Privacy and Data Protection: its application to R&D in Identification Technologies

Raul Sanchez-Reillo

University Group of Identification Technologies – Carlos III University of Madrid
Avda. Universidad 30, 28911 Leganes, Spain
rsreillo@ing.uc3m.es
Recruitment open till March 15!!

- **UNIKENT (UK)**
  - ESR 1: Usability performance assessment of mobile touch-screen behavioural biometrics
  - ESR 2: Continuous and instantaneous authentication using swipe interaction

- **UNIROMA3 (IT)**
  - ESR 3: Template protection in biometric-based mobile scenarios
  - ESR 4: Multibiometric architectures and privacy in a mobile environment

- **WUT (PL)**
  - ESR 5: Countermeasure algorithms against subterfuge in mobile biometric systems
  - ESR 6: Making mobile biometrics more reliable

- **OVGU (DE)**
  - ESR 7: Privacy in nomadic cross-system mobile biometrics
  - ESR 8: User-centric and self-determined privacy management in mobile biometrics

- **UC3M (ES)**
  - ESR 9: Vulnerability assessment in the use of biometrics in unsupervised environments
  - ESR 10: Traceable and comparable evaluation methodology for the usability of biometric systems
• De-identification: when to be used and when it is not possible
• Close relationship between privacy and data needed for human recognition
• Privacy and Data Protection: EU Directive
• How to apply Data Protection Directives R&D in Identification Technologies
  o R&D Requirements
  o Privacy Requirements
  o Recommended Procedure: an example
• The new EU Regulation on Data Protection
De-identification: when to be used and when it is not possible
De-identification

• “Process used to prevent a person’s identity from being connected with information” [Wikipedia]
  o Used widely in fields such as Medicine and Sociological Studies

• It is an ethically-mandatory process in order to avoid attacks against citizens’ privacy
  o By exploiting linking data with a certain user
  o Typical attacks:
    ▪ Marketing campaigns (commercial, political, etc.)
    ▪ Risk analysis (mortgages, employment, etc.)
    ▪ Robbery and/or Identity Theft
  o Typical attackers:
    ▪ Commercial companies
    ▪ Political parties
    ▪ Banks
    ▪ Insurance companies
    ▪ Thieves
    ▪ ... or even the citizen himself!

Use Cases
Data Protection Directive
Privacy vs Recognition
Directive in R&D
The new EU DP Directive
**De-identification: Use cases**

- Theoretically de-identification shall be applied always
  - As soon as data is collected, only that data that is really necessary is stored/processed, removing the rest, including all administrative data

- But sometimes this may not be possible or recommended:
  - When data from the same user has to be collected in different sessions
  - When the working field is focused on recognizing individuals
  - When forensic studies shall be carried out, in particular, within a legal background

- In Biometrics all of them may apply:
  - In many private applications, de-identification may be possible by using pseudonyms and keeping the real correspondence with the citizen identity secretly and securely stored
  - Biometric samples and references may also be used for identity theft if they are obtained by an attacker. Two ways to prevent this:
    - Hashing biometric references (i.e. Template Protection)
    - Working on certified biometric capture device that does not allow Presentation Attacks
  - Whenever it is foreseen the need of a future forensic use of that data, the link shall be disclosed and also the raw data obtained
Close relationship between privacy and data needed for human recognition
Privacy vs. Recognition

• Human Recognition can be performed in three ways (or any combination):
  o What the user knows
    ▪ Knowledge can be changed 😊😊
    ▪ Knowledge is not unique and can be copied 😞😞
  o What the user has
    ▪ The token can be changed 😊
    ▪ The token may not be unique and may be stolen 😞
  o What the user is (i.e. Biometrics)
    ▪ Credentials “are unique” 😊😊
    ▪ Credentials may be public 😞 and spoofed 😞
    ▪ Raw credentials cannot be changed 😞😞

• In any case, there is a moment where the personal data of the user shall be linked with the system
  o At least to achieve the linking of the person to a specific group (e.g. belonging to a sports club)
Privacy vs. Recognition

• The system shall acquire Personal Data (i.e. administrative data)
  o For registering his participation in the system
  o To avoid duplicated entries in the system
  o For future communications between the system and the user
  o For allowing the user to claim his rights

• Biometric Data is a piece of personal data
  o Currently of the same level as administrative data
  o For some modalities, the link between the data and the person can be direct (e.g. face recognition or even signatures)
  o With a direct legal link as is 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:
  o How can the citizen be sure that his personal data is not used outside the claimed purpose?
  o How feasible is for the system provider to use such data for other means?
  o Is it possible for the user to belong to a system without providing his administrative data?
    ▪ Relationship between Privacy and National Security!!
Privacy and Data Protection: EU Directive
### Background

- The Universal Declaration of Human Rights (10 December 1948) include:
  - **Art. 12:** “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Every one has the right to the protection of the law against such interference or attacks.”

- In addition, some countries also declare that:
  - Law will limit the use of computer science to guarantee the honour 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)
  - 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.

**Limits:**
- It 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 has been implemented in laws and regulations in each of the countries of the EU (up to a certain level of coverage depending on the MS).

**Major principles (MS dependent):**
- The citizen is entitled to preserve the full control on his personal data: who is collecting it, for what, and where are they going.
- The data collector shall implement a relevant policy to preserve the citizen right, declaring the filing of that data and the person responsible for keeping the policy.
Important concepts and requirements

• 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:
  - **What is the reason** for collecting that data and **which process** is going to be applied to his data
  - **Who** is collecting the data

• The citizen has the **right to deny** the collection of his personal data.
Important concepts and requirements

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

• Information Right during collection:
  o Existence, finality and recipients
  o Optionality and mandatory character of each collection
  o Consequences in not providing some of the data
  o Possibility of executing the user rights
  o Identity and address of the file responsible

• User Consent:
  o Unambiguous and revocable
Important concepts and requirements

• Levels of Protection:
  o Certain data may be subject to a higher level of protection. For example:
    ▪ Level 1: Administrative data (including Biometrics)
    ▪ Level 2: Health data
    ▪ Level 3: Political, Religious and Ideological data, Race, Sexual Orientation, etc.

• Data Security:
  o No explicit regulation, but the statement of requiring the file responsible to adopt all technical and organisational means to secure the data and preserve privacy
  o Without an active and relevant security policy, data shall not be collected

• Secret Obligation:
  o 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
Important concepts and requirements

• Data Communication:
  o After explicit consent of the citizen

• Access Rights:
  o By the citizen, with no compensation required
  o Once every 12 months

• Rectification and Cancellation Rights:
  o To be applied in 10 days
  o Cancellation shall derive on the blocking to all data related
  o If they have been given to a third party, this request is applicable to the third party
Coding data vs. Anonymisation

• Coded Data:
  o 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

• Anonymised Data:
  o Data which link with the subject is permanently broken, not being possible to re-establish such link through reasonable means
  o Is it possible to give away anonymised data???
How to apply Data Protection Directives R&D in Identification Technologies
R&D Requirements: Biometrics

• Analyse intra-class distribution
  o Collect data from the same subject
  o Several sessions
  o Different use cases
  o Different attitudes
  o Etc.

• Analyse inter-class distribution
  o Collect data from several subjects
  o Same parameters for the intra-class distribution

1. Collect identity data and contact information from the user

2. Deny the existence of duplicated entries
  o Typically you only collect users for their intra-class variability, and use their differences to obtain the inter-class
R&D Requirements: Biometrics

• Analyse influence of certain subject conditions
  o Relevant physical or mental conditions
  o Relevant knowledge background
  o Relevant habits
  o Etc.

• Analyse data stability
  o Sessions with significant time gap
  o Aging effect

3. Collect additional data for further processing
  o Some of this additional data may be subject for a higher level of protection

4. Keep data linked to identity for long periods
**R&D Requirements: Biometrics**

- Analyse robustness against Presentation Attacks
  - Producing attacks by experts, allowing them to deeply examine the subject’s data

- Use data for several iterations of the development phase
  - Or different algorithms

- Acquire a previously acquired DB

- Sell or give the DB to a third party

5. Obtain the permission of the subject

6. Analyse all ethical implications
  - E.g.: Deny the process if there is a relationship between attacker and subject

7. Keep data for long periods

8. Obtain data relevant to our study

9. Establish conditions and regulation of data transfer

---

Use Cases
- Data Protection Directive
- Privacy vs Recognition
- Directive in R&D
- R&D Req.
- Priv. Req.
- Procedure
- The new EU DP Directive

---

http://guti.uc3m.es

COST IC1206 Training School on "De-identification for privacy protection in multimedia content"
Privacy Requirements: Biometrics

1. Collect identity data and contact information from the user
   - Keep a subject database linked with the collected data
   - Assign a code to that subject
   - Use the code for the data reference
   - Keep both sets of data apart and the link table highly secured

2. Deny the existence of duplicated entries
   - Typically you only collect users for their intra-class variability, and use their differences to obtain the inter-class
   - Implement a functionality for the data collector, to ask the database the ID of the subject being collected
     - With a ID/new response
     - No further feedback
3. Collect additional data for further processing
   - Some of this additional data may be subject for a higher level of protection

4. Keep data linked to identity for long periods
   - Analyse the real need of that data and document such need
   - Check the level of protection of each piece of data
     - It may be good to work around the data request, so to not increase the level of protection and still have the needed data
   - A complete and accurate process to allow the cancellation of data shall be implemented
     - 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
Privacy Requirements: Biometrics

5. Obtain the permission of the subject
   - Inform the subject on this kind of actions
     - It may be recommended to state this as an optional consent

6. Analyse all ethical implications
   - Analyse all ethical implications
   - E.g.: Deny the process if there is a relationship between attacker and subject
   - Ensure ethical and legal agreement with attackers

7. Keep data for long periods
   - Consider the possibility of anonymising the collected data
     - 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 our study
   - Study the legal implications of acquiring such DB, and how to keep the claimed conditions

9. Establish conditions and regulation of data transfer
   - Enforce the third party not to distribute the DB, and to react towards all claims for cancellation
     - Will the third party be able to do it?
     - Do you have any power to force this?

http://guti.uc3m.es

COST IC1206 Training School on "De-identification for privacy protection in multimedia content"
Recommended Procedure

- Steps:
  1. Analyse the data needed and its level of protection
  2. Create a remote (not available to the external world) database to store subjects personal data (without samples)
  3. Create web services needed for the study:
     - Insert a subject
     - Delete a subject by ID
     - Ask for the subject's ID by providing a very reduced set of personal data
     - Provide discriminative subject data for collateral data impact (e.g. age group)
     - Provide statistics of the DB (e.g. number of subjects/samples, age/gender distribution, etc.)
  4. Create a local application to enrol subjects in the database
  5. Create a local database for storing subjects’ samples
  6. Create a local application to collect subjects’ samples
  7. Create a local application to process samples and obtain statistics
  8. Create an informed consent form
  9. Declare all the procedure, the file structure, and the security mechanisms to the relevant Data Protection Agency
  10. Start data collection and processing
  11. Termination
Illustration by Example

• Evaluation of Interoperability of Fingerprint Sensors and Algorithms
• 4 semiconductor flat sensors
• 5 algorithms (NBIS + 4 commercial ones)
• 589 subjects
• 6 fingers
• 6 samples / finger
• 2 visits:
  o 1st visit: Enrollment + 1st acquisition session
  o 2nd visit: 2nd acquisition session
• Soft ground truth mechanism to ensure proper collection
• Operator controlled data collection
• Compensation to subjects: 2 cinema tickets once the collection is finished
Step 1: Analyse Data Needed

- ID information:
  - Name
  - 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 of Basic Level of Protection

http://guti.uc3m.es

COST IC1206 Training School on "De-identification for privacy protection in multimedia content"
Step 2 – 5

• Step 2: Create a remote database to store subjects personal data (without samples)
  - A MySQL server with authenticated SSL connection in a computer without access to the external world

• Step 3: Create web services needed for the study
  - Using PHP scripts executed in the same computer where the MySQL database is installed
  - Web server in that same computer with only authenticated SSL connection available

• Step 4: Create a local application to enrol subjects in the database
  - Developed in Visual Studio, in C#
  - Executed in a computer without remote connection to the external world

• Step 5: Create a local database for storing subjects’ samples
  - As a collection of files with particular file naming based on collected data and user code
Step 6: Local App. for Collection

• Acquisition Environment:
Step 6: Local App. for Collection

- Enrolment Phase (operator controlled):

  ![Device Photo](http://guti.uc3m.es/COST IC1206 Training School on "De-identification for privacy protection in multimedia content"/Images/DevicePhoto.png)

- Use Cases
- Data Protection Directive
- Privacy vs Recognition
- R&D Req.
- Procedure
- Directive in R&D
- The new EU DP Directive
Step 6: Local App. for Collection

• Acquisition Phase (2 visits – operator assisted):

- Use Cases
- Privacy vs Recognition
- Data Protection Directive
- Directive in R&D
- R&D Req.
- Priv. Req.
- Procedure
- The new EU DP Directive
Steps 7 – 10

• Step 7: Create a local application to process samples and obtain statistics
  o Provide score tables in CSV to further process in Matlab or Excel

• Step 8: Create an informed consent form
  o Description of the process and the target of evaluation
  o Identification of the File and the File Responsible
  o Information on how to claim for the subject’s Rights
  o Form with the data of the subject already filled in and the ID
    ▪ One copy for the user for future reference
    ▪ Another copy for paper registry
  o Missing parts:
    ▪ Lack of minimum/maximum duration of the data
    ▪ Substitute registry paper with electronic version including biometric signature

• Step 9: Declare all the procedure, the file structure, and the security mechanisms to the relevant Data Protection Agency
  o Through the University, as it is declared as a relay for the Data Protection Agency

• Step 10: Start data collection and processing
  o 4 months of data collection + 3 months data processing

• Step 11: Termination
  o Still analysing the possibilities and impact of anonymisation
The new EU Regulation on Data Protection
New EU Regulation on DP

• Regulation 2016/679 has been launched to derogate the 95/46/EC Directive, improving, updating and extending its scope

• It goes from Directive to Regulation
  o Mandatory to all MS from May 25, 2018

• Some MS are already quite well aligned with this new Regulation, although others will have to really improve their legal environment
  o All MS laws shall be updated to adopt the new Regulation
New EU Regulation on DP

• Major changes:
  o It becomes also applicable to non-EU established institutions, if data is related to EU citizens
  o New categories for personal data
  o Concept of pseudonym
    ▪ Intermediate state between personal data and anonymization
    ▪ But not many details given 😐
  o Additional principles:
    ▪ Transparency
    ▪ Proactive responsibility
    ▪ Privacy by Design
    ▪ Privacy Impact Analysis (PIA)
  o Companies shall/should have a Data Protection Delegate
  o File declaration is no longer required in advance, but an internal registry shall be kept
  o 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
New EU Regulation on DP

• Major changes (cont.):
  o Adhesion to Codes of Conduct is recommended, as well as the establishment of certification mechanisms
  o International data transfer is considered within the new Regulation
    ▪ But not detailed 😞
  o An European Data Protection Council is created
  o Many more requirements about informing the citizen before collecting data
    ▪ Silence is no longer a way of providing consent
  o Additional Rights:
    ▪ 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
New EU Regulation on DP

• Major changes (cont.):
  o Adaptation to the new technologies:
    ▪ Big Data
    ▪ Internet of Things
  o Video-surveillance not clearly defined

• But there are still many concepts, mechanisms and interpretations to be defined 😞:
  o MS laws and regulations will have to cover these gaps
  o This is the reason why we cannot fully defined the impact to our case at this time.

  o We will have to keep an eye on future events!
Thank you for your time!

Questions?

Raul Sanchez-Reillo

University Group of Identification Technologies – Carlos III University of Madrid
Avda. Universidad 30, 28911 Leganes, Spain
rsreillo@ing.uc3m.es