

Ethical and legal constraints on the sharing of personal data

Information security: categories of data

1. Public
2. Internal
3. Confidential
4. Restricted ('highly confidential')

Research data in terms of information security

1. Public
2. Internal
3. Confidential
4. Restricted ('highly confidential')

? ← RESEARCH DATA

Research data in terms of information security

1. Public
2. Internal
3. Confidential
4. Restricted ('highly confidential')

General rule

RESEARCH DATA

Research data in terms of information security

1. Public
2. Internal
3. Confidential
4. Restricted ('highly confidential')

*After the deliberate decision
that it is OK to publish*

RESEARCH DATA

Research data in terms of information security

1. Public
2. Internal
3. Confidential
4. Restricted ('highly confidential')



GENETIC DATA

**Because they are sensitive
personal data**

The ethics of personal data

- The basic right to decide on the collection and processing of your personal data
autonomy; the right of an individual to make his or her own choice
- The sharing/disclosure of personal data can lead to harm
non malice; above all, do no harm

Personal data

'personal data' means **any information relating to an identified or identifiable natural person** ('data subject');

an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

In the context of the General Data Protection Regulation, data of deceased persons are not considered personal data

Sensitive personal data

- Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, **genetic** data, biometric data, **health** data, data on sex life or sexual orientation

The legislation

- Regulation (EU) 2016/679: the 'General Data Protection Regulation'
- Belgian law of 30 July 2018 on the protection of natural persons concerning the processing of personal data

General principles

1. Lawfulness, fairness and transparency
2. Purpose limitation
3. Data minimization
4. Accuracy
5. Storage limitations
6. Confidentiality and integrity

Lawfulness

6 possible legal grounds

1. The data subject has given **consent**
2. Necessary for the performance of a **contract**
3. **Legal obligation** of the data controller
4. Protection of **vital interest** of the data subject or another person
5. Necessary for a task of **public interest**
6. Necessary for purposes of the **legitimate interest** of the controller or a third party

Purpose limitation

- Only process personal data for your particular research purpose
- Processing of data should be reasonable and proportionate for achieving the research goals

Data minimization

- Only use data that are necessary to achieve the objective

Storage limitations

- Personal data may not be kept longer than necessary for your current research and for possible further analyses of data

-> but principle of Open Data resulting in uploading of datasets onto repositories for future use

Rights of the data subject

Transparent information on exercise of rights

1. Right to **information** and access to personal data
2. Right to **rectification**
3. Right to **erasure**
4. Right to **restriction** of processing
5. Right to data **portability**
6. Right to **object**

Limits to the right of erasure

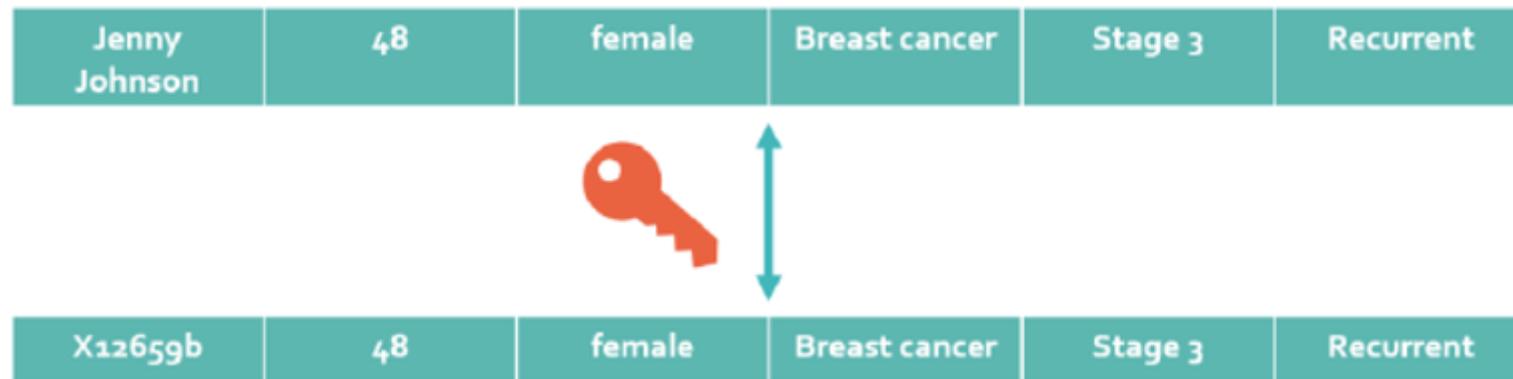
- Right of erasure in the context of scientific research is limited, to the extent that the exercise of the right **would render impossible or seriously impair the scientific research**
 - > Data that have already been generated can still be used.

Obligations of the controller

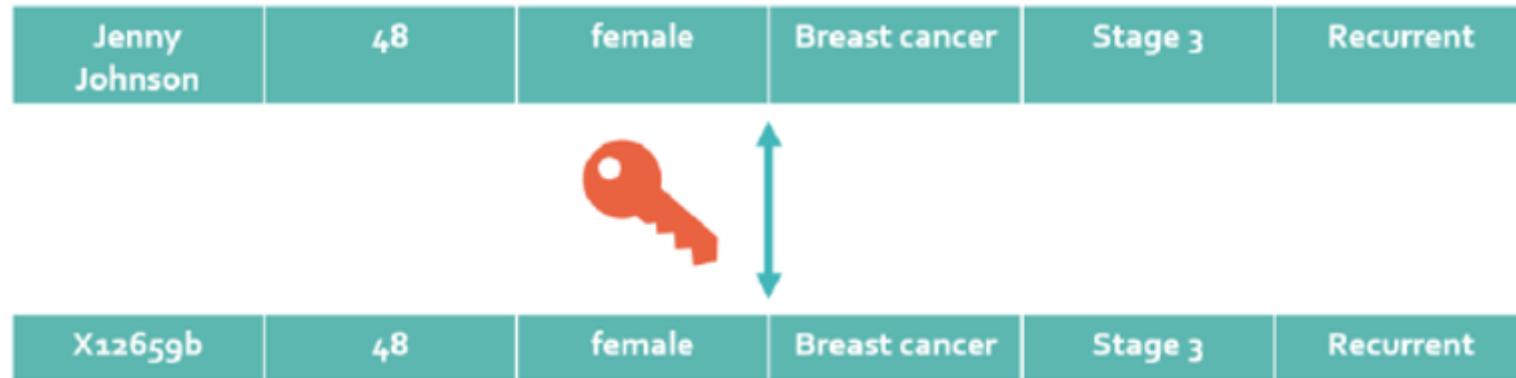
- Implement appropriate technical and organisational measures
- Ensure security of personal data
- Keep records of processing activities -> GDPR processing log
- Notification of personal data breach
- Perform data protection impact assessments where necessary
- Ensure agreements with processors and recipient third parties

Pseudonymous data

- Personal data that has been processed in such a manner that the personal data can no longer be attributed to a specific data subject without the use of **additional information**
 - ▶ The additional information is kept separately and is subject to measures to ensure that this information can only be accessed by privileged persons
 - ▶ Pseudonymization is regarded an important security measure



Pseudonymous data



Not being in the possession of the decoding key does not make a dataset anonymous!

Anonymous data

- information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is **not or no longer identifiable**
 - ▶ Anonymization must be **irreversible**
 - ▶ Methods of anonymization:
 - **Removal** of information that would allow to identify a person
 - **Randomization**
 - **Aggregation**

Identifiability

- To determine whether a natural person is identifiable, account should be taken of all the means **reasonably likely** to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, **account should be taken of all objective factors**, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments

```
graph LR; A[Generate research data] --> B[Process research data]; B --> C[Store research data]; C --> D[Share research data]
```

Generate research data

Process research data

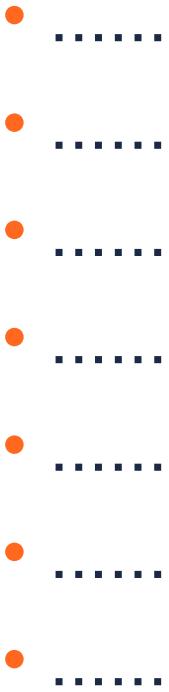
Store research data

Share research data

Manage research data that constitute personal data in a way that is lawful,
and enables you to do what you need and want to do with these data

Consequences of the GDPR for the way you manage your research data

List up possible consequences

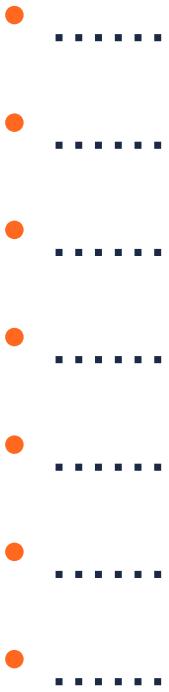


Transparency

- Data subjects must be informed about what their data will be processed
 - They are responsible for processing personal data
- > include information in the ICF
- Management of personal data includes
informing people on what you are going
to do with their data**

You want to inform participants correctly on how their data will be processed

Which information elements would you then take up in your informed consent form?



- The type of data that will be collected
- That data will be coded, researcher will not know the identity
- That processing will be done in a GDPR compliant manner with full protection of privacy
- That coded data may be uploaded into a repository
- That coded data may be shared with third parties (mention recipient if already known)
-

Processing of personal data in scientific research – which legal ground?

1. The data subject has given **consent**
2. Necessary for the performance of a **contract**
3. **Legal obligation** of the data controller
4. Protection of **vital interest** of the data subject or another person
5. Necessary for a task of **public interest**
6. Necessary for purposes of the **legitimate interest** of the controller or a third party

Processing of personal data in scientific research – which legal ground?

1. **The data subject has given consent**
2. Necessary for the performance of a contract
3. Legal obligation of the data controller
4. Protection of vital interest of the data subject or another person
5. **Necessary for a task of public interest**
6. Necessary for purposes of the legitimate interest of the controller or a third party

ICFs in (biomedical) research

- The participant needs to give **explicit consent to participation** to the study
- But the collection and processing of personal data can be done on the legal ground of **public interest**

Entering your project in the GDPR processing log

In Leuven

- Privacy & Ethics questionnaire

Clarification of PRET

You can submit an application for privacy and / or ethical assessment via the [PRET application](#).

You must first agree to the KU Leuven's GDPR code of conduct and apply these rules. You will then receive privacy and / or ethical questions about your research.

Below you will find clarification of the privacy-related questions from the PRET application:

1. [Data controller vs processor](#)
2. [Do you intend to collect new data \(primary processing\) or only use previously collected data \(secondary processing\)?](#)
3. [Categories of data subjects](#)
4. [Large-scale / small-scale processing](#)
5. [What categories of data do you intend to collect or use? Will you collect "ordinary" personal data and/or "special category" data?](#)
6. [Technical and organisational measures](#)
7. [Import/export of data](#)
8. [Will the necessary information be provided to the data subjects or has it already been provided?](#)
9. [Derogations from data subject rights](#)
10. [Lawfulness of processing](#)
11. [Risk assessment: what is a data protection impact assessment \(DPIA\)?](#)

In Gent

The header includes the DMP ONLINE .BE logo, navigation links for My Dashboard, Create plans, Reference, and Help, and contact details for www.ugent.be, Ghent University RDM webpages, and email support.

DMP
ONLINE .BE

My Dashboard Create plans Reference Help

www.ugent.be Send mail to request DMP feedback (attach pdf!)

Ghent University RDM webpages rdm.support@ugent.be

GDPR test

The form displays a list of GDPR requirements with counts and expand/collapse buttons.

Requirement	Count
GDPR (0 / 1)	0 / 29
Collection and processing of personal data (0 / 7)	0 / 29
Categories of personal data & data subjects (0 / 5)	0 / 29
Purpose(s) of the processing (0 / 3)	0 / 29
GDPR responsibility (0 / 3)	0 / 29
Data transfers & categories of recipients (0 / 5)	0 / 29
Retention period (0 / 1)	0 / 29
Risk analysis (0 / 2)	0 / 29
Security measures (0 / 2)	0 / 29

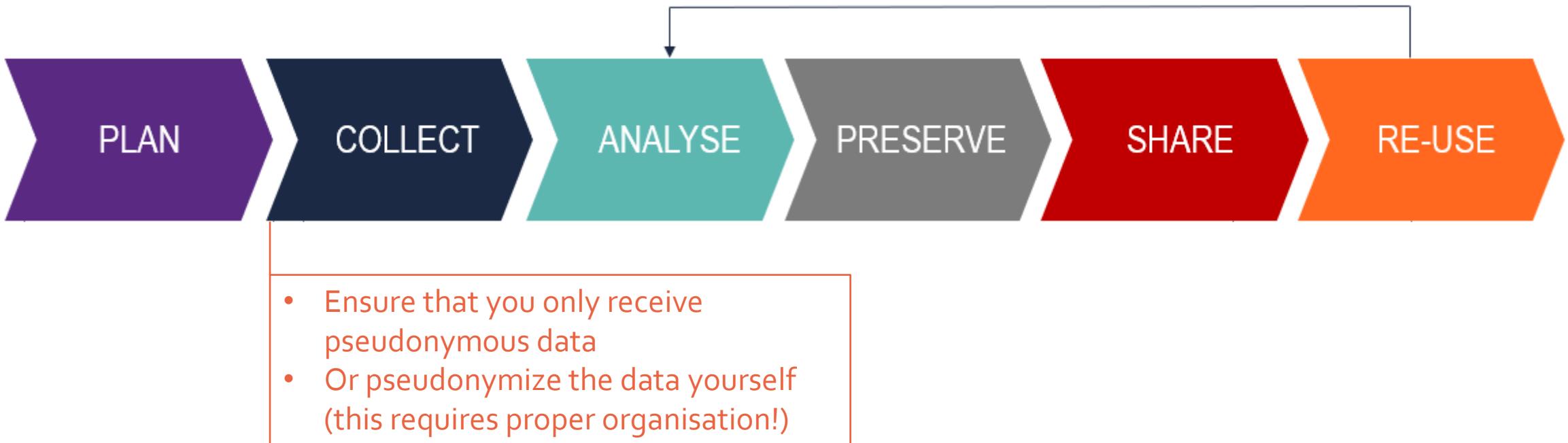
Management of personal datasets in the research life cycle



Step 1. The necessary registrations and approvals to enable you to collect personal data



Step 2: ensure that the researchers only work with pseudonymous data



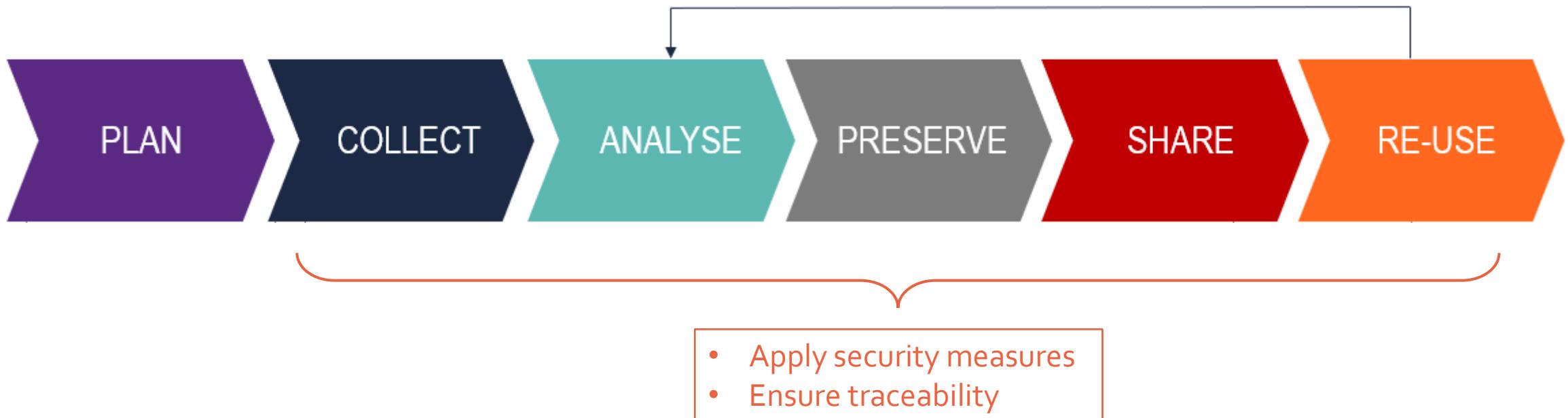
Step 3: Store relevant project-related metadata



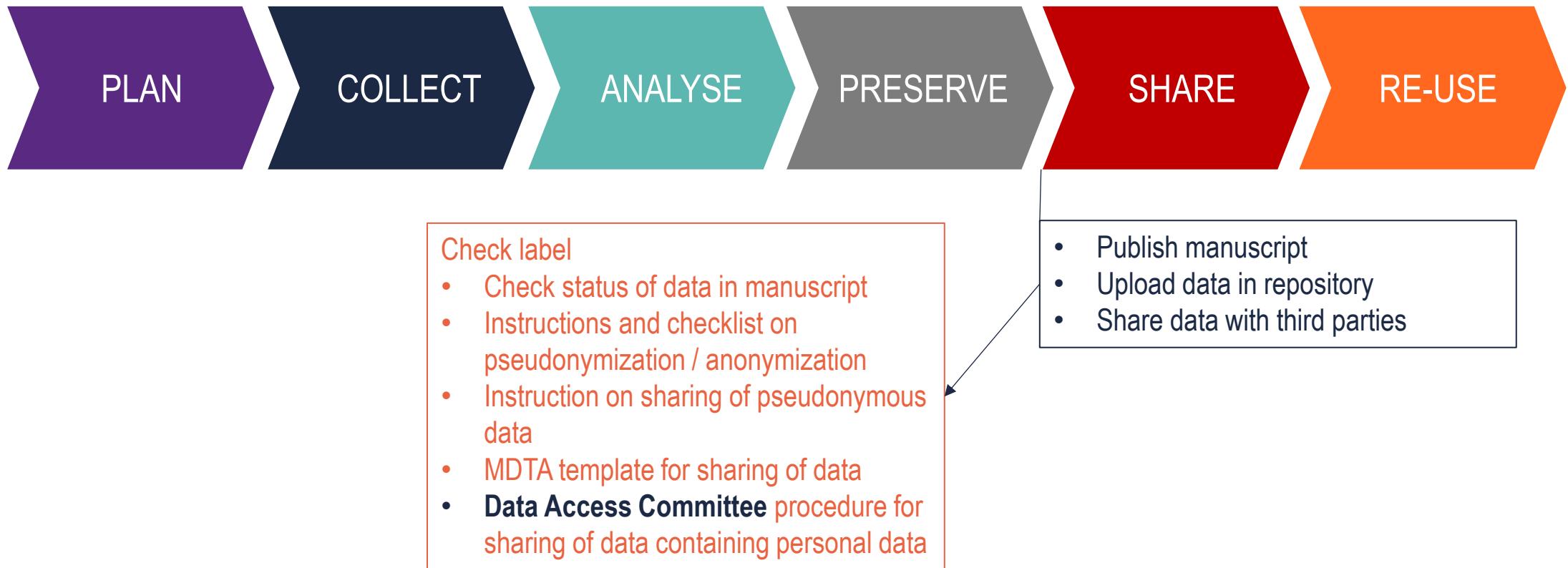
Store project-related metadata: create label

- Relevant info from the ICF or the ICF template
- Scope of research allowed
- Ethical approval
- GDPR processing log ✓
- Link to DTA, if applicable
- Link to DPIA, if applicable
- Third party sharing allowed Y/N

Step 4: ensure that GDPR compliant security is applied during the whole cycle



Step 4: Check metadata and share in a GDPR compliant manner; pseudonymize or anonymize where possible



Uploading of genetic data onto a public restricted access repository – only metadata are publicly visible

- EGA
- dbGAP
- JGA

But also:

- RDR
- ...

Uploading a dataset onto EGA

The screenshot shows the EGA homepage. At the top left is the EGA logo with the text "EUROPEAN GENOME-PHENOME ARCHIVE". To the right are three buttons: "Login", "Register", and "Need Help?". Below the logo is a dark blue navigation bar with four tabs: "ABOUT", "DISCOVERY", "SUBMISSION", and "ACCESS". To the right of the tabs is a search bar with a magnifying glass icon and the placeholder "Search...".

The European Genome-phenome Archive (EGA) is a service for permanent archiving and sharing of personally identifiable genetic, phenotypic, and clinical data generated for the purposes of biomedical research projects or in the context of research-focused healthcare systems.

I want to...



1. Check that the ICF, and where applicable the MTA, that are at the basis of the dataset allow data to be shared with third parties. Uploading a dataset onto EGA means that the dataset is available for third parties and datasets that are not allowed to be shared with third parties should not be uploaded.
2. Prepare the dataset(s) that you wish to upload. Delete columns with information that would not be relevant to share and critically consider whether certain information can be made less specific. For instance you can replace a specific age by an age range.
3. Initiate the submission process at ega-archive.org. If you are not a registered user yet, first create an EGA user account. After validation of the account you should request to obtain the submitter role. Navigate to the submitter portal: [I want to be a Submitter - EGA European Genome-Phenome Archive \(ega-archive.org\)](#)
4. As part of the process to become a submitter you have to upload the signed DPA. Request a copy of the signed DPA with goedele.dedeurwaerder@vib.be. Upload this signed DPA and request the submitter role.
5. **Do not navigate to the DAC Portal, and certainly do not create a new DAC.** There already is a VIB DAC in the EGA portal, with a connected data access policy
6. Follow the steps of the submitter portal. You can find detailed guidance on the EGA website: [Submitter Portal - EGA European Genome-Phenome Archive \(ega-archive.org\)](#).
7. When adding a dataset, EGA will ask for a policy. This is the VIB data policy (provided in annex) with reference number **EGAPoooo1003327**, linked to the VIB DAC (dac@vib.be) with reference number **EGACoooo1003263**.
8. Perform the actual upload of the dataset following the EGA instructions. The files must be encrypted.

In summary

- 1.** Inform data subjects
 1. On what type of data you will collect and what you are going to do with them
 2. On how their privacy is guaranteed
 3. Where they can turn to if they have questions or wish to exert their rights
- 2.** Enter your project in the institution's GDPR log
- 3.** Ensure that you do not receive direct identification information
- 4.** Only give access to the data to those that need access
- 5.** Respect TOM and security measures
- 6.** Do not openly share personal data
 1. Paper should only contain anonymous information
 2. Personal data sets should only be uploaded onto a restricted access repository / database
 3. Sharing of data with third parties is subject to DAC procedure and DTA

When is a dataset anonymous, when is it pseudonymous?

The consequence of pseudonymity

Pseudonymous data	Anonymous data
Subject to the GDPR	NOT subject to the GDPR
Data not allowed to be openly available	Data open (unless contractual agreements or IP analysis determines otherwise)
Data access by third parties subject to Data Access Committee procedure	Unrestricted third party access to open data
Data transfer/access agreement necessary	

The starting point

Data in the hospital patient database

Patient data

Name	Birth date	gender	Date of diagnosis	disease	stage	type	...
Jenny Johnson	10/02/1968	F	10/02/2019	Breast cancer	3	Recurrent	...
Carry Clarkson	11/03/1954	F	11/03/2018	Breast cancer	2	Recurrent	...
Betty Bodson	18/04/1950	F	18/04/2021	Breast cancer	4	Primary	...
Lyn Lynas	20/05/1988	F	20/05/2020	Breast cancer	3	Metastatic	...
Mary Moss	11/08/1972	F	11/08/2021	Breast cancer	3	Primary	...
Samantha Sha	23/09/1976	F	23/09/2019	Breast cancer	2	Metastatic	...
Tina Trump	02/11/1962	F	02/11/2021	Breast cancer	4	Primary	...

The dataset that you get from the hospital

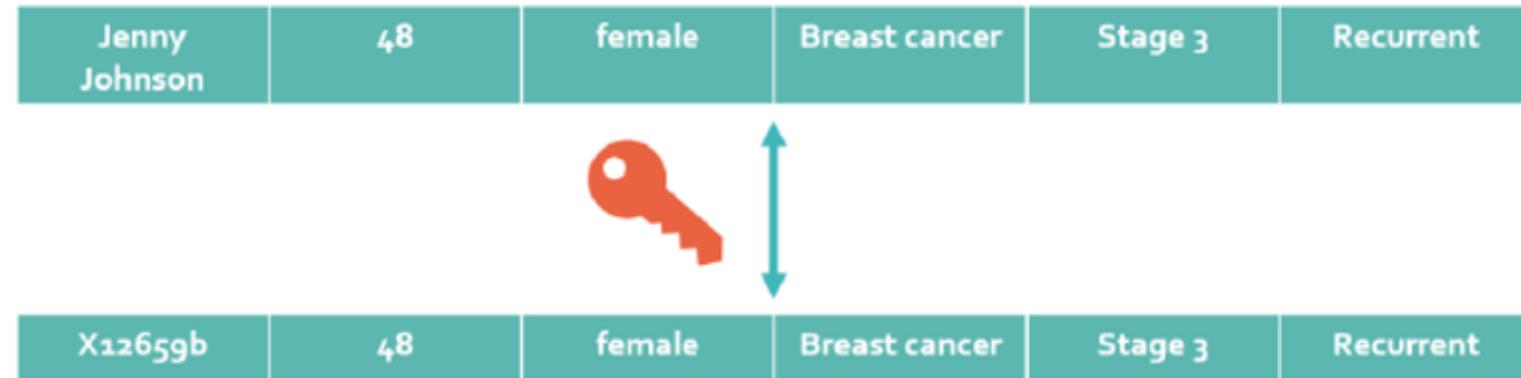
Name	Birth date	gender	Date of diagnosis	disease	stage	type	...
Jenny Johnson	10/02/1968	F	10/02/2019	Breast cancer	3	Recurrent	...



Pseudonymize & minimize

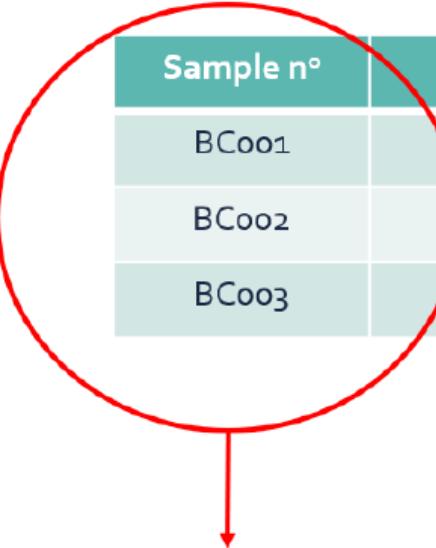
iD	Age	gender		disease	stage	type	
BRC103463x	55	F		Breast cancer	3	Recurrent	

Pseudonymous data



Not being in the possession of the decoding key does not make a dataset anonymous!

Is this dataset anonymous?



Sample n°	Age	Gender	Disease	Stage	Type
BC001	48	female	Breast cancer	Stage 3	Recurrent
BC002	67	female	Breast cancer	Stage 1	Primary
BC003	62	female	Breast cancer	Stage 4	Metastatic

Re-numbered, linking back to hospital sample iD numbers

Sample n°	Age	Gender	Disease	Stage	Type
MetCelloo34	48	female	Breast cancer	Stage 3	Recurrent
MetCello102	67	female	Breast cancer	Stage 1	Primary
MetCello223	62	female	Breast cancer	Stage 4	Metastatic

Randomly attributed new numbers, link with hospital sample ID numbers permanently broken

Sample n°	Age	Gender	Disease	Stage	Type
BC001	46-50	female	Breast cancer	Stage 3	Recurrent
BC002	66-70	female	Breast cancer	Stage 1	Primary
BC003	61-65	female	Breast cancer	Stage 4	Metastatic

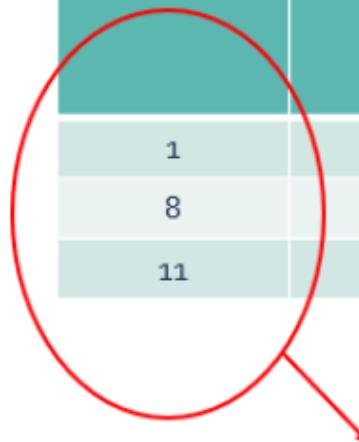


Converted to age ranges

Sample n°	Age	Gender	Disease	Stage	Type	Total read count	Number of reads per gene expressed
BC001	46-50	female	Breast cancer	Stage 3	Recurrent	21456567	xx, xx, xx, xx
BC002	66-70	female	Breast cancer	Stage 1	Primary	98377689	xx, xx, xx, xx
BC003	61-65	female	Breast cancer	Stage 4	Metastatic	65741345	xx, xx, xx, xx

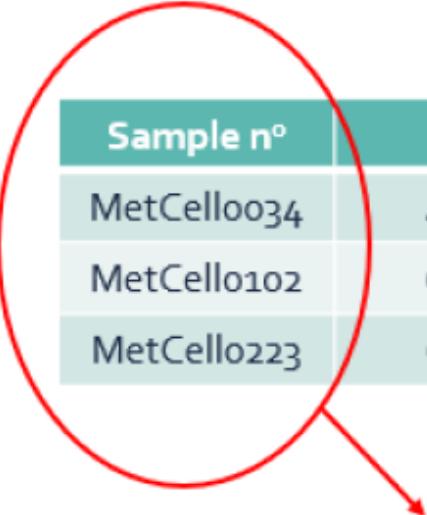
Sample n°	Age	Gender	Disease	Stage	Type	Total read count	Number of reads per gene expressed	Sequence of reads (50bp per read)
BC001	46-50	female	Breast cancer	Stage 3	Recurrent	21456567	xx, xx, xx, xx	TGT..., ATC...,
BC002	66-70	female	Breast cancer	Stage 1	Primary	98377689	xx, xx, xx, xx	TGT..., ATC...,
BC003	61-65	female	Breast cancer	Stage 4	Metastatic	65741345	xx, xx, xx, xx	TGT..., ATC...,

Sequence information included



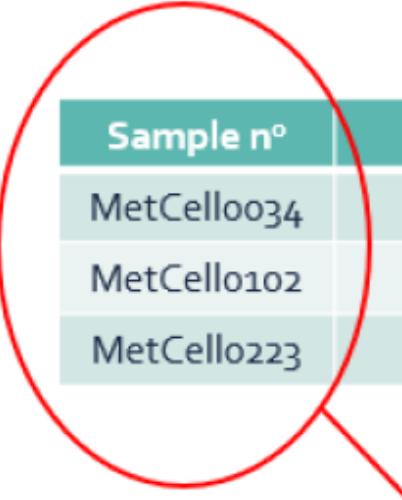
patient n°	Age	Gender	Disease	Stage	Type	Total read count	Number of reads per gene expressed	Sequence of reads (50bp per read)
1	46-50	female	Breast cancer	Stage 3	Recurrent	21456567	xx, xx, xx, xx	TGT..., ATC...,
8	66-70	female	Breast cancer	Stage 1	Primary	98377689	xx, xx, xx, xx	TGT..., ATC...,
11	61-65	female	Breast cancer	Stage 4	Metastatic	65741345	xx, xx, xx, xx	TGT..., ATC...,

Double re-numbered: linking back to sample iD number used during the analysis, which links back to the sample iD number given by the hospital



Sample n°	Age	Gender	Disease	Stage	Type	WGS
MetCello034	46-50	female	Breast cancer	Stage 3	Recurrent	ATCGGT...
MetCello102	66-70	female	Breast cancer	Stage 1	Primary	ATCCGT...
MetCello223	61-65	female	Breast cancer	Stage 4	Metastatic	ATCGCT...

Randomly attributed new numbers, link with hospital sample iD numbers permanently broken



Randomly attributed new numbers, link with hospital sample iD numbers permanently broken

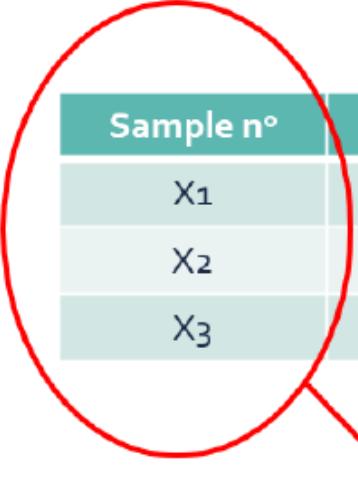
Sample n°	Age	Gender	Disease	Stage	Type	WGS
MetCello034	46-50	female	Breast cancer	Stage 3	Recurrent	ATCGGT...
MetCello102	66-70	female	Breast cancer	Stage 1	Primary	ATCCGT...
MetCello223	61-65	female	Breast cancer	Stage 4	Metastatic	ATCGCT...

When the link with the direct identification information has been permanently broken, one still has to consider the uniqueness of the data!

Sample n°	Age	Gender	Disease	Mutation info
MetCello034	48	male	Dystal Muscular Dystrophy	Mutation X, Y
MetCello102	67	male	Dystal Muscular Dystrophy	Mutation Y, Z
MetCello223	62	male	Dystal Muscular Dystrophy	Mutation X, Z

Randomly attributed new numbers, link with hospital sample iD numbers permanently broken

Sample n°	Birth date	Gender	Length	Weight	Metagenome data	Mutations in specific disease related genes
VDP001	1978-10-11	female	164	72	Gut flora species and their abundance	Mutation A, B, and G
VDP002	1954-04-26	male	186	83	Gut flora species and their abundance	Mutation E, L, R, U, V
VDP003	2001-09-03	female	173	59	Gut flora species and their abundance	Mutation C, D, Z



Sample n°	Age	Gender	Length	Weight	Gut flora type
X ₁	43	female	164	72	Type 1
X ₂	67	male	186	83	Type 2b
X ₃	20	female	173	59	Type 3

Randomly attributed new numbers, link with hospital sample iD numbers permanently broken

Highly unique data are in themselves considered subject to the GDPR

Not considered unique	Considered unique
The sequence of one or a few genes	Genetic fingerprint based on 20 STR regions as used in forensics
Genetic variants specific to tumour cells	Whole genome sequence
Transcriptomic data in the form of total reads and read counts per sequence read	Whole exome sequence
Raw (Gut flora) metagenome	Information on 20 or more optimized SNPs
Metagenome without human sequences	Information on 30 or more statistically independent SNPs
Standard MS-based proteomic data	Information on 50 or more SNPs
De novo AA sequencing data of a set of proteins	Read sequences (50 bp per read) of total RNAseq
Current clinical proteomic data (but this may change in the future)	Proteogenomic data
Metabolomic data	
To be elaborated	To be elaborated

Data not subject to the GDPR, so everything OK?

There may still be **ethical reasons not to share certain data openly**

Is proteomic data personal data?

Is metabolomic data personal data?