### Welcome to workshop 2

**Data Collection** 

### Program

- 9:30 Introduction to participants
- 9:45 Wrapping up the first workshop
- 10:00 Workshop on data collection
- 10:30 Presentations of homework
- 11:00 Brainstorm
- 12:00 Lunch
- 12:30 Putting it together (in groups)
- 15:15 Presentation of results
- 16:00 Bye bye

### Introduction to participants

Give a short description of yourself

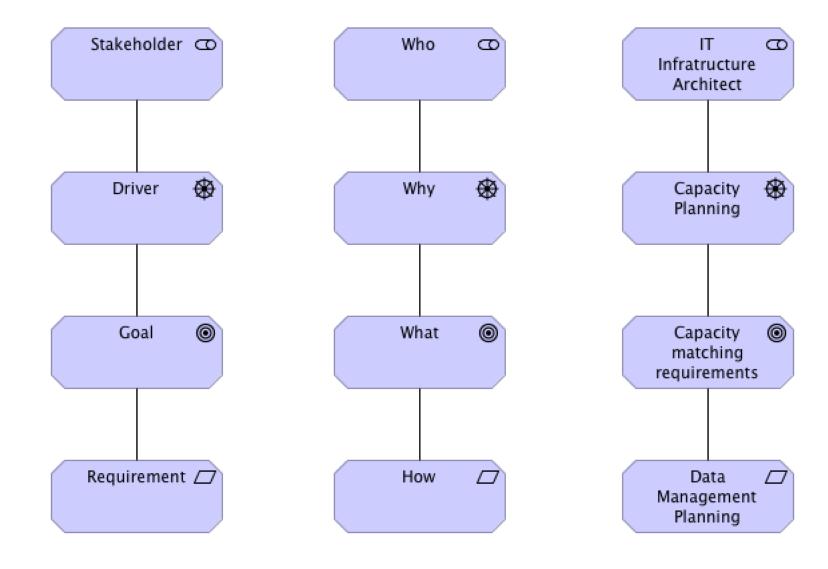
# Wrapping up the first workshop

Bo Nygaard Bai - AAU

### Worshop 1

- Too ambitious on modelling
- Fallback to a simple motivation model
  - Stakeholders
  - Drivers
  - Goals
  - Requirements
- Many stakeholders and requirements where identified
- Preliminary report shared on GitHub
  - https://github.com/Data-Management-by-Design/DMbD

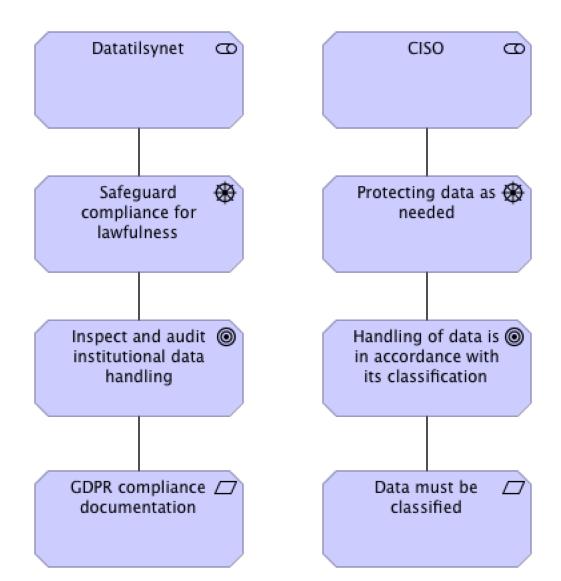
### Motivations model



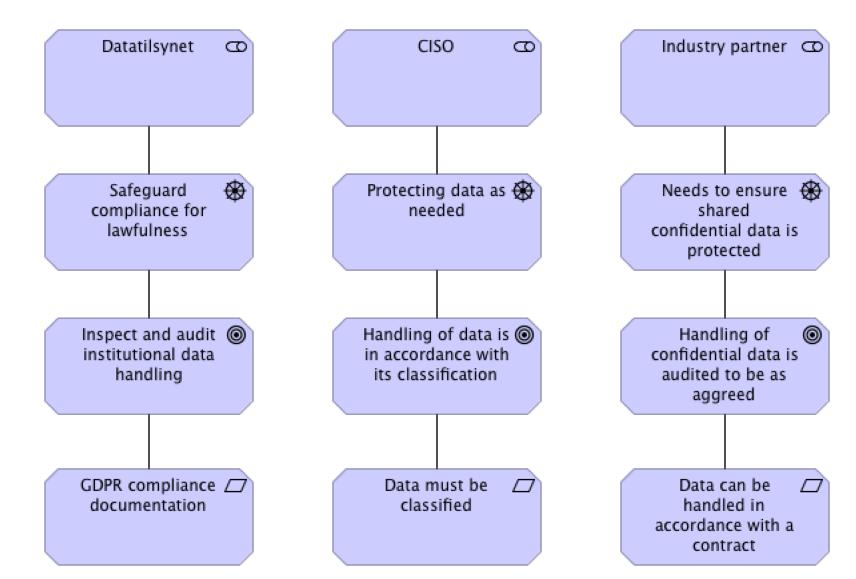
### Arkivering af fortrolige forskningsdata



### Arkivering af fortrolige forskningsdata



### Arkivering af fortrolige forskningsdata



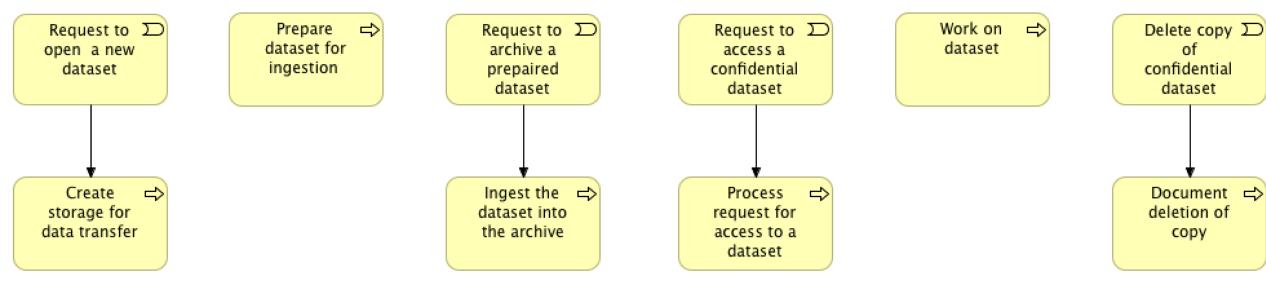
### Research Data Archive

Open Archival Information System ISO-OAIS ISO-STD 14721

**Preservation Planning** Data Management Producer Ingest Research Data Archive Access **Archival Storage** Administraion

Consumer

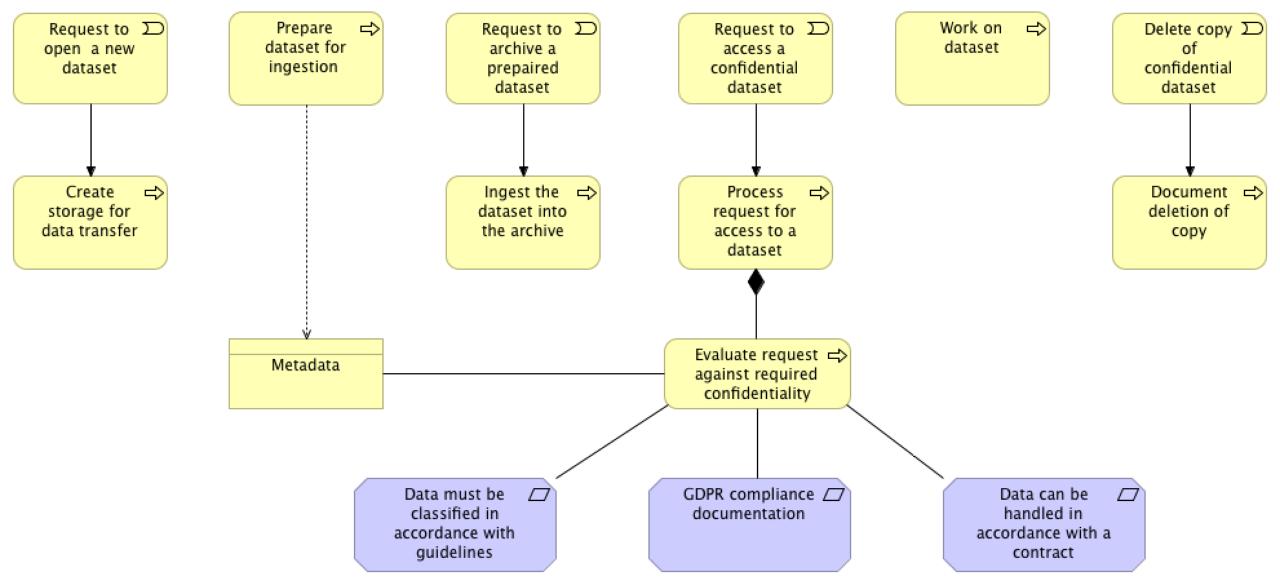
### Hændelser og processer for arkiv



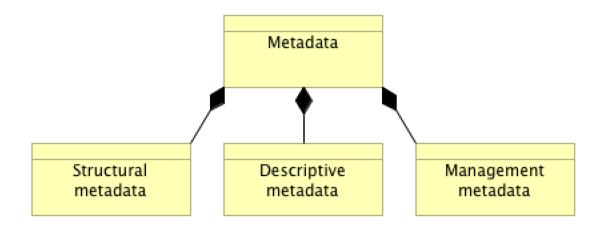
Data must be C classified in accordance with guidelines GDPR compliance 
documentation

Data can be // handled in accordance with a contract

### Hændelser og processer for arkiv



### Krav til metadata

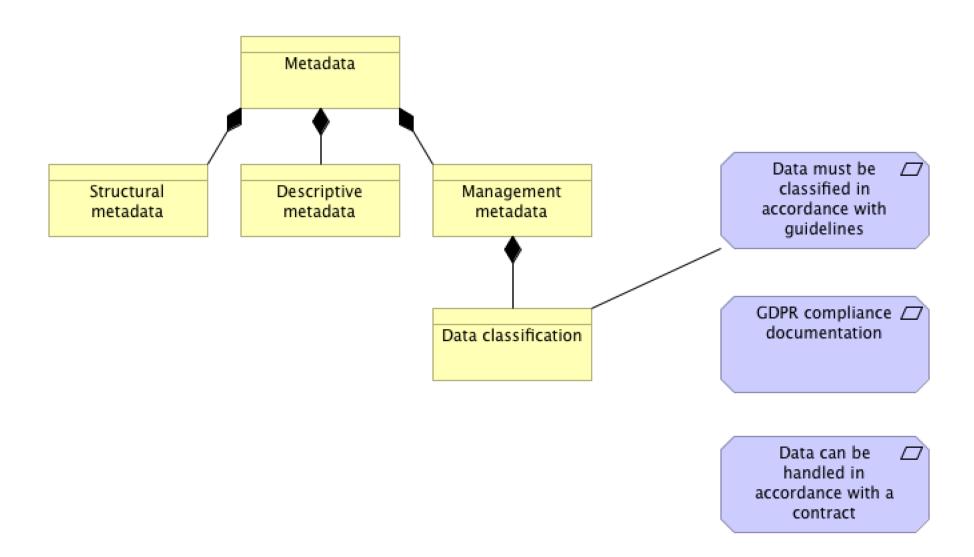


Data must be classified in accordance with guidelines

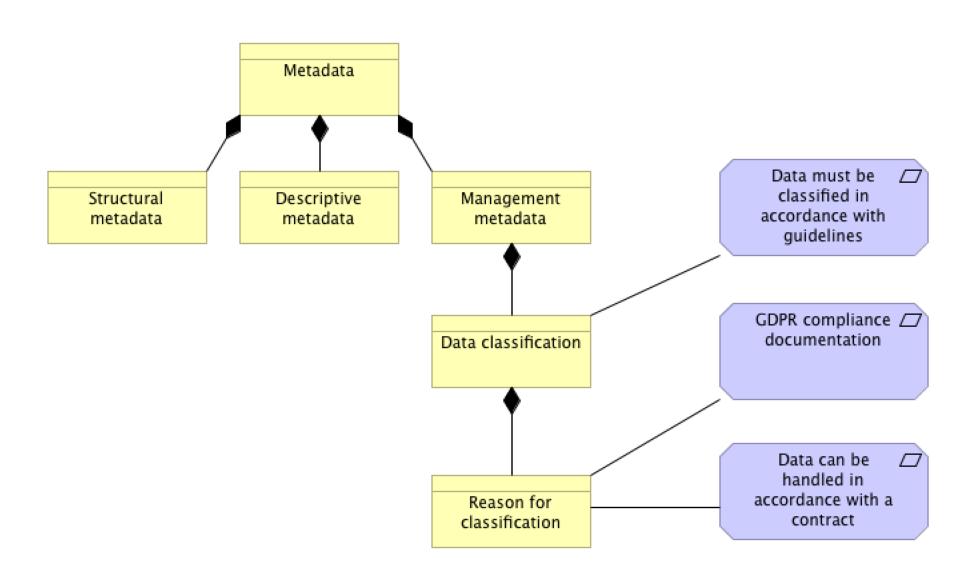
GDPR compliance 
documentation

Data can be Andled in accordance with a contract

### Krav til metadata



### Krav til metadata



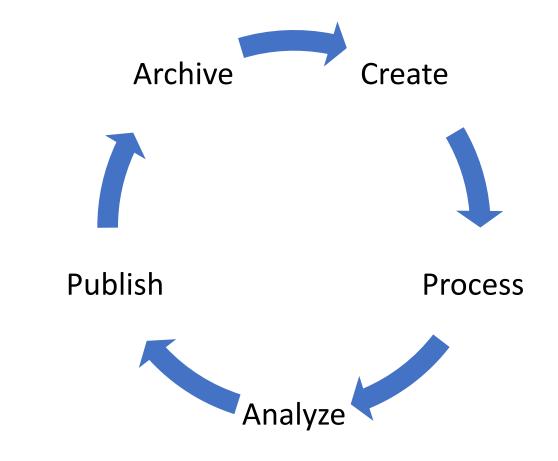
### Wrapping up WS1: Archiving

- Not today!
- We will finish it up after workshop 2 and 3 has made us more experienced working with modelling

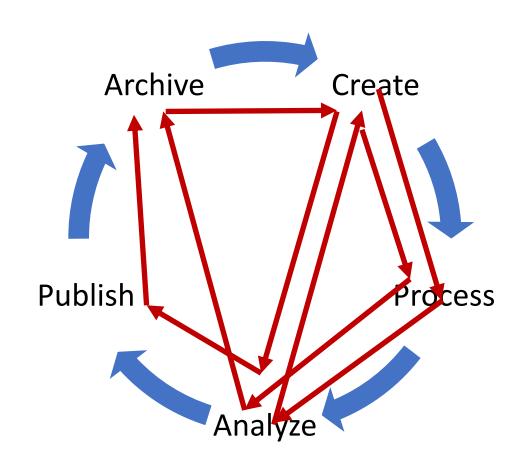
### Data Collection

The subject of todays workshop

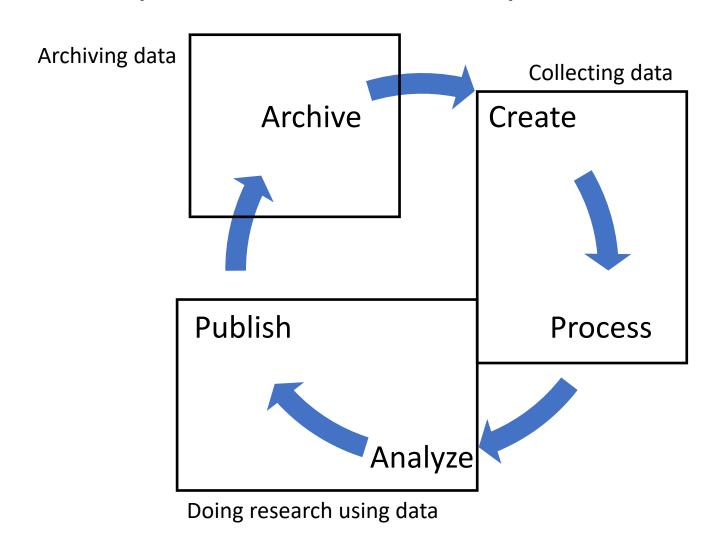
### Data's lifecycle in Research



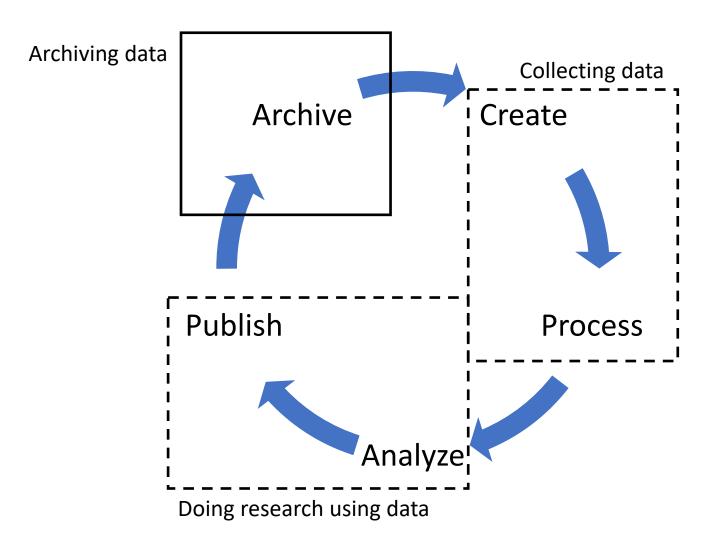
### Data's lifecycle in Research (reality check)



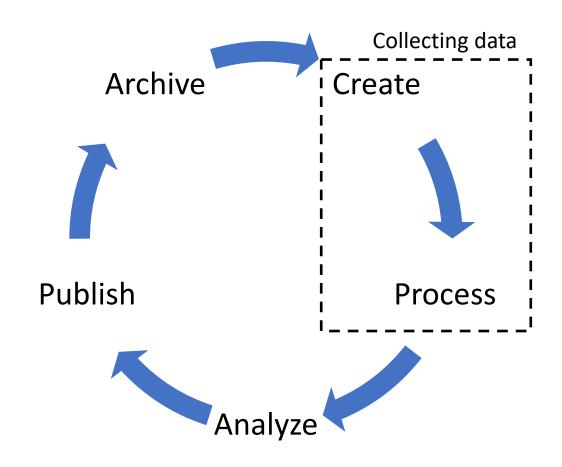
### Data's lifecycle in Workshops



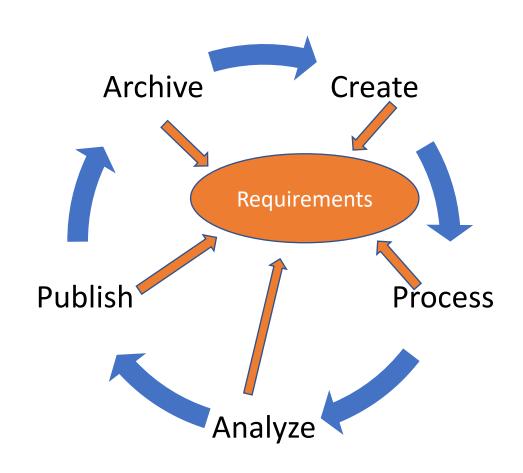
### The lines are blurry (a fact we must accept!)



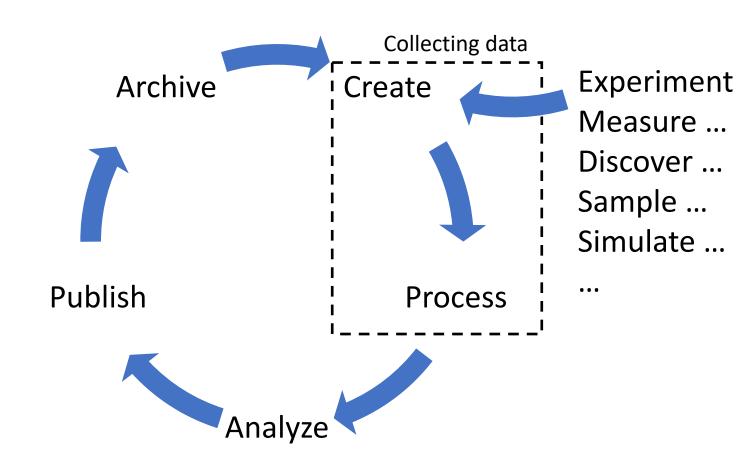
## Focus on requirements to processes supporting digital data collection

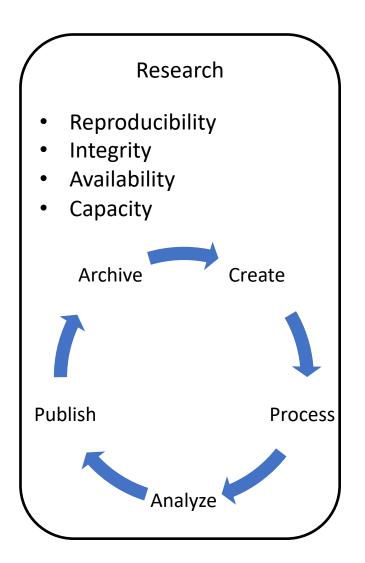


### Requirements may come from all phases



## Lab- and Field- work is probably outside the scope of common IT-infrastructure





#### Administration Research Reproducibility Accounting Integrity Procurement **Availability** Capacity HR **Facilities** Archive Create **Publish** Process Analyze

### Research Reproducibility Integrity **Availability** Capacity Create Archive **Publish** Process Analyze

#### Administration

- Accounting
- Procurement
- HR
- Facilities

#### Control

- Auditing
- CISO
- Datatilsynet

### Research Reproducibility Integrity **Availability** Capacity Archive Create **Publish** Process Analyze

#### Administration

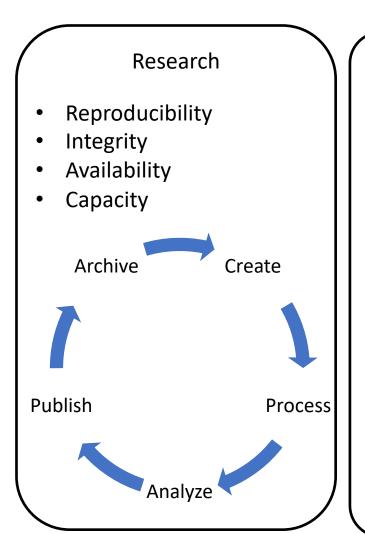
- Accounting
- Procurement
- HR
- Facilities

#### Control

- Auditing
- CISO
- Datatilsynet

#### Society

- National Law
- EU Law
- Rigsarkivet



#### Administration

- Accounting
- Procurement
- HR
- Facilities

#### Control

- Auditing
- CISO
- Datatilsynet

#### Society

- National Law
- EU Law
- Funders
- Rigsarkivet

#### Provider

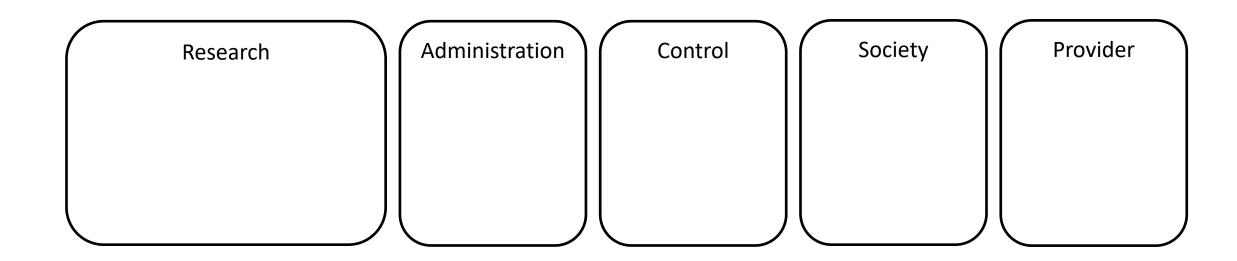
- IT-service
- Library
- PRACE
- ELIXIR
- EOSC
- Azure
- AWS
- Publisher

### Homework Presentations

Speed talking ...

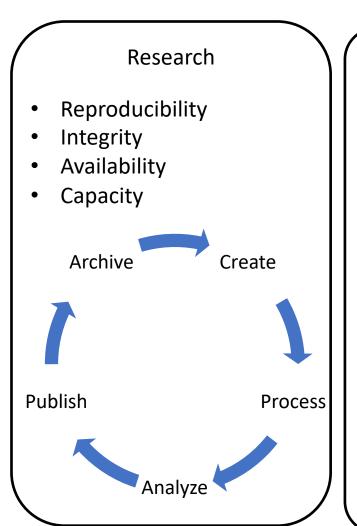
### Presentation of your homework assignment

- Very briefly explain each stakeholder entry
- Place the stakeholder in a category



### Brainstorm session

Stakeholders & Requirements What are we missing?



#### Administration

- Accounting
- Procurement
- HR
- Facilities

#### Control

- Auditing
- CISO
- Datatilsynet

#### Society

- National Law
- EU Law
- Rigsarkivet

#### Provider

- IT-service
- Library
- PRACE
- ELIXIR
- EOSC
- Azure
- AWS
- Publisher

### Categorizing requirements

Data Data format Multidimensional Multimedia Archive Create **Publish Process** Analyze

Metadata

Dublin core

DOI

Technology

Capacity

- Bandwidth
- Latency
- Frequency
- Realtime
- Availability
- Transactions
- Database

**Process** 

• Pipeline

Logging

- Lab-notebook
- Documentation

**Assertions** 

- Integrity
- Preservation
- Certification
- Legality
- License

### Brainstorm!

Missing stakeholders

Research Administration Control Society Provider

Missing requirements

Data Metadata Technology Process Assertions

### Lunch

Half time...

### Putting it together

Producing the common ground for the report ...

### Producing common ground for the report

- Each group works in a new page in the google docs document
- Rate the stakeholders and requirements by approximate importance
- Working from most important/influential to least
  - Fill in goals and drivers that tie a stakeholder to a requirement

### Presentation of results

Speed talking ...

### Bye bye!

Good work!