Mobile Networks and Applications (2021) 26:266–276
https://doi.org/10.1007/s11036-020-01646-9

An Analysis of Blockchain and GDPR under the Data Lifecycle Perspective

Gislaine Parra Freund[1] · Priscila Basto Fagundes[1] · Douglas Dyllon Jeronimo de Macedo[1]

Published online: 29 August 2020

Abstract
The purpose of this article is to present an analysis of the privacy principles of personal data prescribed in the General Data Protection Regulation and the treatment of data in Blockchain technology in its primary version, guided by the phases of the Data LifeCycle. The Data Life Cycle presents the stages in which the data act in a given process and are related to each other, forming a chain of dependence between them. The General Data Protection Regulation, on the other hand, presents privacy principles that contemplate the necessary treatment for data at all stages of its life cycle. This analysis made it possible to identify the influence that the phases of the Data Lifecycle have on the adequacy of the principles of the General Data Protection Regulation and the treatment of the data performed on the Blockchain technology associated with the phases, presenting an approach to lead the process of adapting the technology to compliance with the legislation. In this context, some data treatment options were presented for the phases that presented gaps, making it possible to conclude that the result of this analysis can be used as a support tool to systematize the process of adapting to the law by organizations that use or intend to adopt Blockchain technology.

Keywords Data lifecycle · General data protection regulation · Data privacy · Blockchain

1 Introduction

Technological advancement has enabled greater access to data and with this new ethical issues arise to be addressed within the scope of legislation on this subject. This scenario boosted the approval in 2016 of the European Union General Data Protection Regulation (GDPR), a law that regulates the processing of personal data in European Union countries.

GDPR provides for the processing of personal data, including in digital media, by a natural person or a legal person under public or private law, with the objective of protecting their fundamental rights of freedom and privacy. According to the law, activities for the processing of personal data must observe the following principles: lawfulness, loyalty and transparency; limitation of purposes; minimization of data; accuracy; limitation of retention; integrity and confidentiality; and responsibility [1].

Given the need for new technologies to meet scenarios increasingly complex arises Blockchain. [2] presents Blockchain as being a new architecture for the digital context and points out that the technological revolution demands an architectural change in the hierarchies of networks, since the hyperconnected world has a dark side of access that needs to be considered.

Prior to the Internet, hierarchical architecture was considered secure, since it was more difficult for criminals to access proprietary mainframe systems. However, this scenario has changed in recent years, making traditionally built systems vulnerable to security breaches.

The blockchain concept was introduced by Satoshi Nakamoto in 2008 in a proposal to use bitcoin electronic money. In this context of using Blockchain, the trust of the transaction that was previously concentrated in a financial institution, is now guaranteed by the network components, by the consensus process.

As Blockchain does not use centralized files, if intruders enter the system, the distributed data is unintelligible and they are prevented from causing damage, as it is not possible to take any action on the system without a consensus from the Blockchain network [2].

Although there are discussions in the literature about Blockchain compliance with GDPR and the proposal of models that guide the implementation of technology to comply with the law, research that takes into account the Data Life Cycle (DLC) within this context is incipient.

✉ Gislaine Parra Freund
gislaineparraf@gmail.com

[1] Department of Information Science (CIN), Federal University of Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brazil

According to [3], the DLC comprises the steps in which the data act in a given process and these steps correspond to the phases in which the data are related, forming a chain of dependence between them. There are different proposals to characterize the Data Life Cycles, as well as different stages associated with them. There are countless studies that address problems with DLC, such as [4–8]. From the perspective of Information Science, [9] proposes that the DLC be divided into four phases, collection, storage, recovery and disposal.

Based on the assumption that there are organizations that use the life cycle approach to guide the development of systems, the flow of information, among others, an analysis of the principles proposed by GDPR was carried out and identified in which phases of the Data Life Cycle (DLC) proposed by [9] these principles need to be observed. As well as how Blockchain technology handles data in each of the DLC phases.

This article aims to present the results of this analysis and to encourage new studies involving the GDPR, the Data Life Cycle and the Blockchain, including research involving the proposal of models to be used in the process of adaptation to the GDPR of the organizations that use the Blockchain considering the DLC.

It is noteworthy that this article is an extension of the study presented in [10], in which the analysis was limited to the principles of the Brazilian General Data Protection Law - LGPD and the phases of the Data Life Cycle.

As contributions of this article, the following stand out:

- provide support to organizations assisting them in complying with legislation when the technology adopted is Blockchain;
- promote discussions about other emerging technologies in compliance with regulations aimed at the privacy of personal data from the perspective of the Data Life Cycle;
- provide theoretical framework for future research involving the topics Blockchain, Data Life Cycle and GDPR.

This article is structured in order to present the principles of GDPR, the concepts and definitions about the Data Life Cycle and Blockchain technology, the works related to this study, the methodology used for the analyzes, the results obtained and, for Finally, the final considerations.

2 Background

2.1 European Union general data protection regulation (GDPR)

Globalization and the development of new technologies have demanded greater attention from organizations regarding the security of corporate information, as well as from their customers. Increasingly, businesses and the government are vulnerable to espionage or malicious attacks that culminate in information leakage or misuse, and after some scandals related to data misuse in Europe and the United States, the governments of several countries were obliged to regulate, among other aspects, the access, storage, use and dissemination of their citizens' personal data.

The first discussions about the need for laws to regulate the flow of data and guarantee rights over the use of this data started in Europe in the 1970s and in 1995, the European Union approved the European Data Protection Directive - Directive 95/46 CE of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [11].

As it was conceived in a time before the emergence of the commercial internet and before the diffusion of business models and technologies that make use of the intense use of personal data, this directive had to undergo an update process that culminated in the current European Union General Data Protection Regulation (GDPR).

The GDPR was approved on April 27, 2016 and aims to address the protection of individuals with regard to the processing of personal data and the free movement of such data. The regulation came into force on May 25, 2018, updating and adapting the old European Data Protection Directive to the newest forms of massive use of personal data, such as business models based on big data technologies, artificial intelligence, machine e-learning and Blockchain [12, 13].

According to article 4 of the GDPR, personal data is characterized by being a set of information related to a living person that can lead to their identification and that must be protected regardless of how they are stored or treated, whether in a technological or manual, provided they are organized according to pre-defined criteria [1].

The GDPR in its Chapter 2 Art. 5, determines a set of seven principles that should guide the activities of processing personal data. GPDR, by including such principles, guarantees data subjects the right to request information from public and private bodies on how their data is used, and these have a fixed period of time to meet the data subject's request. The processing activities for personal data must observe the following principles [1].

I. lawfulness, loyalty and transparency - personal data must be subject to lawful, fair and transparent treatment in relation to the data subject, that is, the data must be processed fairly, honestly and be known to the data subject.
II. limitation of purposes - personal data must be collected for specific, explicit and legitimate purposes and these cannot be further processed in a way incompatible with those purposes, that is, the use of the data is restricted to the purpose for which it was made available.

III. data minimization - personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed, that is, limit the collection to the essential data for the offer of the service or product that is proposed.
IV. accuracy - personal data must be accurate and updated whenever necessary. Adequate measures must be taken to ensure that, considering the purposes for which personal data are processed, incorrect data is erased or rectified without delay, that is, have the means to request and execute actions to maintain the accuracy of the data.
V. limitation of retention - personal data should be retained only for the period necessary to exercise the purposes for which they are processed. They can be kept for longer periods, as long as they are treated exclusively for the purpose of archiving the public interest, for the purpose of scientific or historical research or for statistical purposes, that is, the data must be permanently deleted from all means after fulfilling its purpose and / or legal requirements.
VI. integrity and confidentiality - personal data must be treated by adopting the appropriate technical or organizational measures in a way that guarantees their security, including protection against unauthorized or unlawful treatment and against loss, destruction or accidental damage, that is, to implement resources that guarantee the confidentiality and integrity of the data during its life cycle.
VII. responsibility - the person responsible for the treatment is responsible for complying with the principles set out above must prove them, that is, they must implement resources that enable accountability and the possibility to demonstrate compliance with the other principles established in the regulation.

GDPR also presents a series of guidelines regarding the responsibilities for the international transfer of personal data, in defining the role of personal data processing agents (controller, operator and person in charge of personal data processing). Likewise, rules are defined regarding governance, good practices, security and data confidentiality. It is important to note that GDPR has extraterritorial applicability and establishes that the circulation of personal data with origin and destination in countries that are not members of the European Union must follow the same standards and security principles defined in European legislation, a fact that has caused other countries to adapt to these issues and also establish its internal guidelines [1].

It is also noteworthy that the law has other directives in its content, but for the analysis proposed in this article, the seven principles presented here were listed because it is understood that these guide the other items of the law.

2.2 Data lifecycle (DLC)

According to [3], there are several types of life cycles, as well as different stages associated with them. DLC can be designed from different perspectives and at different managerial levels. According to the author, at the level of managerial activity, for example, the life cycle is managed in the domain of the business process, while at the project or system level, the data life cycle is managed in the context of product or service development. This approach is used in several areas, in Computer Science in the context of software development, called Systems Development Lifecycle (SDLC), it is a framework that describes the activities in each stage of a software development project [14].

From the perspective of Information Science, [9] proposes that the Data Life Cycle be composed of four phases, they are:

- Collection: it aims to meet the information needs and it is there that activities related to the initial definition of the data to be used are developed, as well as the planning of how they will be obtained, filtered and organized, identification of the structure, format and means description that will be used
- Storage: where activities related to processing, transformation, insertion, modification, migration, transmission and any action aimed at data persistence in a digital medium are carried out. The focus of this step is to allow data reuse
- Recovery: which is related to the consultation and visualization of data. Its objective is to allow better access and use of them
- Disposal: where the identification of data that is no longer needed and that can be deleted occurs, with a focus on eliminating unnecessary data

The author suggests that the mentioned phases are permeated by six factors (Fig. 1), which:

– privacy that deals with aspects that guarantee the privacy of people or institutions related to the data to be used;
– integration that refers to the identification and use of requirements that will provide data integration with other data;
– data quality related to aspects such as origin, collection mechanisms, physical and logical integrity, among others, to be considered to ensure that the data is reliable and useful;
– copyright related to the respect for the data copyright;
– disclosure related to location and access to data;
– data preservation, related to the preservation of useful data, so that they can be used in the future.

Table 1 presents a summary of the relationships defined and presented in [9], and adapted in [10] involving the four

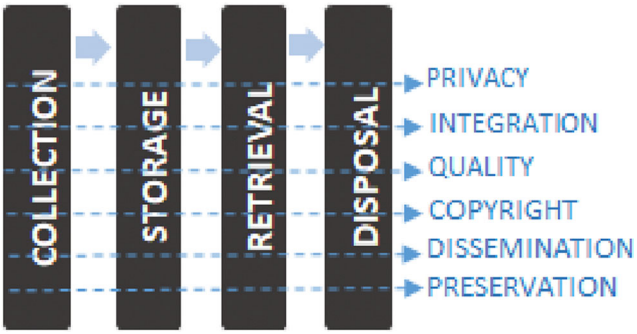

Fig. 1 Data Lifecycle by Information Science

phases of the data life cycle and the six factors that permeate them.

It is important to note that the summary presented was based on the structure defined by [9] and adapted by the authors, which contemplated the factors that permeate the phases of the DLC also in the Disposal phase as they understand that such factors must be considered throughout the cycle and that they can be applied to data wiping.

2.3 Blockchain

Blockchain is a technology that uses distributed storage and cryptographic techniques to ensure reliability in transactions. Proposed by Satoshi Nakamoto in 2008, the concept of Blockchain was first associated with the creation of the digital currency called Bitcoin. Such a proposal involved the use of a peer-to-peer network, based on cryptographic evidence to allow two parties to carry out transactions with each other without the need for a reliable third party, so that the trust in the transaction was provided by the network nodes themselves, without the need for intermediaries [15].

According to [16], Blockchain is a technology underlying Bitcoin. This was the first network that employed Blockchain technology, implemented in 2009 and since then many companies have been exploring the adaptability offered by the technology for business applications. For [17] this fact is justified by the advantages offered by the technology in relation to the existing applications until then, centrally controlled. In his view, the author states that many risks inherent to the current business model are mitigated by paying fees to trusted third parties such as banks, certification authorities, credit card companies, among others.

Since its inception, Blockchain has been constantly evolving and adapting to meet the needs that arise as it is applied in different scenarios. [18] exemplifies the evolution and adaptation of Blockchain technology, classifying it into five generations:

1. bitcoin, resulting from its application in digital currency;
2. Blockchain related to the perception that Blockchain technology could be broken off from bitcoin currency and used for other purposes;

Table 1 Phases x factors involved in the data life cycle process

	Collect	Storage	Recovery	Disposal
Privacy	Collect respecting privacy.	Store the data using access control.	Recover data taking users into account the content to be made available.	Discard the data respecting the right of privacy and request of the "owner" of the data.
Integration	Collect data that can be integrated with other databases.	Store the data considering the form of access and the adoption of a DBMS[a] that allows the integration with other data.	Recover data with the benefits of a good integration that provides greater value in its use.	Discard data observing the consequences of disposal in relation to content derived from integrations made with other data.
Quality	Collect data considering the origin and the collection mechanisms used.	Store data with due regard for its physical and logical integrity.	Recover data with the same quality aspects present in the collection and storage steps.	Discard data by recording information about the elimination process.
Copyright	Collect respecting the copyright.	Store the data by linking the source to obtain the data.	Retrieve the data explaining the usage permissions and how the data can be used.	Discard the data keeping information about its authorship to ensure legal compliance in derivative and/or referenced works.
Dissemination	Collect data to support data findability and access.	Store data providing accessibility means that can be interpreted and easily located.	Recover data considering elements and strategies that allow its location and access through collection processes.	Discard data by observing the impact of eliminating key elements for searches and finding data sets.
Preservation	Collect data so that it can be preserved	Store with the premise that data can be interpreted in the future.	Recover data with the possibility of obtaining the same interpretation at different times.	Discard data from systems considering keeping a copy of deleted data to preserve it if there are unforeseen demands that require it.

[a] Database Management System

3. Smart Contracts already started in the previous generation, with the creation of small programs on the Blockchain to enable their application to different transactions in addition to digital currency;
4. proof of stake regarding the consensus algorithms inserted in the Blockchain to provide greater security in transactions; and
5. from Blockchain scaling (scaled blockchain) Innovation in order to resize the number of computers that are needed to process the blocks and optimize the processing without loss of security and robustness of the technology.

Blockchain changes the concept of trust based on a classic centralized approach to a decentralized structure that provides trust between the parties called "nodes" initially "untrusted" with a consensus mechanism. These nodes maintain the recorded information which cannot be changed or deleted, thus maintaining the history of the transactions carried out. With this, this technology brings many advantages that include provenance, responsibility, traceability and transparency in transactions and provide audit trails that allow the reconstruction of a trajectory of actions within the chain [19].

According to [20] blockchain is a technology that provides auditing due to its nature. In a confirmed transaction, its modification will occur in an identifiable way, a new block will be created to register the desired change and the block which it is intended to modify previously registered, will remain unchanged in the chain.

Blockchain technology combines already consolidated techniques such as Cryptography, Hash Function and Distributed Databases with the aim of keeping data safe [16, 18].

For [2] the decentralized storage with encryption ensures that the data will be incomprehensible in the event of improper access to the network and that the consensus dynamics preclude any action that is not validated consensually by the network nodes. In addition, to guarantee immutability, the Blockchain is composed of a sequence of blocks containing all the operations carried out recorded in the logbook.

Figure 2 represents the sequence of blocks and the use of the hash technique to ensure the immutability of the records made on the Blockchain.

Each block has the data for registration and hashes this content. From the second block onwards, the hash of each block is generated from its data plus the hash of the previous block, thus ensuring the immutability of the data from the previous blocks.

In [21] the Blockchain has attracted the interest of several segments, such as: finance, health, public / government services, real estate, among others and that the reason for this explosion of interest is the possibility that technology offers to operate in a decentralized manner, without the intervention of third parties with the essential security features required in a transaction.

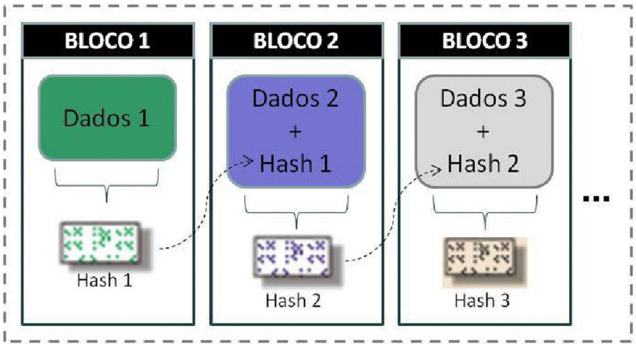

Fig. 2 Blockchain block sequence representation

For the authors [22], the immutability offered by Blockchain favors the use of technology in several fields, as it ensures that a transaction is not violated after being registered on the Blockchain, thus ensuring the reliability and honesty of the transaction. Another point raised by the authors is the fact that storage is distributed and prevents a single point of failure, a fact that also contributes to the reliability of the transaction.

3 Related works

Some studies considered relevant within the context of this work will be presented here. The selected works were the results of research carried out in the Web of Science and IEEE Xplore Digital Library databases.

In [19] presents the results of a systematic review of the state of the art on solutions and search engines for privacy preservation in Blockchain, as well as the main privacy challenges associated with this technology. The autors identified several open research challenges and issues related to privacy-preservation on blockchain, encompassing transaction linkability, crypto-keys management (e.g. recovery), issues with crypto-privacy resistance to quantum computing, on-chain data privacy, usability, interoperability, or compliance with privacy regulations, such as the GDPR. The research concludes that current Blockchain solutions are still far from dealing with privacy challenges holistically and this situation undermines user's rights, such as the right to become anonymous in certain situations, the right to erase data or withdraw consent, thus lessening the realization of a truly privacy-preserving and Self-Sovereign Identity model on blockchain.

In the paper [23] proposes a new Blockchain model with the aim of ensuring the compliance of this technology with the GDPR by handling references to confidential data and using metadata instead of manipulating private data directly within the Blockchain. For that, it was defining a modular architecture that is based on strong cryptographic assumptions that provide the means to guarantee that the right to be forgotten is being well applied. The main contribution of the model

proposal is to allow data providers and consumers to interact with each other using a Blockchain to track all interactions, while at the same time meeting GDPR through smart contracts, but without storing confidential data within the Blockchain. In addition, each entity involved can assume different functions at the same time and, due to the nature of smart contracts, it is possible to obtain greater flexibility without compromising confidential information.

In [24], a decision model for the use of Blockchain technology and its application in two scenarios is presented: 1) processing of a patient's personal health data during the medical treatment process and 2) management of digital identities on the Blockchain. In addition, the article presents an analysis of the conflict between the functioning of Blockchain technology and GDPR. For the analysis, the authors use the three categories of Blockchain (public, private and consortium) and compare with the requirements of the GDPR legislation that deal with the processing of personal data (legality, equity, transparency, Purpose limitation, data minimization, accuracy, storage limitation, integrity, confidentiality, accountability). The results of the analyzes are presented considering two perspectives: existing general conflicts and conflicts resulting from the analysis applied in the two scenarios covered in the article. The authors found conflicts with the principles of precision and storage limitation, in addition to the right to forget and delete the data recommended by GDPR, which were affected by the immutability characteristic offered by Blockchain and concluded that Blockchain in its basic design is not compatible with GDPR and suggest that further studies be carried out to propose structures that establish compliance between them.

In [25] presents the implementation and application of Blockchain technology in a solution used by the Federal Office for Migration and Refugees of Germany, whose procedure Asylum involves several authorities at the municipal, state and federal levels and issues decisions on the application for immigrants. The authors highlight, three potential approaches to ensure that the Blockchain is in compliance with the GDPR: Central Authority, Shared Responsibility and Pseudoanimization and propose a 3-tier architecture with the combination of organizational and technical measures, they are: Blockchain, Integration (services and Dashboard of services) and Back-end. The Blochchain layer is responsible for propagating pseudonym event logs that will provide traceability in log tracks without exposing the personal data itself. The integration layer (privacy services) maps the pseudonym IDs to the specific IDs that will be used by the backend and provide access support based on roles and procedures. In the integration layer (Dashboard), they convert the event log data according to the Blockchain data models. And display the data to the user in a Web browser (according to the user's access right) in the Back-end layer. The authors recommend 3 factors to be interpreted as guidelines for the compliance of Blockchain to GDPR considering the applied context: 1) avoid the storage of personal data on a Blockchain - they must remain stored in systems that allow them to be rectified and deleted; 2) use of pseudo-anonymization in blockchains that process personal data; 3) use of private blockchain with permission when applied in scenarios involving interorganizational workflows. In the authors' view, for the evolution of the theme, the next step is the development of reference standards and architectures that guarantee the interoperability of various Blockchain technologies and solutions.

It is possible to find other studies in the literature in order to present a reflection on the compliance of Blockchain with the principles of GDPR, as in [26] and [27], which demonstrates that the scientific community has raised concerns regarding the service of GDPR in the use of Blockchain technology, however, it is possible to observe new research opportunities in both areas as well as gaps to be filled.

The paper of [10] precedes the study presented here, but the focus differs on the object of analysis. In that, the authors identify the degree of relationship between the principles of the Brazilian General Data Protection Law - GPDL and the stages of the DLC and conclude that the data life cycle model can be used to support and systematize the law's compliance activities, since the principles of the law are related to the stages of the model. It was also observed that the principles of adequacy of data processing for purposes of use and transparency are those that guide the other principles and that, in addition to these, the principles of safety, prevention, responsibility and accountability recommended by GPDL were related to all phases of the data life cycle.

With respect to studies involving the Data Life Cycle associated with GDPR or Blockchain technology, the research carried out in the Web of Science and IEEE Xplore Digital Library databases did not return results that could be considered as works related to this one being presented in this article, thus evidencing the originality of this study.

4 Methodology

The present study is considered a basic research as to nature, as it aims to generate new knowledge, useful for the advancement of science, with no expected practical application regarding the procedures, as bibliographic since it is based on the survey of theoretical references already analyzed, and published by written and electronic means, such as books, scientific articles, web site pages. As for the objectives, it is exploratory in nature, as it aims to obtain a better understanding of the problem to be studied and to promote greater familiarity with the themes, to make them more explicit or to build hypotheses. It has a qualitative approach, since it is concerned with deepening the understanding of a social group, an

organization, among others without considering numerical representativeness [28].

To carry out the analysis and identification of the influence that the DLC steps have on the implementation of the treatment principles regarding the privacy of personal data, the DLC model proposed in [9] and the principles recommended by the GDPR were used. In order to identify such relationships, we tried to answer the following question: What is the influence that the DLC phases have on the adequacy of each GDPR principle?

The influence of the DLC phases on the GDPR principles was identified considering that they can present different levels of performance, as shown below:

- Very Influential: for situations where the phase significantly influences the adequacy of the principle - represented by the symbol ●;
- Influential: for situations where the phase has an influence on the adequacy of the principle, but indirectly - represented by the symbol ○;
- Unidentified: for situations where there was no influence on the adequacy of the principle - represented by the acronym UI.;

In order to identify the treatment of data carried out on Blockchain technology and associate it with the DLC phases, the functioning and functions performed by the technology were evaluated and these were related to the phases. With the view of this and the previous analysis that considered the influence that the phases of the DLC exert in the adequacy of the privacy principles, some treatment options for the data were presented to compose the gaps identified in each phase. The results of the analyzes will be displayed in the next session.

5 Results

5.1 The influence of the DLC phases on the adequacy of the GDPR principles

Table 2 shows the influence of the DLC phases on the adequacy of the GDPR principles. The table outlines the GDPR principles and the DLC steps in the columns.

In order to adapt the principle of lawfulness, loyalty and transparency, it was identified that all phases of the DLC must be observed as they have a relevant influence on their adjustments. When considering that this principle is related to the processing of data for lawful purposes, appropriate to its purpose, which is defined in the consent of the holder, compliance with contracts, compliance with legal obligations by the data controller, legitimate interest and the other requirements defined in the legislation, the treatment of the data must occur in a transparent manner and providing the holders with the guarantee of clear, accurate and easily accessible information in all phases of the DLC.

For the Purpose limitation principle, very relevant points of influence were identified with the stages of collection and recovery. When considering that the collection has the objective of supplying the informational need, identifying and limiting the purpose of using the data in this stage is essential and meets its objective. Likewise, in the recovery stage, the restored data must be restricted to the fulfillment of determined and legitimate purposes. It was also identified that, in order to adapt this principle, the storage step has an indirect influence, however, since only the data necessary to meet the purpose identified in the collection phase should be stored. For this principle, no influence was identified on the Data Disposal stage.

● Very Influential: has a relevant influence on the adequacy of the principle.

○ Influent: has an indirect influence on the adequacy of the principle.

UI: Unidentified Influence.

For the Purpose limitation principle, very relevant points of influence were identified with the stages of collection and recovery. When considering that the collection has the objective of supplying the informational need, identifying and limiting the purpose of using the data in this stage is essential and meets its objective. Likewise, in the recovery stage, the restored data must be restricted to the fulfillment of determined and legitimate purposes. It was also identified that, in order to adapt this principle, the storage step has an indirect influence, however, since only the data necessary to meet the purpose identified in the collection phase should be stored. For this principle, no influence was identified on the Data Disposal stage.

For the principle of data minimization, a very relevant influence was observed with the collection and recovery phases. As this is a principle that advocates limiting treatment to the minimum necessary to achieve its purposes, it is observed that to adjust to this principle, it is necessary to start by minimizing data collection, identifying the data strictly necessary to be collected and similarly those that are essential to be recovered to supply the informational need and availability to fulfill its purpose of existence in the database. It was also identified that in order to adapt to this principle, the storage and disposal phases have influence, but indirectly when considering that the minimum data storage should be adopted and its elimination must meet the minimum necessary retention time and consistent with the purpose of use of the data.

Regarding the principle of accuracy, a relevant influence was identified for its adequacy with all phases of the DLC: in the collection phase by defining elements that enable the perception and validation of the collected data; in the definitions of the storage phase to ensure that the data maintains its physical and logical integrity and guarantees and that it is reliable to the

Table 2 Influence of the DLC phases on the adequacy of the GDPR principles

	Collect	Storage	Recovery	Disposal
Lawfulness, loyalty and transparency (P1)	●	●	●	●
Purpose limitation (P2)	●	○	●	**UI**
Data minimization (P3)	●	○	●	○
Accuracy (P4)	●	●	●	**UI**
Retention limitation (P5)	●	●	**UI**	●
Integrity and confidentiality (P6)	●	●	●	●
Responsibility (P7)	●	●	●	●

original collected; resources made available in the recovery which must reflect these same aspects described for the collection and storage and that it is available whenever necessary to guarantee the holder access and the possibility of repairing them with agility and disposal, so that they can, after the confirmation that the data is incorrect, be erased from all storage locations.

In order to adapt the principle of limitation of retention, a relevant influence was identified in the phases of collection, storage and disposal: in the phase of collection when considering that the purpose of data collection is identified, thus linking the time required to retain them to fulfill their purpose; storage, as it is at this stage that automated measures should or should not be adopted for the deletion of data that are no longer needed and that have expired the retention period and in the disposal that enables the tracking and deletion of data from all databases, whether they be the original, redundancy or backup, ensuring that the retention time is met. Due to the fact that the recovery phase is directly related to data storage, no influence was identified in this phase for the adequacy of the principle - if the data is excluded in the appropriate retention times, it will not be possible to recover them.

When evaluating the principle of integrity and confidentiality, it was also possible to identify relevant influence in all phases of the DLC for its adequacy. It was observed that in order to protect personal data from unauthorized access and from accidental or unlawful situations related to destruction, loss, alteration, communication or dissemination, technical and administrative measures must act from the collection of the data until its disposal.

As for the adequacy of the principle of responsibility, it was observed that all phases of the DLC have a relevant influence on its adequacy. In order for the data controller to be able to demonstrate the adoption of effective measures, capable of proving compliance and compliance with the requirements of personal data protection, these must be employed, practiced and validated at all stages of the cycle and must be implemented with resources that make it possible to evidence its execution.

5.2 Blockchain data processing and DLC

The assessment was carried out considering the specialized Blockchain that processes personal data in its primary version. It was observed that Blockchain technology concentrates all data processing in the storage phase. Therefore, smart contracts can be used to implement parts of the requirements necessary to meet the principles of GDPR in this and other phases of the DLC that do not present any treatment of the data.

Smart contracts are computer codes that are executed on each node of a Blockchain to fulfill the pre-established trades. In the context of this analysis, for example, these negotiations can be initiated by defining the roles of each node. As Blockchain has several nodes in a network, it is necessary to specify what the roles of these nodes will be, qualifying them as controller or operator, since these roles have different functions in the legislation, so that the controls and the level of autonomy appropriately assigned to each of the actors. The purpose of processing is also another important issue to be identified in order to direct all data processing and adapt the technology to its purpose.

Table 3 presents the data treatment that occurs in Blockchain technology linked to the DLC phases and shortly afterwards the analysis performed is described. To support this analysis and the proposals for alternative data processing in each phase, the result obtained in the previous analysis was shown in this table.

In the analysis carried out, during the collection stage, a relevant influence was identified on the adequacy of all GDPR principles: (P1) lawfulness, loyalty and transparency; (P2) limitation of purposes; (P3) minimization of data; (P4) accuracy; (P5) limitation of retention; (P6) integrity and confidentiality and (P7) responsibility.

In the Blockchain, no data processing is performed during the collection phase, that is, if there is no type or category differentiation process, the technology processes all the collected data. Considering that the focus of this work is the necessary treatment of personal data, the collection phase may consider a data classification system that typifies them and allows only personal data to be submitted to the privacy treatments recommended by the legislation.

For this stage, it is observed that an adaptation of the principle of "need to know" is applicable - provided in Chapter III, Section I Art. 18 of Decree 7845 which specifies that - "Access, dissemination and treatment of classified

Table 3 Blockchain data handling x DLC x GDPR

BLOCKCHAIN	CVD	GDPR						
		P1	P2	P3	P4	P5	P6	P7
No treatment is performed.	**Collection**	●	●	●	●	●	●	●
• Stores data in block chains.	**Storage**	●	○	○	●	●	●	●
• Uses distributed storage.								
• All nodes have a copy of the ledger.								
• Change / update of data: a new block is created and the previous one (with the original information) remains.								
No treatment is performed.	**Recovery**	●	●	●	●	UI	●	●
No treatment is performed	**Disposal**	●	UI	○	UI	●	●	●

information will be restricted to people who need to know it...” being its adaptation for “Collection and treatment of personal data strictly necessary to fulfill its purposes”, with questions such as: 1) What data are really necessary for the scope of the system? 2) What is the purpose of data collection? 3) What is the appropriate way to collect such data, considering its purpose? It is worth noting that this is not an activity that is directly related to Blockchain technology, but it is a relevant principle for data processing, and the controller, in a process of surveying the need to identify the informational need and then specify in the technology collecting only the data necessary for its purpose.

In the data storage stage, a relevant influence was identified on the adequacy of the following GDPR principles: (P1) lawfulness, loyalty and transparency; (P4) accuracy; (P5) limitation of retention; (P6) integrity and confidentiality and (P7) responsibility.

In Blockchain technology, all data processing takes place in the storage phase. As described in section 2.3, the data is stored in chains of blocks containing the hash with the link to the previous block and the data of the current transaction, so that the blocks are linked in chronological order, chaining a dependency of the current to the previous ones, that is, a data once recorded in a block, cannot be changed in a later block, without changing the previous ones. It uses distributed storage which each node in the network has a copy of the ledger - the information that passes through the blockchain is visible to all nodes. Some alternatives have been studied to address these issues since, as they are personal data, there are privacy principles to be met. One of them is what has been called “OffChain” which proposes the storage of personal data outside the chain, with restricted access to transaction information only to authorized parties. Another alternative is storage in “SideChain” - the use of a parallel Blockchain for storing personal data with different permissions than the existing one in the primary Blochchain. Smart contracts can also be used to send data to be stored outside the chain or in a parallel chain, since the results of executing the codes of these smart contracts are also recorded on the primary blockchain, thus providing a whole track that allows validating which ones data went to external storage. Smart contracts can also be used to implement a data purge process to meet the required retention time and complementary encryption and hashing techniques can be adopted to ensure data confidentiality and integrity such as additional access control mechanisms to guarantee the confidentiality of the stored data - be it the access control to the data itself, the cryptographic key that makes it unreadable or any other attribute that allows access to the controlled personal data.

In the recovery stage, a relevant influence was identified on the adequacy of the following GDPR principles: (P1) lawfulness, loyalty and transparency; (P2) the purpose limitation; (P3) minimization of data; (P4) accuracy; (P6) integrity and confidentiality and (P7) responsibility.

At Blockchain no data processing is performed at this stage of the CVD. In addition, it was identified that the treatment of data in the recovery phase is directly related to the form of storage and the controls added there. Considering that all nodes have a copy of the ledger, the encryption of personal data can be adopted to ensure that the nodes have access only to parts of the data or is even linked to the access control mechanisms implemented, but all in the storage.

The disposal phase had a relevant influence on the adequacy of the following GDPR principles: (P1) lawfulness, loyalty and transparency; (P5) limitation of retention; (P6) integrity and confidentiality and (P7) responsibility.

Blockchain, on the other hand, does not perform any data processing during the disposal phase. However, it is worth mentioning that Blockchain technology was designed to not allow the data to be erased. In its design, the immutability of the technology works so that the information is not removed from the blockchain. To this end, smart contracts can be designed to revoke access by making data inaccessible by third parties after the expiration of the data retention time, when the data is changed and the updated data is saved in a new block, or for any other reason, it is a solution which does not erase the

data, but does not allow access to it. Another alternative that has been discussed is the use of irreversible encryption, that is, when the data retention time necessary to fulfill its purpose has expired, they can be permanently encrypted so that they cannot be reversed and recovered. In the same way in this option, the data remains stored, but in an unreadable way.

6 Conclusions and future works

The objective of this study was to analyze the GDPR and its principles which aim to ensure the privacy of personal data and to verify if the process of adaptation to them can be guided by the steps of the DLC.

The evaluation considered the data treatment performed on the Blockchain, presenting an approach for adapting a technology to the data privacy principles covered in the DLC-guided GDPR. This study did not aim to assess the compliance of Blockchain technology to the principles of the legislation, but rather a way to lead the adaptation of the technology to that compliance.

It was observed that Blockchain technology concentrates all data processing in the storage phase, considering the model proposed by [9] and that in this context, the phase could also count on a data processing step since some of the treatments perform processing actions.

It can be concluded that the DLC model proposed by [9] and the influence of its steps on the adequacy of the GDPR principles can be used as a support tool to systematize the process of adapting to the law by organizations that use or intend to adopt Blockchain technology.

It can be observed that even in the stages that did not have a relevant influence on the adequacy of all the principles of the legislation, several data treatment actions need to be directed to adapt the technology to the legislation, since the Blockchain in its primary version, does not perform data processing data in the collection, recovery and disposal phase.

When considering the disposal step, for example, which had a relevant influence on four of the seven principles, Blockchain in its conception was designed not to allow the exclusion of the registered data, making it necessary in this step for alternative solutions to treat the data and adapt it to the principles that showed relevant influence.

It is also observed that the treatment definitions in the storage step directly reflect in the recovery step so that the controls to meet the principles that had a relevant influence in the recovery step, must be observed in the storage step.

As future work it is proposed to carry out studies that, considering the influence of the phases identified in this research, present suggestions for concrete actions that can be performed in each phase of the DLC to adapt the technology to the GDPR. In addition to this, it is also proposed to carry out an analysis that contemplates the rights of the holder of personal data guided by the phases of the DLC.

Acknowledgements This research was partially supported by Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES) and the Santa Catarina Research Foundation (FAPESC) by grant Public Note FAPESC N° 03/2017.

References

1. GDPR (General Data Protection Regulation) (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council. https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed 24 April 2020
2. Collins R (2016) Blockchain: A New Architecture for Digital Content, EContent. http://www.econtentmag.com/Articles/Editorial/Commentary/Blockchain-A-New-Architecture-for-Digital-Content-114161.htm. Accessed 20 April 2020
3. Hernon P (1994) Information lifecycle: its place in the management of US government information resources. Gov Inf Q 11(2):143–170. https://doi.org/10.1016/0740-624X(94)90002-7
4. Kreutz D, Malichevskyy O, Feitosa E, Cunha H, da Rosa RR, de Macedo DDJ (2016) A cyber-resilient architecture for critical security services. J Netw Comput Appl 63:173–189. https://doi.org/10.1016/j.jnca.2015.09.014
5. Wallauer J, von Wangenheim A, Andrade R, de Macedo DDJ (2008). A telemedicine network using secure techniques and intelligent user access control. 21st IEEE international symposium on computer-based medical systems, pp. 105-107. https://doi.org/10.1109/CBMS.2008.124
6. de Macedo DDJ, von Wangenheim A, Dantas MA (2015) A data storage approach for large-scale distributed medical systems. Ninth international conference on complex, intelligent, and software intensive systems. Pp. 486-490. https://doi.org/10.1109/CISIS.2015.88
7. Gomes EH, Dantas MA, de Macedo DDJ, Rolt CRD, Dias J, Foschini L (2018) An infrastructure model for smart cities based on big data. International Journal of Grid and Utility Computing 9(4):322–332. https://doi.org/10.1504/IJGUC.2018.095435
8. de Souza IA, Andrade R, von Wangenheim A, and Macedo DDJ (2014). Designing an information retrieval system for the STT/SC. 16th international conference on e-health networking, applications and services (Healthcom), pp. 500–505. https://doi.org/10.1109/HealthCom.2014.7001893
9. Sant'Ana RCG (2016) Data life cycle: A perspective from the Information Science. Informação & Informação 21(2):116–142. https://doi.org/10.5433/1981-8920.2016v21n2p116
10. Freund GP, Fagundes PB, Macedo DDJ (2020) identification of the relationships between the stages of the data lifecycle andthe principles of the Brazilian general data protection act. Lecture notes of the Institute for Computer Sciences, social informatics and telecommunications engineering. (Ed.): DIONE 2020, LNICST 319, pp. 1–10, 2020. https://doi.org/10.1007/978-3-030-50072-6_7 (in press)
11. Voigt P, Von dem Bussche A (2017) The EU general data protection regulation (GDPR): a practical guide, 1st edn. Springer International Publishing, Cham
12. Hoofnagle CJ, Van der Sloot B, Borgesius FZ (2019) The European Union general data protection regulation: what it is and what it means. Information & Communications Technology Law 28(1):65–98. https://doi.org/10.1080/13600834.2019.1573501
13. Mercer S (2020) The limitations of European data protection as a model for global privacy regulation. AJIL Unbound 114:20–25. https://doi.org/10.1017/aju.2019.83

14. Ruparelia NB (2010) Software development lifecycle models. SIGSOFT Softw. Eng Notes 35(3):8–13. https://doi.org/10.1145/1764810.1764814
15. Nayak A, Dutta K (2017) Blockchain: the perfect data protection tool. International conference on intelligent computing and control (I2C2) pp. 1-3. https://doi.org/10.1109/I2C2.2017.8321932
16. Macrinici D, Cartofeanu C, Gao S (2018) Smart contract applications within Blockchain technology: a systematic mapping study. Telematics Inform 35(8):2337–2354. https://doi.org/10.1016/j.tele.2018.10.004
17. Lacity M (2018) Addressing key challenges to making Enterprise Blockchain applications a reality. MIS Q Exec 17(3):201–222 https://aisel.aisnet.org/misqe/vol17/iss3/3/.
18. Gupta V. A brief history of Blockchain. Harv Bus Rev https://hbr.org/2017/02/a-brief-history-of-blockchain. Accessed 20 April 2020
19. Bernabe JB, Canovas JL, Hernandez-Ramos JL, Moreno RT, Skarmeta A (2019) Privacy-preserving solutions for Blockchain: review and challenges. IEEE Access 7:164908–164940. https://doi.org/10.1109/ACCESS.2019.2950872
20. Suzuki S, Murai J (2017) Blockchain as an audit-able Communication Channel. Conference: 2017 IEEE 41st annual computer software and applications conference (COMPSAC), pp. 516-522. https://doi.org/10.1109/COMPSAC.2017.72
21. Christidis K, Devetsikiotis M (2016) Blockchains and smart contracts for the internet of things. IEEE Access 4:2292–2303. https://doi.org/10.1109/ACCESS.2016.2566339
22. Zheng Z, Xie S, Daí H, Chen X, Wang H (2017) An overview of Blockchain technology: architecture, consensus and future trends. 2017 IEEE international congress on big data, (BigData congress), pp. 557–564 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8029379. Accessed 16 April 2020
23. Bayle A, Koscina M, Manset D, Perez-Kempner O (2018) When Blockchain meets the right to be forgotten: technology versus law in the healthcare industry. 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) pp 788–792 https://doi.org/10.1109/WI.2018.00133
24. Zemler F, Westner M (2019) Blockchain and GDPR: Application Scenarios and Compliance Requirements. Portland International Conference on Management of Engineering and Technology (PICMET), pp. 1–8. https://doi.org/10.23919/PICMET.2019.8893923
25. Rieger A, Guggenmos F, Locki J, Fridgen G, Urbach N (2019) Building a blockchain application that complies with the eu general data protection regulation. MIS Quarterly Executive 18(4):263–279. https://doi.org/10.17705/2msqe.00020
26. Millard C (2018) Blockchain and law: incompatible codes? Computer Law & Security Review 34(4):843–846. https://doi.org/10.1016/j.clsr.2018.06.006
27. Moerel L (2018) Blockchain & data protection…and why they are not on a collision course. European review of private law 26(6*): 825-851*. http://www.kluwerlawonline.com/abstract.php?area=Journals&id=ERPL2018057
28. Powell RR, Connaway LS (2004) Basic research methods for librarians (4th ed.) Westport, CT: libraries unlimited