




DIY4U

Open Innovation Digital Platform and Fablabs for Collaborative Design and Production of personalised/customised FMCG

DT-FOF-05-2019 (870148)

Deliverable Report

Deliverable ID	D3.4	Version	V2.0
Deliverable name	Report on how proposed design meets GDPR and other relevant regulations		
Lead beneficiary	EFF		
Contributors	EFF, DCC, IRIS, STELAR		
Due date	31.07.2020		
Date of final version	30.07.2020		
Dissemination level	CO: PUBLIC		
Document approval		31.07.2020	Chandana Ratnayake



The DIY4U project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No. 870148

PROPRIETARY RIGHTS STATEMENT

This document contains information which is proprietary to the DIY4U consortium. The document or the content of it shall not be communicated by any means to any third party except with prior written approval of the DIY4U consortium.



Version	Authors	Date
V.0.1	Alex Butean, Andrei Tara, Ion Ceban	08.06.2020
V.0.5	Alex Butean, Robert Learney, Ignacio Montero, Dominyka Zemaityte	22.06.2020
V.1.0	Alex Butean, Andrei Tara, Robert Learney, Ignacio Montero, Matthias Pocs, Luca Iadema, Ion Ceban, Cristina VasIU, Florian Stoica, Marta Jurado, Elmer Zinkhann, Ivan Pecorari	06.07.2020
V.1.1	Alex Butean, Andrei Tara, Robert Learney, Ignacio Montero, Luca Iadema, Ivan Pecorari, Ion Ceban, Cristina VasIU, Florian Stoica, Matthias Pocs	22.07.2020
V.2.0	Alex Butean, Ignacio Montero, Juan Enríquez	28.07.2020



Executive Summary

Human rights ensure that our basic needs are met; they also guarantee our life, freedom, equality and security. Worldwide authorities issue regulations that are trying to protect human rights against acts of corruptions and abuses of all kinds. In the EU, there are several legal acts enforcing the law: regulations, directives, decisions and recommendations, each of them having different legal consequences. Each organization should be compliant with laws from the territory where they belong, or laws from countries where their users and data come from. In a cross-organizational environment, each organization should understand and respect the other party's laws and regulations that apply to them.

The first step in achieving regulatory compliance is to understand the laws and regulations. In this deliverable, there is an overview of the leading regulations. The General Data Protection Regulation (GDPR) is described in more detail from a legal point of view and also from the implementation perspective. Of course, it is a compressed version of the regulation, and for a more detailed picture, the reader can consult the GDPR law. There are guidelines made by the EU on how organizations should function to be compliant with GDPR; this paper provides details about some of these recommendations.

Based on functionalities described in deliverable D3.3, we specified in the last chapter how the DIY4U platform meets the GDPR. In the DIY4U ecosystem, there are multiple stakeholders; each interaction between them and each personal data processing and storing should be compliant with GDPR. The architecture has been designed based on the principle of "data protection by design and by default". This principle ensures that the architecture is built from the beginning with the personal data security in mind. Access is granted by an access control mechanism that permits particular access for different stakeholders.



Table of Contents

List of Figures and Tables	5
List of Abbreviations and Acronyms	5
1. Introduction	6
2. Objectives	6
3. Terms definitions	6
4. Data perspectives related to regulations	7
4.1 Territorial distribution	8
4.2 Regulations	9
4.2.1 GDPR	9
4.2.2 CCPA	10
4.2.3 HIPAA	10
4.2.4 PCI-DSS	11
4.2.5 CISL	12
4.2.6 PSS	12
4.3 Cross-organisation data operations	12
5. Data security mechanism	13
5.1 Depersonalisation methods	13
5.1.1 Pseudonymisation	14
5.1.2 Anonymisation and anonymous data	14
5.1.3 Anonymisation vs pseudonymisation	14
5.2. Data protection mechanisms	15
5.3 Identity management and data access control	15
6. Technical specifications	16
6.1. How to implement GDPR in software	16
6.1.1 Steps for GDPR compliance	17
6.1.2 Demonstrating GDPR compliance	19
6.2 GDPR and blockchain technology	19
6.2.1 Data responsibility and accountability	19
6.2.2 Data storage	20
6.3 GDPR & DIY4U environment	20
6.3.6 Machine learning	23
7. Conclusion	23

References

25

List of Figures and Tables

Figure or Table	Page
Table 1. Private information in different regulations	7
Figure 1. Data subject rights	9
Figure 2. Data anonymisation levels	13
Figure 3. Checklist for GDPR requirements	18
Figure 4. GDPR in DIY4U platform	24

List of Abbreviations and Acronyms

CCPA	California Consumer Privacy Act
CISL	China Internet Security Law
DLT	Distributed Ledger Technology
DPIA	Data Protection Impact Assessment
DS	Data subject
ECDSA	Elliptic Curve Digital Signature Algorithm
FADP	Federal Act on Data Protection
GDPR	General Data Protection Regulation
HE	Homomorphic encryption
HIPAA	Health Insurance Portability and Accountability Act
IPFS	InterPlanetary File System
LCA	Life Cycle Analysis
LCCA	Life Cycle Cost Analysis
MPC	Multiparty computation
TEE	Trusted Environment Execution



1. Introduction

Efforts for convergence of individual frameworks are underway. However, as of 2020, these efforts are not widely adopted and thus are not of practical relevance. Multiple regions or countries have strict data privacy rules. In the EU, there is the General Data Protection Regulation (GDPR) that is a restrictive regulation regarding data privacy, but some articles contain flexible interpretations; for example, in GDPR, personal data is any data that can uniquely identify a person. In the USA's HIPAA (Health Insurance Portability and Accountability Act) for personal information, there is a specific list of attributes. The US relies on a set of narrow scope industry-specific regulations: HIPAA (health), Federal Trade Commission (e-trade), GBLA (financial). They also have a regulation for business and consumer rights, but it is applicable in California or outside of California for their residents. The CCPA (California Consumer Privacy Act) is similar to the GDPR. However, with a broader scope (more consumer-oriented), it is applied to consumers, devices and households; in GDPR (person-oriented) only the "natural person" is affected. Both of them give rights to access and disclosure, rights to remove data, specific rules for children's data. GDPR is more restrictive with the processing of "sensitive data". At the same time, CCPA does not have such limitations; also, in GDPR public information is considered 'personal data', while in CCPA it is not. GDPR is required for individual business data protection officers, but CCPA is not mandatory.

Japan's Act on Protection of Personal Information (May 2017) is similar to the GDPR, and Japan and EU have reached an agreement on "reciprocal adequacy" of their laws. Also, South Korea's "Personal Information Protection Act" and Thailand's PDPA (Personal Data Protection Act) contain articles and rules comparable to the EU's GDPR; actually, the penalties in PDPA are much higher than the ones in GDPR. This paper will cover privacy regulations and how to be compliant with GDPR. More information regarding the current security and privacy standards like: ISO/IEC 27001, ISO/IEC 27002, ISO/IEC 27701:2019 can be found in deliverable D2.7.

2. Objectives

This deliverable aims to describe how DIY4U meets the requirements for GDPR. The following steps should be reached:

- Regulatory overview
- GDPR definition, rights and obligations required and applicable in the DIY4U ecosystem
- Methods and techniques to implement GDPR in DIY4U platform
- How DIY4U platform design respects the GDPR

3. Terms definitions

Personal data. According to GDPR "Personal data is any information which is related to an identified or identifiable natural person". Identifiers are attributes that allow to directly or



indirectly identify a person: name, location information, identification number, online identification number. There is a subcategory of personal characteristics called "sensitive personal data" which are subject to more restrictive security; it includes genetics, biometric, health data, racial and ethnic origin, political, religious or ideological convictions or trade union membership. The lifetime of personal information is from a person's birth to the person's death; only data that belongs to living persons is under the protection of the personal data law.

Processing. Any operation which uses personal data or sets of personal data.

Controler (GDPR, Art. 7). Any entity which separately or jointly with others, determines the means and purposes of personal processing data.

Processor (GDPR, Art. 8). Any entity which processes personal data on behalf of the controller.

Consent (GDPR, Art. 11). It is the freely given data subject's agreement for the processing of his/her data.

DPO - Data protection officer is a person whose task is to ensure that the processor and controller are compliant with GDPR (can be a company employee or an external person). It is necessary when processing operations require regular and systematic monitoring of data subjects on a large scale (GDPR, Art. 37).

4. Data perspectives related to regulations

Data protection regulation varies across the globe, resulting in a patchwork of requirements which can be a hurdle for international business. Regulations vary in terms of regional scope (which countries, states, or territories enforce it), data scope (what data it covers), and organisational scope (which organisations it affects). Additionally, the restrictions imposed by each regulation can vary significantly.

Regulation/law	Information considered as private according to different regulations
GDPR	Information that can be used to identify a natural person - directly or indirectly
CCPA	Information that identifies, directly or indirectly, with a particular consumer or household.
PSS	Information that can be used to identify a natural person - either independently or in combination with other information
CISL	Personal information (see PSS) and other information with implications to national security
HIPAA	Health information that can be used to identify an individual (a list of 18 identifiers is given)
PCI-DSS	Cardholder data and sensitive authentication data (a list of specific data types covered is given)

Table 1. Private information in different regulations

4.1 Territorial distribution

Europe

The EU relies on the comprehensive legislation **General Data Protection Regulation (GDPR)** which covers all member countries, all organisations (for/non-profit, government, judiciary). It defines private data in a broad sense [1] (that is, any data can be used to identify an individual personally). The GDPR has become a point of reference for economies that are creating new privacy laws. Switzerland's **Federal Act on Data Protection (FADP)** is similar to GDPR.

The European Council has produced **Convention 108**: an international legally binding agreement on data protection. Any country in the world can sign the Convention, and legally binds signatories to enforce data privacy laws satisfying specific requirements - which enshrine the same principles as the GDPR. Convention 108 is not a set of regulations, but a regulatory framework, to enable cross-border data flow.

America

In contrast with the EU, the US relies on a set of privacy law regulations - each applicable in a different industry or type of data. Furthermore, various states may enforce different standards, and regulations may often overlap. The most notable US privacy regulations are:

- **HIPAA** which applies to organisations handling healthcare data originating from the US, and covers the confidentiality, security and transmissibility of healthcare data.
- **CCPA**, a Californian regulation covering companies with extensive data reach which handle personal data of California residents.
- **Federal Trade Commission FTC** fair information practice defines information practices for e-trade.
- **GBLA** regulates how financial institutions explain their information processing and sharing practices to their customers.

The **Brazilian General Data Protection Law (LGPD)**, is mostly aligned with the EU's GDPR.

Asia

Japan adopted the **Act on the Protection of Personal Information (APPI)** one year ahead of the GDPR, they have similar rules, but there are also some small differences; in API notifications in case of data breaches are not mandatory.

Africa

The **African Union Convention on Cyber Security and Protection of Digital** data seeks to provide a unified framework for data privacy and cybersecurity on the African continent. The number of signatories is increasing, but the convention is not in effect yet [2].



4.2 Regulations

4.2.1 GDPR

Information privacy in the EU is defined through the **General Data Protection Regulation GDPR**, a comprehensive regulation which is applicable across industries and countries. The GDPR aims to unify rules about personal data collection, storage, and processing and may be supplemented by country-specific laws. Additionally, several non-EU countries have used GDPR as a paradigm when producing legislation. Notably, the privacy laws of Japan and Brazil enshrine the same core values of the GDPR [3].

The GDPR covers data that can be used to identify individuals from the EU or EEA personally. The GDPR does not give a precise definition of data anonymity. Instead, it considers data anonymous if re-identification is only possible through effortful or nontrivial means [4].

The GDPR applies to all organisations in the EU, and all organisations storing information about EU citizens. Companies with fewer than 250 employees are not required to maintain a record of processing activities under its responsibility, unless “the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects (DS). The processing is not occasional, or the processing includes special categories of data [...] or personal data relating to criminal convictions and offences”.

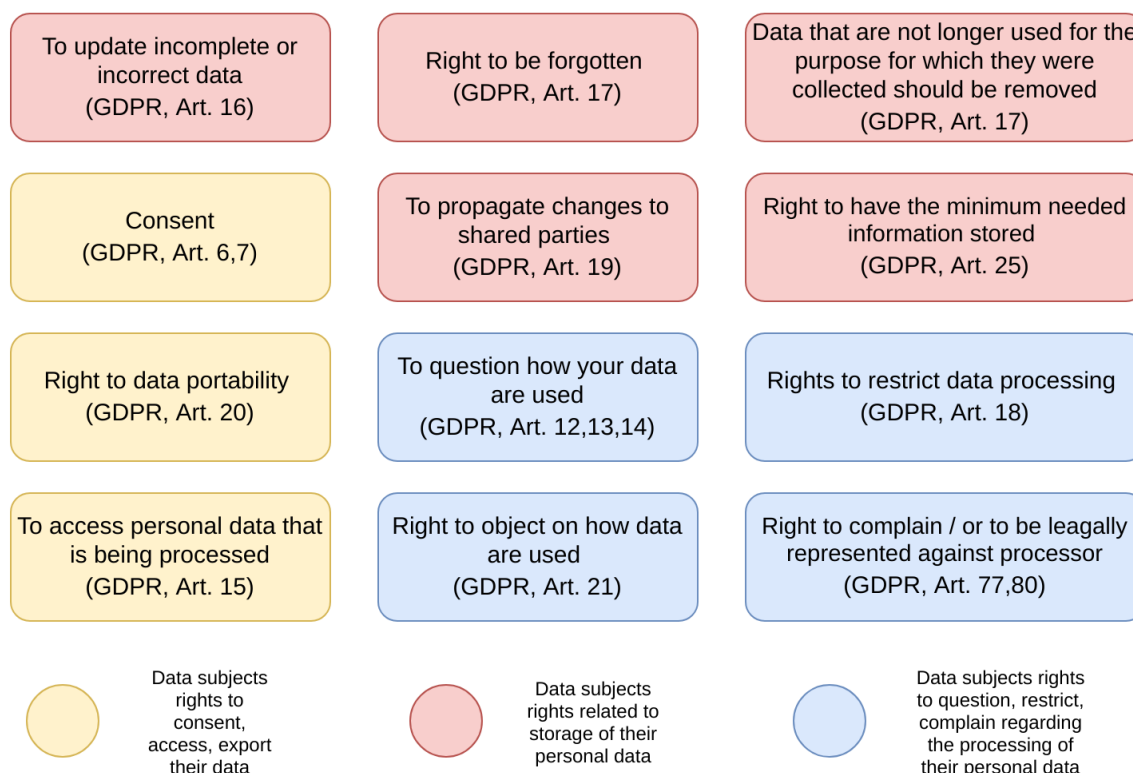


Figure 1.Data subject rights



4.2.2 CCPA

A Californian law, focusing on for-profit organisations that involve data operations. Additional states are adopting similar regulatory frameworks (e.g. New York's law S5462, and Nevada's amendment of the online privacy law). The CCPA covers data that simultaneously satisfies all the following conditions:

- Personally Identifiable Information (PII). PII is data that can identify someone, either as an individual or as part of a household. A non-exhaustive list of the data that qualifies as PII can be found in the CCPA, subsection (o)(1) of 1798.140.
- CCPA does not apply to information that is lawfully made publicly available (e.g. Government census records).
- Data regarding individuals domiciled in California (including Californians on vacation, and short-time residents).

The CCPA covers any for-profit organisation that satisfies two criteria:

- collects personal information of consumers or conducts any form of business in California (including e-commerce).
- has a Gross revenue of more than 25 US\$ million, or collects data on more than 50k customers, or obtains half of its yearly revenue by selling PII data.

Customers have the right to know details on data operations: how data is collected, what data is collected, where the data is stored, whom the data is shared with and for which purpose. Upon request, this information must be provided to the DS within 45 days, and the DS may make such requests twice a year to the same company. The DS must be informed "at or before" the point of data collection. The company should provide a point of contact for clarifications on CCPA compliance. The DS reserves the right to be forgotten (similarly to GDPR), opt-out of marketing usage of their data and prohibit the selling of their data. Finally, the company may not discriminate against clients based on their data preferences. The company should present its privacy policy to customers, update its privacy policy every 12 months, and inform the clients of the updates.

4.2.3 HIPAA

HIPAA is a regulation covering the privacy and security of healthcare data in the US. It applies to 'covered entities' (doctors, dentists, pharmacies, health insurance companies, company health plans) and business associates that are using the personal healthcare information (attorneys, IT service providers, laboratories, analysis services, etc.). HIPAA is an organisation-centric regulation, meaning that it refers to organisations instead of individuals. HIPAA does not cover US citizens who are travelling abroad. In contrast, all organisations handling healthcare data operating in the US are covered - even if they are working with non-US citizen data. Business



associates outside the US still need to be HIPAA compliant when handling HIPAA data (e.g. a data processing centre outside the US working with data from a US hospital).

HIPAA safeguards information referred to as 'protected healthcare information' (PHI) by ensuring that data subjects have adequate ownership of their data. Regarding security, HIPAA prescribes a set of practices including encrypting sensitive data, good password practices, employee training etc. In the context of HIPAA, DSs have the following rights concerning their data:

- They may request the amendment of their medical records to correct errors.
- They can limit who has access to their personal health information.
- They have the right to complain about the unauthorized disclosure of their PHI and suspected HIPAA violations.
- They have the right to be notified of security breaches.

4.2.4 PCI-DSS

The **Payment Card Industry Data Security Standards** is a set of standards aimed at regulating the storage and transmission of sensitive digital payment data (e.g. credit card information) for consumer transactions. PCI-DSS compliance is required by credit card companies to make online transactions secure and protect them against identity theft. Any company that seeks to process, store or transmit credit card data is required to be PCI compliant. This standard enforces a minimum set of necessary conditions that every company and service provider must meet, to protect the data of their customers, and a list of all types of data within the scope of the regulation. The PCI DSS is a data security standard adopted not by governments, but by all major payment card brands. Thus, it is practically a global mandatory standard - since significant card brands have a global reach. PCI-DSS imposes six requirements for data protection:

- Build and Maintain a Secure Network and Systems
- Protect Cardholder Data
- Maintain a Vulnerability Management Program
- Implement Strong Access Control Measures
- Regularly Monitor and Test Networks
- Maintain an Information Security Policy

These requirements are expanded into a set of technical specifications (e.g. use of firewalls, the arbitrary substitution of passwords and card PINs, encryption of cardholder data, using anti-virus and anti-spyware software, restricted personnel access, communicating security policy of employees). Companies are validated for compliance regularly, with a method suited to the volume of transactions they handle (self-assessment questionnaire for small volumes, external audit for moderate volumes, and firm-specific internal auditors for large volume).



4.2.5 CISL

The **China Internet Security Law (CISL)** is focused on protecting critical sectors, such as telecommunications, information services, energy transport, water, financial services, public services and digital government services. The CISL focuses on network operators, and concerns all organisations that manage their own email/data networks within China (referred to as network operators). Network operators are expected to prevent data leaks and theft, report cybersecurity incidents to users and the government, and protect network operators. State agencies are authorised to inspect the networks of companies operating in China - without informing the companies. The Ministry of public security has the authority to search data for prohibited material, perform penetration testing on companies operating within China, and share data with other government organisations. The Chinese authorities have the responsibility to block any information from other jurisdictions that is prohibited by law. Additionally, companies utilising data of Chinese citizens are subject to the data protection rules of CSL.

Notably, CISL imposes more stringent requirements to Critical Information Infrastructure Operators - who process information related to national security, the economy and the public interest. All critical network operators are required to store personal information on servers within China.

4.2.6 PSS

The **Personal Information Security Specification (PSS)** governs the collection, storage and use of personal data by network operators. In the context of PSS, personal data is defined as any data that can be used alone or in combination with other information to identify a natural person. The PSS concerns only particular information about Chinese citizens and focuses on network operators (similarly with the CISL). The restrictions of the PSS are highly similar to the ones found in GDPR - but focus on national security. The most notable difference is that in PSS the concept of 'consent' (e.g. to gather data) is less explicit than it is in GDPR. Specifically, the PSS consent may be implicit (similarly to CCPA).

4.3 Cross-organisation data operations

Data privacy regulations include restrictions that apply to cross-organisational data operations. All the regulations considered in this study define both a legal basis for data exchange between organisations (e.g. fulfilment of a contract according to the PSS). There are several steps for preserving security: data anonymisation, pseudo anonymisation, encryption. However, additional restrictions may arise when the organisations involved in the exchange are subject to different regulations. Such restrictions can be a hurdle to the digital economy, and several international frameworks have been formed to ease business.

For example, the GDPR includes rules for the exchange of data between organisations within the EU, but also includes specific requirements for the transfer of data outside of the EU, one of



which is that the recipient country should have adequate data protection laws [5]. While, for example, Japan meets this requirement, the USA does not [5][6][5]. Private data can only be exchanged between the USA and the EU via the Privacy Shield agreement [7] (a mechanism that ensures individual data protection when transfers are done between Europe - US or Swiss-US).

Other notable frameworks related to private cross-border data flows are the APEC-CBPR - which covers the Asia Pacific region [8], the Convention of 108 - which is drafted by the European Council and enshrines the same principles as the GDPR, and the African Union Convention on Cyber Security and Protection of Digital data [2] - which is not yet in force, but may soon cover the African continent.[9]

Notably, the CISL imposes a different requirement in the cross-border data flows to emphasise national cybersecurity. For example, according to the CISL, data deemed critical to national security [10] (which includes private data) should be cleared by the government before leaving the country. Any significant network operator working within the People's Republic of China needs to store critical information on servers located within China.

5. Data security mechanism

5.1 Depersonalisation methods

The process that transforms personal data so that it cannot be linked to any individuals is called depersonalisation or deidentification. Various regulations have a different understanding of data anonymisation, [11]classify the anonymisation levels, and Figure 2 describes these levels. The identifying data is data that contains identifier attributes for individuals. Codes replace pseudonymised data quasi attributes; this process is reversible. Quasi attributes are attributes that can identify an individual. GDPR, FADP and other regulations do not have a clear list of attributes which are quasi. The identifying data contains the maximum amount of information compared with the anonymised data, which provides less information; there is a trade-off between the amount of data and the anonymity level. GDPR and FADP regulations are more restrictive than HIPAA; here data is considered anonymised if the re-identification is not possible anymore. Or it should take an unreasonable technical effort to achieve it; for HIPAA anonymisation is accomplished if a particular list of identifier attributes is removed. Whenever possible in the DIY4U platform, the personal data will be anonymised. Depending on machine learning algorithms and simulators, it may be possible that some personal data is required. In these cases, pseudo-anonymised datasets should be provided; also, the user consent should be requested for this particular processing, if the consent is not already given.

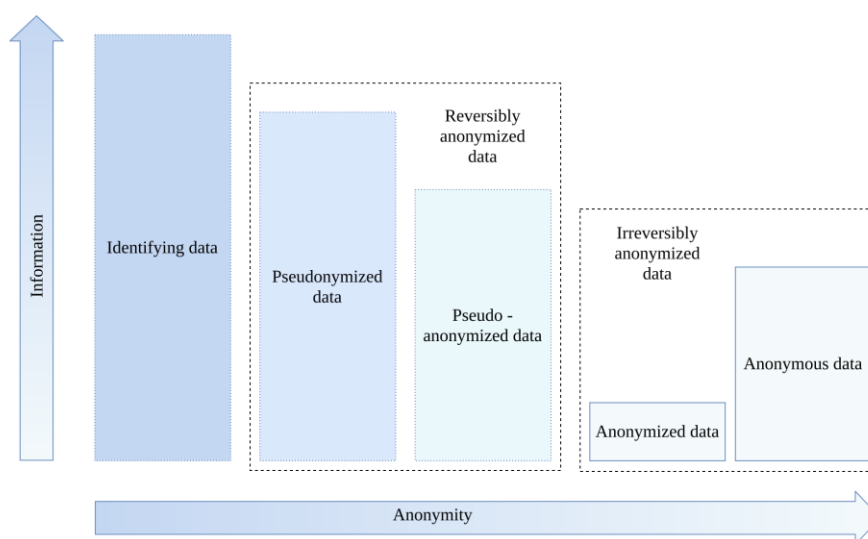


Figure 2. Data anonymisation levels

5.1.1 Pseudonymisation

Pseudo-anonymisation is a risk management measure that removes or replaces the identifiers from the original data set (GDPR Art. 5). Data cannot be connected to a specific data subject, without additional information which should be kept separated. It is a reversible process, and replacing mapping can be used to link processing results with the original data. Pseudo-anonymisation is a **data minimisation** method, and this data remains personal data (GDPR Art.29). In DIY4U pseudonymisation will be used before data analysis when homomorphic encryption and anonymisation cannot be used.

5.1.2 Anonymisation and anonymous data

Anonymised data implies that the resulting dataset cannot be linked to any individual without great effort. **Anonymous data** is data that is collected anonymously or aggregated so that re-identification of individuals is not possible. It may be possible for anonymisation to come in conflict with other regulations, especially for combating terrorism, fraud and other illegal activities [12][13]. In DIY4U platform, anonymisation should be used before data was sent for analysis or if personal data is required in the analysis process, this aspect should be clearly defined in the privacy policy.

5.1.3 Anonymisation vs pseudonymisation

It may be possible to believe that anonymisation is the best mechanism for using data and also for avoiding privacy regulations rules. However, recent studies show the disadvantages of these methods and why they may not be a solution for big data. In [14] they found that using 15 demographic attributes on an anonymized dataset (based on de-identification and splitting), individuals can be identified up to 99.98 %. The result of the study raised concerns about the reidentification of the anonymized data. During the anonymization process, the amount of information from datasets is decreased, and also some methods introduce noise. Because of this, it may be possible for the results based on anonymized data to lack accuracy and utility. Pseudo-



anonymization could be a better approach if data mapping is kept by the same entity who has the data.

5.2. Data protection mechanisms

Instead of depersonalisation, it is possible to protect data and process it in a safe environment; as already specified in D2.7 there are some methods that can be used in order to adhere to regulations and standards regarding privacy.

Hardware protection. Using specialised hardware to ensure security, also known as Trusted Environment Execution (TEE) (Intel SGX [15], TrustZone ARM [16])

Cryptographic protection:

- Data encryption. Encryption methods enforce data protection. It is useful to preserve privacy in data transfer and data storage. Data encryption alone is not enough for preserving privacy in data processing, but is a fundamental method for other cryptographic mechanisms
- Secure multi-party computation (MPC). It is a mechanism that allows the computation of inputs from different sources and keeps the inputs private.
- Homomorphic encryption (HE) [17]. Data is encrypted and shared with an untrusted third party that can process that data without decrypting it. The result is sent back to the first party and decrypted, and it should be the same as the one resulting from processing data over the unencrypted data.
- Zero-knowledge proof. It allows you to determine the truth about a statement without disclosing the underlying data. In [18] EU confirms that zkSnark can be used to ensure data protection and privacy.

The DIY4U platform will use cryptographic protection (encryption, HE, MPC), and implement part of the previous explained mechanism for data protection.

Data protection mechanisms seem to be far better solutions than anonymisation and pseudonymisation because they can use the maximum amount of information from some datasets. For hardware protection, specialised hardware is needed (and sometimes adding extra hardware is not possible and it is costly). The cryptographic solutions have the disadvantage of being slow, they require more processing time and depending on implementation, not all operations can be computed.

5.3 Identity management and data access control

In almost any enterprise applications, there is sensitive information or confidential actions that should be seen only by specific users. Identity management (IDM) or Identity access management (IAM) is a framework that defines roles, policies, grants and revokes access, ensuring that only proper users are allowed to access certain information. The GDPR is focused on protecting users' data; there are identity models that are user-centric and allow users to gain control over their data. The self-sovereign identity model pushes the user-centric model a step further, placing the user in the same central position, but with a more considerable enhancement: users not only control their identity, but also they own it. SSI does not use a trusted authority, but a



decentralised technology. There are implementations where user data is stored on their devices and other ones where data is stored on blockchain smart contracts, or decentralised storages (e.g. IPFS).

Access management can be seen as a framework which grants access to users according to their identity. Each system is based on these four pillars: identification, authentication, authorisation, audit. There are times when they are not entirely separate; a possible definition for them can be found in [19]. Access control is an essential mechanism in information security. It deals with what information is authorised to be disclosed to users, based on their identity. The most reliable access control mechanisms are role-based and attribute-based [20].

6. Technical specifications

6.1. How to implement GDPR in software

Based on [21] and data subject's rights from Figure 1, GDPR compliant software should have the following attributes:

- Availability: Users should have unlimited and unrestricted access to their data. The DIY4U platform uses a distributed technology, which ensures higher availability and resilience compared with centralized solutions.
- Completeness: There should be a history of all processing events related to personal data. DLT has the data immutability property, the past cannot be changed, and multiple servers (nodes in the network) have their copy of the same ledger.
- Confidentiality: Data visibility should be limited to concerned parties. In DIY4U, this aspect will be handled using a control access mechanism, that will grant specific access based on the actors' role in the platform (e.g. the Feedstock supplier should not have access to any recipe)
- Correctness: The recorded information must be accurate. After the majority of the DIY4U's network participants reach consensus, data is stored in the blockchain.
- Immutability: History should remain consistent. Immutability is an intrinsic characteristic of DLTs (underlying technology in DIY4U)
- Integrity: The system should provide protection mechanisms against any malicious and unintentional access and changes. In the case of personal data leakage, users should be informed. The DIY4U platform uses a permissioned DLT's, where access is restricted. Also, cryptographic mechanisms are used for data protection.
- Non-repudiation: Intercommunication with data should not be deniable. In blockchain, after a transaction is marked as complete, no party will dispute the transaction's validity.
- Rectification & Erasure: Users should be able to update their data or remove their data. Data erasure and modification is not usually possible in a blockchain. Because of this, personal data will be stored off-chain, not directly on the blockchain; in blockchain there



will remain references to the off-chain data: hashes, public keys, encrypted data, which will help achieve the erasure, without losing the benefits of DLTs solution

- Traceability: There should be a continuous history of the processed data. The DIY4U activities are paired with blockchain transactions; this will grant an irrefutable transaction history for everything that is stored on blockchain.
- Interoperability: users should be able to export their data. The DIY4U will provide this functionality.

6.1.1 Steps for GDPR compliance

In [22] EU provides guidelines for software design and implementation to achieve GDPR requirements compliance. Figure 3 depicts the required steps to reach an agreement.

1. Lawful basis and transparency

The purpose why personal data is being used should be clearly defined and explained in the privacy policy (GDPR Art. 12). The most used reason for personal data processing is user consent. Still, there are also other reasons (GDPR Art. 6), such as public interest exercised by the authorities, for data subject vital interest, for processing child data, for processing contracts where the data subject is a third party. In the Fablab client, users should agree with the privacy policy. This policy should be detailed (why data is collected, for how long), also fablab clients should allow users to export their personal data.

2. Data security

The best practice is to follow data protection by design and by default (GDPR. Art. 25.). We should always think about data protection in the architectural design, software implementation and data processing. To reach this, the DIY4U platform is using a DLT to ensure data protection. In (GDPR Art. 5) it is depicted how personal data processing should be processed: lawfully, fairly and transparently, benefiting from purpose and storage limitation, based on data minimisation, accurately, with integrity and confidentiality. Even if the DIY4U platform is based on a DLT solution, off-chain decentralised storage is used to achieve data storage limitation.

3. Accountability and governance

Each organisation should ensure that their entire organisation is GDPR compliant (from software to employees that access other personal data). If the personal data is shared across multiple organisations, a processing agreement should be signed [23]. There are 3 cases when a data protection officer [23] is required: a public authority does the processing, it's a large scale regular monitoring or large scale for special data categories.

4. Privacy rights

It refers to the right of data subjects to their data: they gave consent, and have the right to update and remove personal data, to question how the processor used it, to complain against any faulty

processor. A more detailed list of data subject rights that should be respected is depicted in Figure. 1.

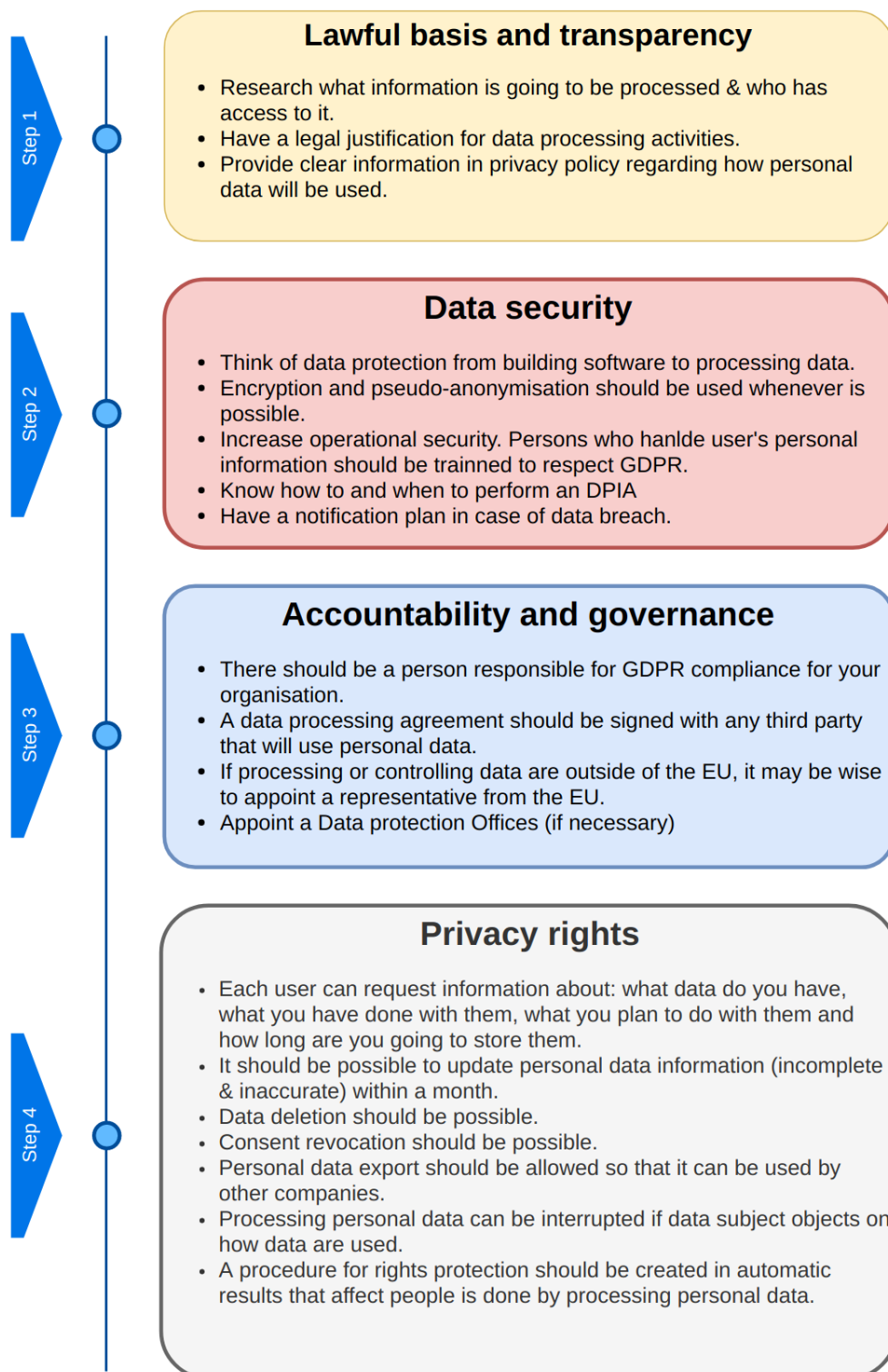


Figure 3. Checklist for GDPR requirements

6.1.2 Demonstrating GDPR compliance

A Data Protection Impact Assessment (DPIA) can be used [24] to prove GDPR compliance. The DPIA is required in the following situations: when new technologies are used, if software tracks people's location or behaviour, if monitoring publicly accessible places on a large scale is done systematically, if sensitive data is processed, if data processing is used for automated decisions about people that could have legal implications, if children's data is processed. As stated in (GDPR Art. 25) having an approved certification as defined in (GDPR Art. 42) will demonstrate the GDPR compliance character. The DIY4U may demonstrate their GDPR compliance by using the DPIA.

6.2 GDPR and blockchain technology

The blockchain technology increases confidence and trust by using a distributed ledger. In [25], there is a comprehensive study of how blockchain can comply with GDPR; in the next lines, are extracted the main reasons that can be deduced from the paper. DIY4U is using a permissioned blockchain to reach a higher level of trust and privacy. Another blockchain characteristic is data immutability (data cannot be updated or deleted) that ensures data protection. Data immutability means that data cannot be changed (updated or deleted); one of the GDPR attributes is the individuals' "right to be forgotten" or update their incorrect or incomplete personal information. It raises the question if blockchain can be GDPR compliant because of its immutability attribute. Also, data minimisation and data purpose of processing data seem to be in antithesis with blockchain technology; once the data is in the blockchain, it is hard to limit the data usage.

In GDPR there is at least a natural or legal person (the data controller) to whom the data subject can address any concerns regarding their rights under these regulations. Blockchain goals are to reach decentralisation, the central data controller may be replaced with a joint-controllership, but there are no clear guidelines regarding this under the law. There is the possibility that some blockchain falls under the "household exemption" (no connection to a professional or commercial activity'); in this case, personal data is not in the scope of GDPR.

Identification in blockchains is usually made based on asymmetric cryptography (pairs of private and public keys); a public key is still personal data (can be seen as pseudo-anonymised data). DLT can use zero-knowledge proof to check statements without revealing the public keys that were involved (e.g. ZCash), other methods like state channels [26] or ring signature [27, 28] can be used to hide the visibility of public keys.

6.2.1 Data responsibility and accountability

The data controller is a single or joint entity (a natural or legal person, authority, or other body) that is responsible for complying with GDPR requirements. It should implement organisational and technical measures to demonstrate that regulatory requirements are fulfilled, should record processing activities (collection, transfers). The controller is allowed to analyse each personal



data processing. There is no consensus on what controller means in a DLT environment because the involvement that is required to be qualified as joint controller is not clearly defined. The joint controller can be anyone who runs a node (because they exert an influence on the processing of personal data) or could be the consortium from a permissioned blockchain (who have greater control over the data). In the DIY4U platform, the GDPR joint controller is the DLT consortium.

6.2.2 Data storage

To comply with GDPR requirements, off-chain storage can be used [27], then linked to the blockchain through hashes. Off-chain storages can be updated or removed, but tied hashes remain on the blockchain; it is not clear from the regulation whether these hashes are still personal data or not. In DIY4U platform data is stored off-chain, using decentralised storage. A solution to achieve storage limitation requirements can be data pruning the old unneeded data, which comes with the disadvantages of losing the full history, but some DLTs regularly do this (e.g. IOTA). Another solution for storage limitations is to explain to the data subject that their data will be used for their transaction and the transactions that will follow. In [29], the ICO recognises that data removal is not possible every time. Instead of this, it can mark data "beyond use" and not access it anymore (but GDPR does not specify that this approach is acceptable for achieving the "right to be forgotten"). Other states admit that erasure of data doesn't necessarily mean the destruction of data. Anonymisation is accepted as a data erasure method. French Data Protection Authority CNIL suggested that the elimination of secret keys together with information from other systems where personal data was processed can be compliant with GDPR. [30].

6.3 GDPR & DIY4U environment

This section describes how DIY4U platform architecture and functionalities are compliant with the EU data protection regulation. In deliverable D3.1 the architecture is defined as using a distributed ledger technology for data storage transport; to be GDPR compliant permissioned blockchain will be used. Also, there are EU recommendations for using encryption whenever possible. The traffic will be done based on HTTPS protocol.

6.3.1 GDPR in DIY4U main services

- Decentralised storage service. The decentralisation features will enhance the data availability by eliminating the central point and also improve the trust in the stored data. Storing data on-chain is costly and will increase the software complexity to achieve GDPR compliance; to ease this, an off-chain storage will be used.
- Identification service. The GDPR main scope is data protection, in a permissioned environment; there is a necessity for identification, authentication and authorisation. Any platform stakeholder will be identified in the system (needed in order to regulate the access control)
- Access control service. A suitable identification and access control mechanism will be used to achieve system privacy and security. The purpose of access control techniques is



to grant preferential access based on the users' roles. Also, access controls can give granular access to user identity so that it can fulfil the minimum information disclosure required by the GDPR. There are communications between parties that will remain private for the other actors of the network. To grant the correct access, the access level in the application should be defined, for each actor. The collection of these access rules outlines the policies. The policy rules are created and updated based on consortium consensus; this will also ensure platform flexibility.

- Data validation service. Validation is mandatory in preserving data accuracy, integrity and consistency. Network participants can reach a consensus regarding any data from the system or outside world.
- Data processing service. Multiple operations can be performed on the Fablab network: adding participants, orders, users, collecting data etc. These services can be reached if the necessary rights to access them are met.
- Ontology service. Ontologies mechanisms are one method to achieve data interoperability; it provides a homogenous standard data format. Reliable data interchange can be used in cross-organisation communication, data export, data storage.
- Cryptographic services. It will expose the needed functionalities for encryption / decryption, signing / signature validation, hashing etc. Some of the cryptographic methods will be available also in the SDK to increase privacy (e.g. encryption / decryption). The cryptographic protection is recommended in the GDPR; it increases privacy and in some cases transforms personal data into pseudo-anonymised data.

6.3.2 DIY4U functionalities

The DIY4U platform will allow multiple operations to run on it with different functionalities, defined in previous deliverables.

Market to customers. Market analytics and offering analytics has the role of analysing the customer profile and needs (based on a quiz) and proposing new types of products or ingredients. To be GDPR compliant, customer data for analytics is collected only with the consent of the customer. The DIY4U must have a granular consent, which means that users should have the possibility to specify for what activity they gave consent (e.g. for product suggestion, but not for overall analytics). If the customer does not agree with analyses, there is no way to use their data (data is encrypted).

Concepts to products. This process is related to managing the formulation rules and other operations that are related to formula (LCA, LCCA computation). The product chemical formulation, product recipe manufacturing process are private information, and the DIY4U platform will control the access to this proprietary information. Only a restricted number of actors are allowed to access them. The formula owner should give consent regarding who can access their formula. Encryption is used to store and transport this private information. Also, the



algorithms for pre-calculating the carbon footprint and product cost should be protected against unauthorised access, otherwise malicious or curious parties can gain knowledge about the formulation.

Customer to cash. Based on customers' preferences expressed in the product customisation quiz, the products are customised. The creation of the customised products is based on a new chemical formula or based on some existent formulas where some parameters are modified. In the entire process, the detergent recipe should not be accessible to the customer, or other actors that did not interact directly with it. Also, the mapping between product features displayed in the quiz and the formulas' parameters should be protected.

There are modules for pre-calculating the carbon footprint (LCA) or pre-calculating the product cost (LCCA); these modules should not have access to the customer's data. Once the customer agrees with the customised product, the process proceeds with placing the order, purchasing the product, and receiving delivery information. The recommendation tools can enhance the customers' experience if they permit the analysis of their historical purchases and preferences.

Distribution of the goods. This process will manage the product delivery process (place, schedule, delivery estimation, (waiting time for in-shop Fablabs). During it, the customer's personal information regarding address, products should be protected against unauthorised access; only the minimum information will be disclosed here. Customers should be able to track and monitor the delivery process regarding their order.

Demand to supply. This module will manage feedstock processes: from requiring feedstock, forecasting the consumption, managing inventory. For forecasting and determining the precise needs of the supplies, the ingredient usage history for specific Fablab should be consulted. A purchase and payment modules should regulate the feedstock orders, packing, reception and payment. This module will contain users' personal information such as: name, address, card number that must be protected from unauthorized access.

6.3.4 GDPR in EFs

Whenever possible, anonymised data should be used; it is an ideal use-case, and sometimes this is not possible to achieve. In these cases, pseudo-anonymised data can be used, but using it will imply that the other articles of the GDPR should be fulfilled. User consent should be requested for any processing on their data. Users should be informed on how their data is being used, for what purposes and for how long. Also, the user should have the possibility to revoke the usage of their data, to request information on how their data has been used, or they can even complain if the processor did faulty processing.

6.3.5 GDPR in LCA & LCCA modules

Life Cycle Analysis (LCA) and Life Cycle Cost Analysis (LCCA) are modules that are able to provide real-time analysis of the product environmental attributes (carbon footprint) and costs. Usually,



these modules will contain only information regarding the product (custom formula, standard formula, feedstock supplies), but not information about the customer. If they are not using personal data information (data about unique individuals), that will not be under the GDPR. If we need to be compliant with CCPA regulation, then these modules should also respect privacy rules. With all these, a data protection mechanism should be enforced to protect intellectual property (the chemical formula) . Also, a security mechanism should be implemented so that different manufacturers do not have access to the competition's detergent recipes in the LCA and LCCA modules.

6.3.5 GDPR in Digital twin

As defined in deliverable D5.8, communication between DIY4U platform and digital twin will be protected based on ECDSA encryption. These modules should provide mechanisms for data privacy: for safeguarding intellectual property, the detergent formula should be encrypted. The raw component of the detergent recipe should not be in plaintext in the digital twin database; processing should be done securely, so that it is not possible to forge access by any unauthorised party. Data minimisation should be applied to respect part of the GDPR requirements. Only needed personal attributes should be forwarded to this module. Also, consent for the simulations that are done should be requested from the user. This module should provide an activity list for action and processing done with the user's data. Furthermore, GDPR defined the right to be forgotten; this module should be able to clean up all information related to a specific user. It should be possible to determine how users' data is used in this module if the users request it (according to the GDPR rules).

6.3.6 Machine learning

As already described in deliverable D2.7 (4.2.2), there are different techniques that preserve privacy in machine learning, most of them based on TEE, multiparty computation and homomorphic encryption.

7. Conclusion

There are benefits to using a decentralisation system: increasing trust, transparency for authorised people, not relying on central authority and resilience to failure. But there are also some technical challenges that we will face during the implementation of the DIY4U platform to implement the GDPR requirements:

- In a decentralised platform, there is no precise specification in the GDPR who can be the controller. We will use the consortium members as join-controllership.
- GDPR right to be forgotten should be respected; there are no commonly agreed methods how this can be achieved in a decentralised platform.



- Storing data in blockchain should be done with caution because of the data immutability property.
- In the case of user personal data breach in any module or part of the platform, the affected persons should be informed. When pseudo-anonymization / encryption is used, it may be hard to remap the correct individual because of missing and not available information in those modules (encryption key, additional details from mapping). As specified in D2.7 (chapter 3.2.2), triggering the review in case of data breach (according to ISO/IEC 27701) may be problematic due to the lack of contact point (that should deliver more information about the problem, consequences, measures taken or planned).

Figure 4 describes how the DIY4U platform meets the GDPR. It is a simplified version of the DIY4U platform, and not all modules and details are visible. The first step to being compliant with the GDPR is to understand the rules, and in previous chapters, it was described what the main requirements for this regulation are. After following the requirements, the next step was designing the architecture with these rules in mind. In the DIY4U platform, personal data is used based on the user's consent; because of this, acquiring consent is a mandatory step. In this paper are also given some small guidelines on what information should be listed in the privacy policy. Based on user consent, the DIY4U platform will use data only for purposes agreed by users and should ensure data privacy against any malicious or curious party that does not have the right to access data. Personal data is collected either when an account is created, or when an order is placed. Data protection is achieved either by personal data anonymisation techniques, or by cryptographic techniques (more details in chapter 5). The DIY4U platform uses a DLT solution that can achieve resilience, secure storage, transparency (for stakeholders with enough access rights over that data) and undeniable history. System functionalities utilise the underlying DLT layer; each functionality module needs to ensure personal data protection during the processing if the processed data cannot be anonymised.

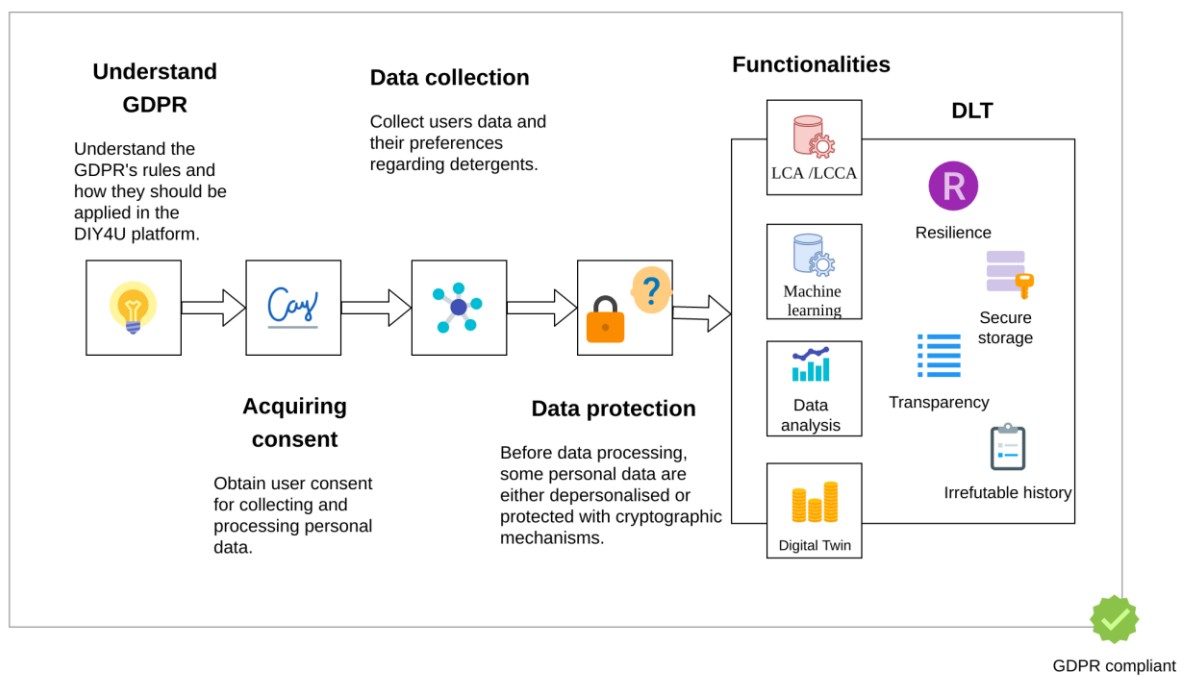


Figure 4. GDPR in DIY4U platform

For maximum privacy, the platform's modules should work only with anonymised information, but the amount of information in the anonymised dataset may be so low that it lacks utility. Decreasing the anonymisation level will also reduce the privacy level, which is unwanted behaviour, so other methods are required. Secure hardware (TEE) ensures a secure computation, but comes with necessities for specific equipment (which will restrict the parties that can use the software), increased costs and also processing restrictions. Cryptographic mechanisms can digest encrypted data (most of them considered pseudo-anonymised) and operate with it, but imply low performances and limited functionality (HE is not a mature technology yet). Building the DIY4U ecosystem will stimulate us to determine the right parameters for the tradeoff between privacy, utility, performance and simplicity (no specialised resources), so that there is a balance between the stakeholders' intents and the users' rights.

References

1. It Governance Privacy Team (2019) EU General Data Protection Regulation (GDPR): An Implementation and Compliance Guide. Itgp
2. Ball KM (2017) African Union Convention on Cyber Security and Personal Data Protection. International Legal Materials 56:164–192
3. Buttarelli G (2016) The EU GDPR as a clarion call for a new global digital gold standard. International Data Privacy Law 6:77–78
4. Gruschka N, Mavroeidis V, Vishi K, Jensen M (2018) Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR. 2018 IEEE International Conference on Big Data (Big Data)



5. (2017) Adequacy decisions. In: European Commission - European Commission. https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en. Accessed 25 May 2020
6. (2017) Adequacy decisions. In: European Commission - European Commission. https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en. Accessed 25 May 2020
7. (2019) EU-U.S. PRIVACY SHIELD. EU GDPR & EU-U.S. Privacy Shield 43–50
8. Sullivan C (2019) EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era. *Computer Law & Security Review* 35:380–397
9. (2019) EU-U.S. PRIVACY SHIELD. EU GDPR & EU-U.S. Privacy Shield 43–50
10. Kosseff J Hacking Cybersecurity Law. *SSRN Electronic Journal*
11. Vokinger KN, Stekhoven DJ, Krauthammer M (2020) Lost in Anonymization — A Data Anonymization Reference Classification Merging Legal and Technical Considerations. *The Journal of Law, Medicine & Ethics* 48:228–231
12. Blockchain concerns - FR. http://www.assemblee-nationale.fr/dyn/15/rapports/cion_fin/l15b1624_rapport-information.pdf
13. Blockchain concerns JP. <https://cointelegraph.com/news/japanese-regulators-discussed-restricting-trade-of-privacy-focused-altcoins-report-says>
14. Rocher L, Hendrickx JM, de Montjoye Y-A (2019) Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 10:3069. <https://doi.org/10.1038/s41467-019-10933-3>
15. Costan Victor And (2016) Intel SGX Explained. *IACR Cryptology ePrint Archive* 2016:1–118
16. Geater J (2015) ARM® TrustZone®. *Trusted Computing for Embedded Systems* 35–45
17. Gentry C, Stanford University. Computer Science Dept (2009) A fully homomorphic encryption scheme
18. EU blockchain report. https://www.europarl.europa.eu/doceo/document/A-8-2018-0407_EN.html
19. Kabay ME (1997) Identification, Authentication and Authorization on the World Wide Web1. An ICSA White Paper
20. Goyal V, Pandey O, Sahai A, Waters B (2006) Attribute-based encryption for fine-grained access control of encrypted data. *Proceedings of the 13th ACM conference on Computer and communications security - CCS '06*
21. Marta Piekarska (Linux Foundation), Michael Lodder (Evernym), Zachary Larson (Economic Space Agency), and Kaliya Young (Identity Woman) (2017) When GDPR becomes real, and Blockchain is no longer Fairy Dust
22. GDPR checklist. <https://gdpr.eu/checklist/>



23. Data protection officer. <https://gdpr.eu/data-protection-officer>
24. DPIA. <https://gdpr.eu/data-protection-impact-assessment-template/>
25. EU Blockchain.
[https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU\(2019\)634445_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf)
26. ETH privacy. <https://blog.ethereum.org/2016/01/15/privacy-on-the-blockchain/>
27. Bacon, Jean and Michels, Johan David and Millard, Christopher and Singh, Jatinder (2018) Blockchain Demystified: a Technical and Legal Introduction to Distributed and Centralized Ledgers. Rich JL \& Tech 25:1
28. Rivest RL, Shamir A, Tauman Y (2001) How to Leak a Secret. Advances in Cryptology — ASIACRYPT 2001 552–565
29. ICO Deleting personal data. https://ico.org.uk/media/for-organisations/documents/1475/deleting_personal_data.pdf
30. CNIL Blockchain. https://www.cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf