Data protection and open research data

Pablo Diaz, FORS / UNIL

Workshop: Safe access to sensitive research data
November 25, 2022, SNF

Legal bases

Anyone who processes personal data must comply with a number of rules and **laws**.

In Switzerland, there are laws at two levels:

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

Two useful questions

Q1: Am I processing personal data?

Q3: Who am I (legal status)?

Q1: Am I processing personal data ?

Processing : 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

Anything you do with data = processing

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: names, addresses, phone numbers, picture, etc. (*direct identifiers*)

But also: original ideas, opinions, quotes, etc. (*indirect identifiers)*

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.

Chinese ‘gait recognition’ tech IDs people by how they walk

By DAKE KANG November 6, 2018

Click to copy

BEIJING (AP) — Chinese authorities have begun deploying a new surveillance tool: “gait recognition” software that uses people’s body shapes and how they walk to identify them, even when their faces are hidden from cameras.

COMPUTER SCIENCE

Artificial intelligence unmasks anonymous chess players

Software that identifies unique styles poses privacy risks

By **Matthew Hutson**

Think your bishop's opening, queen's gambit, and pawn play are unique? A new artificial intelligence (AI) algorithm has got your chess style pegged.

AI software can already identify people by their voices or handwriting. Now, an AI has shown it can tag people based on their chess-playing behavior, an advance in the field of "stylometrics" that could help computers be better chess teachers or more humanlike in their game play. Alarmingly, the system could also be used to help identify and track people who think their online behavior is anonymous.

"Privacy threats are growing rapidly," says Alexandra Wood, a lawyer at the Berkman Klein Center for Internet & Society at Harvard University. She says studies like this one, when conducted responsibly, are useful because they "shed light on a significant mode of privacy loss."

Chess-playing software, such as Deep Blue and AlphaZero, has long been superhuman. But Ashton Anderson, a computer scientist at the University of Toronto and principal investigator of the new project, says the chess engines play almost an "alien style" That required the system to recognize what was distinctive about each player's style.

The researchers tested the system by seeing how well it distinguished one player from another. They gave the system 100 games from each of about 3000 known players, and 100 fresh games from a mystery player. To make the task harder, they hid the first 15 moves of each game. The system looked for the best match and identified the mystery player 86% of the time, the researchers reported last month at the Conference on Neural Information Processing Systems (NeurIPS). "We didn't quite believe the results," says Reid McIlroy-Young, a student in Anderson's lab and the paper's primary author. A non-AI method was only 28% accurate.

"The work is really cool," says Noam Brown, a research scientist at Meta (the parent company of Facebook) who has developed superhuman poker bots. He looks forward to chess bots that mimic Magnus Carlsen, the reigning world champion, and says style-aware AI could transform other computer interactions. "There's a lot of interest in chatbots, where you can have a chatbot that would speak in the style of Albert Einstein or something," he says.


(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 lists provided by the different laws are **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 sensitive data

Contact details	Quote	Picture
First name: Sebastián Last name: Calfuqueo Gender: trans Job: trade unionist	« Today, our developed nations live in opulence, excess and waste, with the consequences that our environment is degrading and the climate is wreaked »	

About anonymisation

- 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 (≠ coding /pseudo).**
- It is therefore generally **safer to assume that we are dealing with personal data.**
- De-identification is a layer of protection

Q2: 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.

Some important questions

Do I need to obtain consent to process personal data?

- Public institutions (such as universities, etc.) generally operate under the principle of **legality** (≠ lawfulness).
- There must always be a legal basis
- Consent is often the only legal basis available for processing **sensitive data.**
- Where **sensitive** data is involved, consent must be **explicit** (oral or written)
- Individuals must be informed of any planned **disclosure** of data (categories of recipients).

Can personal and sensitive data be kept indefinitely?

- Any data collection must have a **purpose** that is identifiable by (communicated to) individuals
- The **purpose** determines the **time limit** for the retention of personal data (once the purpose is fulfilled -> destruction)
- If medium- to long-term preservation is envisaged, it must be **consistent with the purpose of the collection**

⚠ Do not set a purpose that is too restrictive
⚠ Do not promise data destruction too soon

Can personal and sensitive data be shared?

Yes, on condition…

- Personal data may only be used for the purpose for which it was collected.
- Personal data collected for research purposes cannot be used for other purposes (opposite ≠ true).
- When sharing personal data for research purposes, a few conditions must be met:
  - No disclosure of identifying information in publications
  - Destruction of data once used
  - Information of collection

Where to go from here ?

- Only researchers should access personal / sensitive data collected for research purposes (more control)
- Reuse in the framework of a research project that deviates too much from the initial purpose is likely to be problematic (anticipate that)
- Research projects have to be designed to include data sharing as one of their objectives.
- Information and consent must provide for archiving / sharing (in order to avoid systematic information)
- Purpose of the reuse and categories of recipients must be as clear as possible

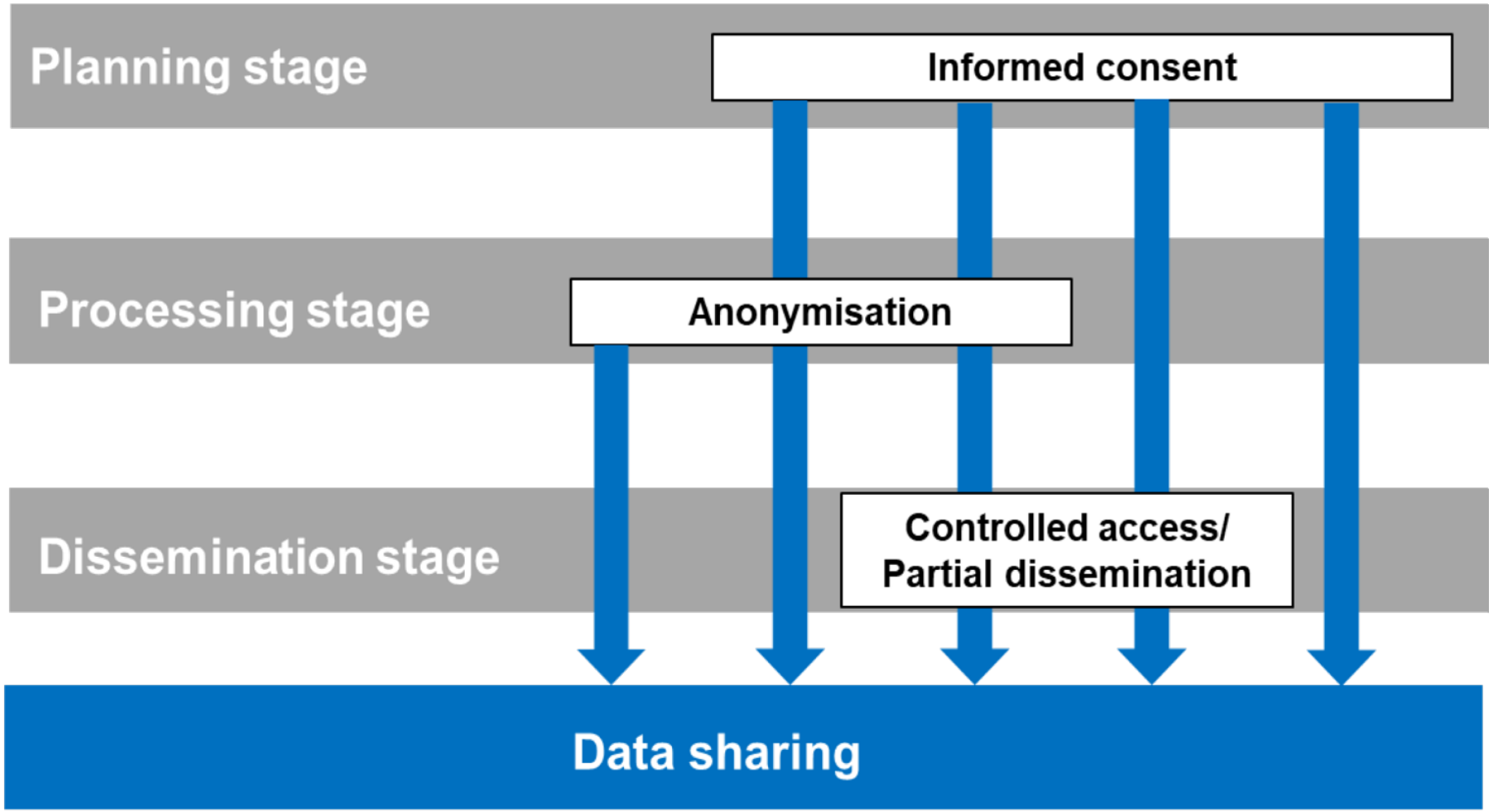
Can personal and sensitive data be shared?

If data is being shared abroad a number of special provisions apply:

- Personal data may only be transferred to a third country if the third country in question 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**

Towards a layered approach



Thank you!