SCIENCE MEETS LIFE

Ethical and legal constraints on the sharing of personal data

13 December 2022

René Custers

Privacy in todays world

The ethics

- The basic right to decide on the collection and processing of your personal data (autonomy)

- The sharing/disclosure of personal data can lead to harm (non malificence)

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;

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 minization
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

VIB

SCIENCE MEETS LIFE

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

SCIENCE MEETS LIFE

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 priviledged persons
  - Pseudonymization is regarded an important security measure
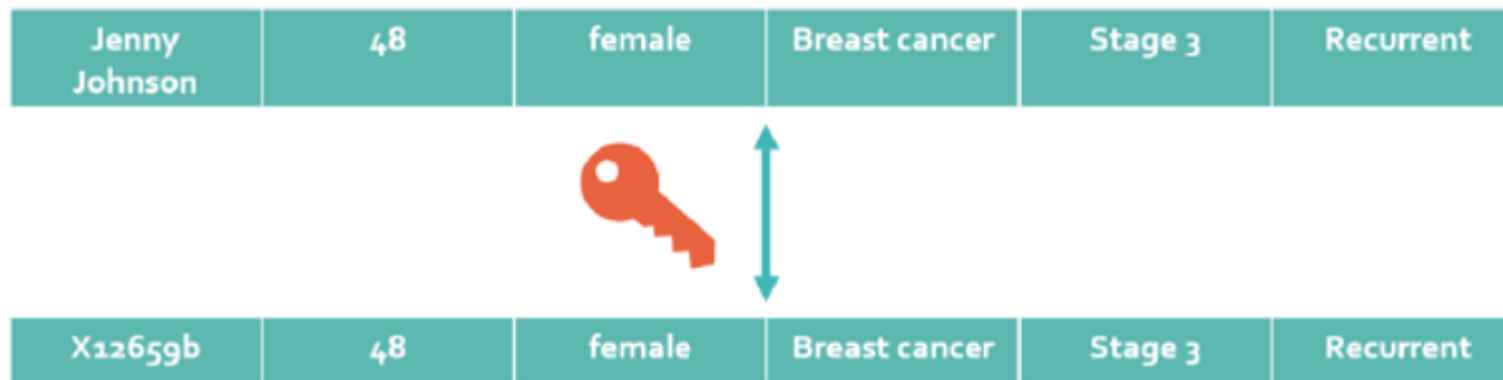
Pseudonymous data

Jenny Johnson	48	female	Breast cancer	Stage 3	Recurrent

X12659b	48	female	Breast cancer	Stage 3	Recurrent

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

VIB

SCIENCE MEETS LIFE

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

SCIENCE MEETS LIFE

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

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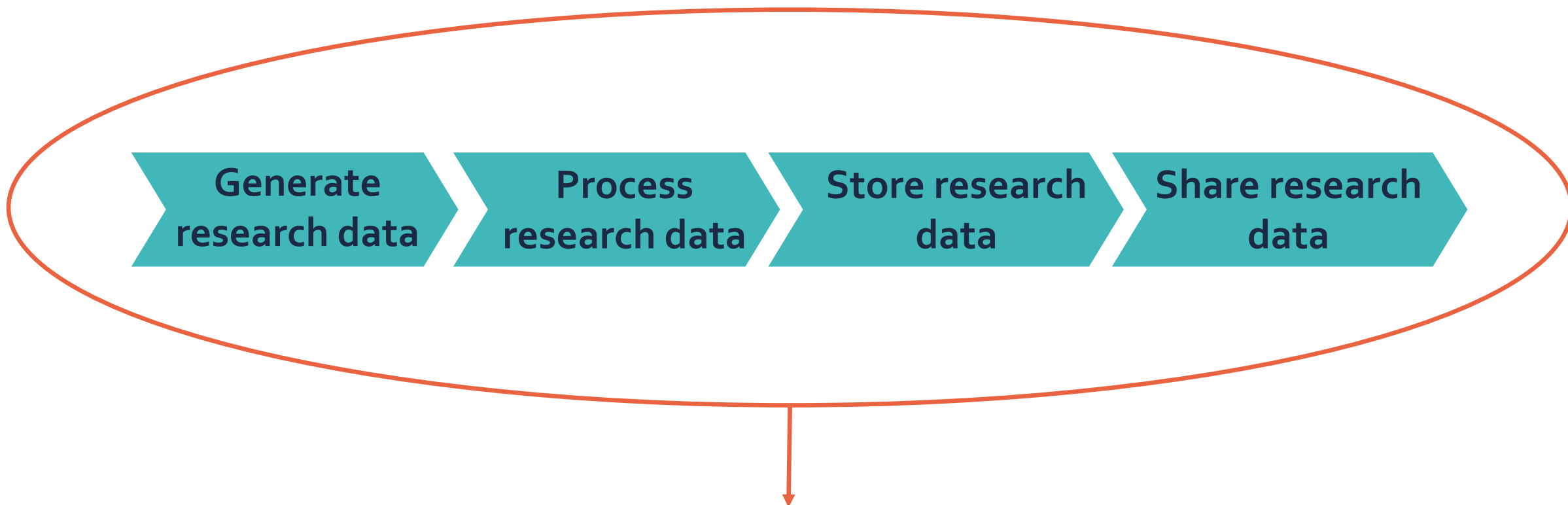
Manage research data that constitute personal data in a way that is lawful, and enables you to do what you need and want to do with these data

- ......
- ......
- ......
- ......
- ......
- ......
- ......

Transparency

- Data subjects must be informed [obscured] w their data will be processed

- They a[obscured] [obscured] [que]stions on the data process[obscured]

-> include information in the ICF

Management of personal data includes informing people on what you are going to do with their data

VIB

SCIENCE MEETS LIFE

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?

VIB

- ......
- ......
- ......
- ......
- ......
- ......
- ......

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

VIB

SCIENCE MEETS LIFE

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

SCIENCE MEETS LIFE

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**

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

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

PLAN → COLLECT → ANALYSE → PRESERVE → SHARE → RE-USE

- Ensure that you only receive pseudonymous data
- Or pseudonymize the data yourself (this requires proper organisation!)

VIB

SCIENCE MEETS LIFE

Step 3: ensure that GDPR compliant security is applied during the whole cycle

PLAN → COLLECT → ANALYSE → PRESERVE → SHARE → RE-USE

- Apply security measures
- Ensure traceability

VIB

SCIENCE MEETS LIFE
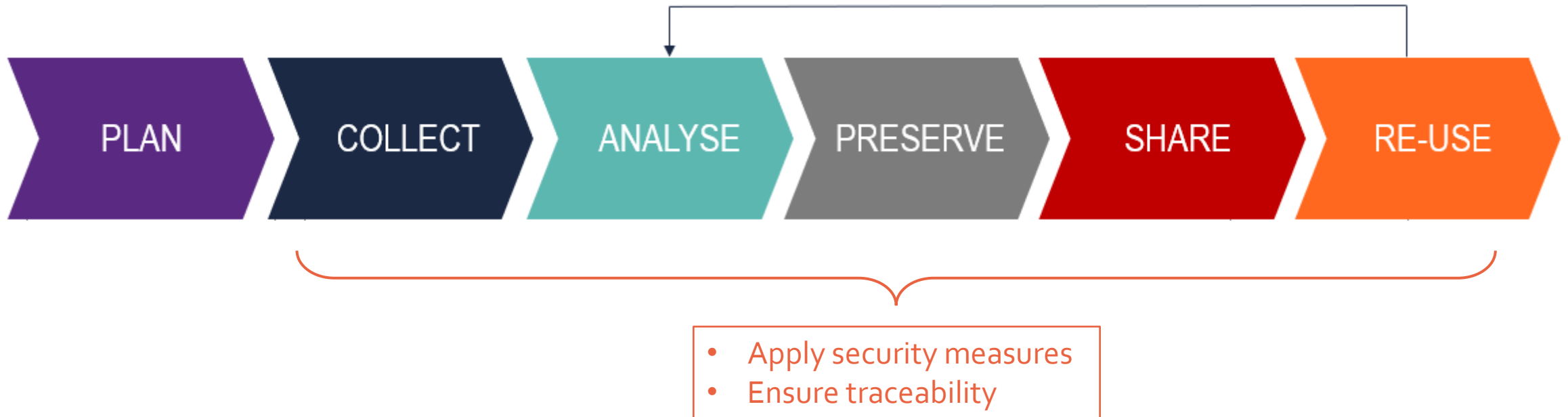
Step 3: Store relevant project-related metadata

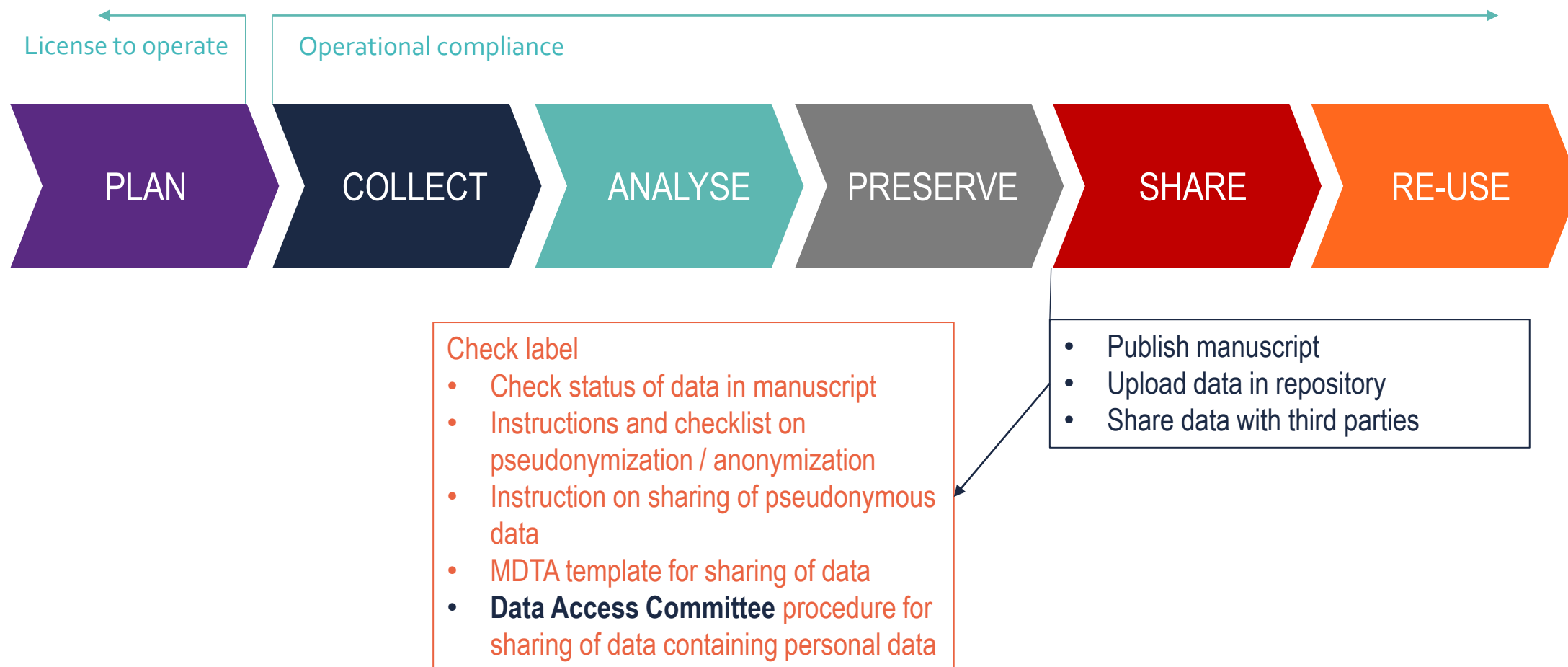PLAN → COLLECT → ANALYSE → PRESERVE → SHARE → RE-USE

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

Ethical, legal services in the context of personal data:
(1) license to operate, and (2) operational compliance

License to operate

Operational compliance

| PLAN | COLLECT | ANALYSE | PRESERVE | SHARE | RE-USE |

Check label
- Check status of data in manuscript
- Instructions and checklist on pseudonymization / anonymization
- Instruction on sharing of pseudonymous data
- MDTA template for sharing of data
- **Data Access Committee** procedure for sharing of data containing personal data

- Publish manuscript
- Upload data in repository
- Share data with third parties

VIB

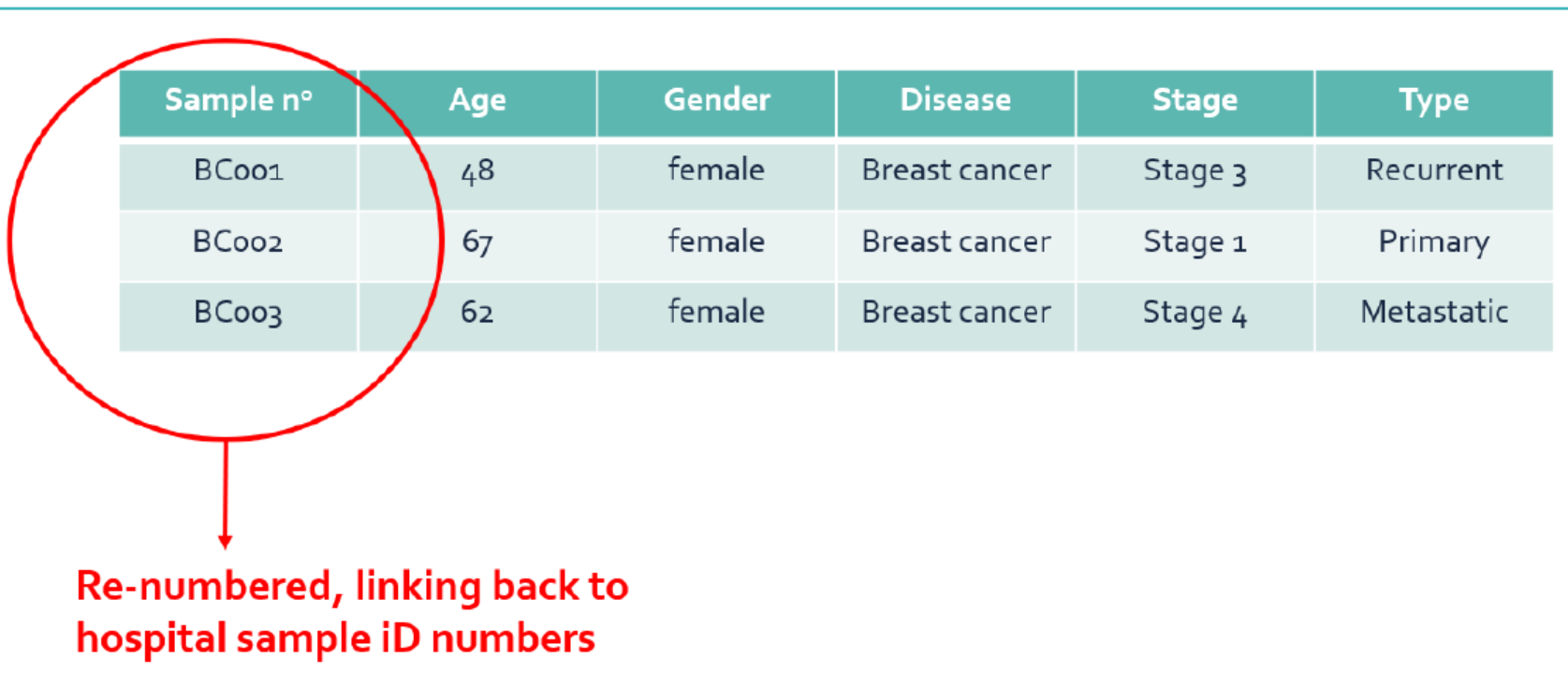SCIENCE MEETS LIFE

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

VIB

SCIENCE MEETS LIFE

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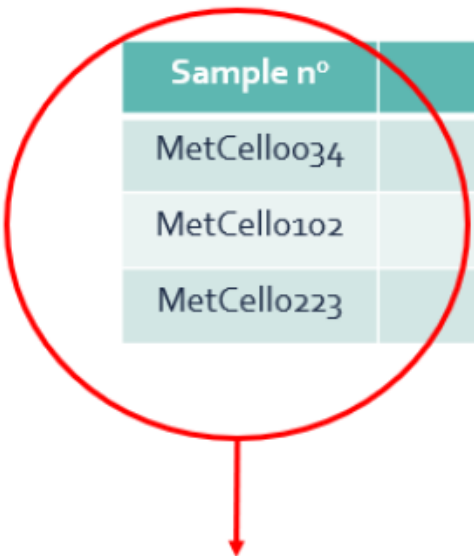
Pseudonymous data	Anonymous data
Subject to the GDPR	NOT subject to the GDPR
Data not allowed to be openly available	Data open (unless contractual agreements or IP analysis determines otherwise)
Data access by third parties subject to Data Access Committee procedure	Unrestricted third party access to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

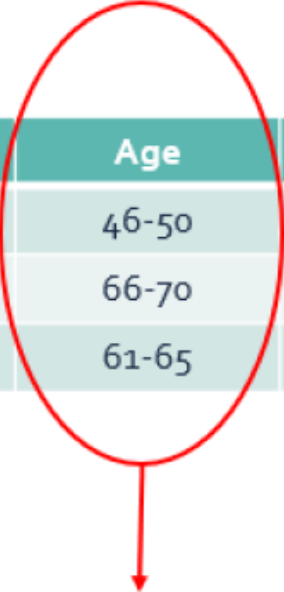| Sample n°	Age	Gender	Disease	Stage	Type
BC001	48	female	Breast cancer	Stage 3	Recurrent
BC002	67	female	Breast cancer	Stage 1	Primary
BC003	62	female	Breast cancer	Stage 4	Metastatic
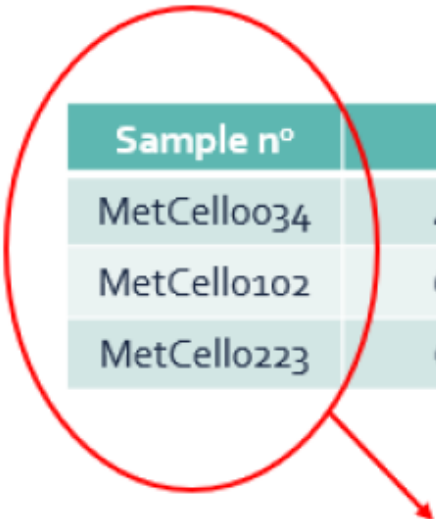
Re-numbered, linking back to hospital sample iD numbers

Sample n°	Age	Gender	Disease	Stage	Type
MetCell0034	48	female	Breast cancer	Stage 3	Recurrent
MetCell0102	67	female	Breast cancer	Stage 1	Primary
MetCell0223	62	female	Breast cancer	Stage 4	Metastatic

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

SCIENCE MEETS LIFE

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

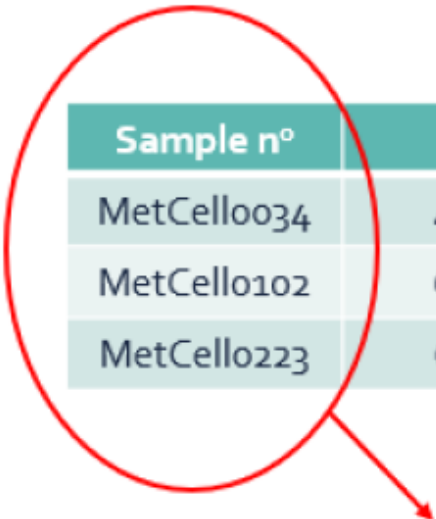| Sample n°	Age	Gender	Disease	Stage	Type	Total read count	Number of reads per gene expressed	Sequence of reads (50bp per read)
BC001	46-50	female	Breast cancer	Stage 3	Recurrent	21456567	xx, xx, xx, xx	TGT..., ATC...,
BC002	66-70	female	Breast cancer	Stage 1	Primary	98377689	xx, xx, xx, xx	TGT..., ATC...,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

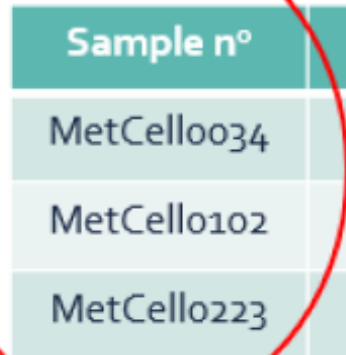| Sample n°	Age	Gender	Disease	Stage	Type	WGS
MetCell0034	46-50	female	Breast cancer	Stage 3	Recurrent	ATCGGT...
MetCell0102	66-70	female	Breast cancer	Stage 1	Primary	ATCCGT...
MetCell0223	61-65	female	Breast cancer	Stage 4	Metastatic	ATCGCT...

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

Sample n°	Age	Gender	Disease	Stage	Type	WGS
MetCell0034	46-50	female	Breast cancer	Stage 3	Recurrent	ATCGGT...
MetCell0102	66-70	female	Breast cancer	Stage 1	Primary	ATCCGT...
MetCell0223	61-65	female	Breast cancer	Stage 4	Metastatic	ATCGCT...

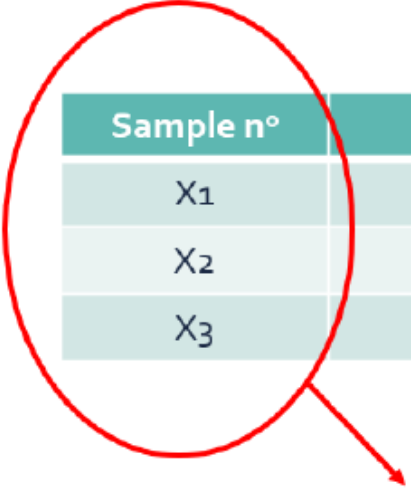**Randomly attributed new numbers, link with hospital sample iD numbers permanently broken**

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
MetCell0034	48	male	Dystal Muscular Dystrophy	Mutation X, Y
MetCell0102	67	male	Dystal Muscular Dystrophy	Mutation Y, Z
MetCell0223	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

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

Is proteomic data personal data?