Materials developed as a part of the *Skills development for research data* project:
https://www.ub.uio.no/english/about/projects/rdm-skills/

Data Management and Protection

A crash course

Agata Bochynska

Open Research and Digital Scholarship Center

University of Oslo Library

What data are you working with?

Mentimeter

71

What is research data management (RDM)?

"**Research data management** concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results and permits new and innovative research built on existing information."

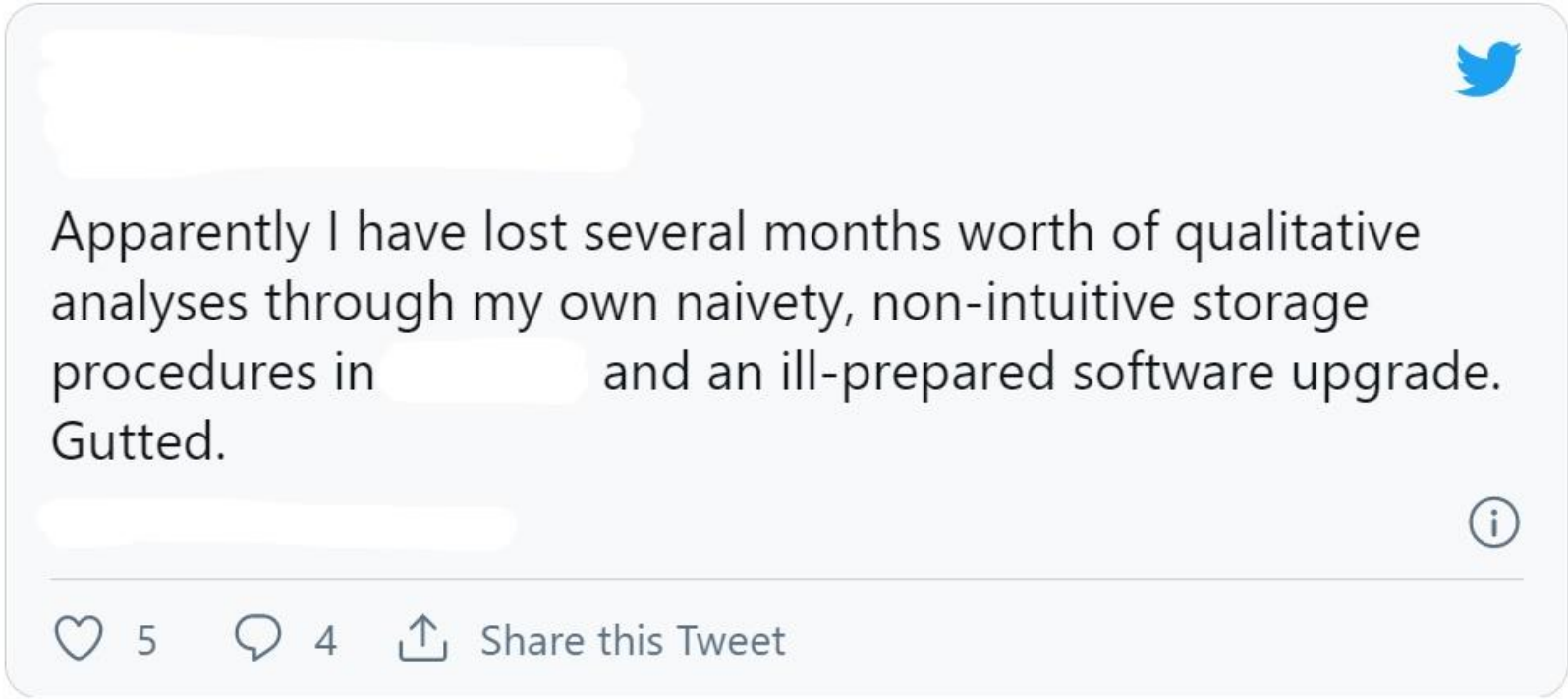
(from, Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. **Available online**)

Source: RDMKit Elixir / https://rdmkit.elixir-europe.org/

What is research data management (RDM)?

Protection
Organization
Documentation (and metadata)
Classification
Short-term storage
Licensing
Sharing
Long-term preservation
Dissemination (publishing)
Reusing

Saves time… and trouble!

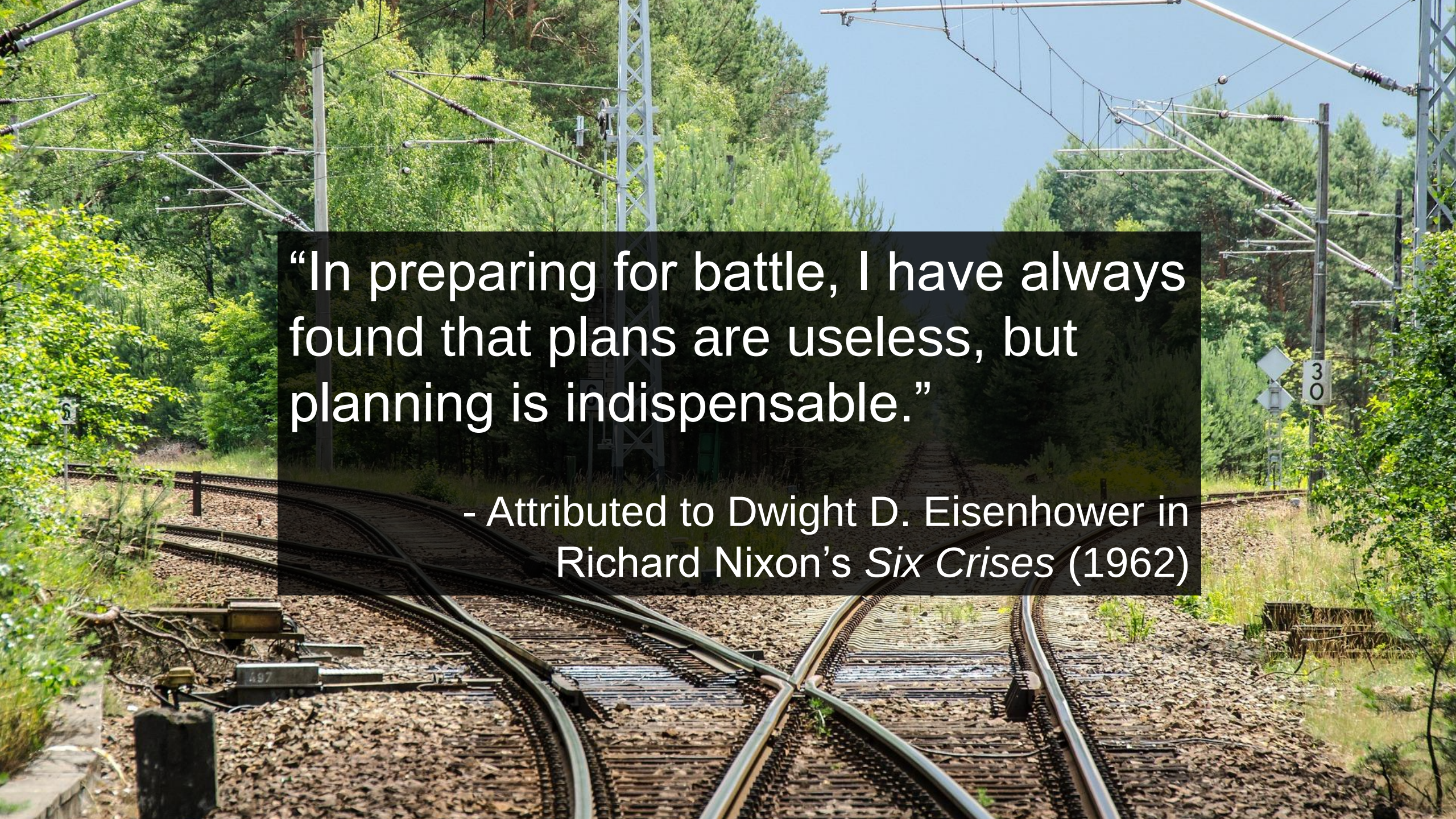


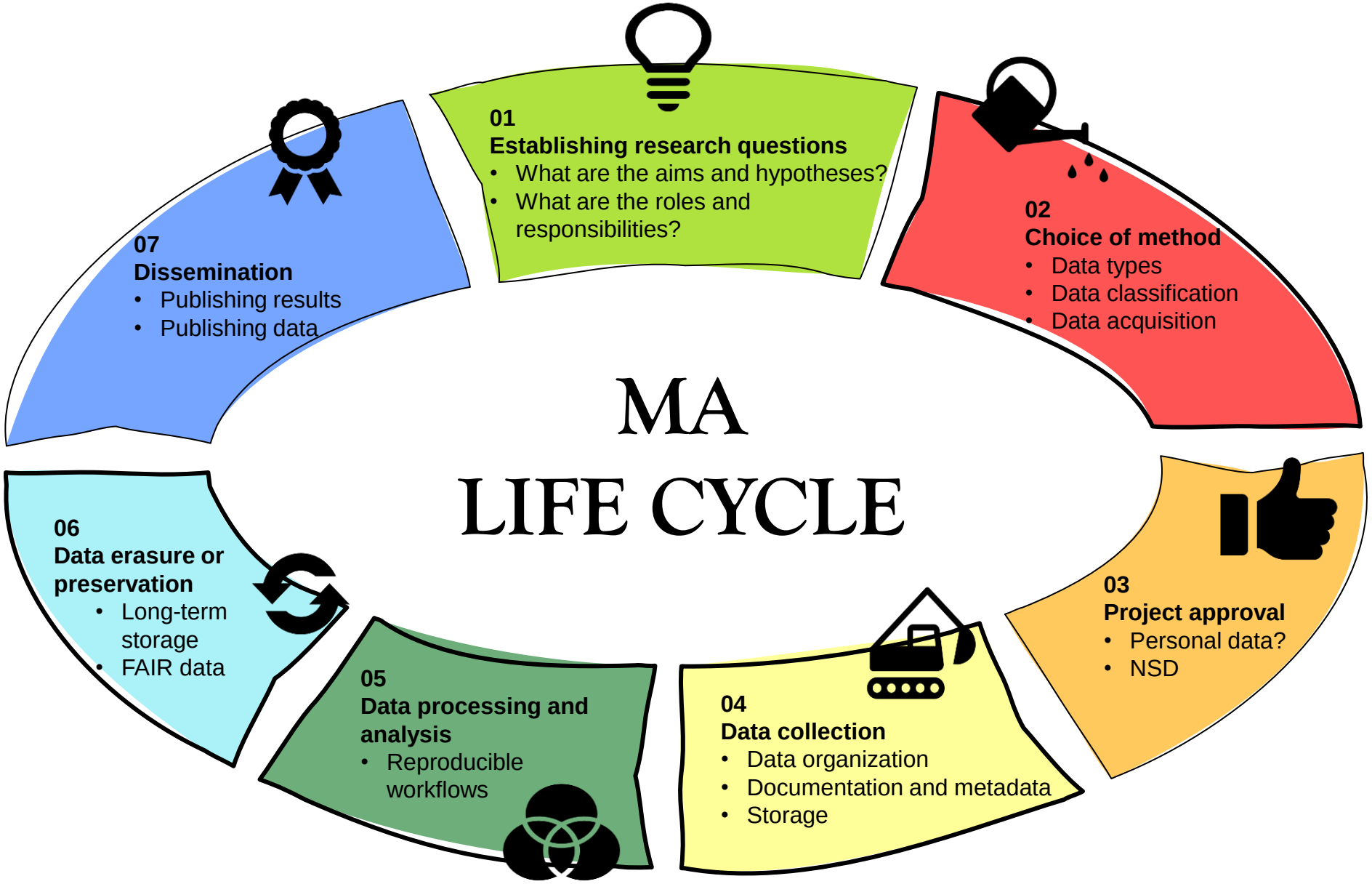
General Data Protection Regulation (GDPR)

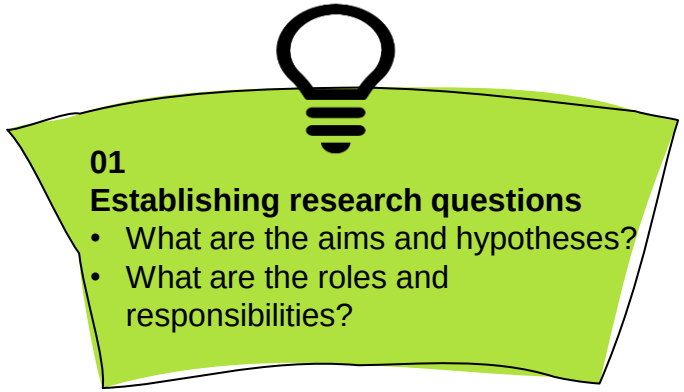
UNIVERSITETET
I OSLO

The Data Management Plan (DMP)

- Is a **living** document that accompanies the research project
- Specifies the **types** of data that will be generated
- Describes how you plan to **manage** your data (organize, document, classify, store)
- Conveys whether and how the data can be **shared**
- **Agreement** between project members
- A **tool** to keep overview over data


“In preparing for battle, I have always found that plans are useless, but planning is indispensable.”
- Attributed to Dwight D. Eisenhower in Richard Nixon’s *Six Crises* (1962)

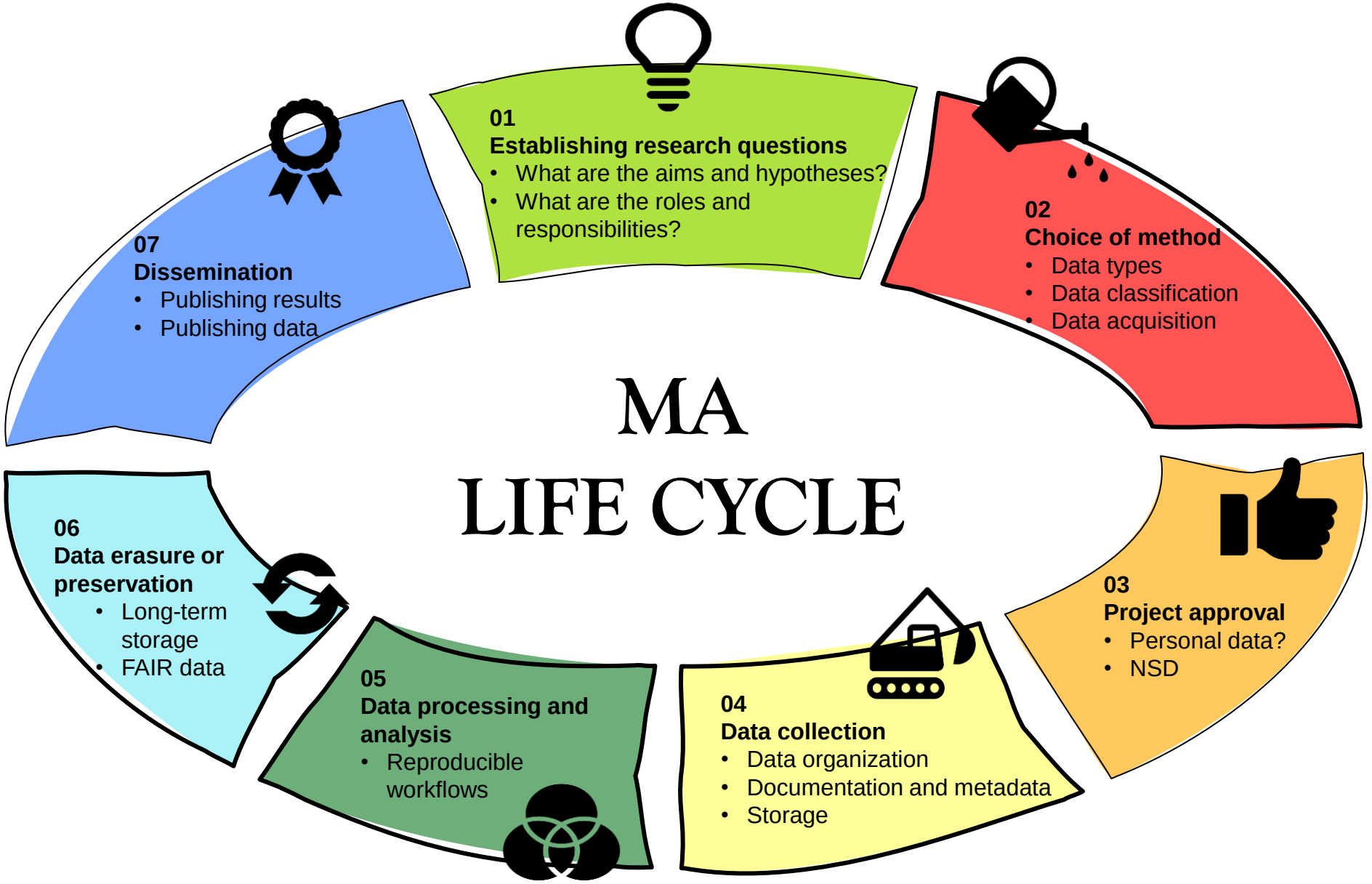

MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data

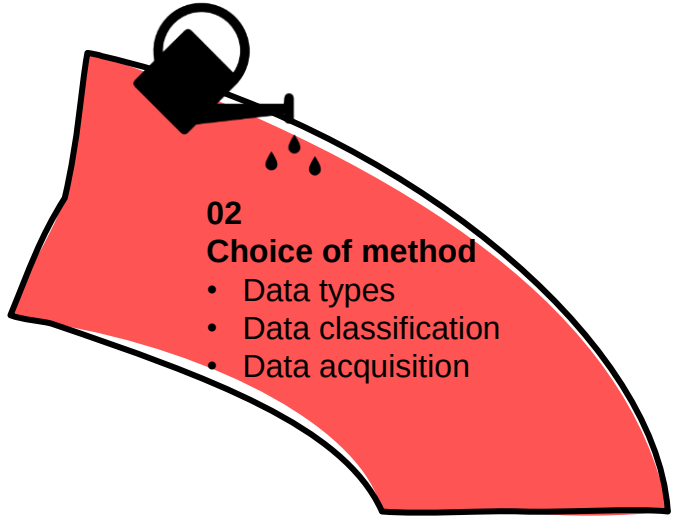


MA LIFE CYCLE

Assigning roles and responsibilities

- What are the roles of each participant in the project?
- Who else outside of the project will be contributing?
- Who is responsible for data management during the project?
- Who is responsible for data management and archiving after the project is finished?
- Who will make sure the data management plan (DMP) is being followed?
- Who will be responsible for updating the DMP and how often?


MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data



MA
LIFE CYCLE

Identify data to be collected

Identify data to be collected

Origin of Data

- Existing data or new data?
- If data is reused: What is the scope, volume, format?
- How are different data sources integrated?

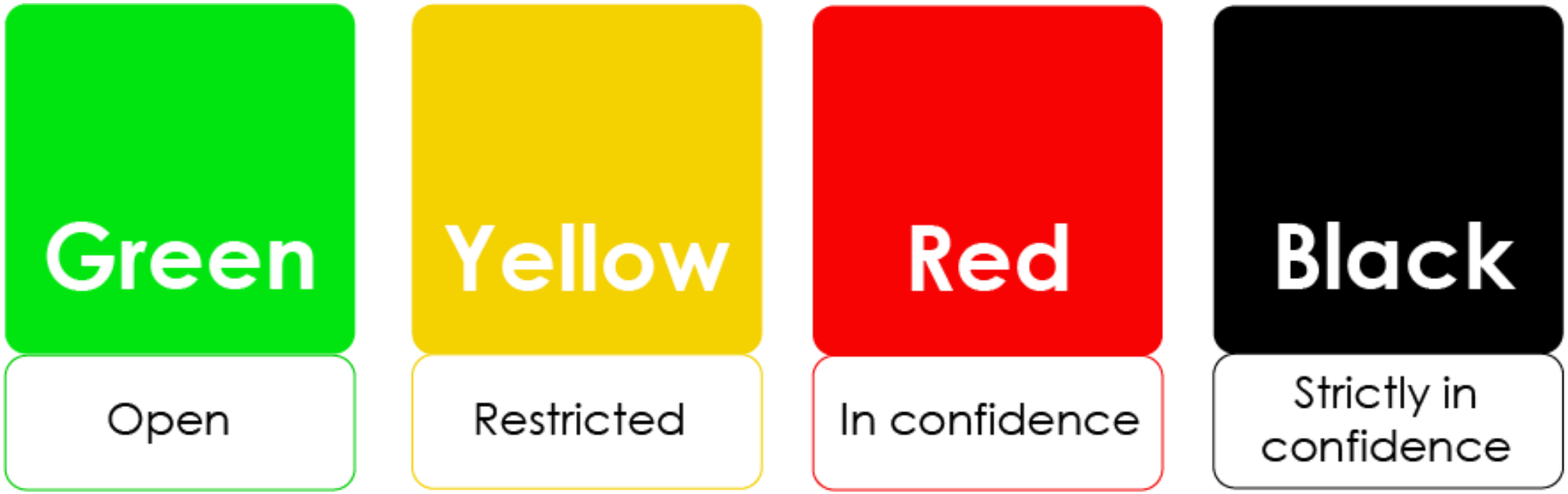
Types of Data

- What type(s) of data will be collected?
- What is the scope, quantity and format of the material?
- What is the total amount of data collected (MB/GB/TB/PT)?

Data Collection

- How will data be collected?
- Is specific software or hardware – or staff required?
- Who is responsible for data collection?
- During which period will data be collected?
- Where will data be collected?

Classify your data



https://www.uio.no/english/services/it/security/lsis/data-classes.html

What is green data?

This class is used when it **does not cause any harm** to the institution if the information becomes known to unauthorized persons.

- Any research that is published openly. Information that is not public, needs to be removed first.
- Teaching material that is not subject to copyright
- Completely anonymous data
- Data without any financial or commercial value
- Statistics

What is yellow data?

This class is used when it **could cause a certain damage** to the institution if the information becomes known to unauthorized persons.

- Documents that you don’t want everyone else to read
- De-identified data where the key is locked away and stored safely away from the data
- Data sets containing minor amounts of non-sensitive personal data
- E-mail attachments that don’t contain information that needs protection
- Unpublished research material
- Texts and pictures under copyright

What is red data?

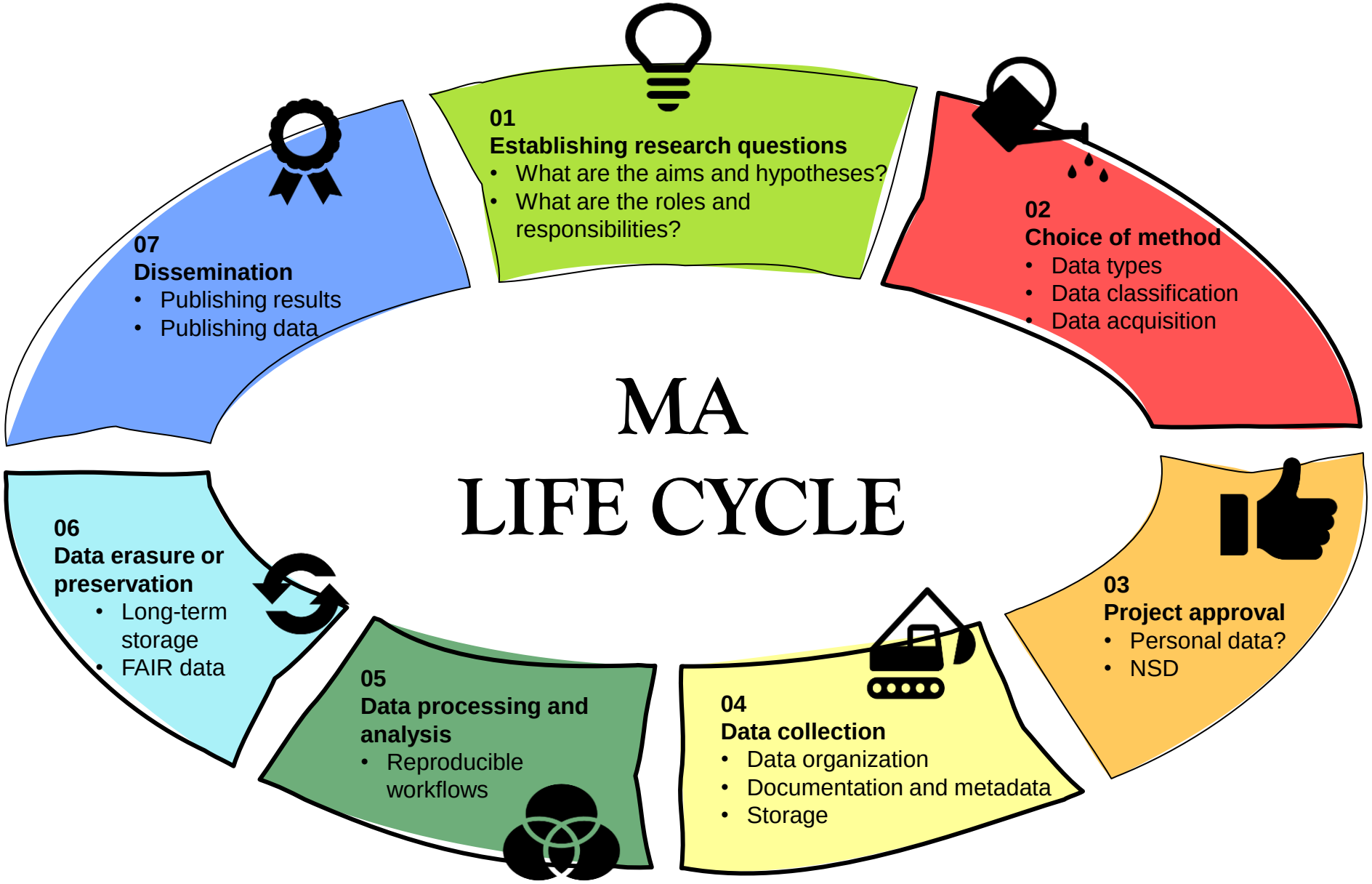
This class is used if it **could cause harm** to public interests, the university or individuals if the information becomes known to unauthorized persons.

- De-identified data where the key is available and stored with the data
- Special categories of personal data
- Personnel files
- Information relating to e.g. safety systems in buildings or IT systems
- Health related data

What is **black** data?

This class is used if it **could cause significant harm** to public interests, the university or individuals if the information becomes known to unauthorized persons.

- Research data that requires extra protection
- Large quantities of sensitive personal data
- Large quantities of health related data
- Research data that is of great financial or commercial value


MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data

MA
LIFE CYCLE

UNIVERSITETET
I OSLO

Notification Form for personal data

On this page you will find useful information about NSD's notification form. You can read about what personal data is, who should send in a notification form, and what you need to have ready in advance.

NSD's notification form is a digital form that you as a researcher or student fill in when you are going to process personal data in a research project. We at NSD then carry out an assessment of the processing you have planned.

We help you find solutions that fit your project and ensure that the planned processing is in accordance with data protection legislation.

By filling out the notification form your institution will also have an overview of the processing, something it is legally required to have.

https://www.nsd.no/en/data-protection-services/notification-form-for-personal-data/

42:54

https://vimeo.com/606519110

Which personal data will be processed?

What are personal data?

What is processing?

Name (also with signature/written consent)

Yes No

National ID number or other personal identification number

Yes No

Date of birth

Yes No

Address or telephone number

Yes No

Email address, IP address or other online identifier

Yes No

Photographs or video recordings of people

Yes No

Sound recordings of people

What is personal data?

Any piece of information about an identified or identifiable physical person.

In other words, any data about living people from which they can be identified.

What is personal data?

General personal data

Special categories of personal data (sensitive data)

What constitutes as general personal data?

- Name, address, age, phone number, e-mail and Norwegian national identity number
- Video- and audio recordings where individuals are recognizable
- Identifying images of individuals
- Contents of email communication
- Content of case documents, investigations or considerations
- Content of exam papers or grades
- IP addresses (that can identify individuals) and logging of activity in computer systems (where the logs can be associated with individuals)

Special categories of personal data (sensitive data)

- health information and health related conditions
- genetic or biometric information which can be used to identify a physical person
- ethnic or racial origin
- political, philosophical or religious perceptions and beliefs
- sexual orientation or sexual relationships
- trade-union membership

What is considered processing of personal data?

- Collection
- Registration
- Storage
- Compilation
- Use
- Transfer
- Publication
- Erasure

Consent

For a consent to be valid and work as a basis for processing, it must be:

- Voluntary
- Specific
- Informed
- Unequivocal
- Given actively
- Documentable
- Possible to withdraw as easily as it was given

If you wish to use the data for a new/different purpose, a new consent must be obtained from the data subject.

Information and consent

When you are collecting personal data for a research project, you usually have a duty of information to the people you are collecting information about. On this page you will gain information about requirements for the information you provide.

Do I have to give information to the sample?

As a main rule, you are obligated to inform the person(s) whose personal data you you are going to collect. This is a fundamental right that is entrenched in Norwegian law. The obligation to inform applies

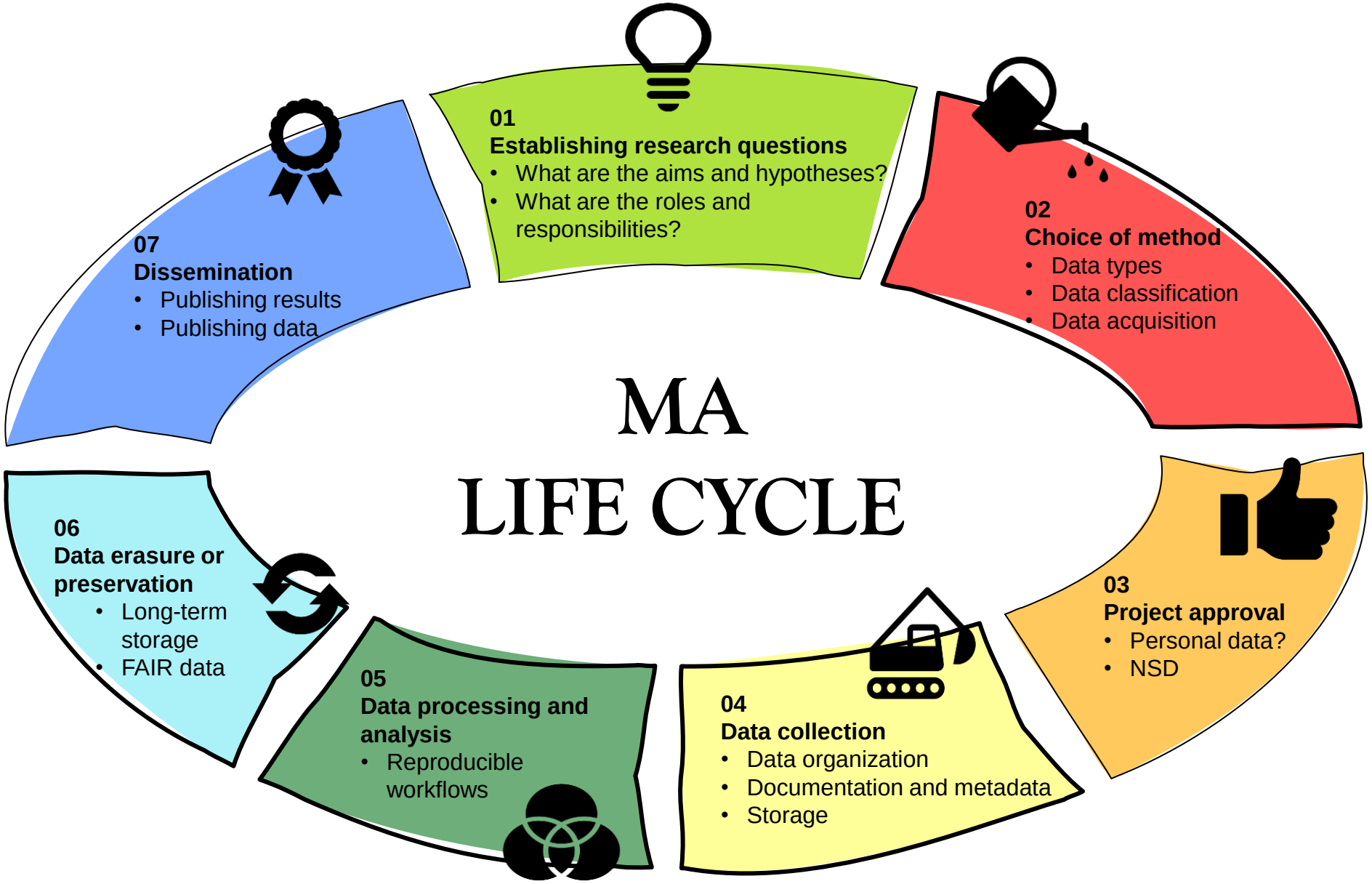
What responsibility do student researchers have for research participants' (respondents or informants) privacy protection?

Reviewing and understanding the rules that govern the use of personal information in research at UiO and routines for processing personal data in research projects.

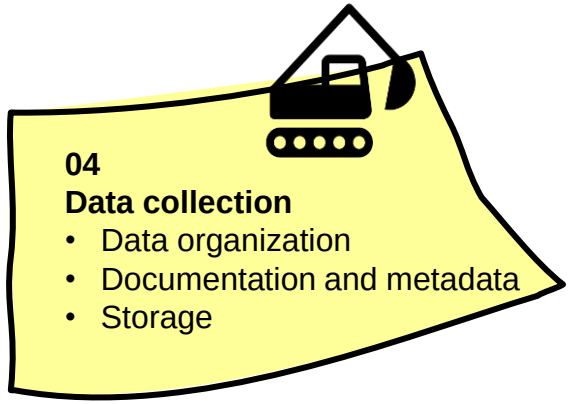
- Following the **rules** for use of personal information in research at UiO.
- Seeking **assistance** from their advisor or coordinator for the work at their department if there is any doubt about what rules apply or how the rules should be interpreted or implemented.

Not beginning collection or other use of personal information before the necessary permissions are in place.

https://www.uio.no/english/studies/examinations/privacy-protection/


MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data

MA
LIFE CYCLE

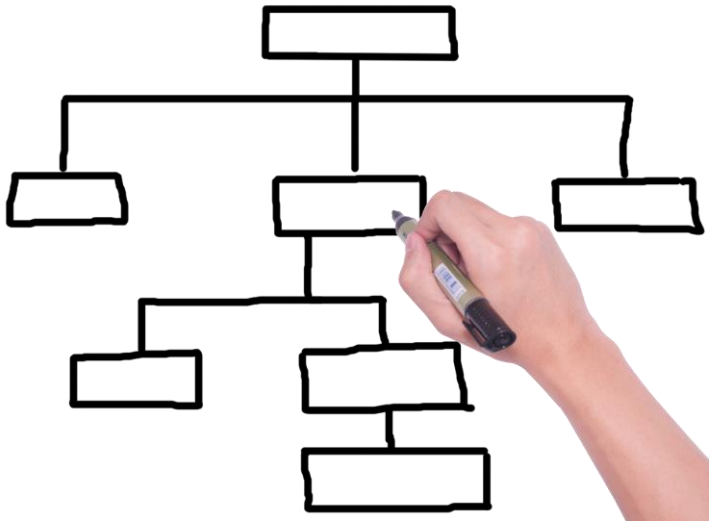


04
Data collection
- Data organization
- Documentation and metadata
- Storage

UNIVERSITETET
I OSLO

How will data be organized?

- Make a structure and stick to it!
- Come to an agreement within your team
- Follow-up and revise as necessary



Questions to answer

- What file formats will you be using?
- How will you structure the folder system and/or database?
- How will you name files?
- How will you take care of older versions and articles?

How will data be documented?

- What information is needed for comprehensibility, reusability, and reproducibility?

- What kind of documentation?
  - Instrument settings and calibration
  - Protocols and scripts
  - Field journals
  - Lab notebooks
  - README files?

README file:

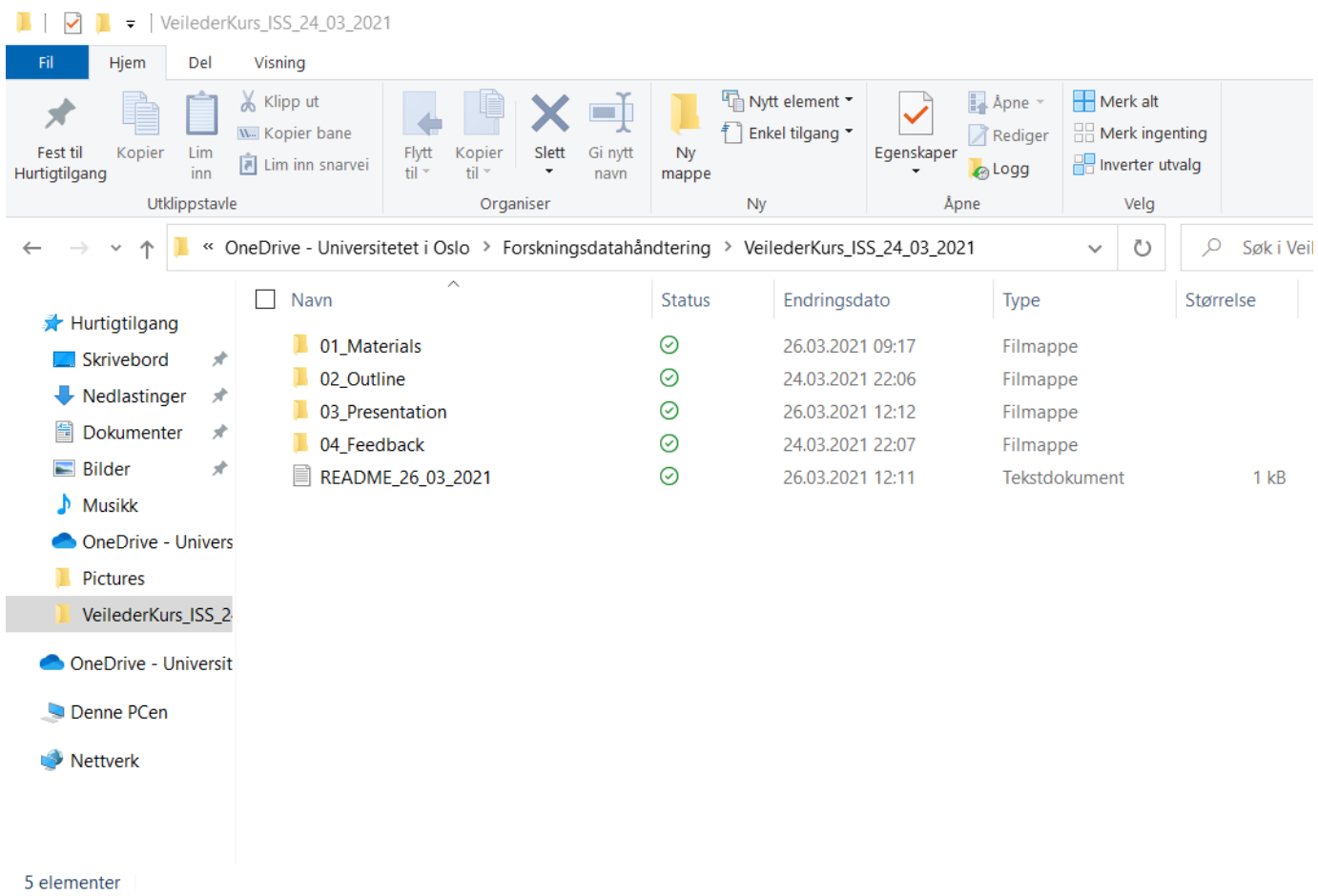
GENERAL INFORMATION

SHARING/ACCESS INFORMATION

DATA & FILE OVERVIEW

METHODOLOGICAL INFORMATION

DATA-SPECIFIC INFORMATION



Where can you store the different types of data?

Where can you store the different types of data?

UiO storage guide:

https://www.uio.no/english/services/it/security/lsis/storage-guide.html

https://www.uio.no/english/services/it/security/lsis/storage-guide.html

Data storage guide

This guide tells you where you can ***store*** and ***process*** information. Click on the headings below to read more about about the different kinds of storage.

Please also see the classification guide for information about the different information categories.

NB! The numbers in the table refers to notes. The notes corresponding to the numbers are in the bottom of this article.

Storing on a Mac, PC or hard drive –

	Open (green)	Restricted (yellow)	In confidence (red)	Strictly in confidence (black)
Privately owned laptop (BYOD)	Yes	11	No	No
Privately owned home computer	Yes	11	No	No

Where can you store the different types of data?

Green data:

- anywhere you want, but ideally **not only** on your private laptop (very vulnerable!)

Where can you store the different types of data?

Yellow data:

- Encrypted memory stick / external hard drive
- Encrypted laptop maintained by UiO
- UiO cloud solutions (UiO OneDrive, UiO G-Suite, …)
- UiO TSD
- UiO storage hotel («lagringshotell»)
- UiO personal research storage
- UiO e-mail; UiO Teams; UiO Zoom; UiO Nettskjema; UiO Canvas, UiO Vortex, UiO Wiki…
- **Not** your private laptop, unless you comply with the guidelines for use of private computer

Where can you store the different types of data?

Red data:

- Fully encrypted disc, memory stick or external harddrive
- Services for sensitive data – TSD
- UiO Personal researcher storage*
- UiO Vortex*
- UiO Storage Hotel*

*see Storage guide for details!

Source: https://www.uio.no/english/services/it/security/lsis/storage-guide.html

Storage Hotel

- Storage service where you can safely store and access up to red data

UNIVERSITY
OF OSLO

More about UiO's storage hotel: https://www.uio.no/english/services/it/store-collaborate/storage-hotel/

Where can you store the different types of data?

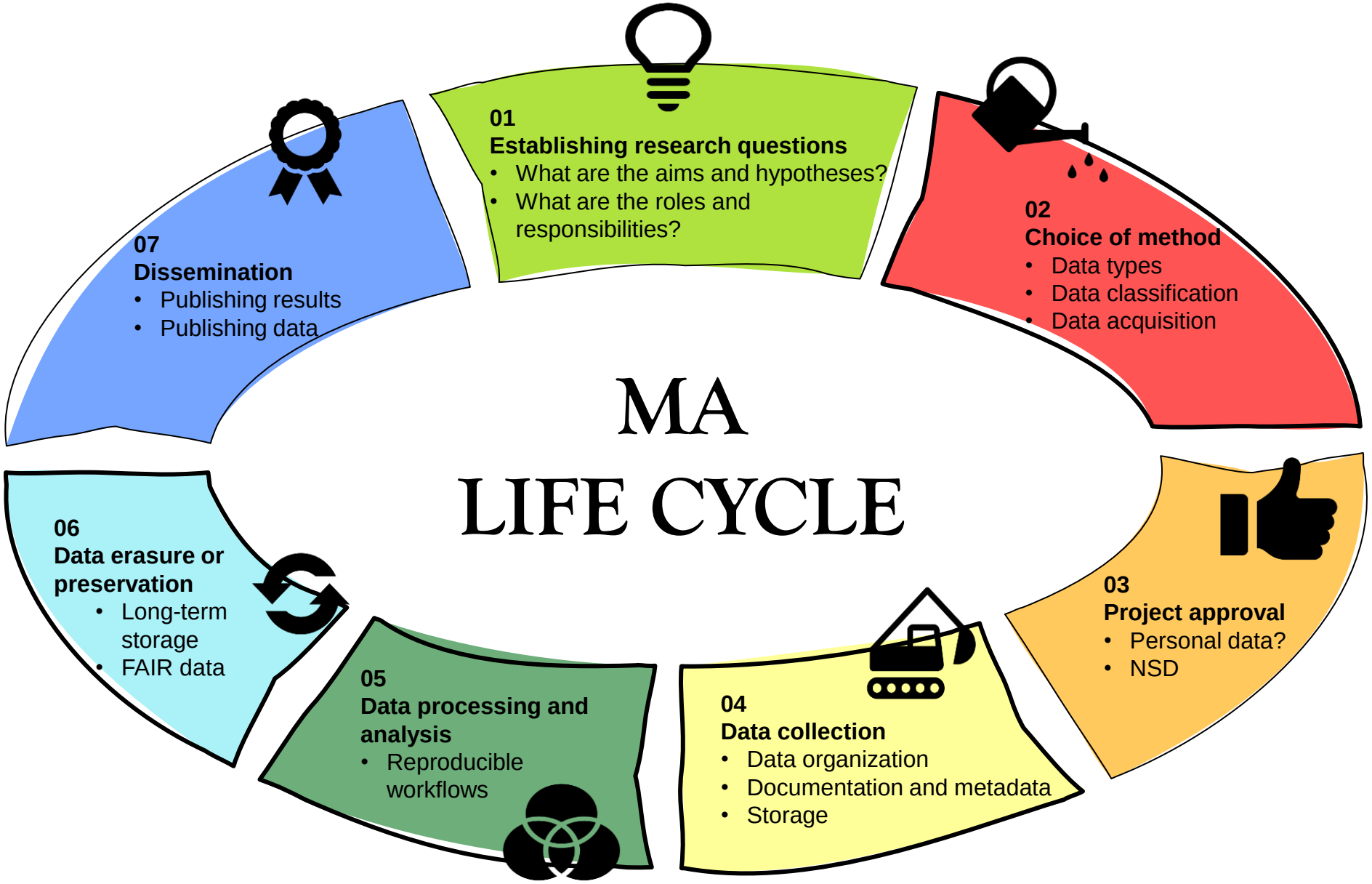
Black data:

Services for sensitive data – TSD

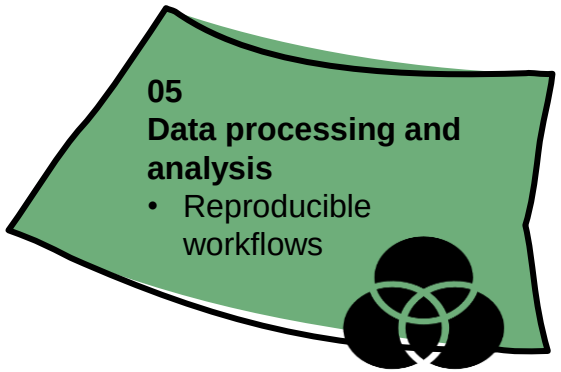
More about TSD: https://www.uio.no/english/services/it/research/sensitive-data/

Some things to remember:

- Use safe storage solutions, where data/files are backed up
- Always keep a copy of your raw data material safely hidden away (like a folder on storage hotel that you never edit)
- Only saving files on your own computer is not safe
- A memory stick or an external hard drive is not safe
- How much extra work would it be for you to recreate your data or your work?

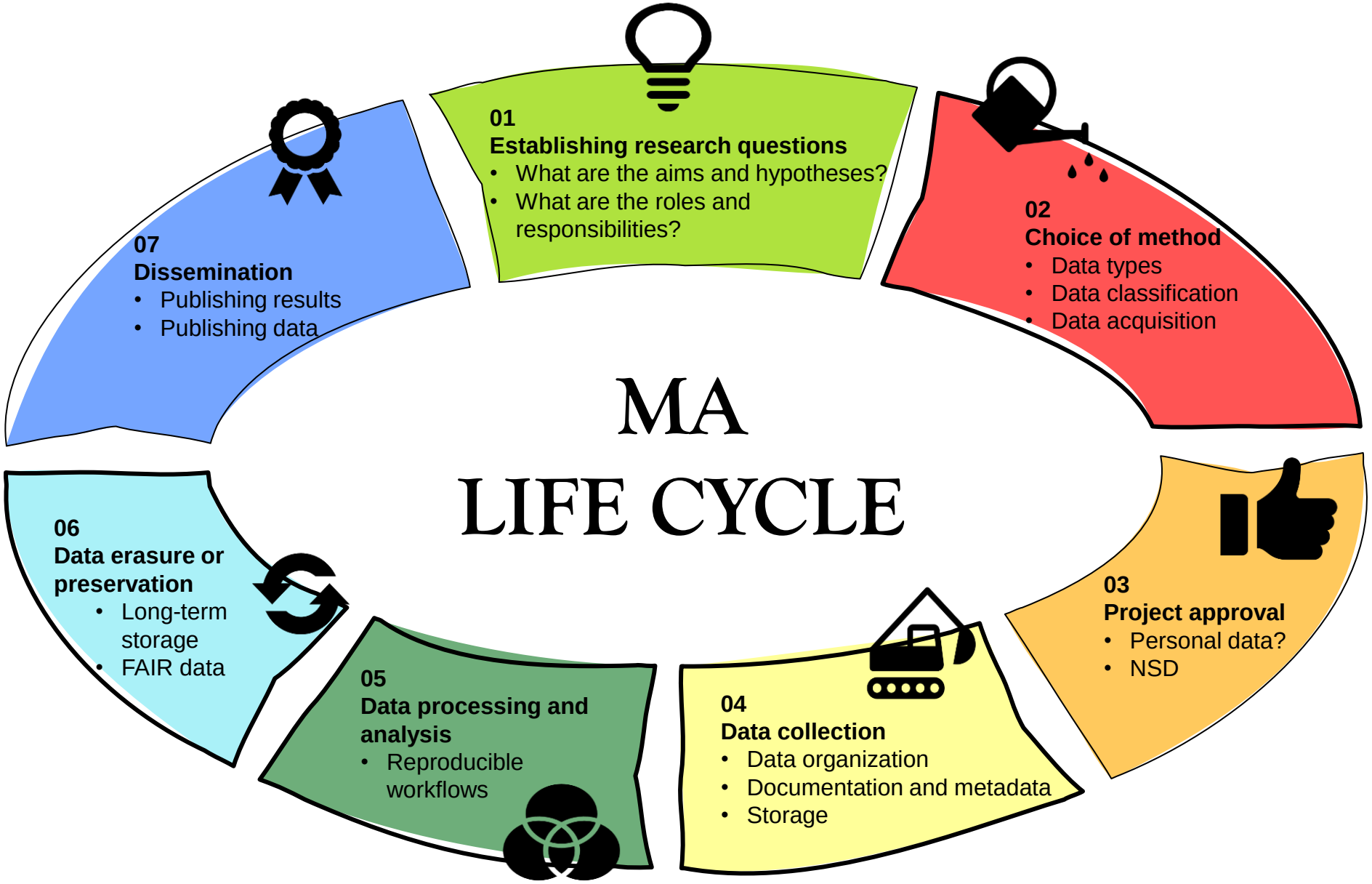

MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data

MA
LIFE CYCLE

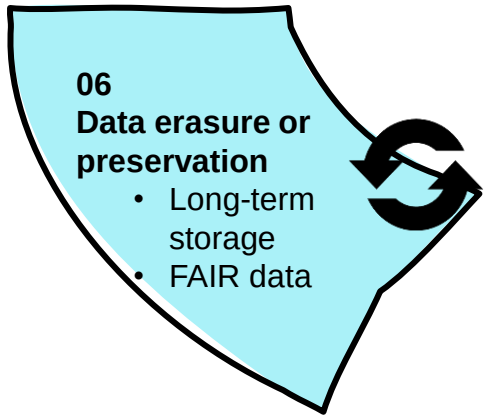


Save all the files!

- “For every result, keep track of how it was produced” (Sandve et al, 2013)
- Save processed data at important stages in the project
- Save processed data and analysis/methods for every table and graph


MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data

MA
LIFE CYCLE



Will the data be stored long-term?

- Discuss with your supervisor how the need for access and backups will be managed and assessed long-term
- What storage/archive solution would be most appropriate?

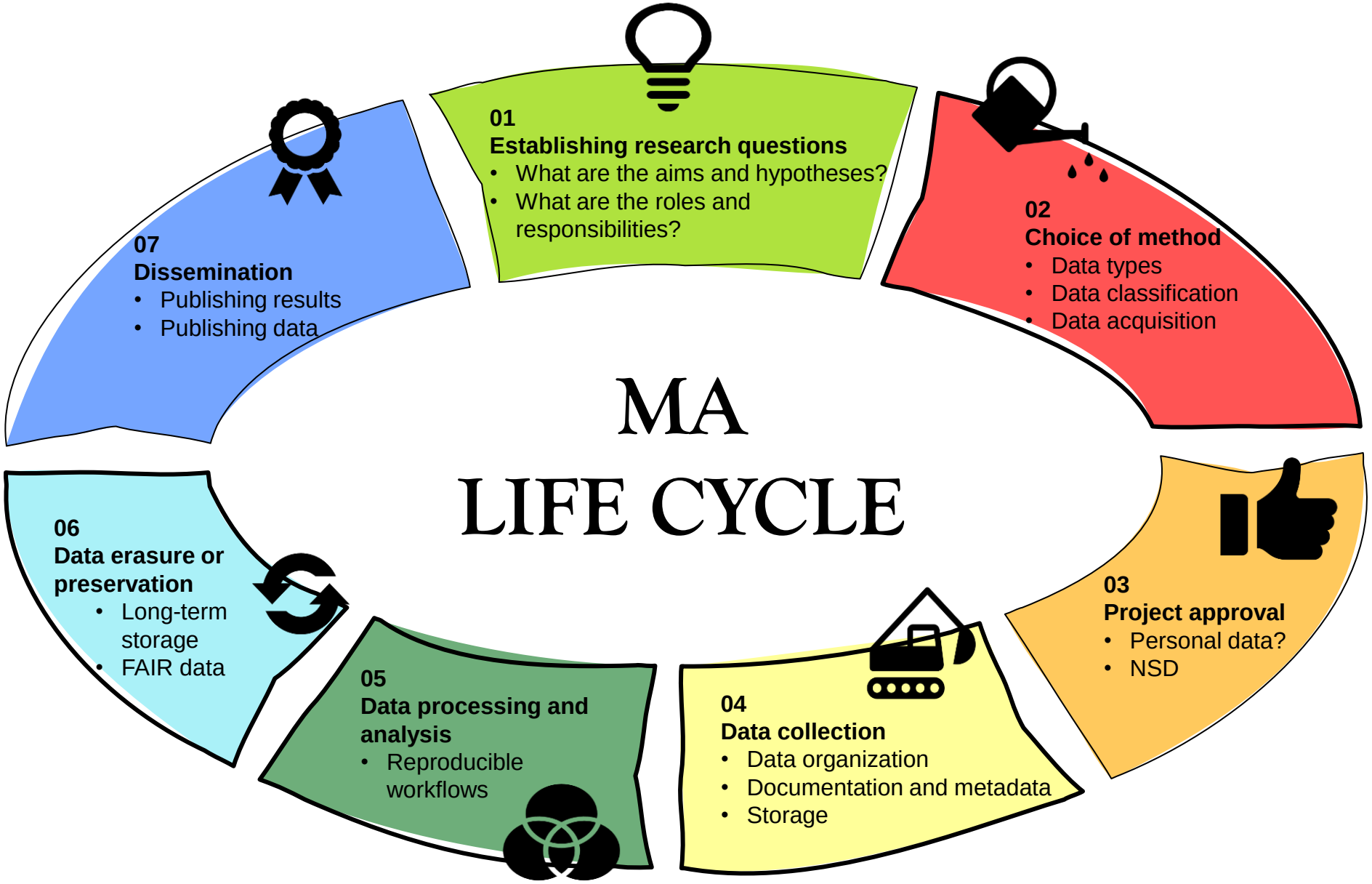
RESEARCH DATA – OPEN BY DEFAULT

Open research / Open science

Open science means transparency and knowledge-sharing in research processes to make knowledge accessible across academic groups, sectors and national boundaries. The concept of open science encompasses the entire research process […].

Source: The Research Council of Norway. Policy for open science 2020

Image by: opensourceway


MA
LIFE CYCLE
01
Establishing research questions
• What are the aims and hypotheses?
• What are the roles and responsibilities?
02
Choice of method
• Data types
• Data classification
• Data acquisition
03
Project approval
• Personal data?
• NSD
04
Data collection
• Data organization
• Documentation and metadata
• Storage
05
Data processing and analysis
• Reproducible workflows
06
Data erasure or preservation
• Long-term storage
• FAIR data
07
Dissemination
• Publishing results
• Publishing data

MA
LIFE CYCLE

Your data vs your thesis

- Not all the data you collect should necessarily be included in your thesis
- You can include data that participants have consented to publishing, or that may be properly anonymized
- You can have personal information in the thesis, but then you HAVE TO specify this when you submit in DUO, so that the thesis does not become openly available.

https://www.ub.uio.no/english/writing-publishing/masters-theses/restricting-access/index.html

Where to get help?

- **Your supervisor!**
- **Your student adviser:** Runar Forsetløkken datahandtering@iss.uio.no
- **Your librarian**
- USIT legal staff and UiO data protection officer: behandlingsansvarlig@uio.no

UiO student pages on data protection:

https://www.uio.no/english/studies/examinations/privacy-protection/

Thank you!

UNIVERSITETET
I OSLO