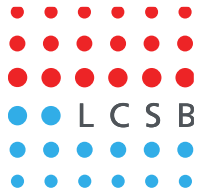




CC BY 4.0

# Data Protection in Biomedical Research

*Pinar ALPER*



*Training on "Best practices in research data management and stewardship"*

*15 June 2021*

**Data Protection  $\subset$  Data Management**

# GDPR basics

- “Principles”
- “Personal data” “Special category/sensitive personal data”
- “Data processing”
- “Pseudonymised data”
- “Supervisory authority”
- “Data protection officer”
- “Processor” “Controller”

# GDPR “Principles”



- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Confidentiality and Integrity
- + Accountability: "The controller shall be **responsible** for, and be able to **demonstrate compliance** with ..."

The GDPR gift that keeps on giving.

# “Data processing”

— Any liaison with the data is “data processing”



“Any operation [...], such as **collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;**[...]” Art. 4 (2)

# “Personal data”, “Data subject”

“Any information relating to an **identified** or **identifiable** natural person (data subject) [...]” **Art. 4 (1)**



“Can be identified, **directly** or **indirectly** by identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to [...] the genetic [...] identity of that natural person” Art. 4 (1)

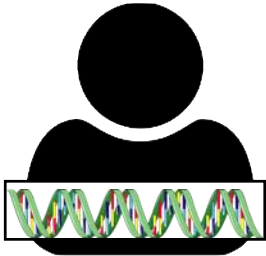
# Identifiers

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, zip code or equivalents except for the initial 3 digits of a zip code if the corresponding zone contains more than 20,000 people.
3. All elements of dates (except year) for dates directly related to the individual (birth date, admission date, discharge date, date of death). Also all ages over 89 or elements of dates indicating such an age.
4. Telephone numbers
5. Fax numbers
6. E-mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan numbers
10. Account numbers
11. Certificate or license numbers
12. Vehicle identification or serial numbers including license plate numbers
13. Device identification or serial numbers
14. Universal resource locators (URL's)
15. Internet Protocol addresses (IP addresses)
16. Biometric identifiers
17. Full face photographs and comparable images
18. Any other unique identifying number, characteristic, or code

*HIPAA Health Insurance Portability and Accountability Act,*

# Genetic data is “personal data”

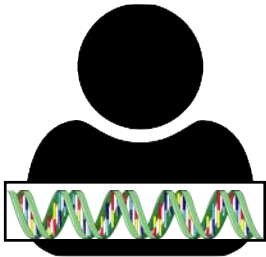
“Can be identified, **directly** or **indirectly** by identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to [...] the **genetic** [...] **identity** of that natural person” Art. 4 (1)



“ **Genetic data** .. personal data relating to the inherited or acquired genetic characteristics of a natural person **which give unique information about the physiology or the health of that natural person**” [...] Art. 4 (13)



# Genetic data is “special category (sensitive) personal data”



“ **Genetic data** .. personal data relating to the inherited or acquired genetic characteristics of a natural person **which give unique information about the physiology or the health of that natural person**” [...] Art. 4 (13)

“... racial or ethnic origin, [...] genetic data, [...], data concerning health ... “ Art. 9 (1)

# “Pseudonymised data”

— is personal data.



... personal data [that] can no longer be attributed to a specific data subject without the use of additional information[...] **Art. 4 (5)**

**Pseudonymisation** is a significant data protection measure!

# “Supervisory Authority”

Perform audits

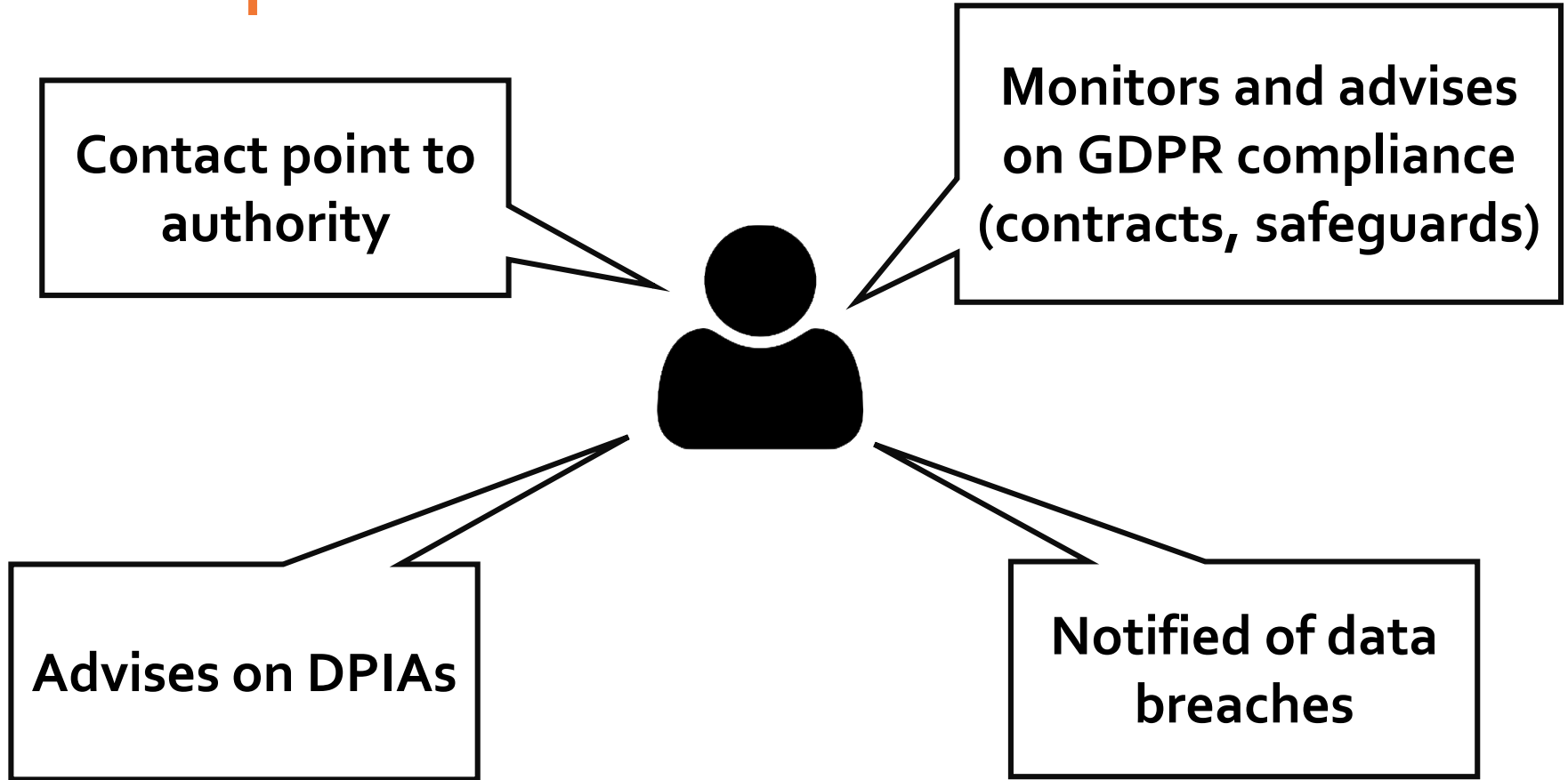
Handle data subjects' complaints

Administer penalties  
e.g. fines, bans on processing



Notified of breaches

# “Data protection officer”



# "Controller" & "Processor"



I determine the purposes and means of the processing of personal data

- Full responsibility for lawful, fair and transparent processing of data



I only process personal data on behalf of the controller

- Only acts on written instructions of the controller
- Decision making limited to technical aspects
- Responsibility limited to security safeguards
- Supports the controller in data protection accountability
- Deletes or hands over data after the project / service

**GA4GH GDPR Brief: Are university-employed scientific researchers 'Data Controllers' for the purposes of the GDPR? (May 2020)**

# Most research consortia are joint-controllers

We'd like to do genome analysis  
on your LRRK2 family cohort  
and compare it with our ...



« (Joint) Controller »

agreed, here's our cohort's genome data



« (Joint) Controller »

# “Legal basis”

- **Consent** — Subject gives permission to controller to process their personal data for one or more processing activities. Consent must be freely given, clear, and easy to withdraw. Consent must be GDPR-compliant, opt-out should be the default option.
- **Performance of a Contract** — The processing is necessary to enter into or perform a contract with the data subject.
- **Legitimate Interest** — This is a processing activity that a data subject would normally expect from an organization that it gives its personal data to do, like website security and optimisation.
- **Vital Interest** — A rare processing activity that could be required e.g. to save data subject's life.
- **Legal Obligation** — The processing activity is necessary for a legal obligation that controller must adhere, such as an information security or employment law.
- **Public Interest** — A processing activity that would occur by a government entity or an organization acting on behalf of a government entity.

# “Subject’s rights”

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision making and profiling



**...so, what does it all mean for research?**

# Research with Personal Data

1. DMP's must address ethics and legal requirements  
Data Protection Impact Assessment (DPIA) is required for certain projects.
2. **Ethics review** is always required for working with human biosamples and data. This is also a measure for data protection.
3. Data acquisition must be **legal**
4. Projects and data need to be **documented**
5. Data should be protected through **technical** and **organisational measures**

# 1 Data Protection Impact Assessment DPIA

# Data Protection Impact Assessment - DPIA

- **DPIA** is a process to help you identify and minimise the data protection risks of a **project**.
- You must do a **DPIA** for processing that is likely to result in a **high risk to individuals**.
- What constitutes high-risk processing is laid out by the European Data Protection Board and the National Authority in our case the CNPD.

# What is high-risk processing

*Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data.*


- GDPR sets out three types of processing, which **always** require a DPIA:
  - Systematic and extensive profiling with significant effects
  - Large scale use of sensitive data
  - Public monitoring

# DPIA- indicators of high risk processing

- using new technologies
- tracking people's location or behaviour
- systematically monitoring a publicly accessible place on a large scale
- processing special category/sensitive personal data
- processing is used to make automated decisions about people that could have legal (or similarly significant) effects
- processing children's data
- processing could result in physical harm to the data subjects if it is leaked

"In most cases, a combination of two of these factors indicates the need for a DPIA. However, this is not a strict rule.... On the other hand, in some cases you may need to do a DPIA if only one factor is present – and it is good practice to do so." <https://ico.org.uk/>

# DPIAs for research projects

- Projects requiring a DPIA should NOT commence their data processing without the completion of the DPIA.
- DPIA process often goes in parallel to the following processes:
  - Ethics Review
  - DMP preparation
- DPIA is a type of risk assessment.
  - If you identify a high risk that you cannot mitigate, you must consult the national authority before starting the processing.
- Performed with support from DPO, Research, Data, and IT Support offices.

# DPIA Responsibilities

	Responsible	Accountable	Consulted	Informed
Top Management		<b>X</b>		
Business owner	<b>X</b>			
DPO			<b>X</b>	
IT department			<b>X</b>	
Processors, where relevant			<b>X</b>	
Data subject representatives			<b>(X)</b>	

Figure 2: RACI matrix DPIA process



# Who performs the DPIA



- Perform DPIA
  - Data protection point of contact
    - At the LCSB this role is assumed by Data Stewards
  - Risk manager
  - Processing owner (Research PI)



- Review & support
  - Data Protection Officer (DPO)
  - Chief Information Security Officer (CISO)

# DPIA process



# DPIA Tool – European Union Agency for Cybersecurity (ENISA)

<https://www.enisa.europa.eu/risk-level-tool/risk>

The screenshot displays the ENISA DPIA Tool interface. On the left, a vertical navigation pane contains six steps: 1. Definition and Context of the Processing Operation (highlighted), 2. Impact evaluation, 3. Threat Analysis, 4. Risk Evaluation, 5. Security Measures (with a 'Help' button), and 6. Export the analysis and the proposed measures. The main content area is titled '1. Definition and Context of the Processing Operation'. It contains two paragraphs of text explaining the step's purpose and providing a reference to the ENISA report 'Handbook on Security of Personal Data Processing'. Below the text are two input fields: 'Processing Operation Description' and 'Personal Data Processed'.

**1. Definition and Context of the Processing Operation**

This step is the starting point of the risk assessment and is fundamental in order to define the boundaries of the data processing operation (under assessment) and its relevant context. In doing so, the organization needs to consider the different phases of the data processing (collection, storage, use, transfer, disposal, etc.) and their subsequent parameters. Specific attention has to be paid to the fact that the analysis below regards a specific processing operation; a data processing system may comprise of more than one data processing operations. The analysis below has to be performed for each processing operation.

An overview of the output and provisional examples on how to describe data processing operations are available within the uses cases (Sections 4,5,6 & 7) of the ENISA report "[Handbook on Security of Personal Data Processing](#)".

**Processing Operation Description**

Descriptive title of the processing operation

**Personal Data Processed**

Type of personal data to be processed (e.g. last and first name, address, social security, etc.)

- One directional wizard like interface
- Risks and measures not re-usable across DPIAs

# DPIA Supporting tool – MONARC

<https://www.monarc.lu>

The screenshot shows the MONARC web application interface. The header displays 'Accueil > LCSB risk analysis'. The main content area is titled '8 risques de l'information'. A table lists various risks with columns for Impact, Menace, Vulnérabilité, Risque actuel, and Risque résiduel.

Actif	Impact			Menace		Vulnérabilité		Risque actuel			Traitement	Risque résiduel	
	C	I	D	Libellé	Prob.	Libellé	Mesures en place	Qualif.	C	I			D
IT organization	-	-	-	Error in use	-	No document base for rules and procedures		-	-	-	-	Non traité	-
IT organization	-	-	-	Forging of rights	-	Logical access authorisations are not checked regularly		-	-	-	-	Non traité	-
IT organization	-	-	-	Denial of actions	-	No definition of responsibilities		-	-	-	-	Non traité	-
IT organization	-	-	-	Theft or destruction of media, documents or equipment	-	Physical access authorisations are not checked regularly		-	-	-	-	Non traité	-
IT organization	-	-	-	Abuse of rights	-	No coordination between the departments concerned before hiring		-	-	-	-	Non traité	-

- Focuses on risks and measures definition and re-use.

# DPIA Tool – CNIL PIA

<https://www.cnil.fr/en/pia-software-20-available-and-growth-pia-ecosystem>

Version 1.1.6

PIA | Privacy impact assessment

The screenshot displays the PIA tool interface. At the top, there is a dark blue header with 'DASHBOARD' on the left and 'Tools' on the right. Below the header, the main content area is divided into several sections. On the left, there is a sidebar menu with categories: 'CONTEXT' (with sub-items 'Overview' and 'Data, processes and supporting ...'), 'FUNDAMENTAL PRINCIPLES' (with sub-items 'Proportionality and necessity' and 'Controls to protect the personal r...'), 'RISKS' (with sub-items 'Planned or existing measures', 'Illegitimate access to data', 'Unwanted modification of data', 'Data disappearance', and 'Risks overview'), and 'VALIDATION' (with sub-items 'Risk mapping', 'Action plan', and 'DPO and concerned persons opin...'). The main content area is titled 'Context' and contains an 'OVERVIEW' section with the text: 'This section gives you a clear view of the treatment(s) of personal data in question. This part allows you to identify and present the object of the study.' Below this, there are two questions: 'Which is the processing under consideration?' and 'What are the responsibilities linked to the processing?'. The first question has a text box containing 'La plateforme Didomi contient les données des utilisateurs du client aux fins d'utilisation de la plateforme.' and a 'Comment' button. The second question has a text box containing 'Didomi est seul responsable des données des utilisateurs et sous-traite l'hébergement.' On the right side of the interface, there is a 'Knowledge base' section with a search bar and a list of items: 'Principle', 'Description du traitement', 'Definition', 'Responsable de traitement', 'Definition', and 'Sous-traitant'.

- Built-in list of measures not configurable via tool

# DPIA content

- Why a DPIA is deemed needed
- Purpose and duration of processing
  - Project partners their activities and GDPR roles
- Standards related to processing
- Scientific and GDPR categorization of data
- Data retention period and justification
- Project data flow diagram
- Platforms used for data processing

## 2 Ethics review

# Ethics review

- All research with human biosamples and data requires valid **consent** and **ethics approval**.
- From the GDPR perspective ethics review is a safeguard that addressed several principles
  - Legal basis, public benefit of the processing
  - Fair and transparent processing
  - Assessment of purpose of use and the data collected (If you want to use data for another research question this would require a new ethics review)



## 3 Legal data acquisition and transfer

# Data acquisition must be legal

- From subjects only with valid information sheet and consent
- From collaborators only with **contract**
- From repository only according to **access policy**

# Data transfers must be legal

- Data has been consented explicitly, allowing international transfers
- To countries with an adequacy decision;

The European Commission has so far recognised [Andorra](#), [Argentina](#), [Canada](#) (commercial organisations), [Faroe Islands](#), [Guernsey](#), [Israel](#), [Isle of Man](#), [Japan](#), [Jersey](#), [New Zealand](#), [Switzerland](#), [Uruguay](#) and ~~the United States of America~~ (limited to the ~~Privacy Shield framework~~) as providing adequate protection.

..... invalidated by the European Court of Justice's Schrems ruling

- Always with safeguards: EC approved contractual-clauses

# Data transfers must be legal

- Data can only be shared with a contract
- Contract should clearly foresee the roles (controller/processor) of research partners wrt personal data
- Contract preparation processes take time...

Standard contractual  
clauses

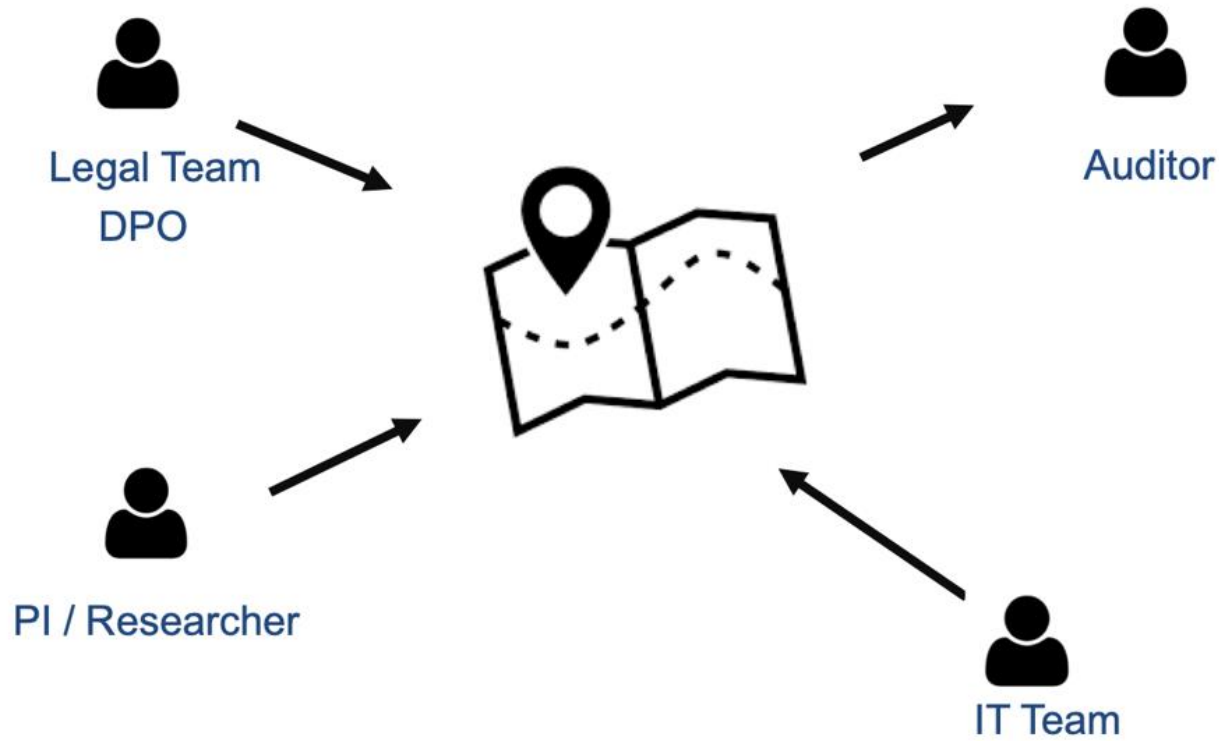


Contract Templates



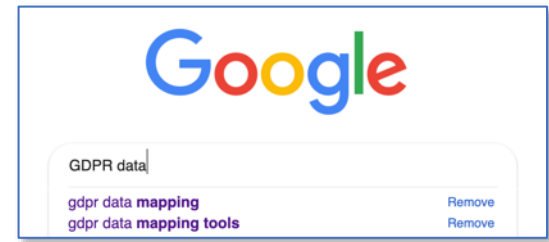
## 4 Projects and data need to be documented

# GDPR data mapping



# GDPR data mapping

- Process-oriented tools
- Good old spreadsheet
- ELIXIR Data Information System - DAISY



A screenshot of the DAISY website interface. The top navigation bar includes links for HOME, ABOUT, DATASETS, PROJECTS, CONTRACTS, and DEFINITIONS. The user name 'JOHN BLACK' is displayed in the top right. The main header features the DAISY logo (a stylized orange flower) and the text 'daisy Data Information System for GDPR compliance' with 'Version 1.4.0' in a small box. Below the header, there are three main sections: 'My datasets' with a list item 'EPIC-DREM Sequencing data' and an '+ Add new' button; 'My projects' with a list of 'GGE-I', 'EPIC-DREM', and 'LUX-Epigen' and an '+ Add new' button; and 'Help' with instructions to use links at the top of the page and a reference to the 'User Guide'.

# Data Information System - DAISY





# 5 Data should be protected through technical and organisational measures

# Data Protection Measures

- Organisational
  - DP Training
  - Procedures/Processes
    - New Project Instigation, Data Breach, Data Classification, Data Breach, DPIA
  - Research Data Policies
    - Storage & Backup, Retention, Deletion, Data Protection
- Technical
  - Encryption, Pseudonymisation
  - Other Infosec measures
    - Access Control, 2FA, Physical Security, Network Security

# Data Breach Reporting

- What counts as data breach?
  - external disk or laptop gets stolen
  - laptop (or server) is hacked
  - data is copied outside policy-allowed realms
  - exchange of data outside policy-allowed channels
  - data is processed without a contract or minimum conditions to ensure data protection responsibilities
  - deliberate/accidental disclosure of data
  - use of real data in demonstrators e.g. use of subject identifying data (e.g. photos) without their consent
- Timely reporting of breach to DPO and Authority is legal requirement



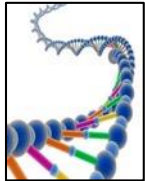
# Data Classification, primary criteria

- **Re-identifiability:** Data contains attributes that permits (potential) reidentification through singling out linking or inference.
  - Singling-out
  - Linking
  - Inference
- **Sensitivity:** Data contain attributes that can potentially cause harm to subject upon re-identification. E.g. disease status

# Data sensitivity, w and w/o pseudonymisation

Peter Pan

Subject X



Genome  
Sequence

Identified  
*Sensitive*

Pseudonymised  
*Sensitive*



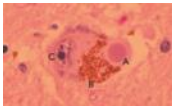
Bare, single  
measurement

Identified  
*Non-sensitive*

Pseudonymised  
*Non-sensitive*

# “Anonymised” data

- Where is the line drawn?

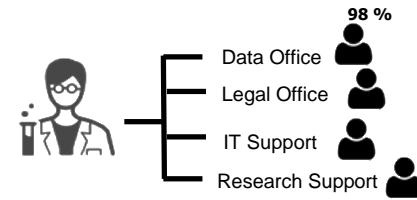
Gender	Age Range	Data
F	40-50	12,34...
M	30-40	

... also

- data downloaded from public repo
- commercially available human data (cellines)
- data that is claimed to be “anonymous” by the provider

The samples for the 1000 Genomes Project are anonymous and have no associated medical or phenotype data. The project holds self-reported ethnicity and gender. All participants declared themselves to be healthy at the time the samples were collected.

# Research with Personal Data



1. DMP's must address ethics and legal requirements

Data Protection Impact Assessment (DPIA) where necessary!

2. **Ethics review** is always required for working with human biosamples and data. This is also a measure for data protection.
3. Data acquisition must be **legal**
4. Projects and data need to be **documented**
5. Data should be protected through **technical** and **organisational measures**

# Practical with DAISY

- Document your research project or the example scenario project
  - Create a **Project** and a **Dataset** record in DAISY
  - Instructions

<https://tinyurl.com/dm-ds-ws-2021-06>

Sheet: Day 2