Italian Blockchain Association

Francesco Rampone*

Data Protection in the Blockchain Environment

GDPR is not a Hurdle to Permissionless DLT Solutions

Ciberspazio e diritto, vol. 19, n. 61 (3 - 2018), pp. 457-20

Abstract

Public keys and hashes are the two fundamental cryptographic solutions commonly used to develop blockchain networks. They are considered almost unanimously pseudonymous data, that is personal data concealed behind an alphanumeric string that, in combination with additional information, can be nevertheless linked to a specific individual. If this were true, the development of blockchain technology would be hurdled by the necessity to comply with GDPR. In this paper, I held that the definition of personal data, albeit in the form of pseudonymous data, set forth in Directive 95/46/EC and today in the GDPR (taking into account the CJEU interpretation and Article 29 Working Party opinion) does not apply to either the public keys or the hashes as they are used in a blockchain. Indeed, they are not used for concealing identities but rather to solve a technical problem (the so-called *double spending problem*) creating trust in a peer-to-peer network. Hence, although they could be (and sometimes are) used to carry out advanced digital forensic searches to track down the identity of the private key holders, they are not actually designed to conduct or allow for such searches and, consequently, they should be considered neither personal nor pseudonymous data.

* IT lawyer. President of the Italian Blockchain Association.

Francesco Rampone

Data Protection in the Blockchain Environment
GDPR is not a Hurdle to DLT Permissionless Solutions

Summary: 1. Introduction. – 2. Personal data in the blockchain. – 3. The public key. – 4. The hash function. – 5. Coinbase and the intentional seeding of personal data in blockchain. – 6. Data controllers in a blockchain environment. – 7. Conclusions.

1. Introduction.

Regulation (UE) 679/2016 (GDPR), not unlike Directive 95/46/CE, follows a centralistic approach, in the sense that it contemplates data processing only in a *vertical dimension* as operations/processes that a person (the controller) performs on personal data, by independently choosing the means and purposes of the processing, and possibly using services from third parties providers (data processors and sub-processors) when appropriate.

Conversely, DLT solutions, and particularly the blockchain that is their most notable expression, unfold along a *horizontal dimension*, presenting an elusive nature that does not seem to fall squarely within the wording of the Regulation[1], perhaps not due to a faulty structure of the latter, but rather due to the intrinsic nature of the decentralized architecture of a peer-to-peer network.

A law, whatever its structure, always contemplates a conduct or an event which allows for the identification of a person who is responsible for harmful consequences of such conduct or event. A blockchain, on the contrary, behaves by synthesis of the will of a number of individuals anonymously interacting with each other, executing the code of open source software that operates on the basis of a P2P protocol. They are often even unaware of the overarching global project in which they are taking part, thus giving rise to what may be defined as an artificial organism[2].

If the blockchain technology is a new stage in the Internet evolution[3], if we believe that decentralized interaction between anonymous users will engender an ever greater number of

[1] For a first introductory analysis on the reconciliation between blockchain technology and GDPR, see S. SATER, *Blockchain and the European Union's General Data Protection Regulation: A Chance to Harmonize International Data Flows*, 2017, SSRN.

[2] J. GARZIK, *Bitcoin, the organism*, TEDx Talk, Binghamton University, New York March 30, 2014, who defines bitcoin as a living organism and his work as software developer as a biologist's research. In the same line of thinking is P. DE FILIPPI, *Blockchain Technology and the Future of Work*, Lift:Lab, Geneva Feb. 11, 2016, who compares DLT to *plantoids*, an artificial proto-life form conceived and realized for the first time by a research team of IIT – Italian Institute of technology (Center for Micro-BioRobotics). More recently, see MATAN FIELD, *The Blockchain Revolution: From Organisations to Organism*, TEDx Talk, Breda Nov. 3, 2016.

[3] According to DON & ALEX TAPSCOTT, *Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business and the World*, Portfolio/Penguin 2016, with the blockchain we are witnessing the second digital revolution after the Internet: «*When decentralized blockchain protocols start displacing the centralized web services that dominate the current Internet, we'll start to see real internet-based sovereignty. The future Internet will be decentralized*» (second conver). Others believe that blockchain is setting about reconfiguring Internet in a completely new fashion; an Internet 3.0, form the access to static pages (1.0), to *user generated* contents (2.0), to the

autonomous organizations detached from the will and responsibility of the natural persons who belong to them[4], then we must conclude that the GDPR in a blockchain environment will soon reach its elastic limit, beyond which we are forced to think of it (and obviously of many other laws) in a completely new way.

Before we get to that point, however, let us ask ourselves whether the use of blockchain technologies, in their permissionless fashion[5], such as bitcoins, are already compliant with the provisions of the GDPR. To do so, we must first of all question the basic cryptographic solutions used in a blockchain, i.e. the public keys and the hashes: are they personal data? If so, the nodes of a peer-to-peer network shall abide with the formal and substantive requirements set forth in the Regulation and imposed upon the data controller.

2. Personal data in the blockchain.

A blockchain, regardless of its design, uses two cryptographic solutions: the RSA algorithm and the hash function[6]. As for the first one, we are all familiar with the asymmetric keys commonly used for digital signatures and certified e-mail. The keys are nothing but two sequences of numbers and letters generated in pairs where the text ciphered with one of them can only be deciphered with the other. This means that by publishing one of the two keys (the so called *public key*) and stating at the same time to be the holder of the other one (the private key), but keeping it confidential, one can prove the origin, and therefore also the authorship, of a message. If it can be deciphered with a specific public key, that means that it was encrypted only by the person who claims to be the owner of the corresponding private key.

As for the hash function, it is a mathematical algorithm that turns a text of arbitrary length into a sequence of numbers and letters of a fixed, pre-defined length such that a minor change in the input (the text subjected to encryption) corresponds with a significant change in the output

construction of a disintermediate network (3.0), as the blockchain actually is (for a review of blockchain evangelists, see ROB MARVIN, *Blockchain: The Invisible Technology That's Changing the World*, in PC Magazine 2017, pcmag.com).

[4] Through the blockchain, we can build Decentralized Autonomous Organizations (DAO) which have their own governance rules and operate in the market through the execution of lines of code (*smart contract*). Since DAOs have neither administrative nor supervisory bodies, they do not belong to any known legal entity standard and their nature is much debated. Furthermore, DAOs do not shield their participant from direct and unlimited responsibility that, however, remains mostly covered by anonymity (also the network of developers of the bitcoin protocol – BitcoinCore – is a DAO). For more details, see R. C. MERKLE, *DAOs, Democracy and Governance*, Cryonics Magazine, July- August, Vol 37:4, pages 28-40.

[5] Public blockchains (also known as *permissionless* or open blockchains) are those where each user may on his own initiative, without the necessary acceptance by the participants of the network, assume the function of a node, i.e. a user that contributes to the operation of the peer-to-peer protocol. There are also private or semi-private blockchains (also known as *permissioned* blockchains) where each member may enter the network upon prior acceptance of the existing nodes or, even though entry is free, upon acceptance of some nodes that are exclusively endowed with privileges or special functions (the Ethereum platform, the first ever for the number of nodes, has a design aimed at creating permissioned blockchain for business).

[6] The mysterious Satoshi Nakamoto in his paper appeared in the web in 2008 described the whole bitcoin protocol functioning essentially as an original mix of these two mathematical instruments: S. NAKAMOTO, *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2008, https://bitcoin.org/bitcoin.pdf. To learn more about how bitcoins work and, more generally, how a blockchain works, I suggest among many: A.M. ANTONOPOULOS, *Mastering Bitcoin: programming the open blockchain*, O'Reilly Media, 2017; ID., *The internet of money*, Vol. 1 e Vol. 2, Merkle Bloom LLC, 2016-2017; DON & ALEX TAPSCOTT, *Blockchain Revolution* cit.. In particular, for clarity and completeness, see J-L. VERHELST, *Bitcoin, the blockchain and beyond: a 360-degree onboarding guide to the first cryptocurrency and blockchain*, Amazon Digital Services, 2017; A. WRIGHT, *Blockchain: uncovering blockchain technology, cryptocurrencies, bitcoin and the future of money*, Amazon Digital Services, 2017; J.B. MORLEY, *That Book on Blockchain: A One-Hour Intro*, Amazon Digital Services, 2017. More broadly, on the *peer-to-peer* phenomenon, although dated, I suggest the still current, visionary and full of food for thought florilegium by A. ORAM, *Peer-to-peer: Harnessing the benefits of a disruptive technology*, 2001, O'Reilly Media, Inc.

(a completely different output indeed). This feature makes it possible to create a "fingerprint" of the text, called a hash or digest, that almost uniquely identifies the text[7] but does not contain enough information to be reconverted to the original text. This is what mathematicians call a one-way function.

At this point, the question becomes: can the public key and the hash, as ordinarily used in a blockchain, be considered personal data? And, if so, under what conditions?

There are several blockchains, some of them specifically designed to protect privacy without reducing efficiency standards[8]. However, these solutions are of no or little interest to jurists because they do not concern the very nature of the data processed, but rather are only aimed to legitimize the processing by adopting sophisticated cryptographic techniques.

3. The public key.

Among those who first delved into the blockchain technology implications under a privacy and data protection perspective, some held that public keys are without exception pseudonymous data[9], even personal data[10]. I believe this is partially wrong, and in my view a stance dictated by the erroneously intended relation between the public key and personal identity of the holder of the corresponding private key. This is probably due to, in part, to the ordinary use we make of the public keys in connection with certified e-mail and digital signature services where they are used just for purposes of identifying the individual who is in possession of the private key.

Since the equation *public key=pseudonymous data* decisively affects any other thought in the field, with this paper, I intend to demonstrate not only that this is not always true, but that it's not true at all when public keys are employed by users and nodes in the ordinary working course of a blockchain.

First of all, it's worth pointing out that a public key is used in a blockchain without openly stating who is the holder of the corresponding private key (unless the relevant holder decides to). Furthermore, a public key is not always associated with a natural person's address. In fact, it may be used by a legal entity or, for example, in the handling and transport of goods by earmarking them with a tag containing a private key and managing the corresponding public key in the blockchain to track the goods along a supply chain[11]. Therefore, the equation *public*

---

[7] For example, the bitcoin blockchain uses the function SHA256 which generates nothing more than an alphanumeric string of 64 characters (32 bits), for a total of different combinations equal to about 2 to the power of 256, a number that competes with all atoms in the observable Universe.

[8] Cryptocurrencies exist, such as Dash, Z-Cash or Monero, which adopt very sophisticated solutions to ensure the non-traceability of transactions. For an exhaustive examination of the techniques and measures available today to protect privacy in blockchain (that is to hinder, through cryptographic and logical solutions, the traceability of individuals who have performed the transactions) see what the founder of Ethereum wrote a couple of years ago: V. BUTERIN, *Privacy on the Blockchain* (2016), at https://blog.ethereum.org/2016/01/15/privacy-on-the-blockchain/.

[9] R.R. Kumar, *Impact of Blockchain Technology on Data Protection and Privacy*, 2017, available at SSRN; P. DE FILIPPI, *The interplay between decentralization and privacy: the case of blockchain technologies*, 2016, in *Journal of Peer Production*, Issue n.7: Alternative Internets, available at SSRN. In Italy, see R. BOCCHINI, *Lo sviluppo della moneta virtuale: primi tentativi di inquadramento e disciplina tra prospettive economiche e giuridiche*, Riv. Dir. Ind., 2017, I, 48. These Authors do not actually qualify bitcoins as pseudonymous data with specific reference to data protection legislation. In fact, they only point out that such data can, through appropriate associations with other information, reveal the identity of an individual who was part of a specific transaction. This is true of course, but it is also applicable to any kind of information and therefore irrelevant to discriminate between personal and anonymous data.

[10] See M. FINCK, *Blockchain and Data Protection in European Union*, 2017, 10, Max Planck Institute for Innovation & Competition Research Paper No. 18-01 (November 30, 2017), also available at SSRN.

[11] The use of *secured element* and *HCE – Host Card Emulation* technology in combination with the blockchain in the next future promises to revolutionize not only the production and distribution chain, but any aspect of our life (for an interesting application developed by IBM and Walmart see *Genius of Things: Blockchain and Food Safety with IBM*

key=pseudonymous data reveals its limits since a public key in a blockchain may represent a legal entity or goods (even immaterial-intangible goods or services).

But even when the key pair is used by a natural person, such an equivalence shows its limits and proves wrong because it is not used for identification purposes as in certified e-mail and digital signature software. In these cases, the public key is used to identify the private key holder to prove the sender's identity and authorship of a document in an act of digital communication. On the contrary, in a blockchain, everybody may perform a transaction without assigning an identity in the network to the creditor or the debtor, and it is not possible to track down the identity of the parties to a transaction without using advanced means of digital forensic and big data analysis, making questionable assumptions (e.g. correspondence between IP address and user), having access to confidential information which may be disclosed only with an order issued by the Authority (e.g. the traffic records preserved by the ISP), or using ordinary means of investigation (e.g. track parcels to their destination). These cases are exceptional and not always effective, and outside of them the public key does not, by itself, fall within the definition of personal data.

At this point, it is useful to read the clarifications provided by Opinion n. 4 of 20 June 2007 (WP136) by Article 29 Working Party (set out in Art. 29 Directive 95/46/CE), which is now more relevant than ever, where detailed criteria are settled to verify whether or not data is to be deemed *personal* within the meaning of EU legislation.

Starting off with the definition of "personal data" provided by Directive 95/46/EC («*any information relating to an identified or identifiable natural person*»[12]), and after providing clarification on the notion of «*information*», according to which everything is information in the essence, the Working Party focuses on the relation between information and natural persons. In other words, the focus is on the meaning of the expression «*relating to*», and proposes a test divided into three steps: *content*, *purpose* and *result*.

As for *content*, information is related to a natural person if the data being processed is immediately perceived as referring to an individual (e.g. a photograph of the person, or his/her medical record). As for *purpose*, information is related to a natural person if the data, even though it may not allow for the immediate identification of an individual, is processed to arrive at such an identification (e.g. data collected by a video surveillance system to allow for the identification of an offender when a crime is committed). As for *result*, information may be considered as related to a natural person if, even though it is not processed for identification purposes, the outcome of the processing is some sort of identification (such as the geolocation of a taxi car fleet to streamline calls resulting in a monitoring of taxi drivers' movements).

The test on the relation existing between data and individual leads to a very broad definition of personal data also because the three elements (*content*, *purpose* and *result*) need not be matched cumulatively, but rather merely alternatively. However, it produces a negative result when applied to non-identifying public keys in the way they are generally used in a blockchain.

As for the *content*, it is self-evident that a public key may not be directly perceived as personal data *per se*; as for the *purpose*, it is equally evident that a public key is not used in a blockchain

and Walmart, February 2017, https://www.youtube.com/watch?v=MMOF0G_2H0A). Other interesting projects for the application of the blockchain technology in this field are *Provenance* (on traceability of consumer products), *Skuchain* (on non-food industry chain) and Blockverify (on luxury products).

[12] The definition comes from Art. 2 of the Strasbourg Convention of 1981 *for the protection of individuals with regard to automatic processing of personal data* that already defined personal data as «*any information relating to an identified or identifiable individual*». The same is in the Directive (Art. 2) and in the Regulation (Art. 4) with the sole replacement of «*individual*» with «*natural person*».

for identification purposes, i.e. to identify the holder of the private key, nor is it ever used to achieve such as purpose. It is instead used to solve a technical problem: the *double spending problem*[13]. Finally, as regards the *result*, even though the public key could theoretically be used to trace back to the identity of an individual, this would be done using the public key in retrospect as an identifier in a transaction concluded by a specific individual. However, this is not as to say that the public key amounts to personal data[14].

To grasp this latter facet concerning public keys (they are not personal data, but at the very most, they could be deemed identifiers. i.e. pieces of information that may be linked somehow to an individual) it is worthwhile to recall the Cooper-Alba case. On July 2013, the famous movie star Bradley Cooper left a hotel in New York and was photographed while taking a taxi whose plate was photographed as well. In March of the same year, a US citizen, invoking the NY State Freedom of Information Law, asked the *Taxi and Limousine Commission* of NYC to be provided with their database containing all the taxi rides (about 173 million) with the route of the taxi, the fare paid and the tips given to the driver. Soon enough, the database was uploaded on the Internet and it didn't take long before the actor's fans cross-referenced the data from the picture (the plate number) with those from the data base finding out not only where Cooper was headed but also how much he tipped the taxi driver. The same investigation was subsequently carried out on many VIPs (often photographed in a public area just when taking a taxi), including Jessica Alba, for purposes of seeking out old photographs published in tabloids around the world and discovering their destinations and habits.

In the Cooper-Alba case, the taxi plate is not personal data *per se*. It does not identify Mr. Cooper, it is the picture that does. Anyone may take a taxi to go from one place to another and, therefore, the univocal link between code and individual, which would make the code a (pseudonymised) personal data, is missing. Likewise, a public key is not personal data either. Like the license plate, it does not identify an individual *per se*, but rather amounts to merely a technical means to make or receive a payment through a digital network.

Moreover, to underline how ephemeral the link between public key and personal identity actually is, consider for a moment the fact that an individual can generate and use a different pair of cryptographic keys for each transaction[15]. We can also conclude transactions by handing around the private key as a means of payment settlement. A transaction, in fact, can also take place off-line with the delivery of a token in which the private keys that give access to the addresses in blockchain are loaded. There are already "coins" in circulation that work in this way[16], and even if they may not turn out to be hugely successful, they demonstrate that a public

[13] Without using asymmetric cryptography, it would be impossible to prevent a user from spending the same virtual currency to make multiple payments, thus reducing confidence in the whole system. Likewise, all blockchain projects use public keys to build trust among users in a network without creating hierarchies and placing trust in the tip of the pyramid (this is why we say DLT are *trustless*).

[14] Even the inventor of Bitcoin said that the network guarantees privacy since the public keys are used in an anonymous fashion, separating the identities of the users from the transactions they carry out, as currently happens on the stock exchange where the volume of the exchanges is known, but not the identity of who buys and sell the shares (S. NAKAMOTO, cit., § 10, 6). This also seems to be the ECB's opinion: «*VC payment transactions do not require the provision of personal or sensitive data, unlike credit card data or passwords in the case of conventional payment methods. In this sense, VC units can be considered to be like cash: whoever possesses them also owns them, removing a source of potential identity*», *EBA Opinion on 'virtual currencies'* of July 4, 2014 (EBA/Op/2014/08).

[15] This can already be done with bitcoins by generating new key pairs for each transaction (see S. NAKAMOTO, cit., 6) or by merging multiple user transactions into one, as does the CoinJoin system, preventing blockchain analytics companies to trace back to wallet holders.

[16] The best-known example of physical bitcoins is Casascius (https://www.casascius.com/). These are coin-shaped metal supports on which the public key is visible (to verify the credit), but the private key is hidden by a holographic

key, at least in principle, is not any more *personal* than the serial number printed on the banknotes we carry in our pockets[17].

In light of the above, the *result* test on the expression «*relating to*» must be carried out restrictively, in the sense that data must not be considered personal merely due to the fact that it may be useful to trace back to the identity of an individual, and to conclude otherwise would lead to the paradoxical result that any data whatsoever, even weather data, would be personal data[18].

Moving onward, let's now take a look at the last part of the definition, «*identified or identifiable*». The law maker's intention is clear: personal data is not only data that directly identifies and individual (like a picture, a name or a surname), but it is also data that can be related, directly or indirectly, to personal data contained in a correspondence list. This means that the logical connection between data and individuals can also be potential, or even arise after the first processing of the data itself. The data is not necessarily personal at the time of collection, but later becomes personal on account of technological advancement or changes in the factual or legal conditions of the person who is processing it. Likewise, a public key, which is not in itself an identifying information concerning an individual (just as the taxi plate is not identifying information), does not start out as personal data. It becomes personal data the moment it is eventually used in connection with other pieces of information (e.g..: telephone number, IP address or the photo of the taxi in the example above) which are either identifiers, or, in turn, linked to identifiers [19].

Moreover, the meaning of «*identifiable*» must be construed strictly. Not everything that is theoretically possible should be taken into account. If so, whatever is not personal data under a narrow and rightful interpretation of «*relating to*», could be potentially construed as personal data under an interpretation of «*indentifiable*» that is too broad, and once again we would run the risk of considering any piece of information that is remotely useful for purposes of spotting an individual, as pseudonymous data subject to the GDPR[20].

---

sticker which is destroyed if removed. The integrity of the sticker is therefore a guarantee of the face value of the coin.

[17] According to R. BOCCHINI, cit., 51, only fiat currency is really anonymous because possession is ownership and no registration of the transaction is required to transfer the title in ownership. However, also for cryptocurrency, possession is ownership but the subject of possession is not a banknote but a private key. Besides, to demonstrate that fiat currency is not substantially different from cryptocurrencies, we may consider that in investigations concerning money laundering, drug trafficking, corruption and bribery, serial numbers of the banknotes are often recorded so as to prove the criminal conduct. In such cases, the banknotes handover involves processing of receiving party's (*accipiens*) personal data, exactly as would happen with cryptocurrencies. Moreover, in the case of common banknotes a certain *identification moment* occurs, namely that of the delivery in which the natural person materially obtains possession of the banknotes. On the contrary, with cryptocurrencies, this is not true even if the *accipiens* only uses an anonymous addresses to which making the credit and immediately afterwards "cleans up" the money by making numerous payments to such addresses or to third parties who receive the payments in good faith thus losing traces of the illicit origin of the proceeds.

[18] Vedi N. PURTOVA, The law of everything. Broad concept of personal data and future of EU data protection law, in 10(1) Law, innovation, and technology, also available on SSRN, who (citing the works of P. OHM, Broken Promises of Privacy, 2010, 57 UCLA Law Rev. 1742 ss.; L. SWEENEY, Simple demographics often identify people uniquely, 2000, available at https://dataprivacylab.org/projects/identifiability/paper1.pdf; P. SCHWARTZ - D. SOLOVE, The PII Problem: Privacy and a New Concept of Personally Identifiable Information, in New York Univerisity Law Quarterly Rev. 2011, 86, 1876), believes that the rate of technological development and the ever-increasing amount of data available for Big Data analysis make anonymity unrealizable (for the weather data example, ivi, § 3.5, 16).

[19] Ivi, 5, «*The same piece of data can be anonymous at the time of collection, but turn into personal later, just sitting there, simply by virtue of technological progress*».

[20] The risk is outlined by O. TENE & J. POLONETSKY, *Big Data for All: Privacy and User Control in the Age of Analytics*, in *Northwestern Journal of Technology and Intellectual Property*, 2013, 11, 258: «*with a vastly expanded definition of*

In a less absolutist approach, on the contrary, the identifiability of an individual must be interpreted in relation to the actual *de facto* situation, i.e. taking into consideration the different degrees of technical, legal or factual possibility of having access to information that allows for such identifiability[21].

Now coming to the public key, it is worth acknowledging from the outset that even though the private key holder may sometimes be tracked down, this would only be possible under extraordinary circumstances, which would require the use of uncommon means and resources, sometimes even unlawful ones. In fact, no "*public key-private key holder*" correspondence list exists, nor could such correspondence be easily obtained in the normal conditions under which the public key is used in a blockchain. The public key, in short, is nothing but a piece of information indicating a certain credit availability (or other asset or right) at a certain time. But it doesn't tell us who the individual is who can spend it. Like the taxi license plate in the Cooper-Alba example, it is just a piece of information which is not personal data on its own. It may become personal data if used in connection with other information from different sources (the photograph and the NY taxi company's database) thus allowing for the tracking down and collection of other information about a specific individual, such as his/her destination on a specific day and his/her propensity to pay tips.

In the wake of the ECJ decision *Patrick Breyer v Bundesrepublik Deutschland*[22], we could be tempted to consider the public key as a dynamic IP address[23]. In both cases, indeed, we deal with a string associated with a single event (the transaction in the blockchain, the PC's login session in the *Breyer* case). However, if we take a closer look, a blatant difference exists. Firstly, in *Breyer* case, they collected and stored the dynamic IP addresses solely to obtain the IDs of the persons who visited the website. This is sufficient to flunk the WP136 test and fall under the definition of *personal data* («*relating to*» read from the perspective of the *purpose* of the processing). The public key, on the contrary, is not processed into the blockchain for such a purpose, but rather to achieve a technical result and, in other words, that of concluding a transaction by resolving, as already mentioned, the *double spending problem*. Yet another difference is discernible. The dynamic IP address used by a device in a specific session is associated with the identity of the user who entered into the contract with the ISP. The latter, in fact, has a list of correspondence that allows for tracing back to the user identity in the long term. The dynamic IP address, therefore, falls under the definition of *personal data* provided by WP136 as regards the identifiability of the individual (§ 3 del WP136: «*identified or identifiable*»). However, this does not happen with public keys. A correspondence list which entangles the keys with IDs does not exist. Even though an occasional correspondence between a key and a personal identity may occur, as obviously would happen in payment situations where the debtor and the creditor know each other, it would be a contingent correspondence related only to a given transaction in progress which could not be extended to other transactions.

PII, the privacy framework would become all but unworkable […] *anonymized information always carries some risk of re-identification, many of the most pressing privacy risks exist only if there is reasonable likelihood of re-identification* […] *many beneficial uses of data would be severely curtailed if information, ostensibly not about individuals, comes under full remit of privacy laws based on a remote possibility of being linked to an individual at some point in time through some conceivable method, no matter how unlikely to be used*».

[21] See P.M. SCHWARTZ - D. SOLOVE, cit., 1876: «*Different levels of effort will be required to identify information, and varying risks are associated with the possible identification of data. To place all such data into the same conceptual category as data that currently relate to an identified person is a blunt approach*».

[22] ECJ Decision (II Section) of October 19, 2016, Patrick Breyer/Bundesrepublik Deutschland (Case C-582/14). According to the Court, accepting the interpretation of "*identifiable*" in WP136, example n. 15, the dynamic IP address is personal data.

[23] This is the opinion of M. FINCK, cit., 13.

Besides, we are dealing with special cases that do not fall within the interpretative confines of *identifiability*.

In conclusion, the notion of «*identifiable*», as used in the definition of *personal data*, must be interpreted in relation to the means that can be «*reasonably used*»[24]. It is not, therefore, a general and abstract notion, but rather a concrete one, that must be read taking into consideration the specific objective and subjective circumstances of the data processing and of the data processor: «*To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments*» (recital 26 of the GDPR).

4. The hash function.

As seen above (par. 2), the hash function is extensively used in the blockchain protocol to compress a set of information into a univocal (quasi-univocal) alphanumeric string which, in turn, refers to transactions (i.e. creation, transmission, modification, or extinction of rights), assets (tangible or intangible) or persons (natural or legal). For example, in an IPRs management service, a creative work cannot be the subject of a transaction if loaded onto the network as it is. The size of the file would hamper the transmission throughout the nodes, thus preventing the formation of a consensus among them. In essence, the nodes could not let the transaction enter a blockchain because it would be too voluminous or because it would render the block not remunerative enough[25]. In fact, although the blockchain is often defined as a decentralized database, it is more similar to a ledger or, better yet, a decentralized register, designed not to store information, but to take note of transactions in relatively small-sized blocks (1 MB in the bitcoin protocol). For this reason, the transactions do not deal with voluminous files, but rather with a hash that, with a handful of letters and numbers[26], can identify the file (and the work along with it) in a univocal and objective way, providing irrefutable proof of its existence and proof that the transfer of the asset took place.

According to WP136, one-way functions, such as the hash function, generate outputs that, due to their unidirectionality (namely, the fact that there is no way to get the original information set knowing only the hash), cannot be considered as personal data but essentially anonymous data (WP136, pp. 18 and 20)[27].

However, the statement of the Working Party is not absolute («*one-way cryptography [...] creates in general anonymised data*», emphasis added). The Party clearly refers to the key-coded data, namely the personal data anonymised with an univocal key whose processing is tantamount to the processing of the personal data itself (WP136, page 18). Whether the key has been obtained applying a one-way function to the data or by randomly choosing a code, is of little importance. The intent of the law is clear: concealing personal data with a code does not allow for circumventing the law whenever the individual to whom the code refers is still in some way identifiable (that is, there is a list of correspondence and it is reasonably accessible).

[24] So reads recital no. 26 of Directive 95/46/CE and, likewise, the same recital of the GDPR.
[25] For each transaction contained in a block, the node that adds it to the blockchain is remunerated with transaction fees. The more transactions the block contains, the more transaction fees the node expects to gain. (v. A.M. ANTONOPOULOS, cit., § 4).
[26] Bitcoin uses SHA256 function whose hash is 256 bits long, equal to 64 bytes (characters) in hexadecimal format.
[27] Even though M. Finck, cit. 11, makes reference to the same paragraph of WP136, she mistakenly attributes a different conclusion to the Working Party, namely that the hash is *pseudonym data*, and since this Author believes that pseudonym data should invariably be considered as personal data, she concludes that a hash is personal data.

Once again, we have to focus on the meaning of identifiability. As previously noted, the Working Party accepts a dynamic notion of identifiability, in the sense of considering it not from an objective standpoint, like an abstract possibility to single out an individual, but from a subjective one, meaning the actual possibility for the data processor to trace back to the identity of an individual. Hence, key-coded data cannot always be considered personal data and it most certainly would not be when it constitutes a real impediment to the reidentification of the data subject, due to the fact that the significant costs and means necessary would not be reasonably likely to be dedicated for such purpose.

Therefore, if we adopt, as we ultimately must, a relative and subjective notion of *identifiability*, indirectly by correlation, we have to apply the same to the notion of *personal data* that relies upon the meaning of identifiability[28]. In such case, key-coded data is personal data only for the owner of the list of correspondence that links the codes and the data subject identities and for those who can reasonably gain possession of it. But key-coded data is not personal data for those who, even though the same data is being processed, do not have the list of correspondence and are not allowed to have access to it (and they are prevented from gaining such access by appropriate means) [29].

At this point, you cannot tell *a priori* whether or not a digest amounts to personal data. This would depend on the actual circumstances in which the data controller operates. We have to consider the right, the means and the real possibilities the controller has to back-trace the list of correspondence and in this manner identify the data subjects. But when we use hash codes and public keys in a blockchain, such a list doesn't exist. Nor is this necessary, nor are hash codes somehow used to trace back to an individual's identity, nor are they effectively used to disguise a data subject's identity. In the bitcoin protocol, for example, the hash function (SHA256) yields the public keys' digests, the digest of each transaction contained in a block and a digest of all the transactions' digests of the block (*Merkle tree*). The latter will then be used to create the header of the block itself that, in turn, will be hashed as a whole to arrive at a single digest that will be part of the header of the subsequent block.

In addition to blockchains conceived to handle cryptocurrencies, other blockchains might be designed to manage other kinds of data, such as IP management[30] as already mentioned, or a network to provide e-government services through digital identity[31], or data sharing and life cycle in IoT's devices. In such cases, the hash function might be used to directly carry out

[28] L. PUTROVA, *The Law* cit., 7.

[29] On this see WP136, example n. 17, and, in particular, where the Work Group, with reference to the coded data processed by third parties, holds: «*In this case, one can conclude that such key-coded data constitutes information relating to identifiable natural persons for all parties that might be involved in the possible identification and should be subject to the rules of data protection legislation. This does not mean, though, that any other data controller processing the same set of coded data would be processing personal data, if within the specific scheme in which those other controllers are operating reidentification is explicitly excluded and appropriate technical measures have been taken in this respect*» (emphasis added). As for the aforementioned technical measures, there are now very sophisticated cryptographic solutions and a smart contract could process personal data and return the desired output without providing access to such data or to the way it has been processed (see *zero knowledge proof* and *black box* solutions, reported by V. BUTERIN, cit.).

[30] On copyright management, two projects are noteworthy: Mycelia and Ujo Music whose progress is actually unclear. But the main sectors in which blockchain technology is being tested are innumerable and range from retail (OpenBazar and OB1), to insurances (Aeternity), to data storage (Storj) to healthcare (Gem and Tieron) and real estate (Ubiquity).

[31] The most advanced case in the world of application of blockchain technology for services to citizens has been implemented in Republic of Estonia that with its *ID-kaarts* is carrying out a national project called *Zero-Bureaucracy* with excellent results (https://www.mkm.ee/en/zero-bureaucracy). This is all the more remarkable if one thinks that the country was part of the dissolved Soviet bloc until 1991. A democracy well known as being very centralistic.

personal data processing. For example, if we use the digest of the public key of a digital signature, this digest would in fact be personal data since there would be a list of correspondence accessible to everyone, enabling them to easily make a link between the public key and the identity of the private key holder. Besides, if the hash is obtained by encrypting a tax code, even in this case the hash itself would be personal data since, although there is no public correspondence list between identity and tax code, the algorithm to obtain the latter from a personal data set (name, surname, date and place of birth) is well known, at least in the great majority of cases[32]. In both examples, however, for a data breach to occur, we would have to know the coding protocol applied to the hash function. In other words, we should know that the hash is obtained by coding public keys or fiscal codes. Hence, it would be relatively easy find out to whom the data corresponds by applying a trial and error process to a specific digest or obtaining the digest form an individual's public key or tax code and then searching for correspondences in the blockchain to access his/her personal data stored therein.

But if the coding protocol were known only to the original data controller and it were not so straightforward as in the example given above, then digests could be used to *indirectly* manage personal data in a blockchain without thereby qualifying it as personal data processing. In this case, indeed, only the data owner would have access to the information necessary for the re-identification of the digests (coding protocol or list of correspondence)[33].

In conclusion, what I pointed out above for the public key holds true for the digest as well: the digest *per se* does not constitute personal data, even under the excessively broad wording in WP136. However, there may be some circumstances in which public keys and digests, because of the way they are used or the risk of actual unauthorized access by third parties, could be considered as personal data or, more precisely, *pseudonymous data*[34].

5. Coinbase and the intentional seeding of personal data in blockchain.

We have seen how the cryptographic solutions that characterize a blockchain do not necessarily involve a data protection issue. However, it is possible to deliberately insert into a blockchain uncoded personal data so that they may be known to anyone in the world without any possibility of removal unless a massive effort in terms of consensus and energy is made.

In the bitcoin blockchain, for example, miners may use a special file in each block header called coinbase to insert messages of any kind. As new blocks "settle" on top of the blockchain, the message/information in the coinbase becomes practically impossible to delete[35]. Furthermore, other blockchains could be designed in such a manner that does not comply with the provisions

[32] In L. SWEENEY, cit., § 2.1, the author shows how he managed to re-identify thousands of individuals by crossing their "anonymous" health data provided by the National Association of Health Data Organizations (NAHDO) with the data from the voters' list of the town of Cambridge in Massachusetts. He was actually able to link sensitive data of patients identified only by postal code, date of birth and sex with their name, address and political affiliation.

[33] H. Chang, *Is Distributed Ledger Technology Built for Personal Data?*, in *Journal of Data Protection & Privacy*, 2018, Vol. 1, n. 4, pp. 5-6; *University of Hong Kong Faculty of Law Research Paper No. 2018/016*, available at SSRN. The Author doubts that adding a pinch of "salt" to the coding protocol is enough to prevent personal data processing.

[34] It is worth pointing out that the GDPR does not provide a definition of pseudonym data, instead it defines the pseudonymisation technique, or the process by which personal data is "hidden": «*so that [it] can no longer be attributed to a data subject without using additional information*» (Article 4.5). Pursuant to the Regulation, therefore, the pseudonym data is born as personal data and is then disguised behind an alphanumeric mask. However, this does not happen either with the public key or with the hash, for they both come out as simple, not personal data.

[35] Accenture created an editable blockchain using a particular function called *chamaleon hash* that allows for the blocks to be modified without modifying their hash (*Accenture to unveil blockchain editing technique*, published on *Financial Times* 2016, available at https://www.ft.com/content/f5cd6754-7e83-11e6-8e50-8ec15fb462f4).

of the GDPR by using, for example, public keys from digital signatures to purposely identify individuals.

In these cases, it is not clear, according to the provisions of the Regulations (and indeed, as mentioned at the beginning of this work, according to any framework of laws), how the data subject could enforce his right to be forgotten through the cancellation of his data, or who should address this request. In fact, it would present an unsurpassable enforcement problem, or rather the type of impediment desired by the crypto-anarchist movement that has sprouted and developed the blockchain technology. From a GDPR perspective (and maybe under any legal perspective, as referred to in the beginning of this article) it is not clear how we should deal with these cases. How may the data subject exercise his/her right to be forgotten? Who would be the recipient of his/her request? An unsurpassable enforcement problem would occur; exactly the kind of impediment envisaged and strongly pursued by the crypto-anarchist lunatic fringe that created and developed blockchain technology.

6. Data controllers in a blockchain environment.

If the public key and the digest are not personal data, then the nodes are not data controllers and must not comply with GDPR provisions, as long as they only perform the *mining* and the network is designed in compliance with data protection laws and best practices[36]. Uploading a blockchain database in whole or in part in order to validate the blocks is not *per se* an action aimed at identifying the public key holders, nor is the identification of such holders a side effect of mining operations or traffic routing. In this context, according to the original crypto-anarchic spirit that fostered this technological revolution, anyone may participate in the construction, strengthening and spreading of the blockchain public network without having to abide by the GDPR and without adopting specific security measures.

The case of wallets and exchanges is of course different. They are ISPs that have intercepted the business opportunities created by a public blockchain and they offer additional services to the network ranging from key security management to cryptocurrencies conversion from legal tenders and among cryptocurrencies themselves, and, in the future, who knows what else. In doing so, they undoubtedly process their customers' personal data and, therefore, like any other service provider in the information society, they must be considered data controllers for all intents and purposes.

7. Conclusions.

In the near future, the digital ecosystem will be populated with peer-to-peer networks designed using blockchain protocol. Some of them will be *public networks*, developed with open source software, granting free access and equal roles to everyone, others will be *private networks*, developed with proprietary software and granting access to selected users only, while still others will be designed with mixed public-private architecture, with special nodes running a piece of the protocol and other users acting as blocks validators. In all cases, the on-line services will be reshaped to adjust to a new business model to exploit the horizontality of the decentralized network and to remunerate the nodes with tokens or newly-minted cryptocurrency.

Not all nodes will be equal in the context of privacy. Some will be dealing with hash digests and asymmetric keys for direct or indirect identification purposes, by having access, or by reasonably having access, to lists of correspondence containing the identities of the data subjects associated

[36] Of a different opinion is M. FINCK, cit., §IV (A), 16, who highlights the difficulties to enforce the GDPR because of the distributed nature of the network and the enormous number of nodes, their multi-territoriality and their continuous changeability.

with them. Others will be dealing with the same hash digests and asymmetric keys for the sole purpose of allowing the network to operate relying on the remuneration for mining, without having access to any list of correspondence, or being expected to have any. The data processed by them, therefore, will not be personal as construed by the Working Party and the ECJ, since it would not be data «*relating to a natural person*», since the data would lacking content, purpose and result, and since the data subjects would not be «*indentifiable*» , given that the reasonable possibility of having resources and information useful for arriving at such identification would be lacking[37].

On the other hand, a broader interpretation of personal data that does not take into account the concrete circumstances of the purpose of the processing and the likelihood of identifying the data subjects, would not only hinder the development of technology, which would be against the interest of citizens and the market, but would also clash with the factual impossibility of enforcing the law that would result in a generalized circumvention of law[38].

The GDPR and the blockchain technology are, therefore, not ontologically incompatible. Designing a blockchain protocol (public or private) in a manner ensuring that the cryptographic keys used would not amount to personal data is possible[39], and the blockchain, rather than constituting a risk for the fundamental rights and freedoms of individuals in terms of privacy, would be in fact a tool that permanently puts into the hands of the data subjects the exclusive possession of and control over their personal data[40].

[37] As R. BOCCHINI (cit., 49) correctly points out, the node that only validates blocks behaves as a *mere conduit ex* art. 14 of the Directive 2000/31/EC. Hence, not only does a node not undertake the data controller duties, but it does not even take the *active hosting* responsibility.

[38] On the "impossible law issue" the anecdote of LL. Fuller is relevant: *The Morality of Law*, in *Yale University Press 1969*, 36-37, cited by N. PURTOVA, cit., 2.

[39] M. FINCK, cit., 27, thinks otherwise. She believes the nodes should interrupt the blocks validation or at least should implement the DPIA pursuant to Art. 39 of the GDPR: «*For the time being, the safest advice for blockchain developers is that transactional data should never be stored on a blockchain. Regarding public keys, the necessary risk-management solutions must be adopted and detailed Data Protection Impact Assessments must be carried out*». This position is not only impractical, but goes against the very functioning of the blockchain network where all blocks can be freely downloaded and consulted by anyone.

[40] G. ZYSKIND - O. NATHAN - A. PENTLAND, *Decentralizing Privacy: Using Blockchain to Protect Personal Data*, report by the *Security and Privacy Workshops IEEE*, 2015 (in https://enigma.co/ZNP15.pdf); the Authors show the architecture of a decentralized system for personal data management.