FORS data management webinar series

Data Protection

Dr. Pablo Diaz

September 30th, 2021

FORS webinar series

1. **Introduction to data management planning**

2. Data protection

3. **Data security** / October 12th

4. **Data sharing** / November 4th

FORS

What is Data Protection?

Data protection is commonly defined as the **laws, rules and strategies** designed to protect people's **privacy**.

- **Privacy** is the ability of an individual or group to seclude themselves or information about themselves, and thereby express themselves selectively.

- In the field of data protection, privacy is generally referred to through the concept of **informational self-determination.**

What is Data Protection?

- DP is an area dominated by **law**
- DP applies when **personal data** are involved
- DP is about protecting **people**

  - Not research institutions
  - Not researchers
  - Not the data themselves (≠ security)

A central issue

- Commercial value of data (new “black gold”)
- Big data
- Greater storage capacity
- Greater computing capacity
- Rise of digital technologies (apps -> traces)
- New analysis / mining tools

The risks of a breach of privacy are high!

Data protection and research

As social scientists, our work often requires the processing of personal data:

- Sociological profiles
- Life story interviews
- Longitudinal surveys
- Etc.

We are therefore **directly concerned** by data protection

Data protection and research

Many of the tools we use to process data from our research participants have privacy concerns:

- Smartphones (cloud backup, geolocation, etc.)
- Tools for online surveys (Survey Monkey, etc.)
- Cloud (Dropbox, etc.)
- Online software (transcriptions, pdf mergers, etc.)

We have to be very **careful**!

Tension between ethics and law

"One day (…) in July 1998, I found myself in the hands of police officers. (…) They seized all my working documents and immediately began to question me about this research and the people I had talked to. I told them that I worked a lot with excluded groups and that such work was based on respect for the anonymity of the people interviewed, and that in this sense it was impossible for me to inform the police. For this research, I made an express commitment to the interviewees, guaranteeing them not to reveal their identity. It was an **ethical** problem for me and a condition for sociological inquiry: I could not give the names".

Sylvain Laurens & Frédéric Neyrat, Entretien avec Pinar Selek : "Je n'allais pas donner les noms, c'est une question d'éthique", in S. Laurens et F. Neyrat, Enquêter de quel droit : menaces sur l'enquête en sciences sociales, Editions duCroquant, 2010, pp. 235-242

Tension between ethics and law

The tension between ethics and law can be felt in several situations:

- Studying criminal activities
- Studying political activities
- Studying child abuse
- *Etc.*

Legal bases

Legal bases

Informational self-determination is guaranteed by a number of **fundamental texts**:

- Universal Declaration of **Human Rights** (art. 12)
- European Convention on **Human Rights** (art. 8)

THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

FACING HISTORY AND OURSELVES

Adopted by the General Assembly of the United Nations in 1948, the Universal Declaration states fundamental rights and freedoms to which all human beings are entitled.

You have the responsibility to respect the rights of others.

We are all born free and equal.

Everyone is entitled to these rights no matter your race, religion, sex, language, or nationality.

Everyone has the right to life, freedom, and safety.

No one can take away any of your rights.

No one has the right to hold you in slavery.

No one has the right to torture you.

You have a right to be recognized everywhere as a person before the law.

We are all equal before the law and are entitled to equal protection of the law.

You have the right to seek legal help if your rights are violated.

No one has the right to wrongly imprison you or force you to leave your country.

You have a right to a fair, public trial.

Everyone is innocent until proven guilty.

You have the right to privacy. No one can interfere with your reputation, family, home, or correspondence.

You have the right to travel.

You have the right to seek asylum in another country if you are persecuted in your own.

Everyone has the right to a nationality.

All consenting adults have the right to marry and to raise a family.

You have the right to own property.

Everyone has the right to belong to a religion.

You have the right to think and voice your opinions freely.

Everyone has the right to gather as a peaceful assembly.

You have the right to participate in the governance of your country, either directly or by helping to choose representatives in free and genuine elections.

You have the right to social security and are entitled to economic, social, and cultural help from your government.

Every adult has the right to a job, a fair wage, and membership in a trade union.

You have the right to leisure and rest from work.

Everyone has the right to an adequate standard of living for themselves and their family.

Everyone has the right to an education.

Everyone has the right to freely participate in the culture and scientific advancement of their community, and their intellectual property as artist or scientist should be protected.

We are all entitled to a social order in which we may enjoy these rights.

Everyone's rights and freedoms should be protected unless they obstruct the rights and freedoms of others.

No State, group, or person can use this Declaration to deny the rights and freedoms of others.

This is a simplified version of the UDHR. For the complete text, visit www.un.org

Legal bases

Swiss Federal **Constitution**:

“Every person has the right to **privacy** in their private and family life and in their home, and in relation to their mail and telecommunications.” (art. 13 al. 1)

“Every person has the right to be protected against the **misuse of their personal data**.” (art. 13 al. 2)

Legal bases

In Switzerland, there are DP laws at two levels:

- Federal (**FADP**)
- Cantonal (**LPrD**, etc.)

In Europe there is a General Data Protection Regulation (**GDPR**)

At the international level 128 countries out of 194 have data protection legislation:

https://unctad.org/en/Pages/DTL/STI_and_ICTs/ICT4D-Legislation/eCom-Data-Protection-Laws.aspx

Legal bases

In order to know which data protection laws apply in the context of scientific research, it is useful to ask yourselves five questions:

Q1: Am I processing personal data?

Q2: Where am I established (institutionally)?

Q3: Who am I (legal status)?

Q4: Where the data collection is done (+ targeting)

Q5: Am I subject to sector-specific laws?

Q1: Am I processing personal data ?

Processing	Personal data	Sensitive data
Any operation with data, irrespective of the means applied and the procedure, and in particular the collection, storage, use, revision, disclosure, archiving or destruction of data. (Art. 3 lit. a FADP	All information relating to an identified person. (Art. 3 lit. a FADP)	Personal data on: 1. Religious, ideological, political or trade-union; related views or activities 2. Health, the intimate sphere or the racial origin; 3. Social security measures; 4. Administrative or criminal proceedings and sanctions. (Art 3. lit. c FADP)

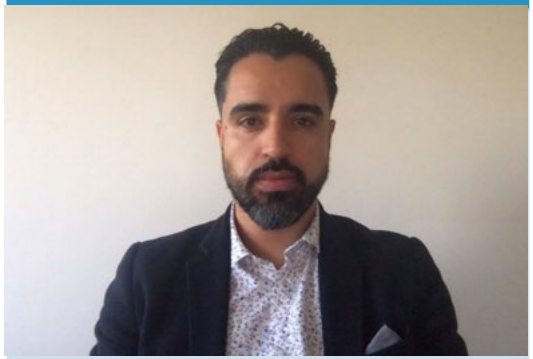
Personal data

"All information relating to an identified person" (Art. 3 lit. a FADP)

Very broad notion: everything that can be related to a specific person is personal data!

- The most common: name, date of birth, address, picture, etc.
- But also: opinions, original ideas, a style of writing, the way of walking, etc.

Examples of personal data

Contact details	Quotes	Picture
• First name: Paul • Last name: Dupont • Phone number: 123456 • Email: paul@dupont.ch	• “I would like to die on Mars. Just not on impact”	

FORS

The way you dance is unique, and computers can tell it's you

Nearly everyone responds to music with movement, whether through subtle toe-tapping or an all-out boogie. A recent discovery shows that our dance style is almost always the same, regardless of the type of music, and a computer can identify the dancer with astounding accuracy.

Studying how people move to music is a powerful tool for researchers looking to understand how and why music affects us the way it does. Over the last few years, researchers at the Centre for Interdisciplinary Music Research at the University of Jyväskylä in Finland have used motion capture technology—the same kind used in Hollywood—to learn that your dance moves say a lot about you, such as how extroverted or neurotic you are, what mood you happen to be in, and even how much you empathize with other people.

Even Anonymous Coders Leave Fingerprints

Researchers have repeatedly shown that writing samples, even those in artificial languages, contain a unique fingerprint that's hard to hide.

RESEARCHERS WHO STUDY stylometry—the statistical analysis of linguistic style—have long known that writing is a unique, individualistic process. The vocabulary you select, your syntax, and your grammatical decisions leave behind a signature. Automated tools can now accurately identify the author of a forum post for example, as long as they have adequate training data to work with. But newer research shows that stylometry can also apply to *artificial* language samples, like code. Software developers, it turns out, leave behind a fingerprint as well.

Rachel Greenstadt, an associate professor of computer science at Drexel University, and Aylin Caliskan, Greenstadt's former PhD student and now an assistant professor at George Washington University, have found that code, like other forms of stylistic expression, are not anonymous. At the DefCon hacking conference Friday, the pair will present a number of studies they've conducted using machine learning techniques to de-anonymize the authors of code samples. Their work, some of which was funded by and conducted in collaboration with the United States Army Research Laboratory, could be useful in a plagiarism dispute, for instance, but also has privacy implications, especially for the thousands of developers who contribute open source code to the world.


(DON’T) WORRY, YOU ARE
UNIQUE!

Sensitive data

Personal data on: religious, ideological, political or trade-union related views or activities; health, the intimate sphere or the racial origin; social security measures; administrative or criminal proceedings and sanctions (Art3. lit. c FADP)

The list provided by FADP is exhaustive (e.g. in Switzerland salary is not considered sensitive data)

That said, depending on the context, almost all data can be considered sensitive (name, photo, job, etc.)

Examples of personal data

Contact details	Quotes	Picture
•First name: Pedro •Last name: Ruiz •Gender: Trans •Job: Trade-Unionist	•“We are now in the process of defeating the radical left, the Marxists, the anarchists, the agitators, the looters, and people who, in many instances, have absolutely no clue what they are doing”	

Sensitive data?

Kayan people (Myanmar)

From Wikipedia, the free encyclopedia
(Redirected from Kayan (Burma))

For the ethnic group from Borneo, see Kayan people (Borneo).

The **Kayan** are a sub-group of Red Karen (Karenni people), Tibeto-Burman ethnic minority of Myanmar (Burma). The Kayan consists of the following groups: Kayan Lahwi (also called **Padaung**, ပဒေါင်[bədàʊ̃]), Kayan Ka Khaung (Gekho), Kayan Lahta, Kayan Ka Ngan. Kayan Gebar, Kayan Kakhi and, sometimes, Bwe people (Kayaw). They are distinct from, and not to be confused with, the Kayan people of Borneo.

Padaung (Yan Pa Doung) is a Shan term for the Kayan Lahwi (the group in which women wear the brass neck rings). The Kayan residents in Mae Hong Son Province in Northern Thailand refer to themselves as Kayan and object to being called Padaung. In *The Hardy Padaungs* (1967) Khin Maung Nyunt, one of the first authors to use the term "Kayan", says that the Padaung prefer to be called Kayan.[1] On the other hand, Pascal Khoo Thwe calls his people Padaung in his 2002 memoir, *From the Land of Green Ghosts: A Burmese Odyssey*.[2]

In the late 1980s and early 1990s due to conflict with the military regime in Myanmar, many Kayan tribes fled to the Thai border area.[3] Among the refugee camps set up there was a Long Neck section, which became a tourist site, self-sufficient on tourist revenue and not needing financial assistance.[4]

Kayan

ကယန်း

The importance of databases / knowledge

For certain information to be linked to someone, a number of conditions must be met:

- There must be the existence of databases that link the information to an identity

  NB: The Internet can function as a giant database

- It is sometimes necessary to know the persons to recognize them (importance of who has access to the information)

Is “public” data personal data?

Sometimes people make their personal data available to everyone. For example via social networks, blogs, websites, the press, etc.

- Even if these data could be qualified as “public”, they are still personal data and must be treated with care.
- It is important to check the privacy policies of the places where data are collected. Some types of processing may be prohibited.

Am I processing personal data?

Yes → **DP laws apply**

No → **DP laws don't apply**

⚠ If you plan to anonymize the data, be aware that they remain personal until you do so (DP laws apply until anonymization is complete).

⚠ Pseudonymized data are personal data.

Am I processing personal data?

- In the social sciences, it is very difficult to have completely anonymous data.
- To be considered anonymous, **all information** that can be linked to an identifiable person must be **permanently destroyed**.
- However, as we have seen, a lot of information can potentially allow the identification of an individual.
- It is therefore generally **safer to assume that we are dealing with personal data**.

Anonymization

Although anonymization is difficult to achieve, removing identifying information is still an important **layer of protection**.

- This layer of protection can be **combined** with access control or individual consent.
- The degree of de-identification should be decided based on an **assessment** of the **likelihood** of re-identification as well as the **risks** associated with disclosure of personal data.

Anonymization

The likelihood of re-identification can be assessed by asking the following questions:

- What types of **direct or indirect identifiers** do my materials contain?
- What **combinations of variables** can allow identification of an individual?
- Can information from **other sources** be linked to the data to make identification possible?
- What are the chances that an acquaintance will access the data?

Risk management

The risks associated with a disclosure of the data are assessed by:

- Determining the potential **impacts** on the lives of the persons concerned (physical, economic, social, psychological...)
- Estimating the potential **severity**, i.e. the magnitude of the risk (negligible, limited, significant, maximum)
- Estimating the **likelihood**, i.e. the possibility that the risk will occur (negligible, limited, significant, maximum)

Q2: Where am I established?

Researchers are **subject to the laws of the country in which their affiliated institution is based.**

- As employees of a Swiss-based institution, we are subject to Swiss law.
- If we were employees of an institution established in France, we would be subject to French law.

This criterion is independent of the place where the data are collected!

Q3: Who am I?

In Switzerland, the “legal status” of the data controller determines which law applies (federal or cantonal).

Private person/company	Federal body	Cantonal body
Federal laws apply	Federal laws apply *EPFL, ETH, FORS, etc*	Cantonal laws apply *UNIL, UNIGE, UNIBE, UNIZH, USI, HES, etc.*

Q4: Where is the data collection done?

What matters here is where the data subjects are when the data are collected as well as the intention to target (or not) specific populations.

- If you **travel** to a country to collect data from people there, the laws of that country apply (whether or not these people are residents).
- When **targeting** populations based on their country of origin/residence, the laws of those countries may apply even if you don't leave your office (online questionnaire). E.g. GDPR.

Q5: Am I subject to sector-specific laws?

Certain sectors of activity are regulated by specific laws that may contain **additional DP requirements**.

- Researchers working on diseases as well as on the structure and function of the human body have their activities regulated by the Human Research Act (HRA).
- Researchers working with federal and/or cantonal statistics are subject to federal and/or cantonal specific laws.
- Etc.

General principles

Main legal principles in Switzerland

Swiss law sets out 6 main principles of data protection that must always be respected (cumulatively):

1. Personal data may only be processed **lawfully**
2. Processing must be carried in a **transparent** manner
3. Data collection must be **proportionate**
4. Personal data may only be processed for the **purpose** indicated at the time of collection
5. Anyone who processes personal data must make certain that they are **correct**
6. Personal data must be protected against unauthorised processing through adequate **technical and organisational measures**.

Data protection in questions

Am I allowed to process personal data?

Swiss law generally allows researchers to process data for research purposes under two conditions:

1. Data must be **anonymized** once the purpose of the project is achieved
2. Results must be **published in a form that does not allow for the identification of individuals**

Compliance with these two conditions provides researchers with a "**research privilege**" that allows them to make exceptions to certain principles:

- Process personal data **without any other legal basis** (important for sensitive data)
- Process personal data **for a purpose other than the original purpose**

Do I need to obtain consent to process personal data?

- Swiss law **does not necessarily require researchers to obtain consent** to process personal or sensitive data in the course of their research activities.
- That said, consent is often **the only legal basis available** to the researcher.

  The ETH Act is the only explicit legal basis that exists.

  "Public" data is usable
- The use of consent is reinforced by the fact that **informing** individuals of any collection of personal data about them is always necessary (even when the data are collected from third parties).

How to ensure that the consent is valid?

To be valid, consent to process personal data must be **free** and **informed**.

- The person should not be **pressured** to participate

  - No hierarchical relationships

  - No excessive compensation

- The person should receive **all the necessary information**

  - Without knowledge there is no consent

  - We have a right to know what we are committing to

How to ensure that the consent is valid?

Regarding the form of consent:

- Consent may be: **oral** or **written**

  This said, it's always useful to have proof

- Where **sensitive** data are involved, consent must be **explicit**

  Simply answering a questionnaire, for example, cannot be considered as consent.

What information do I need to provide?

- The **identity** of the researchers in charge of the project
- **Understandable** statements describing the purpose of the research, the nature and duration of participation, and the research methods
- A **clear description** of the foreseeable risks and benefits of participation
- The **nature** of the data collected and their usefulness
- An **honest and complete description** of the protection/security measures

What information do I need to provide? (continued)

- The guarantee of being **free to decide** not to participate in the project, to withdraw without losing acquired rights and to have the possibility at any time to continue or not to participate
- The right to **access** and **rectify** data
- The existence of any **conflict of interest**
- **Preservation** and **reuse** of data
- Contracts with **third parties**
- The possibility of being informed of the **results**

Exceptions to information

Swiss law provides for a number of restrictions on the obligation to inform:

- It is often possible to restrict or defer information if an **overriding public interest** requires it
- A researcher who wants to collect personal data without informing the persons concerned - or very partially - could do so provided that he/she is able to prove the existence of such interests (= very difficult)
- As soon as the reason for the restriction of the obligation to inform disappears, the controller must provide the information

Deception is easier to justify than covert observation!

Can personal and sensitive data be kept indefinitely?

- Researchers who wish to take advantage of research privilege must anonymize the personal data they process as **soon as the purpose of the processing permits**
- The **purpose** of the processing is a key notion in the sense that it determines the time limit for the retention of personal data
- If medium- to long-term preservation is envisaged, it must be **consistent with the purpose of the collection**. For this reason, participants in a research project should be clearly informed of its objectives and time frame
- Individuals have a right of **access** to their data

Can personal and sensitive data be shared?

- It is possible for a researcher to disclose personal and sensitive data to third parties **for research purposes.**
- If consent is not required for such disclosure, the individuals whose data are disclosed must nevertheless be **informed**.
- If the researcher plans to disclose the data **abroad**, a number of special provisions apply:

  - *Personal data may only be transferred to a third country if it ensures an **adequate level of protection** (the Federal Data Protection and Information Commissioner maintains a list of countries offering such guarantees).*

  - *If the country does not offer an adequate level of protection, contractual **measures** must be taken or consent must be obtained.*

Can personal and sensitive data be shared?

- The opening of research materials containing personal data via archive services (such as FORS) raises a number of challenges (information).
- It may be possible, in some cases, to avoid systematically informing individuals each time data are communicated. This would require that research projects be designed to include data sharing as one of their objectives, e.g. history).
- The purposes of processing (study goals) and the categories of recipients (researchers who would have access to the data) should, however, be as recognizable as possible. A research project that deviates too much from the initial purpose is therefore likely to be problematic. Some control would therefore be required.

Questions?