success of Diem will then depend above all on the extent to which it succeeds in providing forms of expectation, stabilisation and trust with liability and legal protection mechanisms set up specifically in the network, in order to do the best possible justice to the many participants in the network. At least at the beginning of the Diem project, trust will only be granted under legal conditions. This succeeds all the better as corresponding risk assumptions would be legally anchored in the form of liability guarantees. For Diem in particular, it will be crucial to define technical spheres of responsibility within clear boundaries. New liability constructions are required above all where no individually responsible person can be identified, because he or she no longer has sole control over the technical risks in the interplay of powerful artefacts and processes, i.e. where no individual can be expected to take a risk and bear responsibility for it.

124 For this reason, collective risk liability concepts corresponding to the associated network of risk-impacting actors and agents will increasingly come into consideration in the future.

125 What has to be considered then, are models of strict liability of the corresponding risk associations, which are composed, in particular, of the operators as jointly liable according to fixed shares. With its rather complex organisational structure, and as described in the preceding sections, the assignment of operating roles to the actors involved in Diem does not seem straightforward.

126 Overall, at least three spheres of responsibility come into view in Diem's case as far as can be seen from the current state of project planning. Each of these spheres is likely to be associated with different implications under liability law. These include a) *individual liability* of single corporate actors (e.g. designated dealers, VASPs, operators), primarily on the grounds of provable individual misconduct; b) *shared network liability* of validator nodes; and c) *collective fund liability*, governed by Diem Association with Diem Reserve managed by a network of worldwide institutional custodians.

127 Irrespective of how these spheres of risk and responsibility are ultimately structured in terms of liability law, it can at least be stated in general terms that suitable liability models should rely less on individual incentives for lawful behaviour or, in other words, less on negative (monetary) incentives through the threat of damages or compensation for infringing actions.

128 Instead of focusing on acting individuals, liability must also be directed primarily towards the risks of the socio-technical connections in whose interaction infringements of rights occur.

129 Liability responsibility is then no longer primarily based on culpable-causal acts of infringement, but on "infringement structures" that result from socio-technical connections in the sense of "risk associations".[211]

130 In this way, the multitude of damage risks[212] can finally be addressed in a differentiated manner, in particular the possible losses and damage as a result of price or monetary inflation, payment deficits, scarcity of currency, loss of liquidity, but also damages due to violations of the law through data protection breaches, money laundering, criminal financing or fraudulent activities.

131 But even in this respect it remains the case that

208 See Zellweger-Gutknecht & Weber, (2020) (n 181), paras.78ff; for the latest developments see also n 13, 79.

209 Cf Teubner, (2012) (n 184), 24.

210 Ibid 36.

211 For an exemplary legal reconstruction of different forms of risk associations on social networks and trading platforms, see M-C Gruber, 'Legal responsibility of AI in social media and algorithmic trading' in M Jankowska, M Pawełczyk & M Kulawiak (eds.), *AI: Law, Philosophy, and Geoinformatics* (Warsaw 2015), 90ff, 99ff.

212 For an in-depth consideration of these risks, see in particular Zetzsche, Buckley & Arner, (2017) (n 105); cf also DA Zetzsche, RP Buckley & DW Arner, 'Regulating LIBRA: The Transformative Potential of Facebook's Cryptocurrency and Possible Regulatory Responses' (2019) 47 UNSWLRS, 10ff <https://ssrn.com/abstract=3414401>.

state law cannot directly bring about the changes necessary for enforceability in the information technology medium. It cannot determine, control or regulate the normative orders of the *internet of money*. The internet regulates itself. However, this self-regulation of information technology has its limits. In those cases in which, from an internal perspective, seemingly insoluble conflict situations arise between providers and users, so that the necessary trust in the functioning of information technology and even of the institution of information law appears threatened, the specific conflict resolution power of the law is required. Legal enforcement mechanisms ideally provide media feedback that serves the further development in the sense of a constantly improved *compliance by design*,[213] which also includes the well-known concept of *privacy by design* in favour of the users.[214] Furthermore, corresponding enforcement concepts of *liability by design* could be considered, which could be directly inscribed in DLT.

132 The clear allocation of responsibilities to the various actors, as well as the corresponding allocation of liability obligations and their enforceability, can become an existential question for DLT-based crypto tokens such as Diem.[215] To solve this, an identity management system would also be required which would meet legal requirements by technical means. Furthermore, such an identity management system would lay the foundation for technical isolation of the formal content of transaction data from the personal aspect of that data, with individuals and end users given the chance to control what to share, how much and for how long.

E. Identity Management

I. Digital Identification & Representation

133 Digital identification and representation lie at the core of a financial infrastructure operated on DLT, in particular when the primary aim of such an infrastructure is to bring about financial inclusion.

134 Identification is defined as "a process of recognising an entity in a particular domain as distinct from other entities."[216] Identification is seen as an essential process when requesting or accessing a service of any kind. While identity is "a set of attributes related to an entity,"[217] digital identity could simply be defined as "the digital representation of an entity detailed enough to make the individual distinguishable within a digital context."[218]

135 In the absence, and limited uptake of, effective standardisation as well as interoperability among diverse systems, digital identity has continued to be a fragmented development,[219] with pressing issues relating, among others, to security.

136 Digital identity management systems could in principle take various forms among which are centralised, federated, third party identity provider, user-centric and, more recently, self-sovereign identity (SSI). An identity provider[220] is "an entity that makes available identity information." Such information includes not only the creation, maintenance and management of credentials but also the provision of authentication services.[221]

137 With the consumer single sign-on (SSO) identity management of Facebook and its group of social network platforms, the data protection and privacy of end users seemingly remain untested. In the absence of an effective and secure identity management system, Diem, as a digital financial infrastructure, may further aggregate the risk of profiling end users' behaviour online by expanding the scope of reach to payment systems and spending patterns.[222]

138 Facebook's SSO is a common method of authentication of user logins whereby users could utilise their Facebook credentials and connect to other third party service providers. Such a scheme would argu-

213 For a similar concept of "embedded regulation", see DA Zetzsche, DW Arner & RP Buckley, 'Decentralized Finance' (2020) IIEL Issue Brief 02/2020, 51ff <https://ssrn.com/abstract=3539194>.

214 Cf Gruber, (2015) (n 121), 203.

215 Cf Enz, (2020) (n 98), 297.

216 International Organisation for Standardisation (ISO), 'IT Security and Privacy – a framework for identity management – part I: Terminology and concepts' (2019), ISO/IEC Standards No 24760-1.

217 Ibid.

218 International Telecommunication Union (ITU), 'Digital Identity Roadmap Guide' (2018), 4f.

219 See also EU Blockchain Observatory and Forum, 'Thematic Report: Blockchain and Digital Identity' (2019).

220 See n 216.

221 MA Lopez, 'The Future of Identity: Self Sovereignty, Digital Wallets and Blockchain' (2020), LACChain Global Alliance digital identity working group, 16ff.

222 See also Zetzsche, Buckley & Arner, (2019) (n 212), 22ff.

ably[223] increase the risks associated with the creation of a single point of failure. Facebook could therefore be seen as an identity provider with both centralised and third party provider management forms,[224] the latter in the context of Facebook's provision of authentication services through its SSO method.

139 In this respect, the Diem Association is committing itself to a long-term goal of developing and promoting an open identity standard,[225] pointing to *decentralisation* and *portability* of digital identity as prerequisites to financial inclusion and competition.

140 Here, decentralisation would mean that identity data of users, their attributes and identifiers, would be distributed among the running nodes of the Diem DLT network. Portability would mean that credentials and attributes could be moved from one place to another. Neither of these[226] would necessarily imply that users are to maintain effective control over the creation and management of their digital identities and representations. Notably, and in contrast with the open identity standard promoted by the Diem Association, the nodes running on Diem's permissioned DLT network would constitute a rather centralised structure.

141 As previously pointed out, Facebook's Novi as the digital custodian wallet would serve as the main user interface for the Diem network upon which services would be built based on smart contract codes. Novi as a hosted wallet will arguably function as an off-ledger payment mechanism with an obligatory identification system in place, called visual identification. Moreover, only those smart contract codes would be appended on the Diem network that would be pre approved by the Diem Association. The network would allow for pseudonymisation as part of its participation protocol, whereby users would be enabled to hold multiple accounts, which would in return avoid the risk of correlation as to users' activities and profiles. On the other hand, as mentioned, the underlying DLT is set to take the form of a single data structure which would record the history of transactions and states over time, providing for the possibility that all appended data on the network would in theory be visible to all applications.

142 One of the main objectives of Diem Networks as a subsidiary of the Diem Association was the provision of identity management. Furthermore, user interactions on the network would primarily take place through VASPs. These regulated entities would be bound by the travel rule as to beneficiary information disclosures, and would be permitted to record user transactions off-ledger and internally, presumably in their respective central databases.

143 In light of these developments, it would not be far-fetched to take the view that the identity management of Diem may not be of a nature to provide for an effective control in favour of end users as to the creation, management and sharing of their digital representations. Instead, despite the intended application of pseudonymity by the Diem project, going in clear contradiction with the role assigned to VASPs and the travel rule they are bound by, the establishment of correlation between such identifiers and real identities of end users, as well as network participants, would seem inevitable. This rhetoric seems to also tie in well with the growing pressure on social network platforms as to the identification of their users, particularly demonstrated in a recent German higher regional court's decision[227] to authorise Facebook to ban the use of pseudonyms on its platform.

144 Moreover, with regards to the *portability* element of the digital identity standard, put forward as a long-term goal by the Diem Association, such would involve cross-border transactions, including within the EU. Under PSDII,[228] explicit (contractual) consent from payment service users would be in principle required in order to request and obtain access to their transaction data and payment accounts with banks and financial service providers. This would serve relevance to the Diem project, in the context of smart contract code –enabled automated decision making, concerning user transaction data. Under the EU's data protection regime,[229] transaction data would be considered personal data where such information would be attributable to an independent individual. Transaction data could lawfully be processed[230] when necessary for the performance of a contract to which a data subject (payment services user) is a party. Furthermore, lawful processing of transaction data could be justified when necessary for compliance with a legal obligation,[231] laid down

223 See also LH Newman, 'Think Twice Before Using Facebook, Google, or Apple to Sign in Everywhere' (*Wired*, September 2020).

224 See Lopez, (2020) (n 221), 17f.

225 Libra whitepaper v.2.0 (n 32), 25.

226 See also I Allison, 'Buried in Facebook's Libra Whitepaper, a Digital Identity Bombshell' (*Coindesk,* 26 June 2019).

227 See n 170.

228 PSDII (n 145), Art. 64, 66 & 67.

229 See GDPR (n 70).

230 Ibid Art. 6(1)b.

231 Ibid Art. 6(1)c.

by EU law, respectively by the laws of Member States (MS) to which a data controller is subject, among which would be the requirements of the AML regime.

145 As seen, the notion of identity, in particular digital identity, clearly touches upon the legal and regulatory landscape in many respects. In the EU, next to the data protection regime, electronic identification and authentication is regulated under eIDAS[232] which, among others, recognises the use of digital signatures[233] for cross-border electronic transactions. Based on the principle of legally enforceable mutual recognition[234] between MS, eIDAS ensures interoperability by obliging public online services to recognise national electronic identification schemes for authentication purposes. Such has remained voluntary for private online services. With recent developments,[235] which particularly aim at extending the scope of application of eIDAS to the private sector, an EU digital identity scheme (EUid) is set to be introduced. EUid would act as a single sign-on, albeit entirely voluntary, harmonising access to online public and private services, and in principle facilitating anonymous authentication.[236] It is apparent that the relationship between personal identity and authentication[237] mechanisms is becoming increasingly important. In this respect, therefore, any entry appended on DLT would fall under the eIDAS definition of 'electronic document'.[238]

146 Under Swiss law, on the other hand, the Swiss banking sector is subject to compliance with FINMA rules[239] pertaining to the handling of electronic client data in order to ensure confidentiality. Moreover, the Swiss draft eID Act[240] was set to derogate from the traditional issuance of digital identities being conferred to state authorities only, permitting public-private partnership collaborations. In other words, the state would take the role of the issuer and verifier of attributes, whereas the task of authentication of eIDs would be given to the private sector under state supervision. The eID was seen as the key infrastructure element on which further digital services such as, among others, eBanking and eFinance could then be built.

II. A Possible Way Forward: Taxonomy & Basic Definitions

147 The notion of trust as one of the central constituents of almost all industries, including digital financial services, is increasingly transitioning away from purely centralised intermediation by state authorities. As mentioned, in increasingly digitalised societies, the stance of trust as an elementary fact of social life has seen a shift towards augmented reliance on private sector actors.

148 In case Diem is to be eventually rolled out as a private cross-border infrastructure with the alleged aim of ensuring financial inclusion, it is inevitable that the

232 Regulation (EU) No 910/2014 of the European Parliament and of the Council on electronic identification and trust services for electronic transactions in the internal market (eIDAS) (2014) OJ L 257/73; ibid Art. 3(1): "electronic identification means the process of using person identification data in electronic form uniquely representing either a natural or legal person, or a natural person representing a legal person"; ibid Art. 3(5): "authentication means an electronic process that enables the electronic identification of a natural or legal person, or the origin and integrity of data in electronic form to be confirmed".

233 See n 41; eIDAS (n 232) recognises 3 different signatures according to the degree of legal certainty which can be provided. As stipulated in Art. 3(10), (11) & (12) these are 'simple', 'advanced' and 'qualified' signatures.

234 eIDAS (n 232) Art. 6.

235 European Commission, 'Inception Impact Assessment for Revision of the eIDAS Regulation – European Digital Identity (EUid)', Ares (2020) 3899583 <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=cellar:35274ac3-cd1b-11ea-adf7-01aa75ed71a1>; European Commission, 'Proposal for a Regulation amending Regulation (EU) 910/2014 as regards establishing a framework for a European Digital Identity, COM (2021) 281 final.

236 For example, in cases where user identification is not required for the provision of services.

237 For a complete overview of eIDAS see IA Domingo (on behalf of European Commission), 'SSI eIDAS Legal Report: How eIDAS can legally support digital identity and trustworthy DLT-based transactions in the Digital Single Market' (2020) <https://joinup.ec.europa.eu/sites/default/files/document/2020-04/SSI_eIDAS_legal_report_final_0.pdf.>.

238 eIDAS (n 232) Art. 3(35): "electronic document means any content stored in electronic form, in particular text or sound, visual or audio visual recording"; ibid Art. 46 on legal effects of electronic documents: "an electronic document shall not be denied legal effect and admissibility in legal proceedings solely on the grounds that it is in electronic form".

239 KPMG, 'FINMA circular 2008/21 Operational Risks – Banks' (2014), Appendix III, 27.

240 Federal Act on Electronic Identification Services "eID Act/E-ID-Gesetz, BGEID" <https://www.admin.ch/opc/de/federal-gazette/2019/6567.pdf>; note: on the Referendum of 7 March 2021 the Swiss electorate rejected the proposal by 64.4%.

identity management scheme of the project requires a design that would ensure end users maintain an effective and sovereign control over their digital representations on the network.

149 As a result of technological advancements, reliance on third party public or private intermediaries for the provision of verification and validation services, in particular in the context of identity creation and the management of attributes, claims and credentials, could in principle become redundant. Disintermediation in the provision of identity services could therefore place individuals in the driving seat as identity providers.

150 Labelled as self-sovereign identity (SSI), this mechanism could be defined as a "digital movement that recognises an individual should own and control their identity without the intervening administrative authorities. SSI allows people to interact in the digital world with the same freedom and capacity for trust as they do in the offline world."[241] An SSI based identity management would imply a set of inherent principles.[242] These include a) access, b) consent, c) control, d) existence, e) interoperability, f) minimalisation, g) persistence, h) protection, i) portability, and j) transparency.

151 In other words, individuals (and entities) as the sole controllers of their digital identities must have access to their own data, exercise control and agree to its usage. The created identities must be long lived, widely available, usable and transportable. The rights of individuals must be preserved, respectively data disclosure must be minimised and done selectively on a *need-to-know* basis. The systems and infrastructures upon which SSI is built would need to be open and transparent as to their operation and management.

152 In this context, individuals (and entities) as their own identity providers are referred to as principal, subject or holder.

153 Central to the functionality of SSI architecture are decentralised identifiers (DIDs). A DID[243] is defined as a new type of globally unique identifier specification that is portable and rooted in a public source of truth such as DLT, a database, a distributed file system or a similar system. Such specification does not require a centralised authority to create, register, resolve, update or revoke the identifiers.[244] Ownership of DIDs could be authenticated and verified cryptographically, i.e. via digital signatures.[245]

154 As identifiers, DIDs do not carry information about the principal. Every DID is accompanied by a descriptor object known as a DID document or DDO. DDO is a machine readable document containing information about verification keys and proof of ownership of the associated DID, among others. Moreover, DID Methods are mechanisms by which a particular DID and its associated DDO is created and resolved.[246] Notably, DIDs are not always dependent on a DLT protocol for their creation. Depending on method specifications, DIDs could take the form of DLT agnostic, yet in principle interoperable with DLT infrastructures,[247] such as peer DIDs.[248] Peer DIDs could be "created and maintained for an entire lifecycle without any reliance on the internet, with no degradation of trust."[249]

155 Given that the principal or the subject would maintain control over the creation of their DIDs, it is in principle possible that multiple DIDs are generated by one principal or subject for different relationships, in turn providing for correlation resistance in the context of their digital representation.

156 DIDs could technically be created in different formats,[250] namely anywise, pairwise and N-wise DIDs. Anywise DID could be used with an unknown

241 Sovrin.org, 'What is self-sovereign identity?' (2018) <https://sovrin.org/faq/what-is-self-sovereign-identity/>.

242 C Allen, 'The Path to Self-Sovereign Identity' (2016) <http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html>; see also n 221, 26ff.

243 W3C, 'Decentralised Identifiers (DIDs) v1.0, Core architecture, data model, and representations' (2021), Working Draft 20 January 2021 < https://www.w3.org/TR/did-core/>.

244 R Soltani et al., 'A New Approach to Client Onboarding using Self-Sovereign Identity and Distributed Ledger' (IEEE Int. Conf. on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2018), 1131ff.

245 See n 44.

246 See n 243.

247 This process is known as 'grafting'; in other words "...because peer DIDs are globally unique at the moment of creation, their numeric basis will not exist on any other blockchain unless someone copies it there. Blockchain-based DID methods can therefore (redundantly) register a peer DID doc using their own method."

248 W3C, 'Peer DID Method Specification, blockchain-independent decentralised identifiers' (2020), W3C Document 25 August 2020 <https://identity.foundation/peer-did-method-spec/#overview>.

249 Ibid.

250 Ibid.

number of parties, while pairwise DID would be used only between the principal and one other party. In N-wise format, the number of parties could be defined in accordance with a given context.

157 As mentioned, DIDs could be created via different method specifications defined in DID Methods. In order to ensure interoperability among these specifications, certain recent developments are of significance, namely the Universal Resolver[251] tool as a unified interface upon which any kind of DID could in theory be resolved.

158 With respect to the SSI identity management, DIDs are components of a larger picture. Here, claims and credentials play a crucial role as to individuals' digital representations and attributes. A claim[252] is defined as "an assertion made about a subject", and a credential is "a set of one or more claims made by an issuer." Credentials could be verifiable, self-asserted, as well as anonymous.

159 A verifiable credential is a data structure that is "tamper-resistant and cryptographically verifiable." In self-asserted credentials, the issuer is the same as the principal or the subject, whereas verifiable credentials are issued by a trusted third party entity

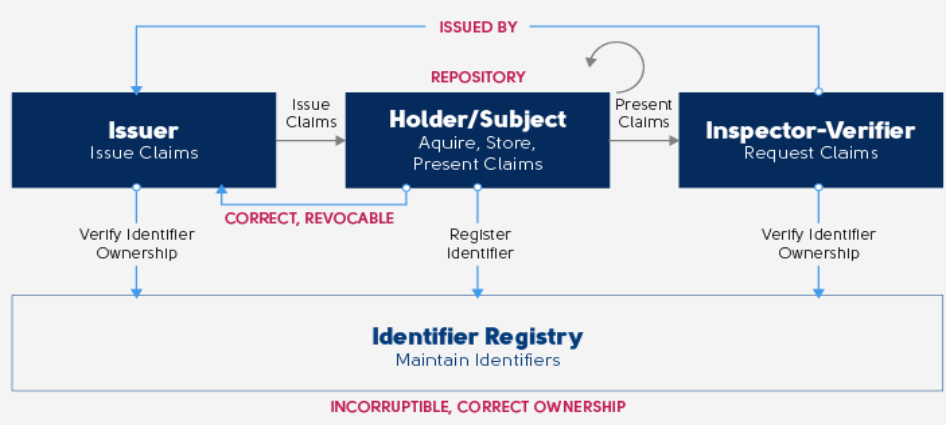


Figure ii: credit: www.luxoft.com/blog

such as a bank or a financial institution. Anonymous credentials[253] refer to data structures created through the means of an algorithmic protocol called zero knowledge proof (ZKP), whereby claims are proven without revealing additional information. This would arguably in return help maintain anonymity by not revealing the underlying identity related data.

160 In a simplified equation, there would be three parties, namely the principal or the subject, the issuer and the verifier. The communication between these parties would be facilitated through software programmes called user agents.[254] Both issuer and verifier are entities mainly responsible for the issuance of credentials requested from them and the reception of credentials presented to them.[255]

161 Verifiable data registry[256] is an underlying system upon which created DIDs are verified and exchanged between parties alongside verification keys and verifiable credential schemas. The Verifiable data registry could be based on a DLT network. Relevantly, a repository is a programme such as a storage vault or wallet which enables the storage of, and secure access to, the verifiable credentials of a principal or subject. Notably, verifiable credentials could be revoked by issuers, respectively deleted by principals or subjects.[257]

162 Digital representation of these actors would be facilitated and secured through encryption schemes such as asymmetric encryption or public key infrastructure (PKI). PKI provides for assignment of key pairs, public and private, to a principal or a subject, with public key being publicly visible and private key remaining under the control of the said principal or subject with which digital signatures would be generated for authorisation and validation

251 M Sabadello (on behalf of DIF), 'A Universal Resolver for self-sovereign identifiers on any blockchain or other decentralised system' (2017) <https://medium.com/decentralized-identity/a-universal-resolver-for-self-sovereign-identifiers-48e6b4a5cc3c>.

252 W3C, 'Verifiable Credentials Data Model 1.0, Expressing verifiable information on the web' (2019), W3C Recommendation 19 November 2019 < https://www.w3.org/TR/vc-data-model/#dfn-credential>.

253 See n 244, 1131ff.

254 See n 252.

255 Ibid.

256 Ibid.

257 Ibid.

purposes. In a PKI infrastructure, the identifier registry is generally managed by a centralised third party such as a certificate authority (CA), who can revoke certificates at any point in time, potentially increasing risks associated with a single *point of failure*. To address this, decentralised public key infrastructure (DPKI)[258] has been developed, whereby the identifier registry takes the form of key value data stores appended on a DLT network or similar systems. DPKI would allow for the principal's identifier to be securely linked to its associated public key.

163 In digital finance, strict KYC and AML requirements make the choice of the identity management mechanism pivotal to the functionality and operation of a given network. Through the means of uniquely assigned DIDs and verifiable credentials, an interoperable and standardised SSI mechanism would facilitate the portability of credentials leading to cost and process efficiency.

164 Furthermore, the personal data protection regimes in the EU, and that of Switzerland, pave the way for a more strict view of digital identity rights of individuals. Here, the principles governing SSI identity management seemingly correspond with the principles introduced by legislation such as, among others, the EU's General Data Protection Regulation (GDPR).[259] These include a) data processing in a lawful, fair and transparent manner, b) purpose limitation, c) data minimisation, d) data accuracy, e) storage limitation, f) data integrity and confidentiality, and more importantly g) data portability, to name a few.[260]

165 Consequently, it is only feasible that a large scale *private* digital financial infrastructure such as Diem implements an effective identity management mechanism, whereby individuals and end users are no longer seen as mere products or an extension of their digital footprints already created elsewhere. Technological developments allow for integration of mechanisms that would in principle limit the ever present collateral damage that is induced on end users by increasing digitalisation in societies.

166 An operational Diem network would be realistic as a complementary financial infrastructure only if its identity management system would provide for the integration of a secure and interoperable SSI mechanism where risks associated with profiling and correlation are minimised and individuals would maintain effective control and confidentiality in relation to their financial and spending behaviour.

F. Concluding Remarks

167 Diem is yet to become formally operational. Any analysis of its technical design and governance infrastructure would therefore need to be solely based on available information to date. Nevertheless, the Diem test network,[261] published late January this year, already documented the interaction of a significant number of addresses with unique identifiers on the network.

168 Diem aims at becoming an alternative worldwide system for digital finance, run and operated on DLT, in order to deliver on the promise of *the internet of money*. A breakdown of Diem's organisational infrastructure revealed that through the bundling of in-house software applications with Facebook's core products, the dynamics of user dependency would inevitably emerge, with Facebook maintaining a certain degree of (indirect) governance and effective control over the project. As argued, the aggregated risk of such a project could render further monopolisation of the data-driven platform economy, potentially leaving its primary purpose as an (efficient) alternative financial system in the cold. Furthermore, due to Diem's anticipated worldwide reach and its projected identity management system, the extent of the network's technological foundation, and its capacity to meet the regulatory obligations of different jurisdictions, in particular in the EU in consideration of the user account data portability facilitated by PSDII, a significant competitive advantage in favour of Diem would then be established. This would be even more prevalent once sCBDCs are introduced positioning Diem in a leading role in the dedicated public-private partnerships.

169 By taking a closer look at the *substance* of the two-fold Diem design and the associated legal implications, it seemed feasible to assume that the design would by definition embed a hybrid nature. Next to regulatory hurdles, as pointed out, the success of Diem will depend above all on the extent to which it succeeds in providing stability and trust with liability and legal protection mechanisms set up specifically in the network. Moreover, an identity management system would need to be in place, effectively meeting legal requirements by technical means. Such a system would then lay the foundation for technical isolation of the formal content of transaction data from the

258 C Allen et al., 'Decentralised Public Key Infrastructure, A White Paper from Rebooting the Web of Trust' (2015) <https://github.com/WebOfTrustInfo/rwot1-sf/blob/master/final-documents/dpki.pdf>.

259 See n 70.

260 See n 244, 1134f.

261 See <https://indiem.info/top#top=balance>.

personal aspect of that data, with individuals and end users given the chance to control what to share, how much and for how long.

170 In other words, for Diem to experience a realistic mass adoption and to serve as a complementary infrastructure to the established monetary systems, it must itself prove to be a constitutive part of the *lex digitalis*. Evolving into the *lex cryptographia*, it will depend on the *pouvoir constituant* of the digital world whether it succeeds in further developing a digital civil constitution in the medium of DLT. Such a constitution, not least with its respective identity management, will determine what human life will be like in a truly *vibrant ecosystem*.

Note: URL links have primarily been accessed within the period of 01.12.2020 - 09.02.2021, excluding those related to Diem's latest developments.

Piercing the Digital Veil
A Case Study for a DAO Legal Framework under Swiss Law

by **Benedikt Schuppli and Golnaz A. Jafari***

Abstract: Blockchain technology is associated with the emergence of decentralised applications such as smart contracts and Decentralised Autonomous Organisations (DAO) as self-governing and software-based agents. The concept of a blockchain-based peer-to-peer vending machine serving both as a testing ground for the design of a marketplace for physical goods and a speculative artefact has been posited and analysed from an economic perspective by a group of scholars at the Center for Innovative Finance of the University of Basel, Switzerland. Building on this particular case study, this paper provides for a legal analysis under Swiss law. In Part A, the economic analysis of the initiative is briefly described. In Part B, the proposed concept is analysed from a de lege lata perspective, taking into account foremost liability questions both from Swiss private law (tort and contractual) and public law (criminal and tax law) perspectives, by building on literature and applicable case law. For this, the authors propose a hypothetical scenario upon which a legal analysis is applied. As a result of the analysis, a conclusion is drawn highlighting the status quo in Swiss legal framework, whereby the authors argue in favour of a possible reform for the purpose of enhancing legal certainty. In Part C, the authors then examine, from a de lege ferenda perspective, the question of whether the Swiss legislative body would require introducing a bespoke legal framework for DAOs. For this, a reference is made to relevant foreign legislation such as the State of Wyoming DAO Bill without essentially taking a comparative approach.

Keywords: Blockchain; DLT; Decentralised Autonomous Organisations; DAO; Legal Personality; Smart Contracts

Recommended citation: Benedikt Schuppli and Golnaz A. Jafari, Piercing the Digital Veil: A Case Study for a DAO Legal Framework under Swiss Law, 12 (2021) JIPITEC 331 para 1

A. Brief Introduction to the Case Study and the Underlying Technology

1 In the economic analysis of the blockchain-based peer-to-peer vending machine concept for the design of a market place for physical goods, namely 'Blockchain Vending Machine: A Smart Contract-Based Peer-to-Peer Marketplace for Physical Goods', the authors propose an autonomous vending machine governed by a public blockchain and smart contracts platform. Set up as a decentralisedautonomous organisation, or DAO, it is set to serve as an open marketplace for physical goods, where anyone can buy and/or sell objects.[1]

* Benedikt Schuppli, Attorney-at-Law, blockchain entrepreneur, member of the Coalition of Automated Legal Applications "COALA"; Golnaz A. Jafari, LL.M., doctoral researcher at Lucernaiuris, University of Lucerne, Switzerland, formerly a research associate at SOCAI, University of Würzburg, Germany, & research fellow at NRCCL, University of Oslo, Norway. Note: URL links were primarily accessed during the period of 1 November 2020 - 30 March 2021, excluding the recent related Swiss regulatory developments.

1 F Schär, K Schuler and T Wagner, 'Blockchain Vending Machine: A Smart Contract-Based Peer-to-Peer Marketplace for Physical Goods' (2020) MPRA Paper Nr. 101733, 1 <https://ideas.repec.org/p/pra/mprapa/101733.html>.

I. Technical Taxonomy

2 In this section, the concepts of 'smart contracts' and 'DAOs' as decentralised applications of blockchain technology[2] are described in some detail. The authors presume that readers possess minimum knowledge as to the underlying technology itself.

1. Smart Contracts

3 The concept of smart contracts was first introduced in 1994, when Nick Szabo, an American computer scientist and cryptographer, wrote an article on contracts as computer protocols that perform independently.[3] At that time, however, computer science had not yet advanced far enough to implement Szabo's new ideas and concepts. From Szabo's point of view, the simplest version of a smart contract is a vending machine.[4] It accepts money in coins and submits goods. Szabo saw the goal of smart contracts as ensuring the fulfilment of standard terms of a contract, such as terms of payment, lien and even enforcement. Smart contracts are set to minimise deliberate, as well as unintended, deviations, and to limit the need for external, trustworthy intermediaries.[5]

4 However, the term smart contract could be misleading in that a smart contract does not constitute a contract in the legal sense *per se*. Moreover, the word *smart* does not delineate *intelligent* but is used in the electronics industry for applications that are capable of connecting, exchanging data and interacting with the user and other applications. A smart contract merely performs what the creator, usually a human, has programmed—with the premise that it does so in a reliable, immutable and deterministic way. Thus, in this paper reference is made to this term denoting digital programmes that are based on a blockchain architecture, self-execute when certain conditions occur, and are therefore in principle self-enforcing and immutable.[6]

2. DAOs

5 DAO is an acronym for decentralised autonomous organisation. The term originated amidst the nascent Ethereum[7] community in 2015. A DAO is essentially a computer software code that is distributed across a decentralised peer-to-peer network and incorporates governance and decision making rules. In other words, it is a form of an organisation that is operated through rules encoded in smart contracts. The purpose behind a DAO is to design a corporate structure that could function and perform actions independently from human hierarchical management. It can implement contractual obligations as well as business logic rules, and hence could be denoted as an *almost* autonomous, transparent and data-driven company. With a DAO, most management and administrative functions and internal processes could arguably be automated, and 'value' in a given context would be distributed among virtual stakeholders via smart contracts.

6 It is worth to emphasise that the independence to perform actions does not *by default* render independence in decision making. A DAO in the end is bound by the governance and decision making rules encoded in smart contracts by—at least for the time being—a human developer or software programmer. In particular, the terms of collaboration between different participants and stakeholders are specified in smart contracts. Once operational, decisions on a

2 The terms "blockchain" and "distributed ledger technology", or DLT, are used interchangeably in this paper.

3 N Szabo, 'Smart Contracts' (1994) <http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/ szabo.best.vwh.net/smart.contracts.html>; N Szabo, 'The Idea of Smart Contracts' (1997) <https://archive.is/wIUOA> .

4 N Szabo, 'Formalising and Securing Relationships on Public Networks' (1997) <https://archive.is/i65kY#selection-17.1-17.59>.

5 Szabo (1994) (n 3).

6 From a technical perspective, "immutability" is not an absolute feature. In blockchain or DLT space, cryptographic immutability is closely linked with the choice of algorithmic consensus mechanisms and the type blockchain or DLT systems would take, i.e. public, private, permissioned, permissionless or hybrid. See RP Dos Santos, 'Consensus Algorithms: A Matter of Complexity?' (2019) in M Swan et al. (eds) *Blockchain Economics: Implications of Distributed Ledgers* (World Scientific), 147-170. In simple terms, "immutability" refers to irreversibility, "a fundamental blockchain property that stems from the fact that transactions cannot be edited or deleted once they are successfully verified and recorded into the blockchain". See E Politou et al., 'Blockchain Mutability: Challenges and Proposed Solutions' (2019) IEEE, 5ff <https://arxiv.org/pdf/1907.07099.pdf>. See also E Landerreche and M Stevens, 'On Immutability of Blockchains' (2018) in W Prinz and P Hoschka (eds) *Proceedings of the 1st ERCIM Blockchain Workshop 2018, Reports of the European Society for Socially Embedded Technologies* <https://ir.cwi.nl/pub/28537/28537.pdf>.

7 Ethereum is a blockchain protocol and smart contracts platform.

DAO would reach finality having passed through a designated algorithmic consensus mechanism.[8] The interaction between the members of a given DAO, who are generally represented by pseudonymous identifiers,[9] would normally take place through the medium of the underlying blockchain, in the form of an interface upon which all actions would be carried out.

7 Originally, the term 'the DAO' was used to describe a specific instantiation of such an organisation, known to be the first DAO of its kind. The DAO was created with the objective of operating as a for-profit entity, a kind of automated investment fund, which would create and hold a corpus of digital assets through the sale of 'DAO Tokens' to investors.[10] DAO Tokens were blockchain-based digital assets that would subsequently be used to fund business ventures.[11] However, in the present case study the term DAO[12] refers more generally to any blockchain-based implementation of such a decentralised autonomous organisation.

II. Case Study: The Economic Analysis in a Nutshell

8 In the economic analysis of the concept, the authors propose a basic architecture for the blockchain-based vending machine, including pricing and fee mechanisms. They also examine potential challenges arising out of the setup. The main purpose behind the physical autonomous marketplace envisaged by the authors is to address counterparty risk associated with virtual assets related to a physical object. A traditional and intermediated way to address such counterparty risk for promises of physical goods is through escrow or custody services. Imitating escrow in function, the authors posit a new type of vending machine that connects the purchase and the delivery of goods atomically via delivery-versus-payment, or DvP.[13] The processes relating to the DvP function are embedded in and executed by smart contract codes in a deterministic fashion. Residing on a *public*[14] blockchain, these contracts form a DAO that controls the vending machine.[15]

9 The basic setup in the economic analysis is described as follows:[16]

> The peer-to-peer vending machine consists of two main elements. First, the machine, i.e., the actual physical vending machine including the required software to connect to the Blockchain and translate the signals received into corresponding actions. Secondly, the DAO in the form of a dedicated smart contract structure on a public Blockchain. The former provides a physical incarnation, while the latter governs the behavior of the machine and controls the logic and conditions of the interactions. It is fully transparent, protected from unforeseen intervention, and open to anyone.
>
> Let us assume that the machine consists of a number of slots. Each slot shows a unique identifier and has a goods compartment with a transparent door that can be locked individually. It also has a display, to assist users in their interactions. The vending machine is located in

8 See also Dos Santos (2019) (n 6), in particular, "In distributed ledgers, similarly, consensus algorithms are the process of the distributed peer-to-peer nodes in the network coming to agreement upon updated states of the ledger per executed transactions. Consensus algorithms are mechanistic and automated. As such, they provide the trustless software mechanism for the automatic execution of blockchain transactions between parties that do not need to know or trust each other".

9 For the definition of the term "pseudonymity" under EU law, see Regulation (EU) 2016/679 (GDPR) OJ L 119, Article 4(5).

10 See also S Polrot, 'Déploiement de The DAO, "mere de toute les dao"' (2016) <https://www.ethereum-france.com/deploiement-du-projet-the-dao-mere-de-toutes-les-dao/>; S Hassan and P de Filippi, 'Decentralised Autonomous Organisation' (2020) <https://policyreview.info/open-abstracts/decentralised-autonomous-organisation>.

11 J Meier and B Schuppli, 'The DAO Hack and the Living Law of Blockchain' (2019) APARJUZ, 33 <https://www.ivr.uzh.ch/dam/jcr:722b55af-b0f7-40de-900f-46a526a93f80/J%20Meier,%20B%20Schuppli,%20The%20DAO%20Hack%20and%20the%20Living%20Law%20of%20Blockchain,%20APARIUZ%202019.pdf>.

12 For more on DAO, see S Polrot, 'Les Decentralized Autonomous Organizations ("DAO"), le future des organisations collectives?' (2016) <https://www.ethereum-france.com/decentralized-autonomous-organization-dao-blockchain/>.

13 Schär et al. (2020) (n 1), 2.

14 Blockchain or DLT systems can take various forms, such as public, private, permissioned, permissionless, hybrid or consortium. The form a system takes has legal and technical implications, in particular regarding governance and participation protocols, as well as the identification of the participants, among other things.

15 Schär et al. (2020) (n 1), 1.

16 Ibid.

a public space. It is easily accessible, meaning that anyone can interact with the vending machine by assuming the role of a buyer or seller.

Buying works in the traditional way. When someone sees a good in the vending machine for which they have a buy interest at the given price, they can buy it instantly. The main difference to a regular vending machine is that instead of buying from a central counterparty, i.e., the vending machine operator, the buyer engages in a peer-to-peer transaction with someone who placed the object in the machine.

Analogously, selling via the vending machine is open to anyone, provided there is a currently unused slot. A seller simply places the goods in the compartment and provides the sales parameters, such as pricing. The machine will then initiate the sale and take over custody by locking the door. Thereafter, no further action is required of the seller, the proceeds are automatically distributed after a successful sale.

Both the buying and the selling process are governed by the DAO's smart contracts. To release goods or lock a compartment, the machine relies on events emitted by the DAO. Also, it is the source for the currently valid parameters on pricing and fees, which are detailed later on.

To increase the autonomy of the DAO, i.e., reducing the dependency on humans the smart contract structure may be designed to cover a multitude of aspects. For the basic setup, however, we propose a lean structure that focuses on the autonomous handling of the core processes of buying and selling goods through the machine, plus a governance mechanism to propose and vote on fee parameter changes or extraordinary events.

While it is possible to interact with the machine directly via smart contract function calls, a simple user interface is proposed to lower the barriers to entry for potential users with limited Blockchain and smart contract knowledge. To provide a basic user experience, the machine has a display to guide through the buying and selling process as well as a button next to each slot to unambiguously indicate which slot the interaction is targeting.

10 The authors of the economic analysis draw the conclusion that the deployment of a blockchain-based vending machine could contribute to the understanding of DAOs whilst building a bridge between digital and physical markets. The authors' proposal is deliberately limited to technical and microeconomic aspects. For a full assessment of the initiative's feasibility, a legal perspective must be added. This is because mainstream economic processes, other than the shadow economy,[17] take place in and are defined by a legal system. The argument for a legal vacuum[18] in which DAOs and blockchain networks exist, as promulgated by some proponents does not withstand further scrutiny.[19] But, as the past has shown, our continental European legal system has been caught off-guard by some of blockchain technology's inventions, and the process of understanding how the said technology could be made sense of and dealt with by our legal system is still underway.[20] In the analysis below, Swiss substantive law is considered from a *de lege lata* perspective to assess the feasibility of the proposed concept.

B. Legal Analysis

11 In this section, questions of liability are raised and put into perspective in the context of Swiss contract and tort law. Thereafter, a public law analysis with a specific emphasis on tax and criminal liability is performed. The authors outline these analyses primarily in the context of a hypothetical scenario.

12 The analyses therefore do not aim to be encompassing or complete but shall rather function as indicators in order to assess the overall readiness of the Swiss legal system, both public and private law, to deal with such a novel market and technology infrastructure. This paper shall not be a contribution to the specific legal assessment of blockchain technology or parts thereof under existing Swiss law for which plentiful publications exist, but shall rather contribute to a policy discussion on the necessity for a bespoke legal framework for DAOs.

13 In a subsequent section, authors also glance through the recent Swiss legislative developments and con-

17 The International Monetary Fund (IMF) defines the term "shadow economy" as comprising "all economic activities that would generally be taxable were they reported to the tax authorities". <https://www.imf.org/external/pubs/ft/issues/issues30/#:~:text=the%20official%20economy.-,What%20Is%20the%20Shadow%20Economy%3F,from%20monetary%20or%20barter%20transactions>.

18 See the term "rechtsfreier Raum" in German.

19 Meier and Schuppli (2019) (n 11), 27.

20 It is noteworthy that novel legislations attempting to deal with blockchain technology, including tokens, smart contracts and, in a few cases, DAOs, have been introduced in jurisdictions as diverse as Malta, Singapore, Germany, Switzerland and Liechtenstein.

clude with a brief analysis as to potential implications for the token economy of the proposed concept.

I. Hypothetical Scenario Applied to the Case Study

14 To assess whether the proposed concept can be adequately utilised given the applicable substantive Swiss law in the areas of contract, tort, tax and criminal law, we formulate a hypothetical scenario in the form of a typical transaction that could occur when making use of the blockchain-based peer-to-peer vending machine.

15 For this scenario, let us assume a DAO maintains an instantiation of the proposed vending machine, which is situated in a common use public area in the city of Zurich, Switzerland. The DAO, hereinafter referred to as 'BVM DAO', is managed by and consists of natural persons who do not know the full identity of each other, but have been coordinating and communicating on an online forum using pseudonyms in order to commission the building of the vending machine, the writing of the smart contract codes and the raising of the corresponding funds for the BVM DAO.

16 For the construction of the physical component of the concept, the BVM DAO has commissioned a construction worker in Zurich who has accepted an upfront payment in the cryptocurrency ether[21] and full payment after successful completion. The smart contract code running the vending machine is written by one of the BVM DAO members, known only by his pseudonym 'BVM Enthusiast', who normally resides—unbeknownst to the other BVM DAO members—in Albania. The BVM DAO maintains the vending machine as a marketplace for physical goods.

17 In order to become a member, a membership fee has to be sent in ether to the BVM DAO wallet. New members must be accepted via a majority vote by the existing BVM DAO members. From the membership fee, the maintenance of the BVM DAO is financed, and an insurance fund is maintained for lawsuits or fines against the BVM DAO or its members for acts committed in the capacity of the BVM DAO. Distributions from the insurance fund are subject to a funding proposal by the liable party and an acceptance by majority voting. Transaction fees earned by the BVM DAO from vendors are evenly distributed among BVM DAO members on a recurring basis. All other functions follow the proposal as described in Section A.II.

18 Slots in the vending machine maintained by the BVM DAO are rented out to vendors. A recurring text on the display of the vending machine reads as follows: by purchasing any goods the buyer accepts the BVM DAO's terms of services which can be found under bvmdaozurich.ch/terrms. As part of these terms, BVM DAO stipulates that vendors of goods in the BVM DAO are vetted for their reputation and the quality of goods sold. Furthermore, an exoneration of liability clause reads as follows: liability of the BVM DAO for any damage incurred out of the use of the BVM DAO is, as permitted by applicable law, excluded. Lastly, under the terms of service it is also reiterated that the use of the BVM DAO and these terms are governed by Swiss law, whereby the courts in London, United Kingdom, shall be exclusively competent to adjudicate any and all disputes arising out of or in connection with the use of the BVM DAO.

19 Vendor X imports and sells luxury goods in one of the slots in the BVM DAO. One such good is a Rolex Daytona.

20 Buyer Y buys a Rolex Daytona from Vendor X. He picks up the released Rolex from the vending machine after the purchase price in ether has been delivered to the smart contract and transferred to Vendor X atomically. Two weeks after the completion of the purchase, Buyer Y starts questioning whether the Rolex Daytona is an original one and whether it has been tested by a horologist. The purchased product turns out to be a counterfeit object. The certificate of authenticity, accompanied with the product, seems to have also been forged. Frustrated with this, Buyer Y seeks action and reimbursement of the purchase price he paid for the fake product. However, Vendor X, the real identity of whom is neither known to Buyer Y nor to BVM DAO, does not react to any contact attempts.

II. Private Law: Contracts & Tort Law

21 In order to assess the suitability of Swiss private law in handling transactions and interactions with the vending machine, the legal relations between involved parties would need to be described. The private law relations between the BVM DAO and Vendor X as well as the Buyer Y shall be taken into account as follows.

21 Ether is the cryptographically generated native currency of the Ethereum platform (see n 7).

- Buyer Y – Vendor X
- Buyer Y – BVM DAO[22]
- Vendor X – BVM DAO

22 It is argued herein that Buyer Y and Vendor X are in a contractual relationship with one another as parties to a purchase contract. The mere fact that the purchase price was paid for in ether does not render it a barter contract, as ether is considered a *private* currency and therefore comparable to fiat currency as a *public* money, exclusively as to its 'means of payment' function.[23]

23 It is furthermore argued that a contractual relationship exists between Buyer Y and the BVM DAO or members of the BVM DAO collectively, given that Buyer Y's implied acceptance of BVM DAO's contractual terms is affirmed by using the vending machine.

24 The BVM DAO is in a contractual relationship with Vendor X, who rents a slot in the vending machine to sell goods. The contract is most accurately described as a lease contract.

25 For the purposes of this hypothetical scenario, the BVM DAO, in the absence of any legal personality, can most accurately be described as a partnership according to article 530 para. 1 of Swiss Code of Obligations (CO), whereby "a partnership is a contractual relationship in which two or more persons agree to combine their efforts or resources in order to achieve a common goal".[24]

26 On the discussion of whether and how smart contracts, in general, can be reconciled with Swiss contract law, in-depth analyses are offered by other, existing publications.[25]

27 Buyer Y entered into a purchase contract with Vendor X. As the purchased good, the Rolex Daytona, turned out to be fake, Buyer Y wants to receive back the purchase price or receive an original product, either from Vendor X or from the BVM DAO.

28 According to article 28 CO, "a party induced to enter into a contract by the fraud of the other party is not bound by it even if his error is fundamental". In order for article 28 CO to apply, a fraudulent behaviour would need to embed certain constituents. These include (i) fraudulent intent, (ii) illegality, (iii) a mistake in motive and (iv) causality between the fraud and the conclusion of the contract.[26] The party acting under fraud is therefore not bound by the contract. By declaring the absence of intent to the counterparty, the contract is nullified *ex tunc*, and the defrauded party may seek restitution for the damage incurred based on articles 62 et seq. CO.[27]

29 Buyer Y entered into the purchase contract with Vendor X based on the assumption that the Rolex Daytona was original, as portrayed in the title of the offer and the corresponding certificate of authenticity. Vendor X acted fraudulently with intent, and no legal justification was offered for such behaviour. Therefore, article 28 CO would apply here awarding Buyer Y to declare the contract void and to seek restitution for the purchase price as well as any additional damage caused by the fraudulent behaviour of Vendor X. Buyer Y must then declare the nullification within one year of learning about the fraud.[28]

30 In general, Swiss contract law has no fundamental difficulties in dealing with the fact that a contract

22 Miners are not taken into account for the analysis at hand due to the slim causal nexus between their actions and potential private law issues for transaction parties. For a more conclusive overview of the potential private law relationships between the parties to a smart contract system, such as the Blockchain Vending Machine, read SD Meyer and B Schuppli, '"Smart Contracts" und deren Einordnung in das schweizerische Vertragsrecht' (2017) Recht, 204 ff, 210 <https://recht.recht.ch/de/artikel/04re0317ver/smart-contracts-und-deren-einordnung-das-schweizerische-vertragsrecht> .

23 Meyer and Schuppli (2017) (n 22), 204 ff, 216.

24 The distinction between a general partnership and a simple partnership is made in favour of the simple partnership, as the former requires a commercial business setup which is not assumed here for simplicity reasons.

25 J Essebier and DA Wyss, 'From the Blockchain to Smart Contracts' (2017) Jusletter <https://jusletter.weblat.ch/juslissues/2017/889/von-der-blockchain-z_5bd3b52a43.html_ONCE&login=false>; Meyer and Schuppli (2017) (n 22), 204 ff; M Eggen, 'Chain of Contracts' (2015) AJP 26 (1), 3 ff <https://boris.unibe.ch/114476/>; A Furrer, 'Die Einbettung von Smart Contracts in das schweizerische Privatrecht' (2018) Anwaltsrevue, 103-115 <http://www.anwaltsrevue.recht.ch/arv/lpext.dll/arv/avarv18/arv0318/inharv0318?f=templates&fn=index.html&2.0&vid=10.1033/Deu>.

26 Swiss Code of Obligations, Short Commentary on Swiss Private Law (3rd edn, Schulthess Zürich 2016), hereafter cited as CHK-Kut CO 1 N1; CHK-Kut CO 28 N3.

27 CHK-Kut CO 31 N2.

28 CHK-Kut CO 31 N1.

was facilitated using a smart contract,[29] even more so when a smart contract was used to hold and release the purchase price for a physical object as a one-time transaction. Given the state of the nullified contract, Buyer Y may seek restitution based on unjust enrichment on the basis of articles 62 et seq. CO. He has a right to reimbursement of the purchase price plus additional damages by Vendor X in ether or in Swiss francs, alternatively.[30]

31 Difficulties instead arise from the fact that BVM DAO allows mutually unidentified persons to enter into transactions, exchanging monetary values for goods. While the DvP part of such a transaction can be aptly handled by the DAO, the post-transaction and settlement lifecycle of a contract cannot. Thus, if Buyer Y reasonably wants to enforce any right arising out of the *consumed* purchase contract, the identity of Vendor X must be accessible to him.

32 As Vendor X is out of reach, Buyer Y may try to seek restitution from BVM DAO as the marketplace provider through which the transaction was enabled in the first place.

33 As detailed previously, BVM DAO claimed to vet Vendors. The question would then arise whether BVM DAO is liable for the damages Buyer Y incurred by trusting the information by Vendor X in light of BVM DAO's claim to vet vendors for the quality of products.

34 As a contractual relationship between Buyer Y and the simple partnership BVM DAO is now affirmed, Buyer Y is in a position to seek restitution from BVM DAO based on contractual damages,[31] specifically for the violation of contractual obligations such as the violated duty to vet Vendors.[32]

35 Buyer Y, therefore, brings claims against the simple partnership BVM DAO via the email listed on their website. As a simple partnership, every member is *jointly* and *severally* liable for the totality of the damage incurred to Buyer Y under the contract.[33] While the law once again is unambiguous here, Buyer Y does not actually have any identifying information on any BVM DAO member. As a result of long internet searches, Buyer Y learns about the true identity, name and address of the BVM Enthusiast pseudonym who resides in Albania. Buyer Y initiates a lawsuit against BVM Enthusiast as a severally liable member of BVM DAO in Zurich. Upon learning of the lawsuit, BVM Enthusiast files for legal funding from BVM DAO's member insurance fund, which is granted by a majority voting. Upon processing the lawsuit, the court called upon in Zurich makes a decision to dismiss the lawsuit without entering into the substance of the case, as the choice of forum in the terms of service on BVM DAO's website was deemed valid, and the case was not characterised as a consumer dispute, which would have established forum in Zurich according to article 32 of Swiss Civil Procedure Code (CPC), given that the product in question is a luxury good.[34] Buyer Y, frustrated with this outcome, leaves the matter be.

III. Public law

1. Tax Liability

36 The introduction of goods into Switzerland triggers value-added tax (VAT). Furthermore, the offering of the blockchain-based vending machine marketplace and the leasing of slots to vendors, which in turn sell goods to buyers, are all acts subject to Swiss VAT.[35]

37 The liable tax subject in the case of the introduction and selling of the goods to buyers is Vendor X. In the case of the maintenance of the marketplace and the charging of a lease fee, BVM DAO as a simple partnership would also be liable for VAT towards the Federal Swiss Tax Administration.

38 As for Vendor X, due to lack of information on his identity and whereabouts, the difficulties for the Federal Swiss Tax Administration to collect VAT become apparent.

29 See, eg, the 2016 position of the Commercial Court of Zurich (HG150136 of the 16.02.2016), Recital 2.3: "Nebst individuell übermittelten Willenserklärungen sind auch solche verbindlich, welche von einem vorprogrammierten Computer automatisch abgegeben werden (sog. 'elektronischer Softwareagent')".

30 For a more in-depth analysis of the status of ether as cash or rather, a good which is bartered, see Meyer and Schuppli (2017) (n 22), 217.

31 If a contractual relationship is denied, tort damages could apply nonetheless. For the differentiation, see CHK-Kut CO 41 N3.

32 CHK-Kut CO 97 N10. It is assumed here that other prerequisites, such as damage and causality, are met, too.

33 CHK-Kut CO 530 V.

34 See Swiss Federal Court Decision BGer 4A_2/2018 for a similar case. <https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?highlight_docid=aza%3A%2F%2F22-03-2018-4A_2-2018&lang=fr&type=show_document&zoom=YES&>.

35 Swiss VAT Act SR 641.2 Federal Act of 12 June 2009 on Value Added Tax, article 21 *e contrario*. <https://www.fedlex.admin.ch/eli/cc/2009/615/de>.

39 As for BVM DAO, the simple partnership where individual members are *jointly* and *severally* liable towards the tax authorities, the case becomes more nuanced.

40 However, given the difficulties of pinning down BVM DAO as an incorporated organisation, as its comprising members are acting with pseudonyms and are spread across Europe, the Swiss Federal Tax Administration will face hindrance collecting VAT from them. In case one member is identified, though, tax liabilities towards the Federal Swiss Tax Administration can be funded via the BVM DAO insurance fund. This will create a moral hazard not to pay VAT before an individual BVM DAO member would be prosecuted for it, which is reminiscent of Planka.nu in Sweden. Here, the case concerns notoriously funded members' penalties via a common insurance fund when they were caught fare-dodging in public transport.[36]

41 In either case, the obscure nature of BVM DAO's individual members, as well as the vendors, would impose enforcement costs on tax authorities while generating loss for them. Furthermore, consumers who are VAT-payers of last resort may still end up paying for uncollected VAT as VAT for which vendors and suppliers are liable is usually priced into end consumer prices.

2. Criminal Liability

42 Let us assume that by importing a counterfeit good, introducing it into the Swiss market and selling it to Buyer Y, Vendor X has fulfilled all the hallmarks of, *inter alia*, criminal offences, fraud according to article 146 of Swiss Civil Code (CC) and counterfeiting of goods according to article 155 CC.

43 In order to prosecute Vendor X for the criminal offences committed, Swiss law enforcement would need to be privy to information about his identity and/or whereabouts. Let us further assume that no such information is accessible to the prosecutors, leaving them only able to investigate accessory criminal liability by BVM DAO. For this, individual BVM DAO members could be held liable in the first degree, and BVM DAO as an organisation would be subject to secondary liability according to article 102 CC.[37]

44 In order for an offender to be held criminally liable under Swiss criminal law, his or her actions—or omissions in the case of a duty of care for inalienable rights of others—must have been causal for the outcome of the offence. To prevent an uncontrollable sprawl of causal relations, legal doctrine has added the prerequisite element of the adequacy of the said causal relation. According to the Swiss Federal Court, the adequate causal connection is to be affirmed if "a behaviour was suitable, after the usual course of things and the experiences of life, to bring about or at least to favour the kind of outcome of the criminal offence as the one that has occurred".[38]

45 While one could argue that a single BMV DAO member's omission to vet Vendor X and the sold goods properly is suitable to favour the outcome of both the fraud and the counterfeiting of goods offences, and therefore fulfilling the requirement of causality, wilful criminal intent is also required for one to be criminally liable as an abettor or an accomplice under either article 146 CC or article 155 CC. Negligent behaviour alone, as it may suffice to establish the kinds of contractual claims against BVM DAO, does not surmount to wilful intent. Correspondingly, wilfulness based on article 12 para 2. CC by knowingly accepting the realisation of the act, the fraud or the introduction of counterfeiting goods as possible, or *dolus eventualis*, is difficult to establish here. This is due to the fact that Vendor X, although not being thoroughly vetted by BVM DAO, has accompanied the Rolex Daytona with a certificate of authenticity. In addition, expecting BVM DAO to verify the said certificate or the product itself with a horologist in order to forego criminal liability would be a stretch.

46 If accessory criminal liability, under articles 146 or 155 CC, of a BVM DAO member were nonetheless established without it being clear who the member was or which member fulfilled in his or her own right all the required hallmarks, BVM DAO could be criminally liable under article 102 CC and therefore be subject to a fine. The make-up of BVM DAO as a pseudonymous group where the identity of members is *deliberately* kept secret and shielded from public view is prone to be deemed "organisationally

36 See <https://planka.nu/om-plankanu/>.

37 See the term "Ersatzhaftung" in German; Swiss Federal Court Decision BGE 142 IV 133, E. 4.1 <http://relevancy.bger.ch/php/clir/http/index.php?highlight_docid=atf%3A%2F%2F142-IV-333%3Ade&lang=de&type=show_document#:~:text=102%20StGB%20ist%20Voraussetzung%20f%C3%BCr,%C3%A4usseren%20Grund%20f%C3%BCr%20die%20Strafbarkeit>.

38 Swiss Federal Court Decision E. 4.1.3, translated from German into English by the authors <https://www.bger.ch/ext/eurospider/live/de/php/clir/http/index.php?highlight_docid=atf%3A%2F%2F138-IV-57%3Ade&lang=de&zoom=&type=show_document>.

defective" and, as such, causal to the inability of the prosecutors to hold one single BVM DAO member criminally liable.[39]

47 However, given the difficulties of pinning down BVM DAO as an organisation, as its comprising members are acting via pseudonyms and are spread across Europe, prosecutors will experience difficulties enforcing the law against them. If the fine for the violation of, e.g., article 155 CC in connection with article 102 CC would be too high, a moral hazard is created to abandon the vending machine and the local jurisdiction, Switzerland, entirely and to set up an alternative BVM DAO elsewhere. If the fine is too low, and it can be paid for with the insurance funds in the BVM DAO wallet, another moral hazard is created to continue to allow criminal behaviour on the BVM DAO marketplace.

IV. Implications of Swiss Regulatory Developments

48 The scope of the present legal analysis clearly excludes the token[40] ecosystem[41] and economic aspects related to the proposed concept.[42] Nevertheless, given the ongoing regulatory and legislative reforms in Switzerland, the authors take the view that these will have implications over the intended token design, in particular when due account is given to its *substance* over its *form*.

49 Currently Switzerland is in the process of reforming existing laws in order to accommodate blockchain systems and to address Decentralised Finance (DeFi) applications. The Federal Act on the Adaptation of Federal Law to Developments in DLT, has already received parliamentary approval, has been implemented partially and is expected to enter fully in force later this year.[43] This Act introduces a number of changes permitting the development of decentralised governance in systems that are aimed at financial transactions.[44] In other words, the reform is set to permit the exchange of asset tokens, among others, as uncertificated securities. This specific category of tokenised rights,[45] defined as uncertificated register securities,[46] and the legal transfer thereof, concerns any right that can effectively be securitised.

50 Under Swiss law,[47] securities in general are certificated or uncertificated securities, derivatives or intermediated securities, which are standardised and suitable for mass trading.[48] Outside of this traditional definition, questions would arise as to whether stan-

39 BGE 142 IV 133, E. 3.1 (n 37).

40 In a technical sense, a "token" stands for a sequence of characters that serves as an identifier for a specific asset, eg, usage rights, participation rights or cryptographically generated currency models such as bitcoin, among others; see A Sunyaev et al., 'Token Economy' (2021) <https://link.springer.com/content/pdf/10.1007/s12599-021-00684-1.pdf>. A token is also a " a digital representation of value on a shared distributed ledger that is owned and secured using cryptography to ensure its authenticity and prevent modification or tampering without the owner's consent". See International Telecommunication Unit (ITU), Technical Specification FG DLT D1.1 (2019), 6 <https://www.itu.int/en/ITU-T/focusgroups/dlt/Documents/d11.pdf>.

41 ITU (2019) (n 40), 6; a "token ecosystem" stands for "a digital system or digital space where participants and users interact and coordinate with each other using tokens".

42 Schär et al. (2020) (n1).

43 Swiss DLT Framework, parliamentary approval (September 2020) <https://www.admin.ch/opc/fr/federal-gazette/2020/7559.pdf>. Note: The amendments to the Swiss Code of Obligations (CO), the Federal Intermediated Securities Act and the Federal Act on International Private Law that are envisaged in the DLT bill have now entered into force from 1 February 2021. These provisions enable the introduction of ledger-based (blockchain-based) securities that are represented in a blockchain or DLT system. The remaining provisions of the DLT bill are foreseen to enter into force on 1 August 2021.

44 The Swiss Federal DLT Act has amended/is set to amend specific laws such as the Code of Obligations (CO), the Banking Act (BankA), the Financial Market Infrastructure Act (FMIA), the Bankruptcy and Insolvency Act (BIA), and the Federal Act on Intermediated Securities (FISA).

45 Swiss Code of Obligations (CO), see Articles 973d – 973i.

46 See the term "Registerwertrechte"; "[U]ncertificated register securities have features largely analogous to traditional certificated securities. Any right that can be securitised also qualifies as an underlying right for uncertificated register securities, including asset tokens and utility tokens" in CMS Law –Now, 'The new Swiss blockchain/DLT laws have been finalised and presumably entre into force early 2021' (15 October 2020) <https://cms.law/en/che/blogs/law-now-blog/the-new-swiss-blockchain-dlt-laws-have-been-finalised-and-presumably-enter-into-force-early-2021>.

47 Financial Market Infrastructure Act (FMIA) (19 June 2015), Article 2(b); see also Financial Services Act (FinSA) (15 June 2018), Article 3(b).

48 For the term 'standardised and mass trading' see: "the instruments are offered for sale publicly in the same structure and denomination, or that they are placed with 20 or more clients under identical conditions" in Financial Market Infrastructure Ordinance (FMIO) (20 November 2015), Article 2.1.

dardised elements such as voting rights could also qualify as securities. It is apparent that definition of a given digital asset in the form of a token as a security would fall outside both certificated and intermediated securities categories, whereby only uncertificated securities and derivatives would serve relevance. Under uncertificated securities category, Swiss law defines three types of rights, such as participation rights,[49] property rights and credits. Traditionally, the only formal requirement[50] for creation of these securities is by keeping a book in which associated details are recorded. With the recent reforms, such a book (or register) can now be created on a blockchain system.

51 On the other hand, a general distinction is made between three token models, e.g. payment tokens, utility tokens and asset tokens.[51] Asset tokens refer to and represent physical assets, company equity, debt and rights such as dividends and interest payments. In their classification, the Swiss Financial Market Supervisory Authority (FINMA) also emphasises on *substance* over *form* of a given design. Essentially, asset tokens are seen analogous to equities, bonds and derivatives, from the perspective of their economic function. An asset token can take the form of a promise, e.g. in future capital flows.

52 For applicability of Swiss financial market and securities laws, an assessment would need to be made as to whether a token would confer claims or rights, such as ownership, in favour of the holder against its issuer or a third party. In addition, the type of the underlying asset referred to by a token, i.e. fiat currency, commodity, real estate or securities, carries importance.[52]

53 Furthermore, for tax purposes in Switzerland, asset tokens are often classified in distinct groups of debt tokens, equity tokens and participation tokens.[53] Concerning equity tokens, an investor's entitlement would refer to a benefit, measured by a certain ratio to profit or liquidation result. In the case of participation tokens, on the other hand, investors would generally be entitled to a proportional share of a certain reference value defined by the issuer. Both equity tokens and participation tokens would be considered as derivative financial instruments in the context of taxation.

54 In the economic analysis of the proposed concept, it is suggested that tradable participation tokens are issued which "entail the right to vote on governance proposals and participate in future cash flows".[54] In order to foster user adoption the vending machine would also distribute "micro participation rights to sellers and buyers through fractions of tokens".[55] Here, "token holders could create proposals and cast votes in proportion to their token holdings," whereby "a combination of cash flows and voting rights in the same token" is argued to "help align interests and incentivise token holders to act in the machine's best interest".[56]

55 In addition, "the right to open a slot in the vending machine is tokenised in the form of a non-fungible token (NFT)",[57] whereby the vending machine would assume a custodian role, with the NFTs bearing redemption rights in favour of the assets placed in the slots. The NFTs are therefore seen to become tradable without the need for the physical displacement of the asset. Crucially, the vending machine is depicted as a "programmable safe deposit box with a large variety of use cases including collateralised loans, smart contract-based implementations of a last will and the issuance of sub-tokens which represent partial ownership of the NFT".[58]

49 Participation rights could either bear financial value or not. In cases where participation rights bear financial value, these are qualified as securities.

50 See n45, Article 973c.3.

51 Financial Market Supervisory Authority (FINMA), 'Guidelines for Enquiries Regarding the Regulatory Framework for ICOs' (February 2018) <https://www.finma.ch/en/~/media/finma/dokumente/dokumentencenter/myfinma/1bewilligung/fintech/wegleitung-ico.pdf?la=en>.

52 FINMA, 'Supplement to the Guidelines' (September 2019), 1-4 <https://www.finma.ch/en/~/media/finma/dokumente/dokumentencenter/myfinma/1bewilligung/fintech/wegleitung-stable-coins.pdf?la=en>.

53 Swiss Federal Tax Administration (SFTA) Working Paper (2019) <https://www.estv.admin.ch/estv/de/home/direkte-bundessteuer/direkte-bundessteuer/fachinformationen/kryptowaehrungen.html>. See also O Favre et al., 'The Virtual Currency Regulation Review: Switzerland' (Schellenberg Wittmer, 2020) <https://thelawreviews.co.uk/title/the-virtual-currency-regulation-review/switzerland>.

54 Schär et al. (2020) (n 1), 6f.

55 Ibid.

56 Ibid 7.

57 Ibid. Note that non-fungible tokens, or NFTs, are blockchain-based assets with unique identification codes designed so that they are not equal and cannot be replicated. These are distinguished from fungible tokens that are substitutable or exchangeable for similar items.

58 Ibid.

56 The proposed design as a participation token could arguably be interpreted as an asset token under FINMA definitions, bearing a derivative character. In other words, these participation tokens could then be considered as uncertificated register securities, whereby the transfer of these types of securities that are exclusively registered on a blockchain system is now permitted under Swiss law.

57 Notably, a form of security under Swiss law, when defined as a derivative or a financial contract, is where the price is set particularly according to a) assets such as shares, bonds, commodities etc., and b) reference values such as currencies, interest rates etc.[59] Also, derivatives are defined as financial contracts whose value depends on one or several underlying assets and which are not cash transactions.[60] These definitions clearly imply that a derivative would require bearing a *price*, which is set according to an underlying asset.

58 The participation tokens in the proposed concept would in principle only assume functionality by means of the underlying (implied) right to claim assets, such as participation in future cash flows and the right to vote on governance of the architecture. These tokens would form the effective embodiment of an uncertificated register security, issued by and subjected exclusively to the rules of the underlying network governed by BVM DAO. In other words, the derivative element would relate to the way value is constituted on the basis of the subject matter. Therefore, these participation tokens seem to bear a *price* given that investors are set to align their interests and to expect profit from the functioning of the vending machine.

59 Once caught under the uncertificated register security, where the electronic register will exclusively be integrated on blockchain systems, a contractual relationship would then need to be established in the form of a registration agreement between the issuer and the holder. Here identification of parties becomes pivotal, with the register potentially taking the role of 'data controller' under the Swiss data protection regime.[61] For the participation tokens to work on such a blockchain-based register, the identity of the holder at each point in time as well as the issuer, who is the obligor of the securitised rights, must be unambiguous at all times. In light of the challenges reflected in the sections above, this aspect furthers the difficulty of embedding an unincorporated DAO structure with pseudonymous members into the legal and financial system.

60 Furthermore, in the proposed tradable NFT as a tokenised right to open a slot on the machine, the latter will act as a trusted custodian of the physical goods aimed at guaranteeing an effective bridge between the on-chain and off-chain spaces. NFTs in general are not considered as securities. Here, NFTs represent deposited assets, i.e. physical goods held in custody by the vending machine. These cannot be considered as standardised within the meaning of a security, discussed above. In this context, the proposed concept also refers to the possibility of the "issuance of sub-tokens which represent partial ownership of the NFT".[62] Division of the NFT into *identical* sub-tokens representing partial ownership, could be argued to constitute a standardised asset, provided that the number of these sub-tokens is higher than 20.[63] The sub-token holders would then be entitled to the right on the partial value represented by these tokens. These tokens could therefore be seen as derivatives, and consequently as securities.

61 Lastly, the assigned role of custodianship to the vending machine which may also act as a "programmable deposit box," would inevitably have implications as to matters related to liability.

V. Concluding Remarks

62 The analysis conducted in the preceding sections highlight that the current Swiss legal framework, both from a private and public law perspective, wrestles with the concept of an autonomously managed and largely pseudonymous marketplace. In addition, authors take the view that the recent regulatory reforms would certainly have an impact on the chosen token design, albeit outside the scope of this study.

63 Law primarily surrounds legal subjects, be it natural or legal persons, and confers rights and obligations to and in relation to them. With the proposed architecture as a blockchain-based peer-to-peer vending machine governed by a DAO, such legal subject is either entirely absent, as shown in the case of tax law, or it is hard to get a hold of, as shown in the section on private law analysis. Furthermore, from a criminal law perspective, establishing the required adequate causal link and wilfulness between the BVM

59 Financial Market Infrastructure Ordinance (FMIO) (25 November 2015), Art. 2.2 (a)(b).

60 Financial Market Infrastructure Act (FMIA), Article 2(c).

61 The new Swiss Data Protection Act (revised FADP) was adopted by the Parliament in September 2020 and is expected to come into force by 2022 <https://www.fedlex.admin.ch/eli/fga/2020/1998/de>.

62 See Schär et al. (2020) (n 1), 7.

63 See n 48.

DAO's actions and the damage caused to the victim seems unrealistic in light of the high threshold Swiss legal doctrine has rightfully levied for criminal liability, specifically for omissions in case of a duty of care.

64 From a private law perspective, the fact that contracting parties have little to no factual recourse in case of a purchase of counterfeit goods is an undesirable state from a public policy perspective, as neither consumer protection, in the wider sense, not within the meaning of article 32 CPC, nor good faith in commercial dealings as a public policy interest, is viably upheld in this scenario.[64]

65 From a public law perspective, on the other hand, the state faces insurmountable challenges in taxing and collecting the taxable transactions involving such architecture. Also, perpetrators of criminal offences, i.e., members of a DAO or unidentifiable associates of a DAO, such as Vendor X, could likely not be brought to justice—an outcome which directly infringes on the public good of legal protection and undermines trust in government.[65]

66 As shown above, with the existing Swiss legal framework, undue burdens are inflicted on market participants interacting with the vending machine on the one hand, and the state as the responsible authority to levy taxes and to prosecute crimes on the other hand. The afore-mentioned undesired results of Swiss private and public law in dealing with the described case study suggests that the existing Swiss legal framework is not adequate, which is why pillars for a new and more adequate framework are discussed below.

C. Policy Agenda

67 As identified in part B, Swiss substantive law currently does not offer a satisfactory framework to deal with novel decentralised market infrastructures such as the one proposed by Schär et al. Individuals interacting with the proposed infrastructure, be it as vendors, buyers or members of the BVM DAO, would face uncertainty related to both private and public law enforcement. Thus, the overall functioning of the legal economy and the rule of law would be infringed upon.

68 The novelty of a DAO such as the BVM DAO that creates these challenges for existing legal tools lies in the propensity for DAO transactions to take place cross-border, between unknown or pseudonymous parties and to be immutable in principle by virtue of the underlying technology. While the concept of enabling untrusted and unidentified parties to transact with one another is appealing from an economic perspective as transaction costs associated with search and counterparty risks can be significantly reduced, both public and private law grapple with it. Law is built upon legal subjects who must be identifiable and known. Furthermore, legal relations such as contracts may span years, in some cases decades, and the law must have tools at its avail to deal with changes in intent or circumstances etc., whereas a smart contract facilitating a single transaction would by default not cater to the dynamic nature of such legal relations.

69 Merely having a DAO register as an existing corporate form, e.g., as a limited liability company (LLC), as has been promulgated by initiatives such as LexDAO,[66] may not adequately tackle the challenges related to DAOs as discussed above. In order to effectively ensure accountability of actors behind a DAO, both human and non-human, such as a form of artificial intelligence, or AI, a "piercing of the digital veil" of sorts, a more in-depth analysis of the nature of DAOs as borderless, fluid and, to some extent, trustless is required.

70 Hence, in our opinion, the *numerus clausus* of corporate and institutional forms under Swiss law does not encompass a solution for the requirements that new blockchain-based organisations, e.g., DAOs, impose. Due to the proliferation of DAOs as novel organisational forms, both participants in these organisations, as well as external persons or stakeholders, such as the market, consumers, and the state itself, have an interest in legal certainty when dealing with them. Next to the creation of new forms of corporations based on blockchain and thus extending the *numerus clausus* of corporate forms, the creation of digital persons as a separate category of legal personality should also be taken into account which would draw implications on the Swiss CC itself, not merely the Swiss CO where corporate forms are regulated. When Swiss legislators created a novel framework for blockchain-based securities, as discussed earlier, for which the amendments to Swiss CO, among others, have recently taken effect,[67] the opportunity to tackle the issue raised herein was missed. Nonetheless, the Swiss Legal Tech

64 P Tschannen, U Zimmer and M Müller, *Allgemeines Verwaltungsrecht* 3 (Aufl. Bern 2009), 489 et seq., 494.

65 On the correlation between trust in institutions and crime, see L Blanco and I Ruiz, 'The Impact of Crime and Insecurity on Trust in Democracy and Institutions' (2013) The American Economic Review 103 (3), 284–288 <www.jstor.org/stable/23469744>.

66 See <https://lexdao.org/#/>.

67 See n 43.

Association advocated for creating a legal framework for DAOs in the process of public consultation leading to the Swiss legislative reform.[68]

71 Therefore, we argue, the legal framework under Swiss substantive law must be amended to deal with the unsatisfactory situation the novel organisational form of DAOs leaves us with. In the words of Max Ganado et al. the task is one of revolutionary proportions:

72 As a practical matter, the collaborative, distributed, and potentially anonymous processes used to create and deploy these code-based governance algorithms have the distinct potential to create an accountability gap between the designers of a DAO and the outcomes of that DAO. All of these points underscore the need to modernize the guardrails of legal personality to accommodate or catch up with the technological revolution of the last decade.[69]

73 In order to achieve this task, Swiss legislators must consider a number of pitfalls to ensure the sensibility of the framework. Of these, the most prominent ones are described, and ways to deal with them are proposed below.

I. Expand Numerus Clausus or Introduce a New Form of Personhood

74 The first question is whether to address the identified challenges by expanding the *numerus clausus* of corporate and organisational forms to include a special DAO form (Approach 1). Alternatively, instead of legislating for a specific technology, to expand the concept of legal personality as a whole to comprise self-executing software-based agents, among others, DAOs, just as legal scholars have expanded the envelope of legal personality to encompass legal persons, centuries ago (Approach 2).

75 Approach 1 was chosen by the State of Vermont in an effort to create a legal framework for DAOs. As a practical matter, the Vermont legislator determined that the autonomous quality of DAOs merited greater safeguards than those of a traditional business entity. Vermont has explicitly accounted for the extension of the *numerus clausus* of corporate forms through the creation of a new entity type, namely blockchain-based limited liability companies (BBLLCs).[70]

76 This solution is seemingly suitable to deal with the regulatory challenges DAOs pose today, as they are by and large managed by humans, whereby the extent to which DAOs are actually governed 'algorithmically,' as suggested,[71] is contested. As such they are different from other non-code-based organisations such as stock corporations in degree, but not in kind. Therefore, introducing a new form of corporation taking into account some of DAOs idiosyncrasies would be a viable medium-term solution.

77 However, at the speed with which AI's capabilities are increasing,[72] Approach 1 may become ineffective and obsolete sooner than one may think.

78 A midway approach, Approach 2, was chosen by the State of Wyoming legislators for the DAO Bill.[73] In recognising the speed at which AI is developing, the Wyoming DAO Bill introduces the concept of an 'algorithmically managed' DAO to deal with the challenge that future, non-human DAOs may pose without touching on the subject of legal personhood for digital software-based agents *per se*. Instead, under this framework non-human DAOs could be legally incorporated. More specifically, it is stipulated[74] that an algorithmically managed decentralised autonomous organisation may only form if the underlying smart contracts are able to be updated, modified or otherwise upgraded.

79 The distinction between algorithmically managed DAOs and non-algorithmically managed DAOs, i.e. 'member-managed', may seem prudent in light of future potencies of AI. However, vesting management powers to a smart contract in the case of 'algorithmically managed' DAOs may prove to be rather problematic in a legal sense. This is because the pseudonymity, or, on rare occasions, the anonymity,

68 Swiss Legal Tech Association, Public Consultation Submission for the Legal DLT Framework, 23 <https://www.swisslegaltech.ch/wpcontent/uploads/2019/06/SLTA_Vernehmlassungseingabe_Version_final_20190624.pdf>.

69 M Ganado et al., 'Mapping the Future of Legal Personality' (2020) MIT Computational Law Report, 2 <https://law.mit.edu/pub/mappingthefutureoflegalpersonality/release/1>.

70 Ibid. See also Vermont Statute on the Blockchain-based Limited Liability Companies, 11 V.S.A. ss 4173<https://legislature.vermont.gov/statutes/section/11/025/04173> and <https://law.mit.edu/pub/mappingthefutureoflegalpersonality/release/1>.

71 Ganado et al. (2020) (n 69).

72 See GPT-3 <https://openai.com/blog/openai-api/>.

73 Wyoming Senate Bill 38, SF0038 Decentralized Autonomous Organizations <https://www.wyoleg.gov/Legislation/2021/SF0038#-408>. Note: the Bill has passed the Wyoming Senate Committee in March 2021.

74 Ibid 17-31-105. (d).

associated with smart contracts in the case of these DAOs would take away the necessary 'safety valve' and could therefore prove not to be sensible from a public policy perspective. Furthermore, the term 'algorithmically managed' is not precise enough and may thus be misleading as many gradations exist on the spectrum of human – AI interaction to which the bifurcated solution in the Wyoming DAO Bill does not cater. Also, algorithms can be found in any process, the law, chemistry or even a recipe. Therefore, building a legal framework around such polysemantic term is far from ideal.

80 Approach 2 is evidently more radical in nature, as it would introduce a new form of personhood as a whole by recognising digital persons next to natural persons and personhood for legal entities, and in some cases, nature bodies such as rivers.[75]

81 If we go back to the origins of legal personhood, it was the great jurist Karl Friedrich von Savigny, influenced by Kant's considerations on legal capacity in metaphysics of morals, who stated that legal capacity could be expanded to encompass something without the single individual, i.e., by artificially construing a legal person. Here, Savigny proposed that legal capacity shall be expanded to artificial subjects, conceived solely via the power of fiction. This subject was called a "legal person", i.e., a person who is assumed to exist exclusively for legal reasons. Thus, according to Savigny, a legal person is an artificially conceived subject capable of owning property.[76]

82 As of today, courts around the world have ruled on the limitations of rights awarded to legal persons and concluded that legal persons are not only capable of owning property but also capable of personality rights such as the right to a reputation and constitutional rights such as freedom of speech.[77]

83 According to article 53 CC, "legal entities have all the rights and duties other than those which presuppose intrinsically human attributes, as gender, age or kinship tributes, as gender, age or kinship, speech".

84 Consequently, a self-executing digital entity such as a DAO could presumably be awarded legal personality with its algorithmic governance and execution of actions, given that it is arguably more self-reliant and is endowed with more agency than, e.g., a stock corporation which needs human agents for every step of the way when forming, communicating and executing decisions and actions. This is highlighted in the essay 'The Advent of Digital Persons' with a concrete example of a highly autonomous DAO:

> In the near-term future, we will face digital entities who act autonomously on a transnational, distributed network and don't always need a physical manifestation or representation to interact with natural or legal persons. They will manage funds, pay humans for labour, possess things and create other entities – independently of third-party involvement. We propose to accept these entities as autonomous, digital persons as they are endowed with no lesser level of autonomy than the legal persons we interact with on a daily basis. A legal entity relies on its organs comprised of humans, such as a board of directors, presidents, secretaries etc. to act as agents in the process of decision-making, and in the execution of these decisions on its behalf. Legal entities are therefore not autonomous agents. It is precisely the characteristic of agency in form of self-execution without the interference or need of a third party that gives digital entities the necessary level of autonomy to be regarded as digital persons. This does not mean that a digital person must be able to execute its will exclusively without a human being or another party. Especially in the analogue realm, a digital person would still need representation through a human surrogate. But given the current technological developments, the digital person can now act directly without human intermediation in e.g. employing humans and paying their salaries through smart contracts as well as autonomously managing its assets, including transactions of programmable funds.
>
> Such is the case with Plantoids: Plantoids are blockchain-based lifeforms that reproduce through the combination of code and human interaction. The goal of a given Plantoid is to raise enough funds to be able to employ a human surrogate that then would produce the Plantoid's offspring. In the example of the Plantoid, technology no longer acts as a tool but as a peer in a direct relationship with natural or legal persons. Similar to a natural person whose mind inhabits a body, each Plantoid consists of comparable components. On the one hand, its physical body in form of an electro-mechanical construction and on the other hand its "soul" – "represented by an autonomous software agent that lives on a blockchain". If the physical body of

75 O Polat and B Schuppli, 'The Advent of Digital Persons' (2018) Future Cryptoeconomics, Vienna, 37 et seq. <https://riat.at/future-cryptoeconomics/>. An insightful analysis of the topic of digital personhood is delivered in G Teubner, 'Digital Personhood? The Status of Autonomous Software Agents in Private Law' (2018) <https://ssrn.com/abstract=3177096>.

76 KF von Savigny, *System des heutigen Römischen Rechts* (1840), 236.

77 Citizens United v. Federal Election Commission, 558 US 310 (2010).

> the Plantoid is destroyed, the autonomous software agent – in the form of a smart contract – continues to live on the distributed network it was deployed on.[78]

85 From this excerpt it can be concluded that the more non-human agency in the form of AI underlies a DAO, the more adequate Approach 2 may prove to be in legislating for DAOs.

II. National Solution to a Borderless Challenge

86 No matter how intricate a legal framework for DAOs is created by any jurisdiction, the effectiveness of such approach would largely be determined by the legal standards in other jurisdictions. Therefore, unitary national or state-level approaches, such as Wyoming and Vermont, are welcomed but a worldwide uniform and standardised solution which would uphold a minimum liability standard and a framework for international cooperation, exchange of information and cross-border enforcement would be needed to tackle the issue effectively. Lessons can be drawn from effective international legal frameworks such as e.g., in the area of money laundering and terrorism financing, as pioneered by the Financial Action Task Force (FATF), an ad hoc membership body organised under the umbrella of the Organisation for Economic Cooperation and Development (OECD). To this end, civil society groups consisting of leading academics, such as the Coalition of Automated Legal Applications, are working on Model Laws, a helpful tool to set legal standards.[79]

87 Therefore, the paper takes the view that any legislative attempt for Switzerland ought to closely monitor and adequately reflect the research and findings from other legislative attempts around the world.

III. Technical Tools to Increase Accountability

88 If we revisit the damage incurred to Buyer Y as well as the state as collector of taxes in part B, it was exclusively pecuniary in nature. This will be the case for most of the challenging cases we can currently conceive of in relation to DAOs which prompt the need for a novel legislative framework in the first place. Therefore, finding a sensible solution to make damaged parties financially whole when the interaction with a DAO has led to damages for which the DAO must then assume responsibility would be of utmost importance.

89 Just as contract law without the tools for international seizure of assets and enforcement of claims would be useless, even the most intricate legal framework for DAOs would be ineffective in the absence of any tools to seize and distribute assets to damaged parties. Hence, we argue, a legal framework for DAOs must include some form of collateral or minimum insurance requirement for DAOs before they may interact with market participants under an effective protection of a legal framework. With this, the characteristics of smart contracts as deterministic and immutable sets of codes can be used to favour consumers and other market participants who are in need of protection.

90 Even beyond a mere collateral requirement, allowing the state some kind of access to smart contracts, via for instance multi-signature function, where the state holds one of the needed cryptographic keys to transfer assets, governing DAOs may—albeit counter to the libertarian cypherpunk ideal—also be a suitable safeguard to ensure protection of market participants when transacting with DAOs.

IV. Concluding Remarks

91 Irrespective of the approach potentially chosen by Swiss legislators, Approaches 1 or 2, it is our opinion that the likes of the Wyoming DAO Bill, despite their apparent shortcomings, could inspire Switzerland to bring further legal certainty to such emerging novel business models and to better integrate existing concepts such as 'unincorporated partnerships' into its legal and regulatory landscape, thereby helping to achieve a fitting framework for these new business and organisation models, which are here to stay, according to our estimation.

92 To sum up, for any legislative effort, it is crucial that participants in the proposed open marketplace concept by Schär et al. are not left without effective legal recourse. To meet this requirement, we take the view that Swiss legislators need to act with bespoke legislative reform, partially taking account of existing legislation in foreign jurisdictions.

93 In the end, the fundamental need for a functioning legal system of knowing "who are you, with whom

78 Polat and Schuppli (2018) (n 75), 37 et seq. For the Plantoid concept art by P de Filippi, see <https://www.forbes.com/sites/katmustatea/2018/01/31/meet-plantoid-blockchain-art-with-a-life-of-its-own/?sh=b754ceb3f641>.

79 Coalition of Automated Legal Applications, the DAO Model Law, (MEDIUM, 18 December 2019) <https://medium.com/coala/the-dao-model-law-68e5360971ea>.

I have to deal with", to put it in the words of Jeremy Bentham, must be catered to in any sensible legislative approach to tackle the issues and challenges raised herein, even with technical tools such as smart contracts at hand. Otherwise, individuals or non-human entities may escape regulatory supervision and legal accountability, a result which may erode trust and legitimacy in the state power as a whole.

Security Implications of Consortium Blockchains: The Case of Ethereum Networks

by **Adrian Hofmann, Fabian Gwinner, Axel Winkelmann and Christian Janiesch***

Abstract: By definition, blockchain platforms offer secure and reliable data exchange between stakeholders without a trusted third party. Private and consortium blockchains implement access restrictions, so that sensitive data is kept from the public. However, due to its distributed structure, only one node with faulty configuration can leak all blockchain data. For our study, we scanned the Internet for misconfigured private Ethereum nodes. Overall, we found 1421 nodes belonging to 621 blockchains that are not one of the large Ethereum-based networks. For our analysis, we chose a diverse sample of networks. Then, we analyzed in-depth 4 different networks with 10 to 20 nodes enabling 800 to over 34 million transactions. We used the exposed remote procedure call interface of nodes to extract the complete transaction history and to gain insights into the actors' behaviors those networks. We used graph visualization tools to picture the networks transactions and to identify stakeholders and activities. Additionally, we decompiled and reverse engineered smart contracts on the networks to infer the purpose of smart contracts, the network, and its participants' roles. With our research, we show how to reveal confidential information from blockchains, which should not be exposed to the public and could potentially include identities, contract data as well as legal data. Thereby, we illustrate the legal and social implications of data leakage by this distributed and supposedly secure technology. In summary, we show that the large attack surface of private or consortium blockchains poses a threat to the security of those networks. The nodes used in this study were not configured according to the Ethereum guidelines and exposed information directly to the Internet. However, even correctly configured nodes provide an excellent target for attackers as they allow them to gain information about a whole network while only breaching one weak point. Lastly, our study discusses whether (private) blockchain networks can reach a consensus without sharing all data between nodes and what data distribution strategies defend best against weak links in the chain.

Keywords: Consortium Blockchain; Privacy; Security; Ethereum; Case Study

Recommended citation: Adrian Hofmann, Fabian Gwinner, Axel Winkelmann and Christian Janiesch, Security Implications of Consortium Blockchains: The Case of Ethereum Networks, 12 (2021) JIPITEC 347 para 1

A. Introduction

1 Blockchain technology has sparked interest in a variety of industries. Even after the initial Bitcoin hype, blockchain as a technology is still regarded to have the potential to drive decentralization and disintermediation. The cryptographic primitives and consensus mechanisms make storing and transferring of data not only secure and resistant against manipulation but also not reliant on a trusted third party.[1] Consequently, many consider the potential of this technology immense and disruptive.

* Adrian Hofmann, Fabian Gwinner, Axel Winkelmann, University of Würzburg, Chair of Business Management and Information System, Christian Janiesch, TU Dortmund University, Chair for Enterprise Computing.

1 Satoshi Nakamoto, 'A Peer-to-Peer Electronic Cash System' <https://bitcoin.org/bitcoin.pdf> accessed 22 January 2021; Sarah Underwood, 'Blockchain beyond Bitcoin' (2016) 59 Communications of the ACM <https://dl.acm.org/doi/10.1145/2994581> accessed 22 January 2021.

2 Most commercial blockchain applications rely on a private or a consortium blockchain. The purpose of this sort of blockchain is only to allow a select group of participants to read or write data from or to the ledger. Customer-focused solutions, such as the Diem[2] cryptocurrency, use this approach to keep customer transaction data private[3]. However, depending on the protocol's configuration, blockchain nodes share data with every other node on the network. The distributed nature of blockchains makes them more failsafe and resistant to manipulation. Attacks such as 50+1 percent attacks and selfish mining, therefore, are well researched. However, with each additional node that joins the network, simultaneously its attack surface for data theft increases. This implies that, even for large networks, only one misconfigured node can leak the whole blockchain data to malicious actors. In business contexts, information about internal structures can be leaked to competitors. For private use-cases, information about the individual transaction structures can give deep insights into personal behavior and contain the most sensitive information.

3 To assess the severity of a data breach on one node of the network, we conducted a study to determine how information can be extracted and visualized to gain as many insights into a private blockchain as possible. Thus, our study reverse engineers parts of blockchain networks to gain the necessary information. Reverse engineering a system is typically used to infer how an underlying mechanism works. The difficulty of reverse engineering systems is determined by the number of their components and the interdependence of their components as well as the number of their settings.[4] For our work, we chose the Ethereum platform as a framework and a popular part of the blockchain universe. Inspired by the Internet Census[5], our approach relies on data reverse-engineered from a security issue in a faulty configuration of Ethereum. Starting there, we conducted four small case studies on different implementations of the Ethereum platform to identify stakeholders and mechanisms of these networks. Building on this, we want to address the following research questions (RQ) in this study:

RQ1: Which methods and tools are required to reverse engineer Ethereum networks?

RQ2: How much information can be extracted from consortium blockchains with one misconfigured node?

4 Our paper addresses managers, lawmakers and scientists who are interested in a more technical evaluation of the security of private blockchains. In this paper, we contribute methods used in the process of reverse engineering, as well as the results of the evaluation. Additionally, we provide the insights we gained from the reverse engineering of blockchain networks and the implications they provide for the adoption of the technology. The rest of the paper is structured as follows: In the next section, we lay the foundations by discussing relevant literature and previous work. We then introduce the methodology as well as the data we used for the analysis. The following chapter contains our main research results, by first providing an overview of the technological side of the market and then a detailed analysis of four different blockchains and their use. The final chapter summarizes and concludes the research.

B. Foundations and Related Work

5 In its very basics, the blockchain is a distributed ledger of transactions autonomously managed by a consensus mechanism. Technically, it can be pictured as a growing chain of linked blocks, from where its name originates. The blocks of a blockchain are stored distributed by the participants, the so-called nodes.[6] This distribution also brings the advantage that no single party could manipulate already stored data and that the storage is resilient against outages of nodes. The blocks of a chain consist of a block header and a list of transactions. In the Ethereum blockchain, each transaction has one sender and one recipient. Today, it is possible to not only store transactions in the blockchain, but also data objects and small programs, which is how (smart) contracts are implemented.[7] In Ethereum, this is often used to realize user-defined tokens. There are many smart contract-based tokens, often standardized by

2 Formerly known as *Libra.*

3 'White Paper | Diem Association' <https://www.diem.com/en-us/white-paper/> accessed 22 January 2021.

4 Seungwoon Lee, Seung-Hun Shin and Byeong-hee Roh, 'Abnormal Behavior-Based Detection of Shodan and Censys-Like Scanning' (2017) 9. *ICUFN* <https://ieeexplore.ieee.org/document/7993960> accessed 11 January 2021.

5 'Internet Census 2012' <http://census2012.sourceforge.net/paper.html> accessed 11 January 2021.

6 Nakamoto (n 1); Roman Beck and others, 'Blockchain Technology in Business and Information Systems Research' (2017) 59 Bus. Inf. Syst. Eng. <https://link.springer.com/content/pdf/10.1007/s12599-017-0505-1.pdf> accessed 11 January 2021.

7 Kevin Delmolino and others, 'Step by Step towards Creating a Safe Smart Contract: Lessons and Insights from a Cryptocurrency Lab' (2016) vol 9604 Lecture Notes in Computer Science <https://doi.org/10.1007/978-3-662-53357-4_6> accessed 22 January 2021.

Ethereum Request for Comments (ERC) standards, which define their characteristics and interface.

6 Given all transactions in a network, naturally, a graph can be built to model the interactions of the participants. The nodes of this graph do not necessarily have to correspond to the nodes of the blockchain network and must not be confused. One physical node of the network could, for example, host multiple Ethereum accounts and therefore represent several nodes in the transaction graph. Additionally, the nodes of the transaction graph can be smart contracts as well. There has been a lot of prior research on the technical analysis of blockchains. This research strongly focuses on large public blockchains, analyzing the transaction structure of public blockchains and the usage patterns therein. First analyses were used to deanonymize Bitcoin users.[8] In the early years of blockchain, it was still possible to dissect the whole transaction graph of the first cryptocurrencies.[9] Due to Bitcoins' transaction structure, it was necessary to apply advanced heuristics to reconstruct and analyze the user graph of the Bitcoin network.[10] There have been fewer studies on the public Ethereum networks.[11] These studies could only link nodes if Ether (the currency of the Ethereum networks) were sent. To consider all transactions, it would be necessary to include the additional network structure that is built by interacting with smart contracts. Studies researching transaction networks of ERC-20 tokens partially deconstructed those structures.[12] Interaction networks within smart contracts can be researched in a similar fashion.

7 The limited existing research regarding the programming interface (JSON-RPC) of a network focuses mostly on the possible attack surface it provides, such as stealing mining reward and denial-of-service attacks,[13] or the use of blockchain-based applications.[14] So far, we could not find any studies that use this interface to map transaction networks or reverse engineer the users and use-cases of private blockchains.

8 In contrast to other security or software engineering related topics, we focus on extracting knowledge for a more research-driven goal. Therefore, our motivation was led by the "Internet Census" of 2012, where the authors used a security vulnerability to create the first full "map" of the internet. Several researchers used this as a foundation, regarding the provided knowledge as well as the used methods, to get insights in other technologies or security-related issues.[15]

C. Materials and Methods

9 To answer our research questions, we used a multiple case study approach. The case study research design consists of the study's *questions*, its *propositions*, *units of analysis*, the *logic linking of the data to the propositions*, and the *criteria for interpreting the finding.*[16] We already posed the research questions in the introduction of this paper. As units of analysis, we chose the block headers and transaction data, as well as the network node data for different blockchains. To identify potential blockchains for a more in-depth analysis,

8 Fergal Reid and Martin Harrigan, 'An Analysis of Anonymity in the Bitcoin System' [2013] Security and Privacy in Social Networks <https://doi.org/10.1007/978-1-4614-4139-7_10> accessed 22 January 2021.

9 Dorit Ron and Adi Shamir, 'Quantitative Analysis of the Full Bitcoin Transaction Graph' [2013] Financial Cryptography and Data Security <https://doi.org/10.1007/978-3-642-39884-1_2> accessed 22 January 2021.

10 Damiano Di Francesco Maesa, Andrea Marino and Laura Ricci, 'Data-Driven Analysis of Bitcoin Properties: Exploiting the Users Graph' (2018) 6 International Journal of Data Science and Analytics <https://doi.org/10.1007/s41060-017-0074-x> accessed 22 January 2021.

11 Wren Chan and Aspen Olmsted, 'Ethereum Transaction Graph Analysis' (2017) 12th International Conference for Internet Technology and Secured Transactions 498; Andra Anoaica and Hugo Levard, 'Quantitative Description of Internal Activity on the Ethereum Public Blockchain' (2018) 9th IFIP International Conference on New Technologies, Mobility and Security 1.

12 Friedhelm Victor and Bianca Katharina Lüders, 'Measuring Ethereum-Based ERC20 Token Networks' (2019) vol 1159 Lecture Notes in Computer Science 113; Shahar Somin, Goren Gordon and Yaniv Altshuler, 'Network Analysis of ERC20 Tokens Trading on Ethereum Blockchain' (2018) IX Unifying Themes in Complex Systems 439.

13 X Wang and others, 'Attack and Defence of Ethereum Remote APIs' [2018] IEEE Globecom Workshops 1.

14 Chaehyeon Lee and others, 'Blockchain Explorer Based on RPC-Based Monitoring System' [2019] IEEE International Conference on Blockchain and Cryptocurrency 117; Kyungchan Ko and others, 'Design of RPC-Based Blockchain Monitoring Agent' [2018] International Conference on Information and Communication Technology Convergence 117.

15 John Heidemann and others, 'Census and Survey of the Visible Internet (Extended)' [2008] ISI-TR-2008-649; Lee, Shin and Roh; (n 3).

16 Robert K Yin, *Case Study Research and Applications: Design and Methods* (Sage publications 2017).

we first created an overview of the Ethereum platform landscape.

10 To do so, we used Shodan, a search engine for Internet-connected devices. We searched the search engine by the query "port:8545" for Ethereum nodes with an active RPC interface. We additionally searched for the string "Ethereum RPC enabled" but considered the results nearly identical.[17] We exported the 3,042 found IP addresses and metadata from Shodan in CSV format. Each IP address represents a node in an Ethereum blockchain network, with an exposed RPC interface. Technically, this gives everyone the possibility to not only extract data from the whole blockchain but also to manipulate the node. It should however be noted that each node in our dataset is for some reason not configured according to the official recommendations, as the RPC interface should never be exposed openly to the internet. Therefore, we only cover blockchains where at least one node was not configured properly.

mechanism to check how valid our data was and how representative our sample of blockchain nodes was.

Our final overview dataset consists of 2,063 active Ethereum nodes, of which 1421 nodes are used in 621 unique blockchain networks and 622 nodes are connected to the Ethereum main network. The network size of the entire Ethereum main network is at the time estimated at 6,900 nodes according to ethernodes.org.[18] As a result, our dataset covers about 9 % of the Ethereum main network. Additionally, we compared how many nodes of the mainnet[19] are operated in different countries and arrived at a very similar distribution, as shown in Figure 1. We did this estimation with other known networks, such as the various Ethereum test networks, which we extracted from an open-source repository for known networks.[20] We arrived at similar results, which lets us conclude that our dataset covers the overall landscape of the Ethereum platform comprehensively.

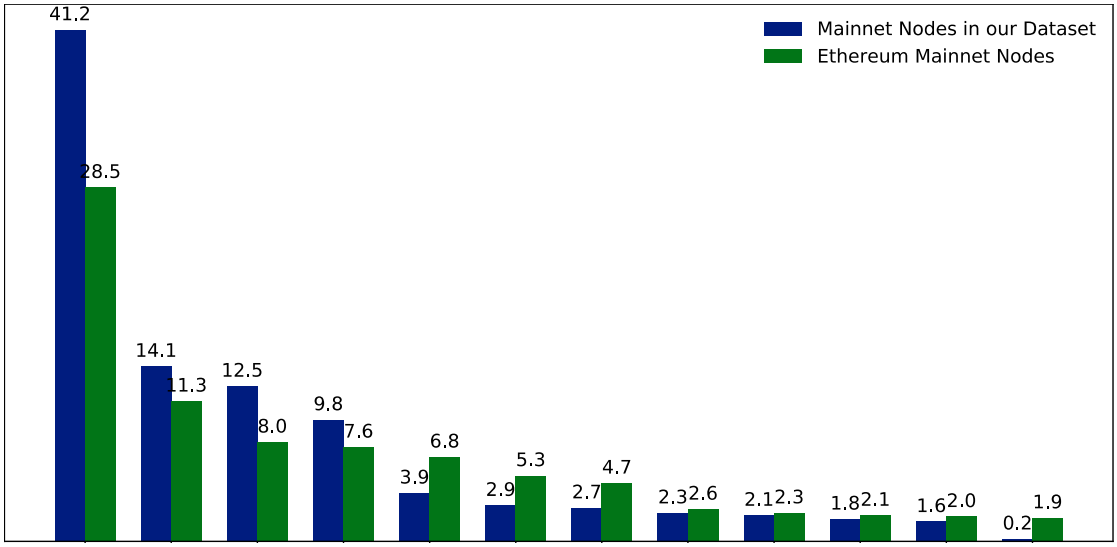


Figure 1: The Distribution of the Mainnet Nodes in our Dataset Compared to all Mainnet Nodes

11 To build our overview dataset on the operation of nodes, we queried the RPC interface of each of the 3,042 nodes. We extracted the chain version, genesis block (i.e., the first block of a blockchain), and information on whether the node was mining or not. To determine the age of each blockchain, we additionally queried the second block of each chain. We decided not to use the timestamp provided in the genesis block since it often provided a zero value in the timestamp. For nodes that are running on the Ethereum main network, we also queried block number 1,920,000 at which the chain splits into Ethereum and Ethereum Classic. We used this as a

12 We used the final overview dataset to provide high-level insights into the Ethereum landscape. Additionally, we used this data to identify potential candidates for our case studies. We chose the blockchains according to the number of active nodes,

17 'Ethereum RPC Enabled - Shodan' (shodan) <https://www.shodan.io/report/VwRYVIqq> accessed 11 January 2021.

18 'Clients - Ethernodes.Org - The Ethereum Network & Node Explorer' (bitfly gmbh 2021) <https://ethernodes.org/> accessed 11 January 2021.

19 Mainnet refers to live blockchain where tokens are in use.

20 Sebastian Gerske, 'GitHub - Ethereum-Navigator/Atlas: The Single Source of Truth for All Ethereum Networks.' <https://github.com/ethereum-navigator/atlas> accessed 11 January 2021.

length, and age of the blockchain as well as the distribution of nodes. The goal was to get a diverse set of blockchains to study and draw generalized conclusions. For the chosen blockchains, we extracted account holders for each node and the complete blockchain record of transactions. To identify usage patterns, we used social network analyses on the transaction networks to identify commonly used smart contracts. We extracted and decompiled the smart contracts with the Panoramix decompiler[21] to find out what their role in the blockchain is. While this is a state-of-the-art approach, the decompilation of Ethereum contracts is still in an experimental stage and does not guarantee success. Therefore, we were not able to decompile and analyze all relevant smart contracts. We summarize the overall data extraction process in Figure 2. The mix of source code analysis and social network analysis allowed us to reverse engineer use cases and interaction patterns with the blockchains, and hence provide a suitable way to investigate the proposition.

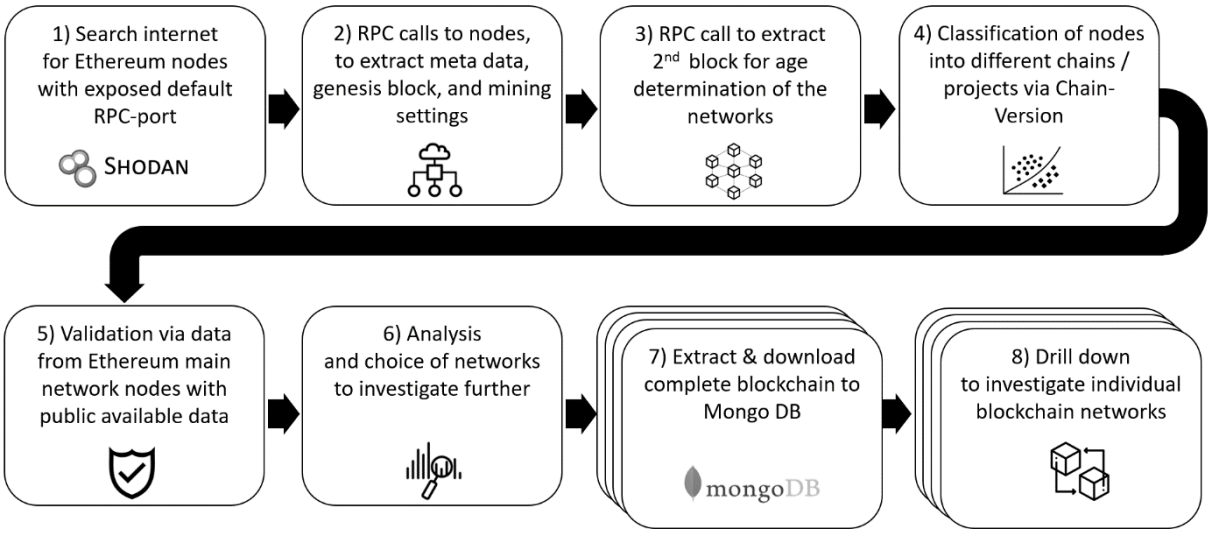


Figure 2: Overall Data Collection Process

D. An Analysis of Business Blockchains within the Ethereum Landscape

13 The primary analysis of this paper consists of two parts. First, we describe the overall landscape of the Ethereum protocol using the overview dataset. From there, we can draw the first conclusions, before providing a more in-depth analysis of four case studies for Ethereum-based blockchains.

21 eevm, 'Panoramix' <https://github.com/eveem-org/panoramix> accessed 11 January 2021.

I. Mapping out the Ethereum Landscape

14 To get an overall view of the Ethereum Landscape and map our findings, we analyzed the metadata from the collected dataset. For further analysis, we have chosen different dimensions, which contribute to our overall goal and give us first useful insights in the Ethereum universe to determine the potential case study candidates later.

15 As a first dimension, we analyzed the hosting of the different nodes. Figure 3 (left)shows that almost 75 % of all nodes are hosted by major hosting or cloud providers. With over half of all nodes, the big cloud providers Amazon, Digital Ocean, Microsoft, Google, and Alibaba are claiming a large piece of the Ethereum hosting. This shows that the Ethereum technology shows great potential for business adoption since the cloud setup process is a fast solution to get started. It is an advantage over other technologies, which currently rely on specialized mining hardware that is not widely available.

16 We were surprised by the large share of cloud providers since one of the main advantages of blockchain applications is its distributed topology that affords the technology security and resilience advantages. These advantages are strongly mitigated, when the majority of nodes use the same hosting provider or same data center.[22] To use the full potential of decentralization, blockchain nodes should be

22 Xiaoqi Li and others, 'A Survey on the Security of Blockchain Systems' [2017] Future Generation Computer Systems 841; Deepak Puthal and others, 'The Blockchain as a Decentralized Security Framework [Future Directions]' (2018) 7.2 IEEE Consumer Electronics Magazine 18.

hosted on-premise. We assume to see a smaller share of cloud providers in the dataset, once the technology is more adopted.

17 As another dimension, we analyzed the country where the nodes are operating. This analysis should give us a picture where most of the Ethereum projects are implemented and may be used as a hint in which country the technology receives most attention. However, since the nodes are mostly cloud-based, this metric can be skewed. Additionally, because nodes of the same chain can operate in different countries, it was not possible to normalize our analysis.

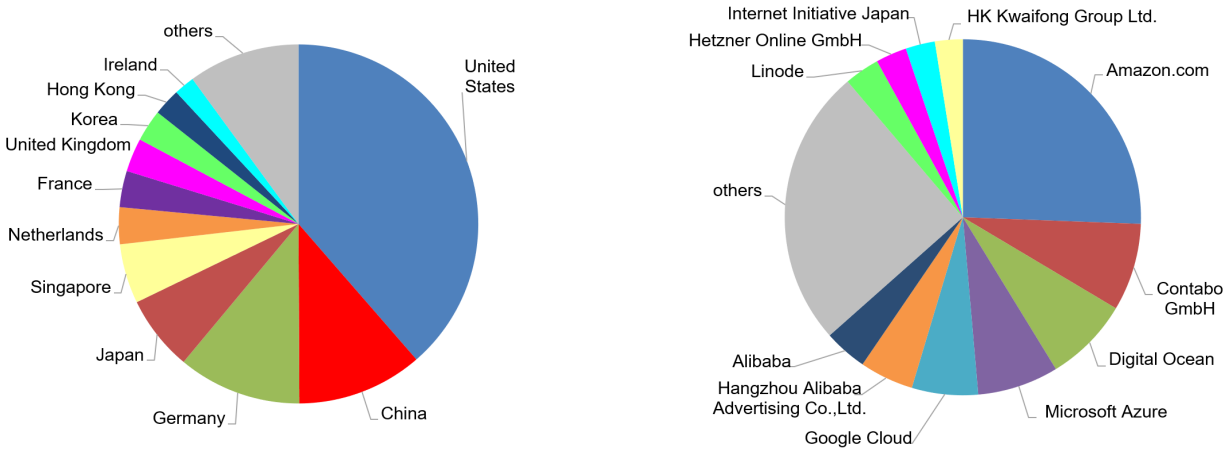


Figure 3: Distribution of Nodes per Hoster (left) and per Country (right)

Instead, we have decided to include all nodes in this distribution (Figure 3 (right)) to give a weighted analysis of origin. Therefore, blockchains operating with more nodes increase the respective share of a country. With this knowledge, the chart becomes an activity analysis, showing which country is more active and may have advanced further in the process of adopting Ethereum technology. Yet from this point of view, it is not possible to determine if there are more projects or just networks with more nodes that determine the share of a country. To determine the state of the different chains and thereby to gain knowledge about the phase in which these projects are, we analyzed the length of the different chains. Figure 4 (left) shows that there are many very short chains. After analyzing and exploring some random samples of these short chains, it showed that these were purely test setups, either with only some test data, partly with less than ten transactions or even completely empty. Extracting information form these projects does not advance this study, and, therefore, we did not consider them in our analyses further. To achieve better knowledge of potential chains, which we could use for further analysis, we analyzed the age of the different implementations. Figure 4 (right) shows the distribution of age, based on the first block. That the initiation of most chains was less than a year ago leads to the conclusion, although the technology is not new anymore, that either projects implementing it are still in an experimental state or that only projects in an early stage still have misconfigured nodes.

18 To consolidate our findings, we put the length of chains in relation to their age, illustrated in Figure 5. Newer but longer chains are either configured with a shorter time per block (block time) or represent fast-growing chains. Older but shorter chains were more mature blockchains such as the Ethereum main- and testnets as well as other public Ethereum-based projects. There is a visible forming of "beams" originating from the lower right corner. All networks on the same beam have the same configuration for the block time. There seem to be only a few main variants for this configuration, which could indicate that many of the private Ethereum networks only use a few boilerplate projects as setup. Considering just the distribution and the aggregation of a line in the center, we assume these represent chains with the default configuration. Additionally, increasingly short block times (indicated by a strong negative slope) are introduced in the last years. This could be either due to the need for higher transaction throughput and lower latency or due to the increase in computation power and network speed. A common criticism of the blockchain technology is the high computational overhead and the resulting lack of performance.[23] Blockchains running at a lower block time are less performance-intensive and are less likely to become out of sync. Additionally, when using the proof-of-work consensus mechanism, shorter block times indicate a lower difficulty,

23 Kim, Soohyeong, Yongseok Kwon, and Sunghyun Cho, 'A Survey of Scalability Solutions on Blockchain' [2018] International Conference on Information and Communication Technology Convergence 1204.

and therefore, a higher risk of double-spending attacks in the network. However, since most private blockchains are not based on this mechanism, we do not research this phenomenon further in this paper.

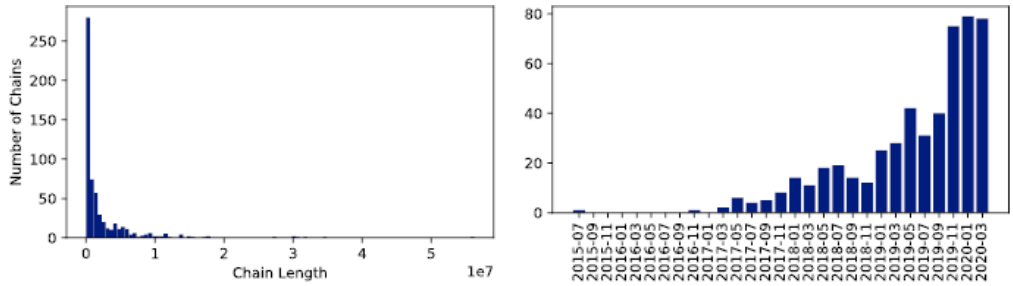


Figure 4: Distribution of Blockchain Length (left) and Number of Networks over Time (right)

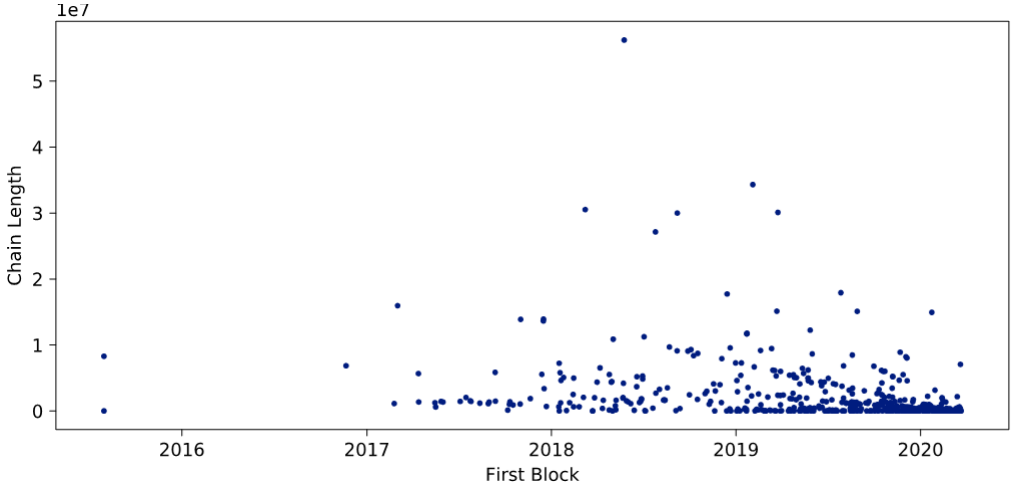


Figure 5: Blockchain Length in Relation to Age

II. Detailed Analysis of Consortium Blockchains

19 As shown in the previous section, most of the networks are either not mature enough to research or are inactive. We identified many blockchains with only one active node and some networks with less than ten transactions over the last two years. For our case studies, we chose four blockchains, that all have more than ten active nodes as well as more than 1 million blocks. Additionally, we excluded the large public blockchains, like the Ethereum mainnet and the various public test networks. Table 1 summarizes the networks chosen for analysis.

Table 1: Blockchains for Case Studies

Case	Network ID	First Block	Length	Number of Nodes	Number of Transactions
1	10	2019-11-03	1,400,000	16	29,000
2	1337	2019-10-22	7,500,000	20	804
3	2894	2018-11-04	3,200,000	13	2,700,000
4	159	2019-08-18	10,500,000	19	34,000,000

1. Case Study 1: Network ID 10

20 We chose the first blockchain we analyzed for its unique properties. It uses the chain version 10, which could indicate that it uses the Quorum variant of Ethereum. Quorum is being developed by JP Morgan Chase as a blockchain, particularly for financial transactions, and offers additional features for this purpose. The Quorum protocol is designed as a permissioned or private blockchain.[24] The analysis of the transactions revealed an unusual transaction graph. Only 102 addresses were creating a one-to-one pairing of senders and receivers as displayed in Figure 6 (left). More precisely, half of these addresses only sent transactions to a single address, and the other half received transactions from a single address. In all following graphs, accounts are colored blue and smart contracts are colored red. The width of the edges indicates the number of transactions sent from one node to another.

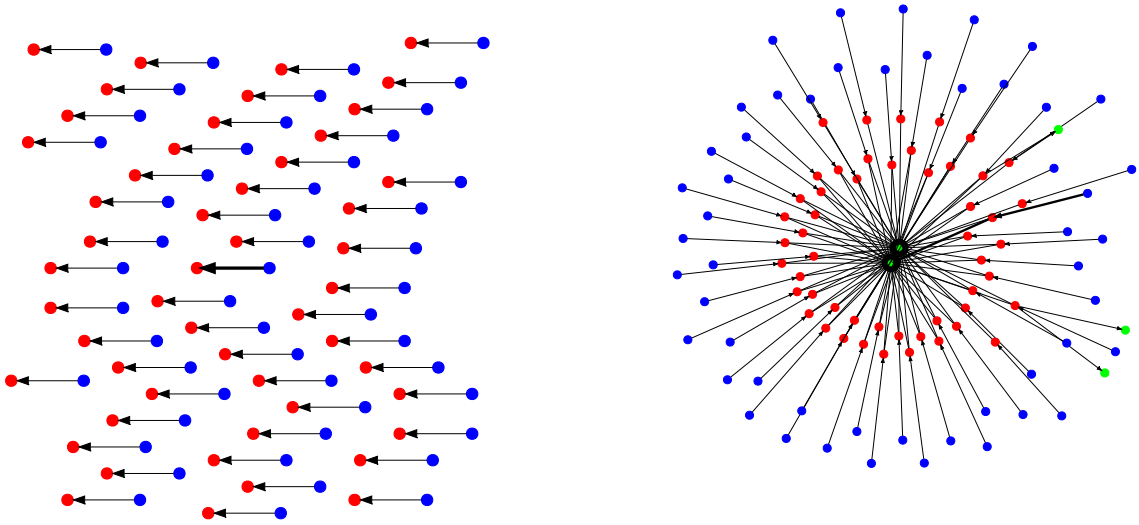


Figure 6: Complete Graph without (left) and with Proxy Contracts (right)

21 This structure led to the assumption that the receivers are all smart contracts with a single user each. We hence queried the nodes for the contract code of the addresses, downloaded, and decompiled the code. The contract provided 22 public functions, most of which are used to manage ownership and access to the smart contract. However, the transactions called only one of those functions named *execute*, which takes two parameters as input. The first parameter is an address of the contract, which the call is delegated to. The second parameter are the parameters of that contract call. This means that the smart contracts, we identified initially, are so-called proxy-contracts that are used to call other contracts. We expanded the transaction graph by the contracts that were called by the proxy contracts. We show the resulting full transaction graph in Figure 6 (right).

24 JP Morgan Chase, 'Quorum Whitepaper' <https://github.com/ConsenSys/quorum/blob/master/docs/Quorum Whitepaper v0.2.pdf> accessed 11 January 2021.

The added contracts are colored in green. It can be seen that there are two very central contracts that contain the actual logic, and that every user interacts with. Unfortunately, we were not able to decompile these contracts, and therefore were unable to find out what the purpose of this blockchain network is. However, the overall structure lets us assume that the centralized contracts only accept calls from the proxy contracts and that the proxy contracts are used to manage user access. It should also be noted that the calls to the smart contract are not associated with any cost. Normally deploying or calling a smart contract would cost the user gas[25], which is paid for in Ether. However, the accounts all have a balance of zero Ether and there are no transaction fees in this network. This, along with the fact that the central smart contracts were too complex to decompile, could imply that the developers test a novel use-case that exceeds the current computational limits of standard Ethereum configurations.

22 From a social network perspective, the graph seems very decentralized. Since each user interacts with only one proxy contract, which in turn interacts with at most two other contracts, the out-degree centrality of the nodes is equally distributed between the users. It should be noted that one user sent 87.6 % of all transactions. Additionally, we examined how many blocks were mined by each individual miner. With 85.4 % of all blocks, we do not consider this a secure network, since this miner has over 50 % of mining power.[26] With this much power for one node, it should be reevaluated if a centralized solution could be a better alternative.[27] However, if the network is indeed only a test setup, the security implications are not as important.

2. Case Study 2: Network ID 1337

23 The second blockchain we identified exhibits a different kind of centralization. While the nodes are distributed all over the world, they are all hosted in the Microsoft Azure cloud. This centralization to a single provider gives a single entity immense power over the network, since it could completely shut down all nodes or simply block access to the nodes on short notice.[28]

24 Furthermore, we noticed that many contracts deployed on the blockchain use smart contracts developed by Ambisafe[29]. Ambisafe offers a blockchain quickstart platform that lets users easily build a blockchain by using preconfigured modules. We identified an EToken2 contract, which offers advanced token functionality but is compatible with the ERC20 interface. Additionally, we identified contracts for identity management (ERC725) and claim management (ERC735). Again, we found proxy smart contracts, but in this case, they were not for access management, but they made contracts upgradeable.

25 The overall network structure looks distributed, as shown in Figure 7 (left). There is one centralized node that interacts with a lot of smart contracts. Approximately a third of these contracts are EToken2 contracts. Each of these contracts corresponds to a contract deployed by the same address that allows transfers of EToken2 to ICAP addresses. These are addresses that are compatible with the IBAN bank account numbers. Another very central node is the smart contract in the upper cluster. This smart contract is a claim management contract. While this looks like the architecture of a decentralized exchange, there is little to no interaction of different accounts with each other, either direct or via smart contracts. Figure 7 (right) shows the transaction graph with a dot layout[30], which indicates that the transactions all flow in only one direction. In addition to this unidirectional transaction flow, the root node holds an overwhelming majority of Ether with approximately 10^{32} Ether. In comparison, the second largest account holds 18.7 Ether, while most accounts hold less than one.

26 We conclude that this is an experimental setup that is used for testing or demonstration purposes only, or possibly a network that is currently being built and the funds are being distributed to the nodes according to their needs.

25 Gas measures the amount of work of miners to include transactions in a block.

26 Nakamoto (n 1).

27 Karl Wüst and Arthur Gervais, 'Do You Need a Blockchain?' [2018] Crypto Valley Conference on Blockchain Technology < https://doi.org/10.1109/CVCBT.2018.00011> accessed 11 January 2021.

28 Primavera De Filippi and Smari McCarthy, 'Cloud Computing: Centralization and Data Sovereignty' (2012) 3.2 European Journal of Law and Technology 1.

29 'Ambisafe | Making Financial Markets Universally Accessible.' (Ambisfe) <https://ambisafe.com/> accessed 11 January 2021.

30 John Ellson and others, 'Graphviz—Open Source Graph Drawing Tools' [2001] International Symposium on Graph Drawing < https://doi.org/10.1007/3-540-45848-4_57> accessed 11 January 2021.

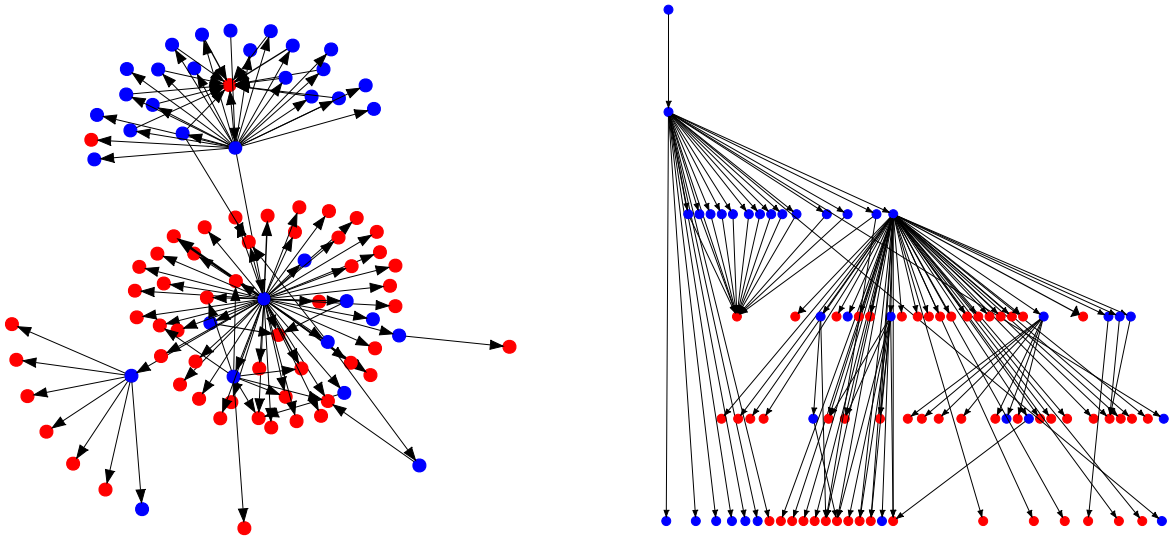


Figure 7: Transaction Graph in Neato Layout (left) and Dot Layout (right)

3. Case Study 3: Network ID 2894

27 The first insight of our analysis was that there are no smart contracts deployed in this network. This means that the transactions transfer Ether. In fact, the transactions in the network carry on average 2,176.3 Ether.

28 The overall transaction graph is much larger than the previous blockchain. The network consists of 15,489 addresses. This size makes it too complex to display completely. Therefore, we chose the representation of the graph as an approximation in Figure 8 (left) by only displaying edges where there were more than 1,000 sent transactions with the corresponding nodes. The second representation we chose was a transaction graph that only displays those transactions that have data attached in addition to the transaction value, as shown in Figure 8 (right). We could not identify what this data represents since the data seemed to be in the form of arbitrary numbers not correlated with the transaction value. However, there were three different types of numbers: small numbers between 1 and 256, medium numbers around 10^6, and extremely large numbers in the order of magnitude 10^{56}.

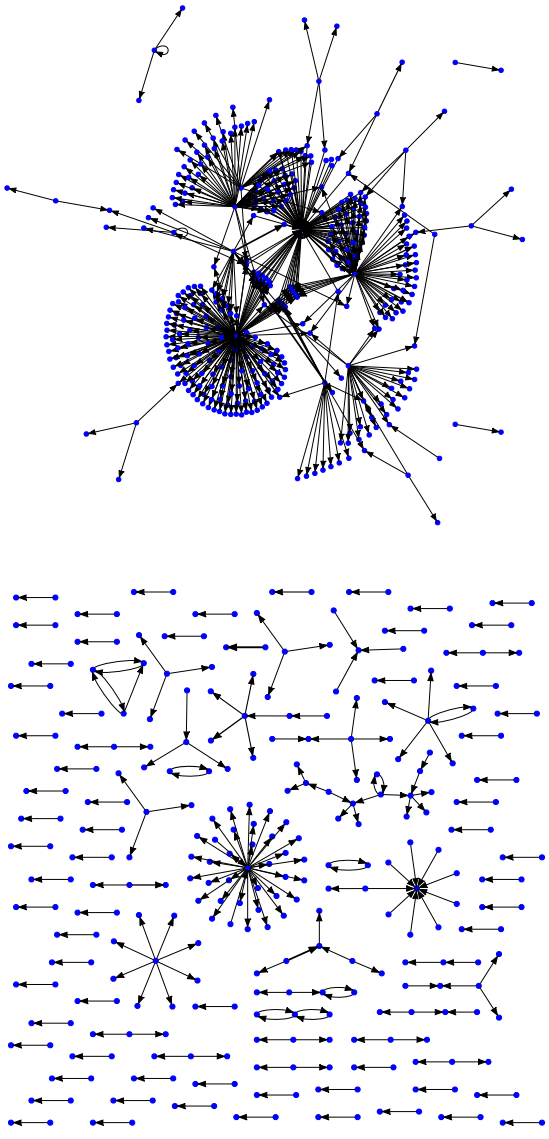


Figure 8: Transaction Graph with nodes with more than 1,000 Transaction (top) and with attached data (bottom)

Even though the number of nodes is much larger than other networks, the graph is much more centralized. Figure 9 (top) shows the indegree and chad to use a logarithmic scale due to the massive differences in centrality. These differences could be as a result of an initial token distribution process. Additionally, the distribution of mining power is not distributed equally either. Figure 9 (bottom) shows that two miners mined a disproportionally large share of the blocks. While this might not be an immediate problem, if those two miners cooperate, they could overrule the rest of the network. Finally, the distribution of Ether is unequal among the nodes, but it is not nearly as unequal as seen in the previous case study. A large portion of the nodes have one to 10^8 Ether, but the majority have less than one. The centralized transaction network and mining, as well as the unequal distribution of Ether, are phenomena that can be seen in large public blockchains, in particular because larger networks tend to centralize. This network, despite its use as a pure accounting network, is the most used network in our dataset.

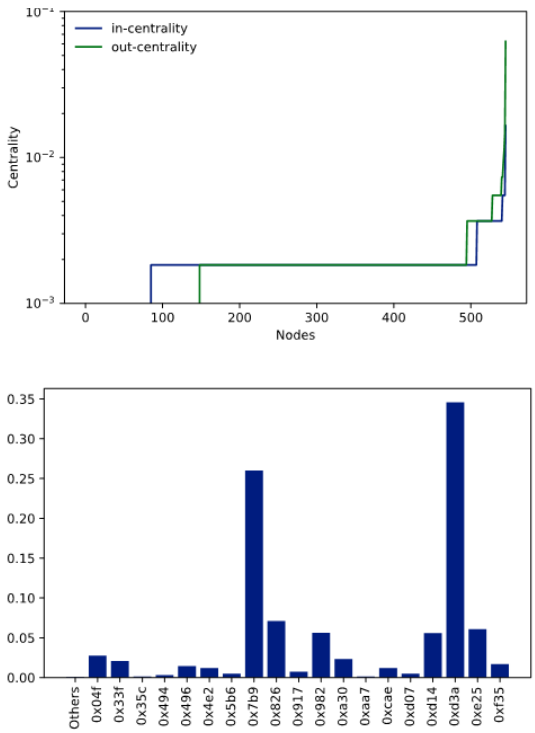


Figure 9: Centrality Scores per Node (top) and Share of Mined Blocks per Miner (bottom)

4. Case Study 4: Network ID 159

29 Our last case study concerns a network that has a massive number of transactions. Since it was launched, the network has about 20 % of the public Ethereum mainnet transactions. The Ethereum mainnet is used by thousands of users. However, we noticed a very centralized contract in the network, as shown in Figure 10 (top). We identified it as a TomoChain BlockSigner smart contract[31], which is used as an alternative consensus mechanism. In fact, all smart contracts we identified are used for this mechanism, and the transactions therein are not relevant to the actual transaction network structure. Therefore, we also analyzed the network structure of the remaining network separately as shown in Figure 10 (botom). The resulting graph only considers 895 transactions.

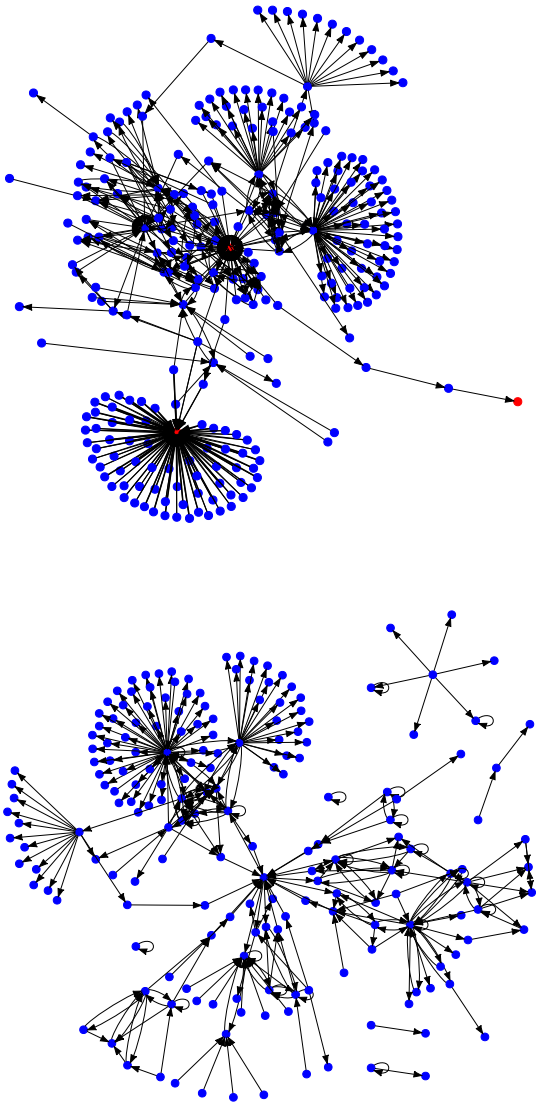


Figure 10: Transaction Structure with (top) and without Smart Contracts (bottom)

30 This transaction graph is not fully connected. There are some small islands with unidirectional transactions. The main island consists of a few larger clusters of outgoing transactions. Again, this could indi-

31 TomoChain~R&D~Team, 'OmoChain: Masternodes Design-Technical White Paper Version 1.0' (tomochain Pte. Ltd. 2018) <https://tomochain.com/docs/technical-whitepaper--1.0.pdf> accessed on 11 January 2021.

cate an initial token distribution process. Since this network is not as old as the previous network we analyzed, it could show much more activity in the future and build a similar transaction graph. Since a smart contract handles the block generation process, we could not easily identify the miners of the blocks, and hence could not analyze the distribution of mining power.

31 Upon further investigation through the IP addresses of the nodes, we found out that the network is connected to the Caelum Project, which is not accessible anymore. It is described as a decentralized storage solution, to secure digital crypto assets[32] with inheritance functionalities.[33]

E. Conclusion

32 Past research on blockchain security has focused mainly on the prevention of fraudulent transactions. However, with the rise of private and consortium blockchains, data privacy has become another important topic, lacking extensive research. Against this backdrop, in this paper, we analyzed the exploitation potential of misconfigured private blockchains. Our approach consisted of reverse engineering actual implementations of the Ethereum platform for individual use-cases to analyze the transaction structure and smart contract implementations, to gain insights into the usage patterns and stakeholders of the networks.

33 In our first research question, we asked, which methods and tools are required to reverse engineer Ethereum networks. Our approach consisted of using a port-scanning dataset and enriching it with additional data that the listed nodes provided. Using social network analyses and source code analyses, we additionally conducted small case studies on selected networks. The social network analysis proved to give useful insights into the actual usage of the network but fell short of revealing the whole structure without the source code analysis of the smart contracts. The smart contract analysis was a very successful approach for some networks, while for others, we could not retrieve the source code of the smart contracts by decompiling them. The main improvement we would suggest for future research would be a "magical" decompiler that can retrieve the original commented source code from Ethereum bytecode. Additionally, it should be checked whether some of the analyses can be automated, to give a quick overview of all networks fast and not rely on analyzing them step by step.

34 Our second research question was how much information can be extracted with only one misconfigured node. We could identify that our approach is not able to paint the full picture of the networks but can give valuable insights. For some networks, we could link IP addresses and specific smart contract structures with publicly available data to get insights of stakeholders. For other networks, we had to rely on the transaction structure and could only identify entities by their cryptographic addresses. Especially for Ethereum networks, each node holds a full copy of the ledger. Therefore, all analyses were based on a maximum of available data. In further research, other structures such as the Hyperledger project should be examined, where the network is segmented into channels. Here, attacking only one node should only provide partial information about the network and would hence call for more elaborated analysis techniques.

35 Due to the availability of data, our research focused on organizational entities rather than individuals. However, the results indicate that for our analysis of the data from an analytical point of view, it does not matter whether the data is of organizational or personal nature. Network structures and agreements can be derived or inferred be it the one or the other. Therefore, we think that the results can be transferred to blockchain networks comprising end users sharing personal data. Thus, our study also raises the very relevant question as to whether (private) blockchain networks can reach a consensus without sharing all data between nodes and what data distribution strategies would defend best against weak links in the chain that exposes private information of individuals.

36 Our dataset consists of over 621 unique blockchain networks, of which we were only able to analyze four for more detailed insights. The process of retrieving and analyzing the entire blockchain for many networks is extremely time consuming, but we are sure that analyzing a larger portion of it would give even better insights into information extraction processes. Overall, improving the systems and tools needed for the reverse engineering as well as a full analysis for the network information, can therefore be future work.

37 The research provided us with an exciting puzzle that is still not assembled completely. We, therefore, hope that the approach is adopted for other

32 Crypto assets are "a new type of asset recorded in digital form and enabled by the use of cryptography that is not and does not represent a financial claim on, or a liability of, any identifiable entity". European Central Bank, 'Crypto-assets – trends and implications' <https://www.ecb.europa.eu/paym/intro/mip-online/2019/html/1906_crypto_assets.en.html> accessed on 11 January 2021.

33 'Caelum Project' <https://web.archive.org/web/2020*/www.caelumproject.io > accessed on 11 January 2021.

blockchain technologies such as Hyperledger or even other unrelated technologies to improve current tools.

This work has been developed in the project PIMKoWe. PIMKoWe (reference number: 02P17D160) is partly funded by the German ministry of education and research (BMBF) within the research program "Industrie 4.0 – Kollaborationen in dynamischen Wertschöpfungsnetzwerken (InKoWe)" and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

JIPITEC

Journal of
Intellectual Property,
Information Technology,
and Electronic Commerce
Law

www.jipitec.eu