This is a postprint version of the following published document:

Sanchez-Reillo, R., Ortega-Fernandez, I., Ponce-Hernandez, W. & Quiros-Sandoval, H. C. (2019). How to implement EU data protection regulation for R&D in biometrics. *Computer Standards & Interfaces*, 61, 89–96.

DOI: [10.1016/j.csi.2018.01.007](https://doi.org/10.1016/j.csi.2018.01.007)

© 2018 Elsevier B.V. All rights reserved.

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

How to implement EU data protection regulation for R&D in biometrics

Raul Sanchez-Reillo*, Ines Ortega-Fernandez, Wendy Ponce-Hernandez, Helga C. Quiros-Sandoval

Carlos III University of Madrid, University Group for Identification Technologies (GUTI), Av. Universidad, 30, Leganes, Madrid 28911, Spain

ARTICLE INFO

Keywords:

Anonymization
Biometrics
Coding
Data Protection
Databases
Privacy
User consent

ABSTRACT

Biometrics R&D has to deal with personal data. From the Universal Declaration of Human Rights, privacy of a human being shall be protected, and this is addressed in different ways in each region of the world. In the case of the European Union, Data Protection Directives, Laws and Regulations have been established, and interpreted in different ways by each European Member State. Such a diversity has pushed the European Union to generate an improved regulation that will be mandatory from May 2018. Biometric R&D shall not only comply with the current Directive, but also has to adapt its work to the new Regulation. This work is intended to describe the situation and provide a recommended procedure when having to acquire personal data. The recommended procedure is illustrated by the implementation of a Biometric Data Acquisition Platform, used to acquire fingerprints from nearly 600 citizens using different sensors.

1. Introduction

Data Protection Regulations are essential to guarantee the privacy of citizens, in particular in current society, where our personal data is extremely exposed to a large number and variety of attacks and abuse. This was initially included in the Universal Declaration of Human Rights, back in 1948, in its Article 12. But this article has evolved differently in different parts of the world. One of the regions where this article has been strongly articulated is the European Union (EU), where in 1995 a Directive was issued as to regulate data protection in all EU Member States (95/46/EC) [1]. Although the ideas behind such Directive were very clear, the Directives were understood and implemented differently in each Member State. The evolution of Information Technology, and the growing number of cases of identity theft, as well as personal data abuse, have led to the EU to develop a new version of this Directive, but in this case not as a Directive, but as a proper Regulation. This regulation, referred as 2016/679 [2], will be of mandatory application from May 25, 2018.

This kind of regulations, and in particular the new one, create major challenges to activities where personal data has to be considered [3]. In particular, Biometrics require to acquire personal data from citizens, as to both, develop the technology and evaluate the results achieved [4]. Such evaluation does not only include performance, but also security and

usability, and not only in prototypes, but also in final products. The relationship between Biometrics and Data Protection has been studied in depth in the last decade [5,6]. But special emphasis has to be applied to R&D activities, where biometric and other personal data has to be acquired, stored and kept. In addition, the transmission of such personal data to third parties is under major controversy, which also make selling or acquiring databases a serious legal problem.

The present work presents the problem of privacy and data protection when related to activities requiring the recognition of human beings. In order to better understand the problem, the current data protection directive from the EU will be studied in Section 3, introducing the most important concepts and requirements. From such a knowledge, the application of such directive to Biometrics R&D will be shown in Section 4. This section will also include the proposal of a recommended procedure, which will be illustrated by explaining the particular case of a major evaluation of fingerprint technology. The new coming scenario, created by the new EU Regulation, will be analysed in Section 5, leading to proposed future work and conclusions.

2. Privacy vs. recognition

Human Recognition can be performed in three different ways, or any combination of those three:

* Corresponding authors.

Email addresses: rsreillo@ing.uc3m.es (R. Sanchez-Reillo); inortega@ing.uc3m.es (I. Ortega-Fernandez); wponce@ing.uc3m.es (W. Ponce-Hernandez); hquiros@ing.uc3m.es (H.C. Quiros-Sandoval)

- By **what the user knows** (e.g. passwords). The advantages of this method are based on the lack of additional infrastructure, plus the possibility of changing such a knowledge, and therefore updating the credential. On the other side, knowledge can be forgotten, or even easily copied by eavesdropping.
- By **what the user has** (e.g. cards). Again the token used for recognition can be changed, although that will require some additional cost. But also the token can be stolen or lost.
- By **what the user is** (e.g. Biometrics). The credentials are expected to be unique (if the discriminative power of the algorithm is high), but they cannot be changed. If Biometrics is well developed and deployed, it can be a comfortable way to identify the individual. But credentials are usually publicly available (e.g. by taking a photo), and they may be spoofed. Therefore Presentation Attack Detection (PAD) mechanisms should be in place.

In any of these cases, the credentials are linked to the real identity of the subject within the system. As identity is a piece of personal data, both the identity and the link shall be protected, as to preserve citizens' privacy. Therefore many applications involve the acquisition and use of Personal Data (i.e. administrative data), for several reasons:

- For registering the participation of the citizen in the system
- To avoid duplicated entries in the system
- For future communications between the system and the citizen
- For allowing the citizen to claim his rights

Obviously, any system using a Human Recognition technique shall integrate a set of security mechanisms so as to provide the best conditions to preserve the privacy of the system. Therefore the evaluation of the security of the recognition procedure [7] should be mandatory.

In addition, it has to be considered that Biometric Data is a piece of personal data, which is physically linked to the user. Currently it is considered as of the same level as administrative data (e.g. name or address). In addition, for some modalities, the link between the data and the person can be direct (e.g. face recognition or even signatures). Such direct link may be considered as an evidence by a trial court.

As there is a direct relationship between Biometrics and Personal Data, there is a huge concern about privacy:

- How can the citizen be sure that his personal data is not used outside the claimed purpose?
- How feasible is for the system provider to use such data for other means?
- Is it possible for the user to belong to a system without providing his administrative data?

So in few words, if we need to recognize human beings, but handling such data may attack his/her privacy, how can this be handled? This question led to the definition of data protection laws and directives, such as the ones above mentioned.

3. Personal data protection: the EU directive

The present section will describe the current EU Directive, as a background to understand the changes that the new regulation will bring.

3.1. Background

Article 12 of the Universal Declaration of Human Rights (10 December 1948) reads as: *"No one shall be subjected to arbitrary interference*

with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks". In addition, some countries also declare laws that limit the use of computer science to guarantee the honor and personal privacy of citizens.

With that in mind, some countries developed laws and acts regarding the automatic process of personal data (e.g. LORTAD in Spain in 1992 [8]). These laws should always be sustained by a regulation stating the rules for the treatment of the automated files using personal data.

In 1995, the EU approved the 95/46/EC Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1]. It shall be clear that such directive only applies when data processing is automated or when data is stored in a structured file so that access to such data is simplified.

The Directive is not a law, but a strong recommendation to Member States. Therefore such Directive was implemented in laws and regulations in each of the countries of the EU. Depending on the Member State, the level of coverage of the Directive changed, achieving laws of different strength levels. Some researchers from some Member States provided strategies to support the local Data Protection law with security mechanisms [9].

In any case, there are two major principles behind the Directive:

- The citizen is entitled to preserve the full control on his personal data: who is collecting them, for what, and where are they going
- The data collector shall implement a relevant policy to preserve the citizen rights, declaring the filing of that data and the person responsible for keeping the policy.

3.2. Concepts and requirements

Going more in depth on the Directive, the following concepts and requirements are defined:

- The citizen is the one **deciding which part of his personal data is to be provided**.
- The citizen has the **right to declare his consent** towards the data collection act.
- The citizen has the **right to be informed** about **who** is collecting the data, **what** is the reason for collecting such data and **which** processes are going to be applied to his data.
- The citizen has the **right to deny** the collection of his personal data.

The Directive also defines the concept of Data Quality, including rules such as:

- Data should be relevant, adequate and non-excessive.
- Data cannot be used for a purpose different to the one declared when being collected.
- Data shall be maintained exact and accurate.
- Data shall be cancelled when no longer needed.

The Information Right that the citizen has during collection of his data, allows to him to ask about the existence, finality and recipients of the data, the optionality and mandatory character of each data collection, if there are some consequences in not providing some of the data, and how he can execute his rights. Additionally he has to know the identity and address of the file responsible.

In order to allow the traceability of such Information Right, a user consent shall be signed, being extremely important that such user consent shall be unambiguous and revocable.

Certain data may be subject to a higher level of protection. Typically 3 levels are defined, from a basic protection to the strongest one. For

example:

- Level 1: Administrative data (including Biometrics)
- Level 2: Health data
- Level 3: Political, religious and ideological data, race, sexual orientation, etc.

The Security of the data collected is implicit within the Directive, but there is not an explicit regulation. The only statement is that the file responsible is required to adopt all technical and organisational means to secure the data and preserve privacy, and the instruction that without an active and relevant security policy, data shall not be collected

The file responsible and any person taking part in the collection and/or processing processes, shall keep professional secret about the data and the organization of such data, even after the termination of relationship with the file responsible.

Data can only be transferred to a third party after the explicit consent of the citizen.

The Access Rights are property of the citizen, who can claim them with no compensation required. In order to reduce citizens' abuse, they can only claim for his access rights once every 12 months.

The Rectification and Cancellation Rights shall be applied within 10 days of the receipt of the request. Cancellation shall derive on the blocking to all data related to that user. If data have been given to a third party, this request is applicable to the third party.

3.3. Coding vs. anonymization

A very important concept is the definition of coded data and anonymised data. Coded data is data that is not directly linked to a subject, as the link has been substituted by a code that can be reversed under highly secure conditions (typically having to contact the file responsible).

On the other hand, anonymised data is data which link with the subject is permanently broken, not being possible to re-establish such link through reasonable means.

Data may not be anonymised always. For example, in the case of biometrics, the face image will always have a direct link with the user, easily found by any other person. Some other biometric modality could be anonymised (e.g. fingerprints), as it is not trivial for any person to detect the identity of a person by the inspection of the fingerprint image.

These two concepts have led to further controversy, as it is not clear whether anonymised data could be transferred to a third party or not. Another concern is if the file responsible should have the explicit consent of the citizen in order to anonymise his data.

4. How to apply the EU directive in biometrics R&D

In order to determine when and how to apply the Directive in the field of Biometrics R&D, it is necessary to study the requirements first. Such technical requirements will derive into a set of privacy preserving actions. This section will cover such study, and will result in a set of recommended actions, which will also be illustrated in a real case.

4.1. R&D requirements

Researching in Biometrics imply the need of obtaining several biometric samples for several users. As it has been said, biometric data is considered a piece of personal data, and that will derive on a set of requirements, as such data will have to be protected, and, at the same time, that data will have to be linked to other personal data, such as administrative data. A set of 9 requirements can be defined, as detailed in Table 1.

Table 1
List of data requirements for biometrics R&D.

Req. no.	Requirement	Reason
1	Collect identity data and contact information from the user	R&D in Biometrics require a deep analysis of the intra-class distribution, which means the need of collecting data from the same subject in several sessions, different use cases, different attitudes, etc. Therefore that subject will have to be contacted for attending new collection sessions.
2	Deny the existence of duplicated entries	In addition also the inter-class distribution, shall be analysed. This means to collect data from several subjects, with the same parameters used for the intra-class distribution. While performing this, particular care has to be taken in order not to mislabelled samples, and to avoid a subject to be collected with two different labels.
3	Collect additional data for further processing. Some of this additional data may be subject for a higher level of protection	Another aspect of interest is to analyse the influence of certain subject conditions, such as relevant physical or mental condition, relevant knowledge background, relevant habits, etc. That may increase the amount of personal data to be acquired from the user.
4	Keep data linked to allow identification during long periods	It is essential to study biometric data stability, and for this it is needed to acquire data in multiple sessions separated a significant time gap. This is also applicable when analysing the aging effect. Therefore the 1st requirement is enhanced in duration of the maintenance of the link between the data under study and the contact information.
5	Obtain the permission of the subject for allowing research on Presentation Attacks	When researching on the robustness against Presentation Attacks, some experts shall produce such attacks, allowing them to deeply examine the subject's data. This is a very sensible action which may derive on some subjects to deny its participation under such circumstances.
6	Analyse all ethical implications on the execution of Presentation Attacks (e.g.: deny the process if there is a relationship between attacker and subject)	The person producing the attacks may achieve a high level of accuracy in spoofing the identity of the real subject under study. This may derive in the potentiality of the attacker of using such accuracy outside of the evaluation.
7	Keep data for long periods	Most of the times it is needed to use the same data for several iterations of the development phase, and this may imply longer periods of time for keeping the personal data, and even the possibility to contact some of the subjects for further study.
8	Obtain data relevant to your study from a third party	As the acquisition of a DB is a time and money consuming process, sometimes it may be better to acquire a previously captured DB which may fit your study.
9	Establish conditions and regulation of data transfer to third parties	Once a new biometric DB has been captured, the interest in distributing it may arise, either by selling or by donation.

In a biometric R&D project some or all of these requirements may appear. The project will have to determine which of these requirements apply, and act consequently following the recommendations from the next subsections.

4.2. Privacy preserving actions

Each of those 9 requirements derive a set of privacy preserving actions. These preserving actions can be shown in Table 2. This preserv-

Table 2
List of data requirements for biometrics R&D.

Req. no.	Requirement	Preserving actions
1	Collect identity data and contact information from the user	<ul style="list-style-type: none"> - Keep a subject database with all its contact information and further administrative data - Assign an ID code to each subject in that database. Such code shall be unique to the experiment, not derived from any external ID number (e.g. never use the passport number as such a code) - Use only such ID code for referencing the data under study. - Store the data under study in a separate database, and if possible, in different computers using different technologies.
2	Deny the existence of duplicated entries	<ul style="list-style-type: none"> - Keep the link table between both databases in a highly secured way - Implement a functionality for the data collector, to ask the database for the ID code of the subject being collected, but submitting other piece of personal data (such as name or surname, or the passport number). - Such functionality shall only return the ID code, or the confirmation that such subject has not been registered in the system. - If possible, perform also a low-level biometric authentication for double checking.
3	Collect additional data for further processing. Some of this additional data may be subject for a higher level of protection	<ul style="list-style-type: none"> - Analyse the real need of the data to be collected - Check the level of protection of each piece of data - It may be good to work around the data request, so not to increase the level of protection and still have the needed data (e.g. in the declaration of potential illnesses that may impact the data collection, a yes/no answer may be enough, instead of storing the specific illness, which will raise the level of protection).
4	Keep data linked to allow identification during long periods	<ul style="list-style-type: none"> - A complete and accurate process to allow the cancellation of data shall be implemented. - In case of cancellation, all data from that user shall be deleted, not only the link. - Determine at the very beginning the minimum duration of the data in the informed consent.
5	Obtain the permission of the subject for allowing research on Presentation Attacks	<ul style="list-style-type: none"> - Inform the subject on this kind of actions. - It may be recommended to state this as an optional consent item.
6	Analyse all ethical implications on the execution of Presentation Attacks (e.g.: deny the process if there is a relationship between attacker and subject)	<ul style="list-style-type: none"> - Ensure ethical and legal agreement with attackers.

Table 2 (Continued)

Req. no.	Requirement	Preserving actions
7	Keep data for long periods	<ul style="list-style-type: none"> - Consider the possibility of anonymising the collected data as soon as possible. - Once all data collection has been done, it may be anonymised by breaking the link and (recommended) changing the user codes.
8	Obtain data relevant to your study from a third party	<ul style="list-style-type: none"> - Study the legal implications of acquiring such DB, and how to keep the third party requested conditions.
9	Establish conditions and regulation of data transfer to third parties	<ul style="list-style-type: none"> - Enforce the third party not to distribute the DB, and to react towards all claims for cancellation - Establish the legal conditions for a potential violation of this request. Consider that such conditions shall be at least as hard as the ones your legislation may impose you (e.g. monetary fines).

ing actions are considered to be the minimum acceptable ones, being possible to enhance them if a previously risk analysis imply further protection.

4.3. Recommended procedure

With all these requirements, a procedure has been designed to comply with them. This section provides such procedure which is based on 11 steps:

- 1) Analyse the data needed and its level of protection. This is stated in the Directive, as you should fit the personal data captured to your needs.
- 2) Create a remote (not available to the external world) database to store subject personal data (without samples). By keeping the database isolated from the external world, the risk of non-authorized access to the database is minimized. Obviously the room where that database is to be located shall have the logical and physical access conditions to minimize a direct attack to the infrastructure.
- 3) Create web services needed for the study. Web services can be defined as “an application or data source that is accessible via a standard web protocol (HTTP or HTTPS). Unlike web applications, web services are designed to communicate with other programs, rather than directly with users” [10]. Those other programs may be running in systems connected either through a public network, or through private ones. In this work, the use of private network is considered, as to not ease external attacks. By using web services, direct access to the database is forbidden and only the authorized actions are allowed. The actions to be developed are:
 - i) Insert a subject.
 - ii) Delete a subject by ID.
 - iii) Ask for the subject ID by providing a very reduced set of personal data.
 - iv) Get discriminative subject data for collateral data impact (e.g. age group).
 - v) Get statistics of the DB (e.g. number of subjects/samples, age/gender distribution, etc.).
- 4) Create a local application to enroll subjects in the database. By using local applications instead of web-based ones, the risk of the external world to tamper with the communication protocol or the data is also minimized. Even if access is gained through a VPN, as only the local application is authorized, a security breach will have a nearly null impact.

- 5) Create a local database for storing subjects' samples. This database is separated from the one holding the registry information, because that will be the database to be accessed by processing applications (e.g. evaluation application). Such processing application shall not have access to the registry information. In case a security breach is found in those processing applications, at least the administrative information from the test crew is kept secret.
- 6) Create a local application to collect subjects' samples. Again, the use of local applications (i.e. not remote ones) is encouraged.
- 7) Create a local application to process samples and obtain statistics. Even though this application is expected to be executed only within the laboratory restricted area, its number of executions, as well as the very long execution times, require that security is implemented with extreme care.
- 8) Create an informed consent form. Again this requirement is coming directly from the Directive. But as it will be seen in the following section, due to the new Data Protection Regulation, this consent form shall be written in a really serious way to avoid ambiguity and "by default" acceptance.
- 9) Declare all the procedure, the file structure, and the security mechanisms to the relevant Data Protection Agency. This will depend on the particular regulation of the Member State where this is being executed. With the new Regulation, this step is relaxed in terms of timing constraints, but strengthened in terms of procedure and mechanisms documentation.
- 10) Start data collection and processing.
- 11) Termination, either by deleting or anonymizing the data.

4.4. Illustration by example

The above recommended procedure has been applied in a medium-high scale evaluation of fingerprint systems. In particular, the evaluation was called "Evaluation of Interoperability of Fingerprint Sensors and Algorithms", and a partial public report is available in [11]. Data coming out of this experiment can be found in [12]. This kind of data collection can be found in other works and projects (e.g. EU FP7 INGRESS project, which can be followed in <http://www.ingress-project.eu/>, in which they collect large fingerprint databases). The main characteristics of the example shown in this work (i.e. the one in [11]) are:

- 4 semiconductor flat sensors
- 5 algorithms (NBIS [13] + 4 commercial ones)
- 589 subjects, 6 fingers/subject, 6 samples/finger
- 2 visits:
 - 1st visit: Enrolment + 1st acquisition session
 - 2nd visit: 2nd acquisition session
- Soft ground truth mechanism to ensure proper data collection (deny duplicated entries and finger exchange)
- Operator controlled data collection
- Compensation to subjects: 2 cinema tickets once the collection is finished

The 1st step was to analyse the data needed, resulting in the following:

- ID information:
 - Name and Surname
 - ID document number
- Contact information:
 - Phone
 - E-mail

- Demographics:
 - Gender
 - Age group
 - Habituation to IT
 - Habituation to Biometrics
- User condition:
 - Lack of fingers
 - Skin issues that may impact acquisition (YES / NO – no further detail)

All data is considered of Level 1, i.e. Basic Level Protection.

The 2nd step was to create a remote database to store subject personal data (without samples). A MySQL server has been used, with authenticated SSL connection in a computer without access to the external world. The SQL script for the creation of the database is given in [12].

Once the database was created, the 3rd step was to create all web services needed for the study, using PHP scripts executed in the same computer where the MySQL database is installed. For such a task, a web server in that same computer was installed, with only authenticated SSL connection available.

The 4th step was to create a local application to register subjects in the database. This was done in Visual Studio, using C# as a programming language. Such application is executed in a computer without remote connection to the external world.

With the administrative registry application created, an additional database was created for storing subjects' samples. Such database, for operational purposes, was created as a collection of files with particular file naming based on collected data and user code (also detailed in [12]). This was step 5. Fig. 1 shows the architecture of the whole acquisition platform, where the 2 servers are shown: the MySQL database for the users' registry, and the file server for the coded samples acquired. These 2 servers are implemented in different computers with different server systems and even operating systems. They are only connected internally to the Acquisition Points, which have the biometric sensors installed.

The 6th step was to create the local application for data collection. Fig. 2 shows a screenshot of one of the main forms during the acquisition process, where the user is identified by his code, the finger to be presented is shown, as well as the attempt and sample of the acquisition. Also the photo of the sensor to be used is provided, as well as the fingerprint image acquired. These two images are hidden in Fig. 2 due to confidentiality and data protection restrictions.

The 7th step was to create a local application to process samples and obtain statistics. Scores for mated and non-mated comparisons are provided in CSV format, to further process data in Matlab or Excel. The CSV files provide only the comparison score, and the name of the coded fingerprint files compared (e.g. a mated comparison to contribute to the intra-class distribution) or to different fingers or users (e.g. a non-mated comparison to contribute to the inter-class distribution). An excerpt of each CSV file can be found in [12].

The step number 8 was the creation of the informed consent form, describing the process and the target of evaluation, identifying the File and the File Responsible, and providing information about how the citizen can claim for his Rights. The form was provided to the user with its data already filled in and only requiring his signature. One copy was given to the user for future reference, and the other signed one is kept for our records and stored securely.

Step number 9 was to declare all the procedure, the file structure, and the security mechanisms to the Spanish Data Protection Agency, which was done through the University where the authors work, as it is authorised as a relay for the Spanish Data Protection Agency.

With all the previous steps finished we started the data collection and processing, which took 4 months of data collection, plus

3 additional

Fig. 1. Acquisition platform architecture.

Fig. 2. Screenshot of the acquisition application.

months for data processing. The result of this work can be partially examined in [11].

The last step is the Termination, which is based on the anonymization of data, by randomizing the ID number of each user. The randomization also re-built the database by changing the date and time of the creation of the samples (i.e. user with number X should not always be recorded before the samples from user number X + 1). Further clues on anonymization are given in [14].

5. The new EU regulation: general data protection regulation (GDPR)

Regulation 2016/679 [2] has been launched to derogate the 95/46/EC Directive [1], improving, updating and extending its scope. The most important change is that it is no longer a Directive, but a Regulation, and will be mandatory to all Member States in May 25, 2018. It is currently known as the General Data Protection Regulation, or its much more used acronym GDPR.

The Regulation takes the most protective interpretation of the Directive, and updates some parts to both, improve the protection of the Eu-

European citizen, and to also ease the accomplishment of some of the bureaucracy.

Some Member States are already quite well aligned with this new Regulation, as their interpretation of the Directive was also high protective. But there are other Member States which will have to re-do all their current laws, regulation, and procedures.

The most important changes that the new Regulation introduces are:

- It becomes applicable to non-EU established institutions, if data is related to EU citizens.
- Creates new categories for personal data.
- Introduces the concept of pseudonym, as an intermediate state between personal data and anonymization, but no many details are given, so this is creating some controversy. For the time being, it is considered as a synonym of coded data.
- It regulates additional principles, such as Transparency, Proactive responsibility, Privacy by Design and Privacy Impact Analysis (PIA).
- Companies shall/should have a Data Protection Delegate.
- File declaration is no longer required in advance, but an internal registry shall be kept.
- If a Security Breach or Incident is found, it shall be declared immediately to the Data Protection Agency. If the breach or incident impact subjects, they shall also be informed.
- The adhesion to Codes of Conduct is recommended, as well as the establishment of certification mechanisms.
- International data transfer is considered within the new Regulation, although it is not highly detailed.
- An European Data Protection Council is created.
- Additional requirements about informing the citizen before collecting data are introduced. The most important one is that silence is no longer a way of providing consent, i.e. the consent shall always be explicit.
- It defines additional Rights, such as Deletion of Data Right, Right to Forget, Process Limitation Right, Data Portability Right, and the right of not being subject of a decision taken only by automatized means.
- It is adapted to the new technologies, such as Big Data and Internet of Things.
- One of the issues that has been most criticized is the lack of a clear definition of Video-surveillance.

Unfortunately there are still many open concepts, mechanisms and interpretations. These gaps will have to be covered by Member State laws and regulations, which may again create different environments in different Member States, although they are considered to bring a minimal impact among EU Member States.

5.1. Practical impact of the GDPR in R&D projects

Even though there are still some gaps to be solved, the current definition of the GDPR shows that the conservative procedure proposed in clause 4.3 is fully compliant with GDPR. But there are some additional comments to be made.

One of the most important topics in order to avoid further problems, is to force the use of pseudonyms (i.e. what has been called coded data along this work), and whenever possible, and as soon as possible, to anonymise such data. If data used for the study are anonymised, then it could be used without further concerns. But it has to be clear that certain biometric modalities may not allow a full anonymization. In order to better understand this, let's recall the definition of anonymized data: "... data which link with the subject is permanently broken, not being possible to re-establish such link **through reasonable means**". This means that, for example, if the project handles iris recognition data, it may be considered nearly impossible to perform such re-establishment of the link, as

for that, you should be able to access the iris recognition database of the whole population where the database was captured.

In practical terms, this is impossible for iris, and for many other biometric modalities. Even for fingerprint, although this modality is under continuous discussion in this regard. The people defending this possibility show examples like the following. If a database has been captured by a certain university using people not only from the university but from other external people, in order to re-establish the data links you should be able to access the fingerprint database of all citizens within the city that university is located. In some countries, this kind of database does not exist, and in those countries where the database exists, it is under the custody of the Police, and therefore not being able to be accessed **through reasonable means**.

Unfortunately, for some biometric modalities this is not the case. Practical examples are face recognition and handwritten signature. When the database handles raw data from these modalities, by simple overlooking such data, many people can be recognized, at least within the environment where the database has been acquired. This overrides the concept of anonymity, as defined by the GDPR. In other words, by taking a look at one of the facial samples (i.e. the face photograph), people from that same university will know who the person is, and therefore the link between the ID assigned and the person. In the case of handwritten signature, and in particular in certain societies where the signature is just the written name of the user, you can determine the identity of the person easily by looking at the signature graph in the case of static signature, or reconstructing the signature graph by reading the temporal signals in a dynamic signature record. In these cases, it is recommended to keep raw data secured, and work with non-reversible processed data. This presents the inconvenience of having data which is algorithm dependent.

One of the good news about GDPR is that is no longer needed to seek for the Data Acquisition Approval before starting the data collection. This means an important save in time as the document submission, plus the response from the relevant Data Protection Agency, typically took quite a long time. So now you could start the data collection as soon as you're ready. This great advantage has a major drawback for the institution where the collection and personal data handling is going to take place. The reason why there is no need to seek for the approval of the Data Protection Agency in advance, is because the responsibility is pushed to each institution, who has to establish a whole set of policies in accordance to GDPR to ensure the protection of personal data. Most of the institutions will have to do this in one way or the other as, at least, they will have to handle the personal data of their employees. The institution will have to audit the accomplishment of such policies at a regular basis, and therefore the group that will acquire, store and process personal data, will have to comply with those internal policies, and suffer regular audits. Regarding the recommended procedure, step 9 will not be longer needed, although it may be substituted for any kind of internal requirements within the institution.

Last, but not least, a major issue has to be handled. In most R&D projects, data collection and/or data processing is handled by several partners. The big question is how to handle this cooperation when personal data is involved. The first recommendation to be made is to minimize the number of partners handling such data to the strictly needed for the project aims. This is always a good practice for limiting the risks. Then considering those partners that need to share data, we have to analyse the legislation that apply to them.

Let's consider the simplest case: when all partners belong to the same Member State. In such a case, the same legislation applies to all of them, and they all have to commit to it. The only additional need for the project is to establish a Partnership Agreement, in which the partners involved are clearly identified, as well as the people responsible for the management of the data and the control of the data protection policies. Those same policies will force that the data exchange will be done in the

most secured possible way (e.g. ciphered and authenticated using cryptographic algorithms) and, if possible, with data fully anonymised.

The good news comes again with the GDPR for the case of those partners belonging to the European Union. In such a case, all partners are covered by GDPR, so this case becomes the same one as if the partners belong to the same Member State.

But when some partners are outside of the European Union, then two possibilities may appear [15]. The first (and most convenient one) is if the third country is either adhered to the EU GDPR, or its applicable legislation is equivalent to it. In such a case, the previous case is to be followed, except that the Partnership Agreement shall also include provisions for the possibility of a change in the third-country legislation, as well as evidences of the equivalency between the third-country legislation and the GDPR.

The second possibility, when such equivalence in legislation cannot be found, can only be solved by the use of contractual clauses and binding corporate rules. Such clauses shall replicate the requirements of the GDPR, and to provide practical consequences (e.g. high monetary fines) in the case of the third-country partner violating the agreement. The use of certifications (if available) as to proof the data protection capability of the third-country partner is allowed by the GDPR, which can help in writing those contractual clauses. Anyway, the strongest recommendation in case of cross-border data transfer is to do it always in a fully anonymized way, not transferring ever data which may have neither a direct or indirect link to the specific individual.

6. Future steps and conclusions

Currently all Member States are working in adapting their laws and regulations to the current EU Regulation, but there is still a long way to go, as to get final documents. For the time being, some organizations (e.g. [16]) are starting to divulgate the new EU Regulation and assessing on the potential impact that may have.

In addition each economic sector is trying to influence in the writing of the new laws, as to minimize the impact of its companies and institutions. It is expected that some Member States will have to modify their laws only lightly, as their previous interpretation was extremely conservative. But some others will have to re-define them completely.

This work has taken such a conservative approach, and has defined a procedure to acquire Biometrics DBs, in order to either work on R&D or in performing independent evaluations. This procedure is also in compliance with the GDPR, and authors have also provided an analysis of the impact of the GDPR in those R&D project involving several partners.

Anyway, the Biometrics sector shall also be aware of future events, when trying to solve those aspects of the new Regulation that are not well or fully defined.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.csi.2018.01.007.

References

- [1] European Parliament, Directive 95/46/EC 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, Off. J. Eur. Communities (1995), [Online]. Available: <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046>, [Accessed: 08-Aug-2017].
- [2] European Parliament, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data," 2016. [Online]. Available: <http://eur-lex.europa.eu/eli/reg/2016/679/oj>, [Accessed: 08-August-2017].
- [3] F. Massacci, M. Prest, N. Zannone, Using a security requirements engineering methodology in practice: The compliance with the Italian data protection legislation, *Comput. Stand. Interfaces* 27 (June (5)) (2005) 445–455.
- [4] R. Sanchez-Reillo, R. Alonso-Moreno, B. Fernandez-Saavedra, Y.-B. Kwon, Standardised system for automatic remote evaluation of biometric algorithms, *Comput. Stand. Interfaces* 34 (5) (2012).
- [5] E. Kindt, Biometric applications and the data protection legislation the legal review and the proportionality test, *DuD* 31 (March) (2007) 166–170, August 2003.
- [6] E. Kindt, Best practices for privacy and data protection for the processing of biometric data, *Security and Privacy in Biometrics*, London: Springer, London, 2013339–367.
- [7] A.J. Palmer, Criteria to evaluate automated personal identification mechanisms, *Comput. Secur.* 27 (December (7–8)) (2008) 260–284.
- [8] Jefatura. del Estado, Ley Orgánica 5/1992, de 19 de octubre, de regulación del tratamiento automatizado de los datos de carácter personal, *Boe* (1992), [Online]. Available: <https://www.boe.es/buscar/doc.php?id=BOE-A-1992-24189>, [Accessed: 08-August-2017].
- [9] C. Pounder, Security and the new data protection law, *Comput. Secur.* 17 (January (2)) (1998) 124–128.
- [10] TechTerms, "Web service definition," *Tech Terms Comput. Dictionary*. [Online]. Available: https://techterms.com/definition/web_service, [Accessed: 28-December-2017].
- [11] B. Fernandez-Saavedra and R. Sanchez-Reillo, "Performance testing evaluation report of results," 2015. [Online]. Available: http://idtestinglab.uc3m.es/data/_uploaded/Publications/PUBLICREPORT_589_users_Fingerprint_v1_1_release.pdf, [Accessed: 08-August-2017].
- [12] R. Sanchez-Reillo, I. Ortega-Fernandez, W. Ponce-Hernandez, H.C. Quiros-Sandoval, Data for a fingerprint biometric evaluation compliant to EU data protection regulation, *Data Br* (2018), no. submitted.
- [13] NIST, NIST biometric image software, *NIST Biometr. Image Softw.* (2015), [Online]. Available <http://www.nist.gov/itl/iad/ig/nbis.cfm>, [Accessed: 08-August-2017].
- [14] M. Jänel, Decision support for releasing anonymised data, *Comput. Secur.* 46 (October) (2014) 48–61.
- [15] A. Myers, "Top 10 operational impacts of the GDPR: part 4 -k cross-border data transfers," 2017. [Online]. Available: <https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-4-cross-border-data-transfers/>, [Accessed: 14-December-2017].
- [16] ESYS, "El Reglamento General de Protección de Datos de la UE: Una perspectiva empresarial."