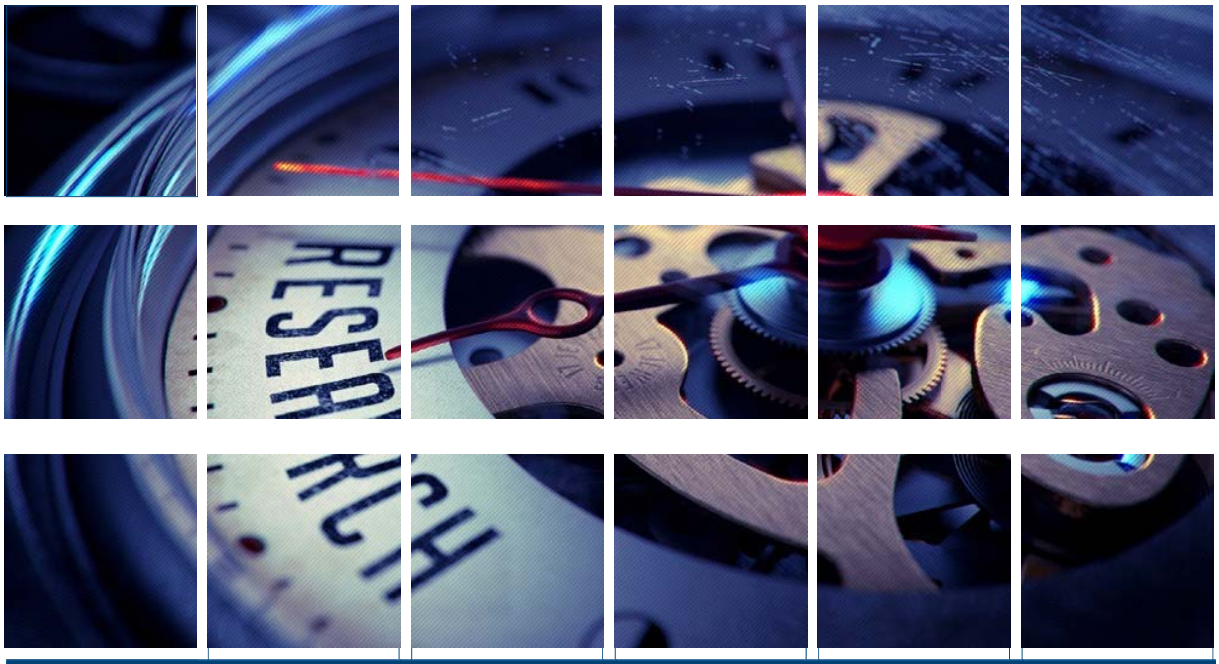
European University Institute

Guide on



GOOD DATA PROTECTION PRACTICE IN RESEARCH

Table of Contents

General Disclaimer, Scope and Purpose of the Guide

This guide is mainly a compilation of guidance offered primarily by the EU-funding authorities, as well as by the independent EU Advisory Body on Data Protection and Privacy. This information is assessed in the context of the ***EUI's Data Protection Policy****.*

The Guide is intended to provide an overview of the major concepts of data protection and privacy in research. It aims at raising awareness for these concepts in the academic community and at assisting applicants with the preparation of their project proposals.

The goal of this guide is also to provide researchers with a handy tool to guide them through the daily work on their research project.

The relevant notions and measures should be studied carefully. Requests for review by the Ethics Committee should be sent early, especially in case a first Screening Ethics report by the EU-funding authorities (such as the European Research Council) has identified ethics issues in relation to personal data. If a copy of ethical approval for the collection of personal data is requested by the EUI's Data Protection Officer, the researcher should also fill- in the sample Notification Form (sample annexed to this guide). Likewise, detailed information will be provided on the procedures that will be implemented for data collection, storage, protection, retention and destruction and confirmation that they will comply with EUI as well as national and EU rules, when applicable.

The information contained in the Guide is of a general nature only. It is not intended to address the specific circumstances of any particular individual or project but rather informs about the main aspects of data protection in the context of research carried out at the EUI. It does not provide binding legal advice.

This Guide should be consulted in parallel with the EUI's Data Protection Policy as well as with the ***EUI Code of Ethics in Academic Research****.*

The first edition of the Guide was prepared by Ms. Polyxeni Melidou (former Data Protection Officer at the EU) in May 2016.

The document remains open to adaptation, invention and modernisation whenever legal, pragmatic or technological developments make room for it. Comments and suggestions for improvement are welcome at: data_protection_officer@eui.eu

Dr. Günter Wilms

Data Protection Officer (DPO),

European University Institute

March 2017

1. INTRODUCTION

Privacy and data protection are fundamental rights which need to be protected.

Privacy **can mean different things in different contexts and cultures**. It can be the right to be left alone, but it can also be something positioned at the interface between private life and public life. It entails a dynamic relationship between private persons in different situations and different degrees of interaction. It is crucial to respect the privacy of research participants, but there are also other relevant rights the researchers should consider. It is therefore important to detail the purpose of the research according to the different understandings and legal definitions of privacy. For example in "covert research," researchers should take into account the meanings of public and private in the contexts they are studying. Covert observation should only proceed if researchers can demonstrate clear benefits of the research, when no other research approach seems possible and when it is reasonably certain that no one will be harmed or suffer as a result of the observation.

Data protection is meant to guarantee one's right to privacy. **Data protection refers to the technical framework and security measures designed to guarantee that personal data are safe from unforeseen, unintended or malevolent use**. Data protection therefore includes *i.a.* measures concerning access to data, processing, communication and conservation of data. Also measures to assure the accuracy of the data can be included in a data protection strategy.

In the context of research, privacy issues arise whenever data relating to persons are collected and stored, in digital form or otherwise. **The main challenge for research is to use and share the data, and at the same time protect personal privacy**.

In order to ensure respect for data protection and privacy, the EUI has adopted a Data Protection Policy that shall be respected by all EUI members and which is inspired by the EU data protection rules.

As a source of further reference, the EU Data Protection Directive contains a number of key principles for the handling of personal data[1]. This Directive provides the framework for the regulation of data protection and privacy issues in the Member States of the European Union. When the planned research includes collection and processing of data carried out in a EU-Member State, applicants need to identify the applicable local or national legal requirements and the competent authorities to provide any necessary authorisations, if applicable.

[1] In 2018 the Data Protection Directive will be substituted by the General Data Protection Regulation, adopted by the Council and the Parliament in April 2016. Among other issues, the new legislative framework will imply matters such as: secondary processing of data, the right to be forgotten, big data, cloud computing, web and social media mining and monitoring tools, etc.

2. PERSONAL DATA

What are personal data?

In legal terms, **“Personal data”** means:

1) **any information** relating to an **identified or identifiable natural person** referred to as ‘data subject’; an identifiable person is one who can be
2) **identified directly or**
3) **identified indirectly**, in particular by reference to an identification number or to one or more factors specific to his or her physical, psychological, mental economic, cultural or social identity.

1) The term **“any information”** calls for a wide interpretation. From the point of view of the nature of the information, the concept of Personal data include any sort of statements about a person. It covers **"objective"** information, such as the age of a data subject. It also includes **"subjective"** information, such as opinions or assessments.

2) Concerning **"directly"** identified or identifiable persons, **the name of the person is the most common identifier. Therefore in practice, the notion of “identified person” refers often to the person’s name.**

3) As regards **"indirectly"** identified or identifiable persons, **this category typically relates to the phenomenon of "unique combinations"**, whether small or large in size. In cases where *prima facie* the extent of the identifiers available does not allow anyone to single out a particular person, that person might still be “identifiable” because that information combined with other pieces of information will allow the individual to be distinguished from others. To establish whether a person is identifiable, one needs to consider whether it could be singled out using all the means likely reasonably to be used either by the Data Controller or by any other person to identify the said person.

Some characteristics are so unique that someone can be identified with no effort ("present Prime Minister of Spain"), **but a combination of details on categorical level (age category, regional origin, etc.) may also be pretty conclusive in some circumstances, particularly if one has access to additional information of some sort.**

While identification through the name is the most common occurrence in practice, a name may itself not be necessary in all cases to identify an individual.
This may happen when other "identifiers" are used to single someone out.

Examples of potential identifiers: physical characteristics, pseudonyms, occupation, address etc. or any combination of these.
However, an individual is not regarded as ‘identifiable’ if it involves excessive effort to identify them.

Considering the format or the medium on which that information is contained, the concept of personal data includes information available in whatever form, be it alphabetical, numerical, graphical, photographical or acoustic, for example. It includes **information kept on paper, as well as information stored in a computer memory by means of binary code, or on a videotape**, for instance. This is a logical consequence of covering automatic processing of personal data within its scope. In particular, **sound and image data qualify as personal data from this point of view, insofar as they may represent information on an individual**.

On the other hand, it is not necessary for the information to be considered as personal data that it is contained in a structured database or file. Also information contained in free text in an electronic document may qualify as personal data, provided the other criteria in the definition of personal data are fulfilled.

E-mail will for example contain "personal data"!

Awareness Q #1: Will any type of personal data be used and/or stored within the framework of the research?

Awareness Q #2: What kind of human participants/data are involved within the research?

2.1 Indicative categories of human participants:

- Patients;
- Healthy volunteers (related to health research);
- Volunteers (for surveys, etc…);
- Workers' (e.g.: research lab personnel…);
- Participating researchers' list;
- Children;
- Vulnerable adults;
- Others … special population groups? Developing countries? etc.

2.2 Categories of data used:

Previously collected data (their sources and usage history)

The content of the data set needs to be specified and copies of appropriate authorisations need to be provided according to the legal requirements of the area where the research is planned to take place.

What is Processing of Personal Data?

In legal terms, "processing of personal data" means: "any operation ... which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction".

In a nutshell: anything you do with personal data is considered as processing. Here are two examples:

- You create a mailing list or a list of participants.
- You manage a database.

Awareness Q # 3: Are you familiar with the defined rules and procedures governing the use and disclosure of personal data?

These rules and procedures are outlined in the Institute's Data Protection Policy complemented when necessary, by local privacy and data protection laws that exist in the country of data collection.

Further explanations can be found in guidance documents and in model operational documents as to ensure that researchers can implement the procedures on how to manage personal data and are familiar with these rules and procedures. For instance, this will include the principle that consent is required from the research participant before any such data can be disclosed.

Awareness Q # 4: Which is the applicable legal framework in case data are processed in more than one jurisdiction and a transfer of data from/outside the EU?

When data will be processed in more than one jurisdiction, the researcher must provide detailed information regarding the applicable legal framework in the countries where data collection or processing in general is going to take place.

As a matter of law and principle, researchers shall comply with Data Protection legislation in the Member State where the research as a whole or parts of it will be carried out.

In view of the status of the EUI as an international organisation, if the research is carried out at the EUI's premises, the applicable data protection framework should be the EUI's Data Protection Policy, complemented when necessary by local privacy and data protection laws.
For research and data collection carried out under a different jurisdiction (EU Member States, non-EU countries or in the frame of any other international organisation which has its own Data Protection Regulation/Policy), researchers should make sure they comply with the respective Data Protection and Privacy Requirements (including prior-authorisations and notification requirements to National Data Protection Authorities/ local Data Protection Committees).

3. BASIC PRINCIPLES OF PROCESSING OF PERSONAL DATA

Whenever personal data are being processed, you must keep in mind certain principles and conditions.

"Data quality" is the aim and this is achieved when the data processed are:

- adequate, relevant and non-excessive (e.g. by minimising collected information/database fields);
- accurate and where necessary, kept up to date;
- processed fairly and lawfully;
- processed for specified, explicit and legitimate purposes and not further processed in a way incompatible with these purposes;
- processed in line with data subjects' rights;
- processed in a secure manner;
- kept for no longer than necessary for the purposes for which the data was collected or for which it is further processed.

Tip to ensure compliance with necessity and proportionality principles:

- Design proper data retention and deletion plans already at proposal stage;
- Consider automated deletion of certain types of data during the carrying out of research and introduce a data storage scheme for data kept after the project is completed.

Moreover, you need to take adequate precautions when personal data are transferred to third parties to fulfill "data quality".

The purpose for which the data were collected or further processed determines the length of time for which the data should be kept. Once the data are no longer needed they should either be deleted or kept in anonymous form if they serve historical, statistical or scientific uses.

N.B.: Personal data collected by the EUI for research purposes can be processed only for the scientific objectives for which they were collected (Art.15 of the EUI's Data Protection Policy).

In cases of secondary processing of research and scientific data previously obtained for other research purposes can be used in so far as they are not incompatible.

4. SPECIAL CATEGORIED OF PERSONAL DATA (commonly called "SENSITIVE DATA")

This is a delicate matter

Article 7 of the EUI's Data Protection Policy indicates some categories of data that are more sensitive than others and require special treatment, **personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and data concerning health or sex life**.

Examples of Sensitive Personal Data:

- Membership in a religious or political group;
- Sexual orientation;
- Health-related records (e.g. patient records, biographic data, medical photographs, diet information, hospital information records, biological traits and genetic material);
- Criminal records or legal justice investigations and proceedings;
- Circulation records such as visas; residence or various geographic recordings such as GPS satellite localization recordings.

As a general rule, the processing of such personal data is prohibited. However, Article 7 of the EUI's Data Protection Policy does allow it to be processed in specific circumstances. The most common in research is upon the **data subject's *explicit* consent.**

If you intend to process sensitive personal data in the course of your research, or if there is a possibility that sensitive personal data may be processed (unanticipated sensitive data), this will impact upon the conditions you will need to satisfy to carry out that processing lawfully, the justifications you may need to provide to the research ethics committee, and the uses to which the research data and research outputs can be put.

Awareness Q #5: Are all sensitive data that are planned to be collected really focused on the research question and are they relevant for the research?

You will need to explain the reasons behind the proposed data collection: data from different sources should not be amalgamated without making sure that this action is legally possible, especially in cases where a data set might contain information that identifies individuals and information.

5. INFORMED CONSENT

With your permission

Informed consent: declared one of the most pivotal principles in research ethics, a comprehensive informed consent is a crucial requirement in research.

Appropriate use of consent: Consent is sometimes a weak basis for justifying the processing of personal data and it loses its value when it is stretched or curtailed to make it fit to situations that it was never intended to be used in.

The use of consent "in the right context" is crucial.

Absence or weakness of the elements for valid consent creates vulnerability and, in practice, weakens the position of data subjects. It can lead to legal challenges and liability claims against the researcher.

Definition in the EUI's Data Protection Policy:

"the data subject's consent' shall mean "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed"

Main aspects of Informed Consent:

1. ***"… freely given …"*** Consent can only be valid if the data subject is able to exercise a real choice, and there is no risk of deception, intimidation, coercion or significant negative consequences if he/she does not consent.

In that regard, the potential participant must be given sufficient information in order to be able to make a choice of whether or not to participate that is based on an understanding of the risks and alternatives in an environment which is free from any coercion;

2. ***"… specific …"*** To be specific, consent must be intelligible: it should refer clearly and precisely to the scope and the consequences of the data processing. It cannot apply to an open-ended set of processing activities. This means in other words that the context in which consent applies is limited and that blanket consent without specifying the exact purpose of the processing is not acceptable.

Consent must be given in relation to the clearly identified aspects of the processing. It includes notably the kind of data which are processed and the purposes of the processing.

Consent refers to reasonable processing which is reasonable and necessary in relation to the purpose. It is generally sufficient to obtain consent only once for different operations as long as they are covered by the data subject's reasonable expectations.

3. ***"… informed …"*** Prior information (appreciation and understanding of the facts and implications) is a precondition for valid consent.

The individual concerned must be given, in a **clear and understandable manner, accurate and full information** of all relevant issues such as:

- nature of the data processed;
- purposes of the processing;
- recipients of possible transfers;
- rights of the data subject;
- consequences of not consenting to the processing question.

Quality, Accessibility and Visibility of information are key requirements!

4. The participant's consent must be evidenced. The participant needs to agree that her/his data will be used for a specific research scope and must be aware of the meaning of such use.

Examples of important aspects that researchers must take into account:

- the power relationship between the researcher and research participants;
- the vulnerability of the population under study;
- the impact of the research results on individuals and communities, with particular emphasis placed on avoiding stigmatization and discrimination.

All relevant aspects from this list should be given thorough consideration in research protocols.

When writing a research proposal you should make sure that:

- You show a detailed understanding of the nature of the information that should be provided to the potential participants.
- Write in a way that will be understandable to potential participants; their decision should be based on free will – i.e. the participant's decision not to participate in a survey should not lead to any negative consequences or the impression thereof.

Tip: the most convenient way to show this is to produce a draft information sheet and attach the informed consent protocol (use the EUI's model consent form) to the application.

- Pay close attention to the way research participants are approached.

Caution: In case consent is obtained from family or community leaders only to approach individuals, this should not substitute an individually obtained consent. In case of people not able to consent (e.g. children), parents' or legal representatives' consent and children`s assent are necessary.

If research consists of fieldwork, obtaining informed consent might not be a onetime event, but should rather be regarded as an ongoing process, which might evolve differently from what was anticipated before beginning research.

Consent should be renegotiated if the inquiry moves in an *unanticipated new direction*. The methodological limitations of gaining "fully informed" consent has to be made clear at the outset.

- When seeking to obtain individual written consent from research participants, take into account the cultural and ethical norms of the population(s) under study.

Tips to address special circumstances:

- In case a written consent does not respond to the ethical norms of those studied, provide alternative ways of obtaining consent (such as recording the oral consent, the presence of witnesses, all procedures used must be documented).
- In case of participants who for any reason are not fully capable of understanding and expressing their will, replace the informed consent by other sufficiently equivalent protective measures.
- In case of observational studies, obtain reasonable informed consent from all participants, and approval from the gatekeepers before the beginning of the study; if individuals cannot be identified, then individual consent might be sought after the study is finished.
- Depending on the nature of the study, observation of people in a completely public environment might not require consent, but researchers have to demonstrate that their study would in no way alter the usual behaviour of those under scrutiny and that their privacy would be respected.

What about children and vulnerable adults?

If you are involving children in your research **you should devise appropriate strategies of informing them** about their participation (for example by using audio or video materials, posters, flyers, suitable to their age and understanding). Children who are capable of forming their own views should be granted the right to express their views freely in all matters affecting them, commensurate with their age and maturity.

Awareness Q #6: Do you have the necessary legal permission(s) to obtain and process the data?

- If data is directly gathered from individual study participants, is the planned informed consent system effective?
- Informed consent for the proposed project will be required, even if personal data has been collected in the frame of previous research projects: If data from a previously gathered set - either by the applicant or from another project or person – are used, does the initial informed consent cover this *complementary* use of the data, or does the applicant have to obtain a completely new informed consent for the proposed study. The researchers need to discuss these options along with the Institute's Local Ethics Committee and/or the Data Protection Officer or other competent body.

Secure data storage so as for the data not to become accessible to unwanted third parties and to be protected against disaster and risk.

Consider the following:

1. Where is the data stored? Data must be stored within a secured environment. If stored electronically, this must be a machine, or set of machines located in a physically secured environment – with controlled access - as well as technically secured.

2. On which hardware type is the data stored: paper, disk, removable device? Considering the nature of the data used in the project: what will the adapted security processes to be followed? How do they guarantee confidentiality of data?

3. Who has access to the data? Can the data be accessed by any third party? Can the data be copied by any third party?

4. For how long will the data be stored, accessed? What will happen to the data after the end of the study: duration of storage should be justified. Destruction at the convenience of the researcher would appear insufficient, unless clearly stated in the consent form and the approval of the local competent authority. Such deletion of data should be defined as irreversible, or reversible.

5. If stored on a machine, is the storage machine/server equipped with:

- Wifi
- Bluetooth
- USB drive
- On the whole, devices that might ease data duplication of circulation.

6. What data backup policies and processes will be implemented?

Answering these questions will help you assess the data protection and privacy risks within the project and therefore provide a state of the art risk management policy.

6. DATA SECURITY

Handle with care

To process data in a secure manner you must:

- take the appropriate technical and organisational measures to prevent any unauthorised act with regard to the data;
- make sure that no one will access, read, copy, alter, use in any way or process the data unless he/she is authorised to do so according to clear access rules;
- organise the processing in a way that gives you the best possible control and allows for tracking of the procedures followed;
- if someone is processing data on your behalf, you need to choose someone able to guarantee secure processing.

These requirements are provided and further outlined in Articles 10 and 18 of the EUI's Data Protection Policy.

In practical terms, these measures could result in:

6.1. User authentication

The way to verify the identity of a user:

- One-factor: *"something a user knows"*, e.g. a strong password.

Key aspects of a strong password:

- length (the longer the better);
- a mix of letters (upper and lower case), numbers, and symbols;
- with no ties to your personal information, and no dictionary words)

- Two-factor: *"something a user has"*, e.g. a signed digital certificate in a smart card;

6.2. Access control

A mechanism to allow or deny access to certain data:

- Based on predefined user lists and access rights, e.g.:

- who can access what
- type of access (read, write, delete, etc.).

- Based on the functions of each user within the project;
- Role based – attribute based.

6.3. Storage security

Storing data in a way that no unauthorised party is able to access it:

- Operating system controls (authentication & access control).
- Use of passwords to access electronic files (e.g. use the text editor function to save a document password-protected).
- Local encrypted storage (enable the full disk encryption, enable the file system, enable the text editor encryption).
- Database encryption: turning data into a form that are unintelligible (for anyone not having access to the key).
- Consider that your storage concerns are equally important if your data are on your local PC, your portable storage device or in the cloud storage!

6.4. Communication security

Protecting data when transferred via communication means:

- Encrypted communication (SSL/TLS); (e.g. use web services whose URL starts with "https://" and not only http://.
- Firewall systems and access control lists (e.g. make sure the firewall service is enabled on your PC).
- Anti-virus & anti-malware systems.

6.5. Other IT technical controls:

- Back-ups: necessary for the availability of the systems and information;
- PC configuration: security-aware settings at user level (e.g. installing security updates, anti-virus protection, local back-ups, blocking of certain software installation, etc.).

Awareness Q # 7: How will the collected personal data be securely accessed?
Secured access policy needs to be worked out and clearly specified. It needs to be proportional to the risks involved and the sensitivity of the data, and must clearly state the type of processes - such as password protection, encryption, "need to know basis" principles (i.e. : only the users that need to access the data will be allowed to do so),- that will be implemented.

Awareness Q #8: How will the data be securely stored: data structure and format?
Data structures such as databases need to be specified - if applicable, it should be specified that identification data will be encrypted and strictly separated from sensitive data such as health data

– It should also be specified how the unforeseen data added during the research such as incidental findings will be treated.

Awareness Q #9: How will the data be securely stored: location & hardware?

Conservation methods need to be specified. A non-WAN connected computer server or HARD disk should be preferred. Data should not be stored on a memory stick or other easily lost/accessed media.

7. DATA TRANSFER

Sharing while caring

"Data transfer" would normally imply at least the following elements:

- The communication, disclosure or otherwise making available of personal data from the researcher to a third party regardless of the medium, **including but not limited to movement across a network, physical transfers, transfers from one media or device to another, or by remote access to the data**;
- Conducted with the knowledge or intention of the researcher that the third party - recipient will have access to it.

The concept includes: **"deliberate transfers"** and **"permitted access"** to data recipients.

Transborder flows of personal data means the movement of personal data across national borders by any means, including access of data from outside the country where collected and use of cloud technologies for data.

Article 13 of the Institute's Data Protection Policy disallows transfer of personal data to a third party (including country or territory outside the European Economic Area) ***unless that third party ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data****.*

For intra-EU transfers…

The recipients of the data shall have in place adequate safeguards for the protection of privacy compatible with the applicable EU Data Protection Directives or Regulations (e.g. Directive 95/46/EC or any subsequent amending acts).

Article 13 requires also the following specific safeguards:

- the data shall be necessary for the legitimate performance of tasks covered by the competence of the recipient or
- the recipient establishes that the data are necessary for the performance of a task carried out in the public interest or subject to the exercise of public authority or
- the recipient establishes the necessity of having the data transferred and if there is no reason to assume that the data subjects legitimate interests might be prejudicated.

It is the responsibility of the project beneficiary/principal investigator/researcher to conduct a specific adequacy assessment of the data protection system of the recipient.

Special precautions need to be taken when personal data is transferred to countries outside the EEA that do not provide EU-standard data protection.

Without such precautions, the high standards of data protection established by the Institute's Data Protection Policy and those applied also at EU level would quickly be undermined, given the ease with which data can be moved around in international networks.

Data transfer to the US

Data transfer to the US deserves special attention.

It is doubtful whether US legislation provides a standard of data protection comparable to the one in the EU[1].

Although the EU Commission and United States agreed in July 2016 on a new framework for transatlantic data flows, the EU-US Privacy Shield[2], recent developments put those achievements into question. The Executive Order on Enhancing Public Safety in the Interior of the United States waters down the protection granted by expressly excluding "persons who are not United States citizens or lawful permanent residents" from privacy policies[3]. This order risks to undermine the protection granted by the "EU-US Privacy Shield".

For those reasons it is necessary to carefully evaluate on a case by case basis whether personal data can be transferred to the US.

If personal data are to be transferred from one jurisdiction to another, make sure this is done in such a way that it meets the data protection requirements in both the origin and the destination jurisdictions.

Research participants must know what is happening with their personal data and this must be either explained verbally or provided in some written format or document that research participants have agreed to – i.e. via their consent which is recorded as evidence that they have agreed.

Tips concerning data transfer:

- If data processing is outsourced, remove personal data where practicable and as far as **possible** so that only a pseudonymous ID number is used to link individual-level data with participants' identities.
- Assess the level of protection afforded by a third country or international organisation **in the light of all circumstances surrounding a data transfer operation or set of data transfer operations**.

Awareness Q # 10: How will data transfer be monitored?

Transfer of data outside the EU needs to be identified and specified. The handling process should be specified. Data transfer (between whom and whom) within the project, especially with partners from non-EU countries (developed and/or developing countries) must be given special care due to the variety of legal and administrative standards. This is because both the EUI's Data Protection Policy as well as EU legislation requires that the transfer of data outside Europe to be undertaken only to places where there is a local assurance that the level of data protection is compatible/ at least equivalent to that of the EU area. Researchers need to consider this aspect not only between institutions and companies and the like, but also within companies and the research partnership across geographical borders.

---

[1] Judgment of 6 October 2015 in Case C-362/14, Schrems v Data Protection Commissioner, press releases available under: http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf ; full text: http://curia.europa.eu/juris/document/document.jsf;jsessionid=9ea7d0f130d69f148306a4c04a91b2edde46ac640262.e34KaxiLc3eQc40LaxqMbN4PaheQe0?text=&docid=169195&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=495630

[2] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.207.01.0001.01.ENG&toc=OJ:L:2016:207:FULL

[3] https://www.whitehouse.gov/the-press-office/2017/01/25/presidential-executive-order-enhancing-public-safety-interior-united

8. ANONYMISATION

Anonymisation techniques make personal data unidentifiable ("further processing").

One of the big advantages of anonymisation is to allow research that would otherwise not be possible due to privacy concerns.

Example: Collecting immigrants' data on their immigration experiences could lead to an added-value in research on irregular migration, but could also seriously infringe people's privacy and put them at risk of prosecution by the authorities as well as persecution by human smugglers.

A solution can be to remove direct identifiers such as names, birth dates, and addresses, although this might not be sufficient to avoid that the data can be traced back to individuals.

An **effective anonymisation solution** prevents all parties from singling out an individual in a dataset by:

- Linking several records within a dataset (or between several separate datasets).
- Inferring any information in such dataset.

Food for thought:

- Removing directly identifying elements in itself is not enough to ensure that identification of the data subject is no longer possible.
- Additional measures are usually necessary to prevent identification, depending on the context and purposes of the processing for which the anonymised data are intended.

Example: If a person is described as "a man" the anonymity set size is three and a half billion, but if he is described as "a middle-aged Dutchman with a beard" it is maybe half a million and if he is described as "a middle-aged Dutchman with a beard who lives near Cambridge" it might be three or four.

The principles of data protection do not apply to data rendered anonymous in such a way that the data subject is no longer identifiable.

Common Anonymisation Techniques

1. ***Randomisation as…***
   a family of techniques removing the link between the data and the individual. If the data are sufficiently uncertain then they can no longer be referred to a specific individual.

2. ***Generalisation as…***
   an approach consisting of generalizing, or diluting, the attributes of data subjects by modifying the respective scale or order of magnitude (i.e. a region rather than a city, a month rather than a week).

Food for thought:

Generalisation can be effective to prevent singling out but does not allow effective anonymisation in all cases. Make sure you devise specific and sophisticated quantitative approaches to prevent linkability and inference.

3. Pseudonymisation as…

a hybrid technique referring to the process of disguising identities by replacing one attribute (typically a unique attribute) in a record by another**.**

Main characteristics:

- Natural person still likely to be identified indirectly but only under pre-defined circumstances;
- When used alone it does not result in an anonymous dataset;
- It reduces the linkability of a dataset with the original identity of a data subject;
- It is a useful security measure but not a method of anonymisation.

Factors influencing its effectiveness:

- stage at which it is used;
- its level of protection against reverse tracing;
- the size of the population in which the individual is concealed;
- the ability to link individual transactions or records to the same person, etc.

Useful Tips:

- Use random and unpredictable pseudonyms.
- Make sure the number of pseudonyms possible is so large that the same pseudonym is never randomly selected twice.

Example of pseudonimisation: Key-Coded data.

Information relates to individuals that are earmarked by a code, while the key making the correspondence between the code and the common identifiers of the individuals (like name, date of birth, address) is kept separately.

Main Characteristics:

- If the codes used are unique for each specific person, the risk of identification occurs whenever it is possible to get access to the key used for the encryption.
- If the codes are not unique, but the same code number (e.g. "123") is used to designate individuals in different towns, and for data from different years (only distinguishing a particular individual within a year and within the sample in the same city), the controller or a third party could only identify a specific individual if they knew to what year and to what town the data refer.

Ultimate Goal: Anonymous data

Anonymous data as…

any information relating to a natural person where the person cannot be identified, neither by the data controller nor by any other person, taking account of all the means reasonably likely to be used.

Anonymised data as…

anonymous data that previously referred to an identifiable person, but where that identification is no longer possible.

Food for thought:

Carry out a case-by-case analysis to assess:

- Whether the data be considered as anonymous or not;
- Whether the information can allow identification of an individual;
- Which means are likely to be used for identification.

This is particularly relevant in the case of statistical information, where despite the fact that the information may be presented as aggregated data, the original sample is not sufficiently large and other pieces of information may enable the identification of individuals.

Awareness Q # 11: Have you considered anonymity and confidentiality?

- Clarify whether the data will be anonymised (link to the individual will be destroyed) or coded (the data will be reversible).
- Explain how you will ensure data protection and how any link to the research participants will be handled if not fully anonymised.
- Insist that participants in surveys/experiments use their initials (not ID session numbers or full signatures) when they sign their consent forms, and not simply suggest that they do so.
- If the data will not be anonymised, explain why you cannot anonymise the data (e.g. you need to recontact the participants or do follow-up in case of long-term study). If the data will be coded, describe the coding system, and who will have access to the code. Confirm that it cannot be traced back to individuals unless essential for the study.
- Use data for statistics only after anonymisation techniques have been applied.

List of main reference documents

- Ethical Review in FP7 - Data protection and privacy ethical guidelines (European Commission);
- Ethics for researchers - Facilitating Research Excellence in FP7 (European Commission);
- Guidance Note for Researchers and Evaluators of Social Sciences and Humanities Research (European Commission);
- Guidance How to complete your ethics self-assessment (European Commission, Directorate-General for Research & Innovation, Version 5.0, 15 March 2016);
- Opinion 05/2014 on Anonymisation Techniques (Art. 29 Working Party);
- Opinion 4/2007 on the Concept of Personal Data (Art. 29 Working Party);
- Opinion 15/2011 on the Definition of Consent (Art. 29 Working Party);
- Data Protection and research guidance note (London School of Economics);
- Guidance on Data Protection, Confidentiality and Records Management (University of Sussex);
- Data Protection Checklist (ESOMAR);
- Introduction to Data Protection (EDRI).

ANNEX: SAMPLE NOTIFICATON FORM TO BE SUBMITTED TO THE DPO WHEN SEEKING DATA PROTECTION CLEARANCE IN THE CONTEXT OF AN ETHICS REVIEW

European University Institute	**PROTECTION OF PERSONAL DATA** **NOTIFICATION OF PROCESSING OPERATIONS - EUI**

REFERENCE:	
Title:	**[Please insert the reference number and the title of the project.]**

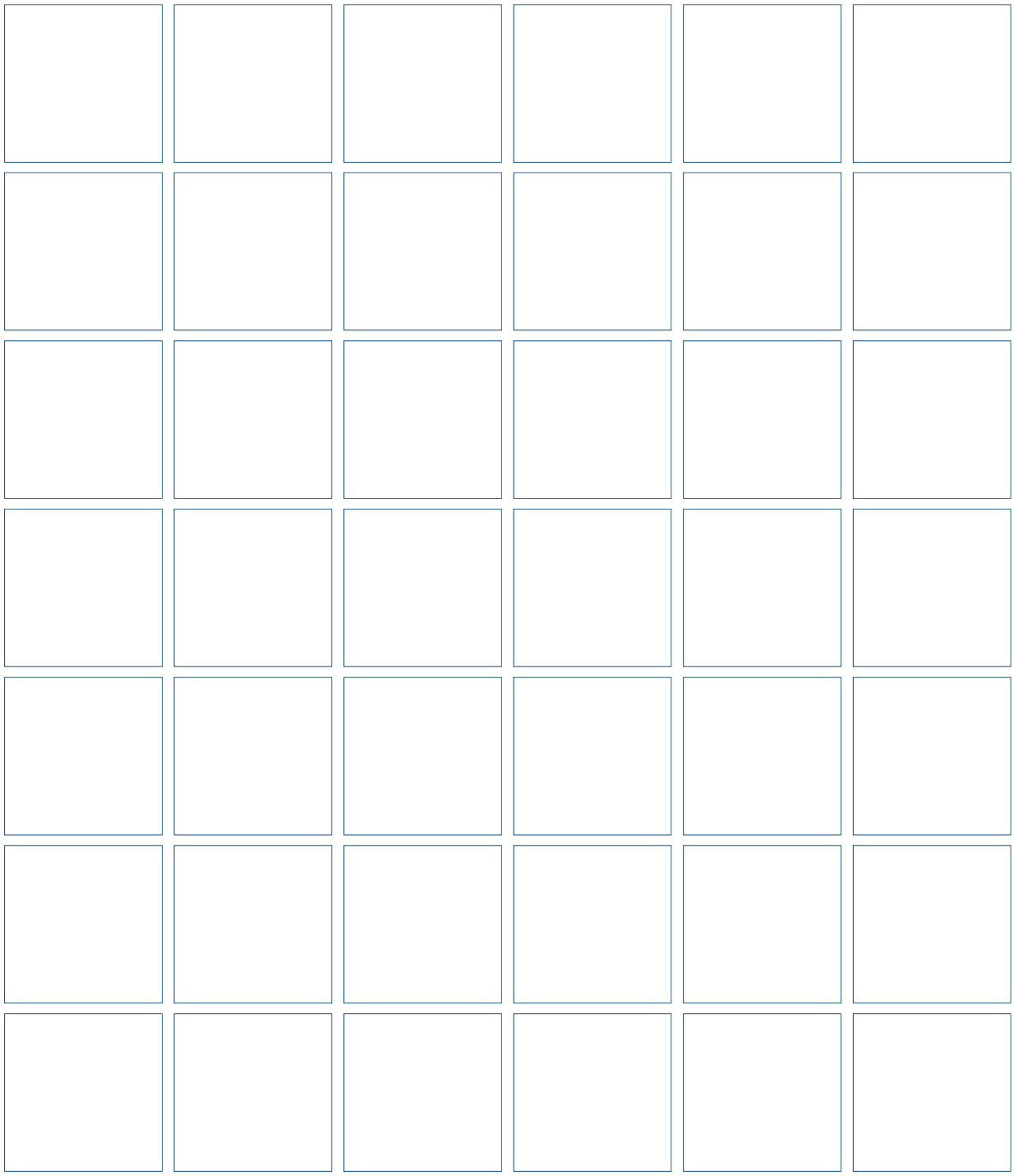
1. Processing	
1.1 Name of the processing	[Please insert the title of the project.]
1.2 Name and First Name of the Data Controller	[Please insert the name of the Principal/Research Investigator who is the "master" of the personal data and determines the purposes and means of the processing of personal data and ensures data protection compliance].
1.3 Name and First Name of the Processor(s)	[Please indicate the names of any other natural or legal person that may process them. If processors can be categorised into groups please refer to them by groups and not necessarily by name, otherwise indicate their names.]
1.4 Lawfulness of Processing	[You may collect and process data only if and insofar as it is really necessary for your research. Collecting personal data (for example, on religion, sexual orientation, race, ethnicity, etc.) that is not essential to your research may moreover expose you to allegations of 'hidden objectives' or 'mission creep-- i.e. information being collected with permission for one purpose and being used or made available, including online, for another reason, without additional permission'.]
1.5 Description of the processing (i.e. what you do with personal data and how)	[Please describe the means of collection and processing of personal data: recording, notes etc. and give detailed feedback on any methods used for tracking or observing participants. Indicate also if a copy of notification/authorisation for tracking or observation is required. "Processing of personal data" means any operation (or set of operations) which is performed on personal data, either manually or by automatic means.

	This includes i.a.: − collection (digital audio recording, digital video caption, etc.) − recording − organisation and storage (cloud, LAN or WAN servers) − adaptation or alteration (merging sets, amplification, etc.) − retrieval and consultation − use − disclosure by transmission, dissemination or otherwise making available (share, exchange, transfer) − alignment or combination − blocking, deleting or destruction, etc.] Processing covers normally any action that uses data for research purposes (including if interviewees, human volunteers, patients, etc. are not actively included in the research). Data may come from any type of research activity (ICT research, genetic sample collection, personal records (financial, criminal, education, etc.), lifestyle and health information, family histories, physical characteristics, gender and ethnic background, location tracking and domicile information, etc.).
1.6 Types of Data Subjects	[Please indicate any types of data subjects involved in the processing operations of the project]
1.7 Data Fields	[Please list concretely the categories of personal data that you collect (e.g. names, age, country of origin, etc.)]
1.8 Rights of data subjects	[Art.12 DP Policy Data subjects have the following rights: a) to obtain a confirmation whether or not their data are processed, and information on the categories of data that are being processed, in what ways, and for what purposes as well as the recipients or categories of recipients to whom the data are disclosed b) to obtain communication of their personal data undergoing processing and of any available information as to their source, c)to knowledge of the logic involved in any automated decision process concerning him or her d) to the rectification of inaccurate or incomplete personal data,

	e) to the erasure of data of which processing by the EUI is unlawful, f) to block the processing of data of which they contest the accuracy, until accuracy is checked. Please indicate which rights are relevant to this project and how data subjects will be able to exercise them. E.g.: Participants will be free to withdraw at any time without justification, and their data will be deleted upon their request. In order to ensure that the data can be deleted if necessary, while maintaining anonymity, each participant is attributed a, pseudonym, and the key to these pseudonyms will be password-protected and available only to the principal investigator.]
2. Detailed procedures	
2.1 Details on the procedures that will be used to identify/recruit research participants.	[E.g.: Through a consultative process using gatekeeper (through NGOs providing support to the participants); snowballing (through referral from one participant to another); personal contacts (through the researcher's own contacts from his contextual knowledge of the country and place where the research is conducted), etc.]
2.2 Details on the informed consent procedures that will be implemented.	[Please give further details on implementation procedures for each category of data subjects, if appropriate. Specify which of them are integral to the recruitment process, and which will also be implemented as an ongoing process. Please give information about the written record of consent In case of children/minors and/or adults unable to give informed consent, indicate the tailored methods used to obtain consent. According to the H2020 Guidelines, if the consent cannot be given in writing, for example because of illiteracy, the non-written consent must be formally documented and independently witnessed. Please explain if and how the oral consent will be documented and if and how it will be recorded. In case it won't be recorded please justify. If deception is going to be used for another type of data subjects retrospective informed consent should be obtained and participants must be debriefed. Deception requires strong justification and appropriate assessment of the impact and the risk incurred by both researchers and participants.
2.3 Measures taken to prevent the risk of enhancing vulnerability/stig	[Please indicate any such measures (e.g.: providing anonymity, use of pseudonyms, non-disclosure of audio-visual materials, voice records, etc.)]

Matisation of individuals/groups	
2.4 Safeguards taken to preserve the data subject's identity.	[Please provide details on identifiable data subjects and the measures taken to avoid direct or indirect identification e.g. by use of pseudonyms.] (An identifiable person is one who can be identified directly or indirectly, in particular by reference to identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural or social identity). E.g.: At no time will names be disclosed in audio recording and published material. Pseudonyms (a reversible system of coding will be used in order to be able to re-contact participants in case of necessity) will be used in all documentation, and any circumstantial information that may reveal the identity of participants will be concealed in published matter. Any residual information that could jeopardise the anonymity of participants will be destroyed at the end of the project. This procedure will be clearly explained to participants during the recruitment process].
2.5 Re-use of data publicly or non-publicly available.	[Please indicate whether the project is concerned with such data. E.g.: Research is qualitative and will produce data through processes such as interviews, and by tracking people during the course of time. The public/private distinction cannot be usefully applied in this case.]
3. Technical and organizational security measures (Data Security) [Secured access policy needs to be worked out and clearly specified. It needs to be proportional to the risks involved and the sensitivity of the data, and must clearly state the type of processes – such as password protection, encryption, "need to know basis" principles (i.e.: only the users that need to access the data will be allowed to do so), - that will be implemented.]	
3.1 Storage medium	[Please indicate any methods considered for data storage. E.g.: Each transcription will be identified by a pseudonym while the informants' names will be stored in a separate file to ensure security of

	personal data. All of these files will be password protected. Data collected will be stored in a secure database in the EUI, to which only the project team members will have access, in order to prevent any possible misuse (e.g. data mining, profiling). E.g.: Temporary storage (on site): The transcribed interviews and field observations will be stored electronically and 'Password-protected by the researcher in each field site. The researcher will make regular back-up copies of these files, 'which will be stored offline on the hard drive of their laptop as well as in an external hard drive (stored in a different location). Audio recordings will be stored securely in the researchers' own home. Conservation method needs to be specified. A non-WAN connected computer server or HARD disk should be preferred. Data should not be stored on a memory stick or other easily lost/accessed media. Data structures such as databases needs to be specified - if applicable, it should be specified that identification data will be encrypted and strictly separated from sensitive data such as ethnical data, etc. It should also be specified how the unforeseen data added during the research such as incidental findings will be treated.]
3.2 Retention Period	[*Personal data shall be stored for no longer than it is required for the processing purposes for which it was collected (*Art.6 of EUI's DP Policy). Please indicate how you will comply with this requirement and when exactly personal data will be destroyed. Please indicate if anonymised data will be kept for longer period.]
4. Data Transfer	
4.1 Inside EUI.	[Please indicate if any Personal Data is transferred to recipients inside the EUI and to which person or category of persons/legal entities and for what purpose.]
4.2 Outside the EUI, inside the EU/EEA	[Please indicate if Personal Data is transferred outside the EUI but inside the EU/EEA area and to which person or category of person(s)/legal entities and for what purpose.]
4.3 Outside the EUI, outside the EU/EEA	[Please indicate if Personal Data is transferred outside the EUI and also the EU/EEA area and to which person or category of person(s)/legal entities and for what purpose.]
5. Complementary information	[If necessary]



Second edition - Printed in Italy in March 2017