

Historisch-Kulturwissenschaftliche Informationsverarbeitung Woche 2

Software development

(with contributions from Christian-Emil Ore, Jon Holmen, and other colleagues at the Unit for Digital Documentation, University of Oslo and from Martin Dörr and Stephen Stead, CIDOC-CRM SIG)



The media server

What are the requirements for a system for media objects?

- Upload
- Storage
- Metadata
- Presentation
- Long term preservation

Will focus on images but equally relevant for other media types



Upload operations

- Connect to the storage
- Find a logical place to put the data
- Submit metadata
- Establish a stream connection
- Upload bits
- Check result
- Get a receipt and an identifier back
- *The client may be:*
 - *a human*
 - *a computer programme*



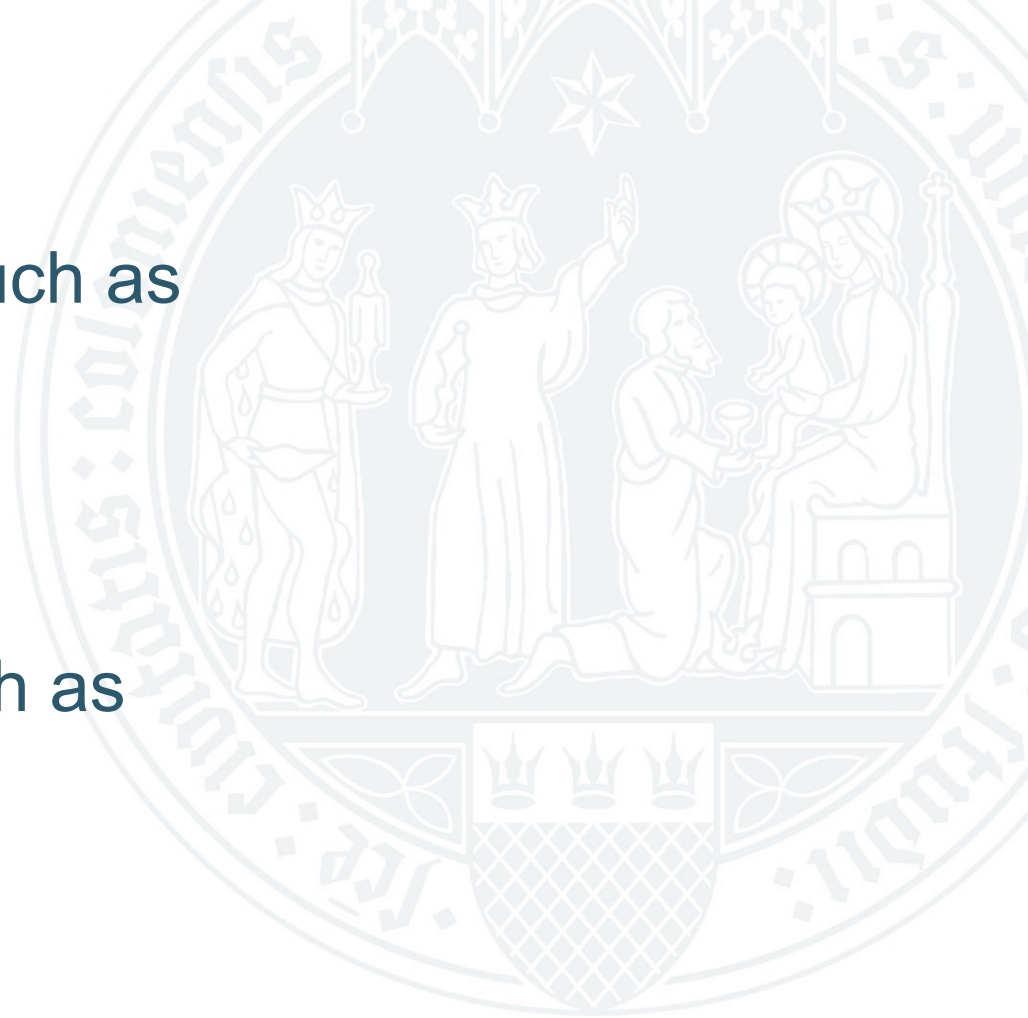
Storage operations

- Receive a request with metadata
- Return an ID for the stream
- Receive the stream
- Receive further metadata
- Enter metadata into database
- Store file based on stream on disk
- Establish link from database to disk file



Metadata

- Technical metadata such as
 - file format
 - shoot date and time
 - size
 - location and direction
- Content metadata such as
 - motive
 - classification
 - date and time
 - source
 - location



Presentation

- User interfaces for
 - searching
 - listing metadata and/or thumbnails
 - delivering images in different formats
 - protecting images
 - ordering images
 - payment
- Different platforms
- Different user groups
- Different contexts



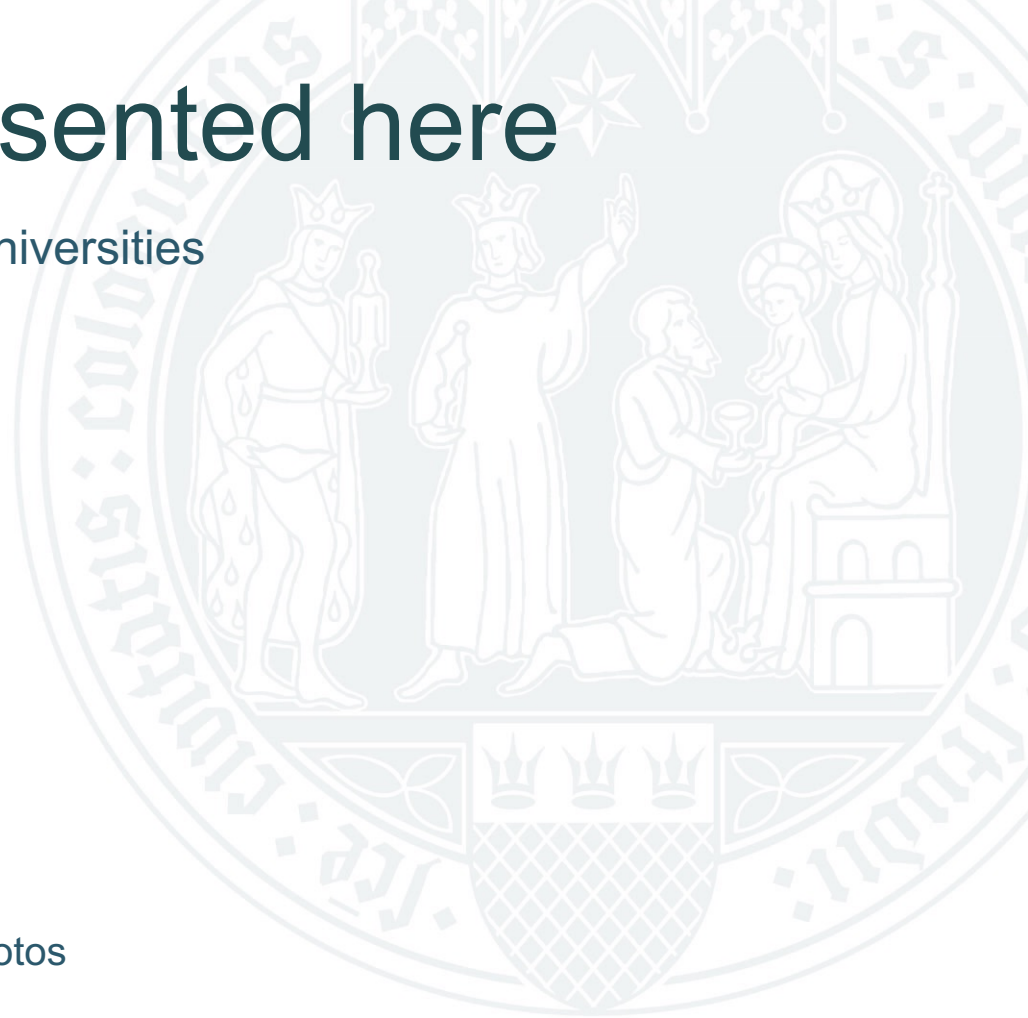
Long term preservation

- Make sure the data survives for the future
- *Long* term – not just 10 or 30 years
- Preservation
 - bitstreams
 - meaning
 - context
 - usability
- Technology
- Administration
- Politics

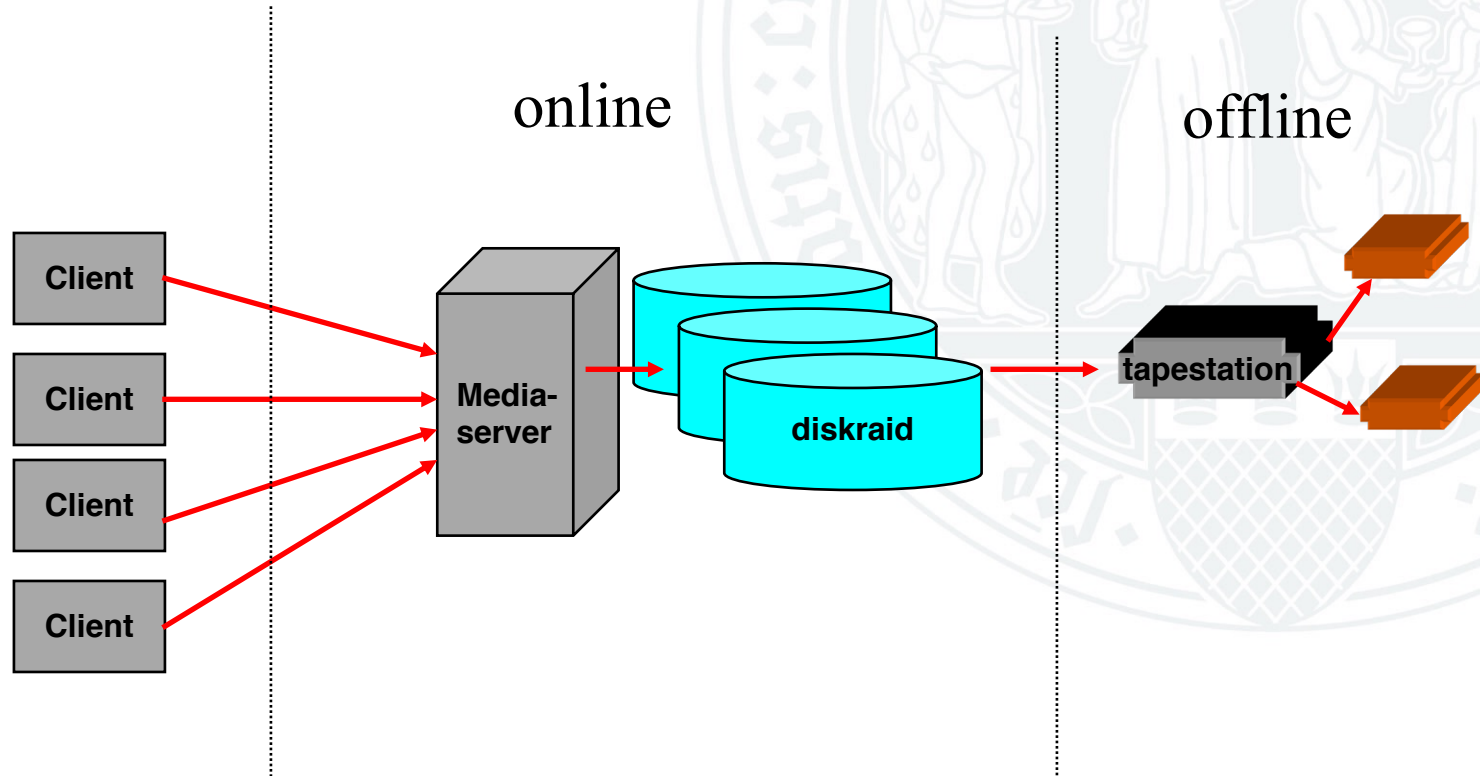


The system presented here

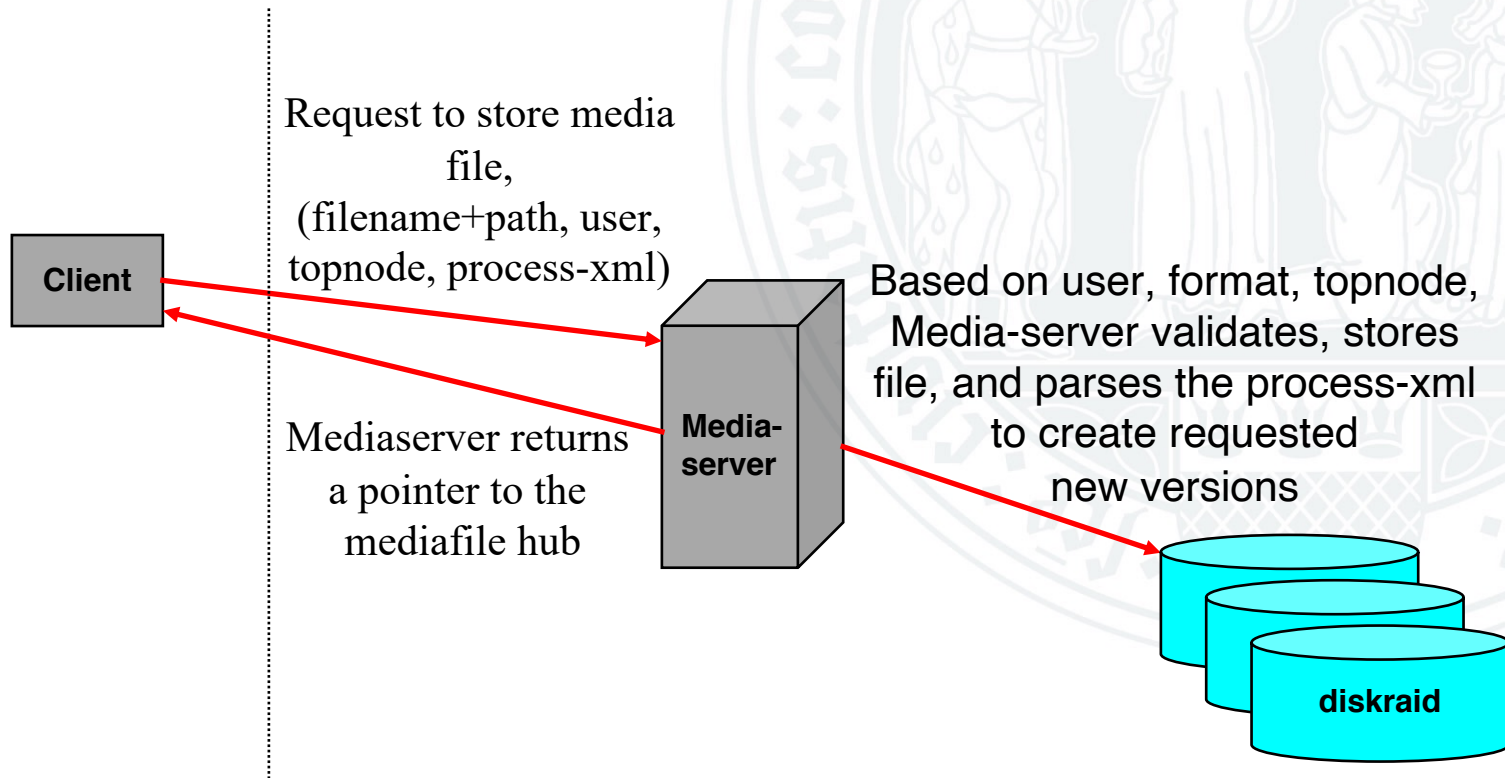
- Image collections at Norwegian universities
 - University history
 - Art history
 - Cultural history
 - Archaeology
 - Natural history
 - ...
- Document archives
 - archaeology
 - dialectology
 - ...
- Sum:
 - 1-2 000 000 traditional digitised photos
 - 3 000 000 document facsimiles
 - *(figures a few years old)*



Overall architecture



Data flow

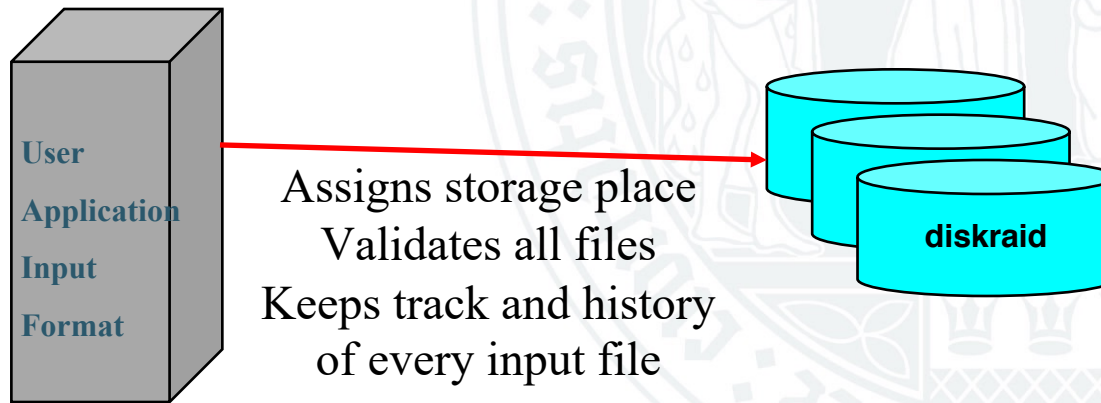


User applications

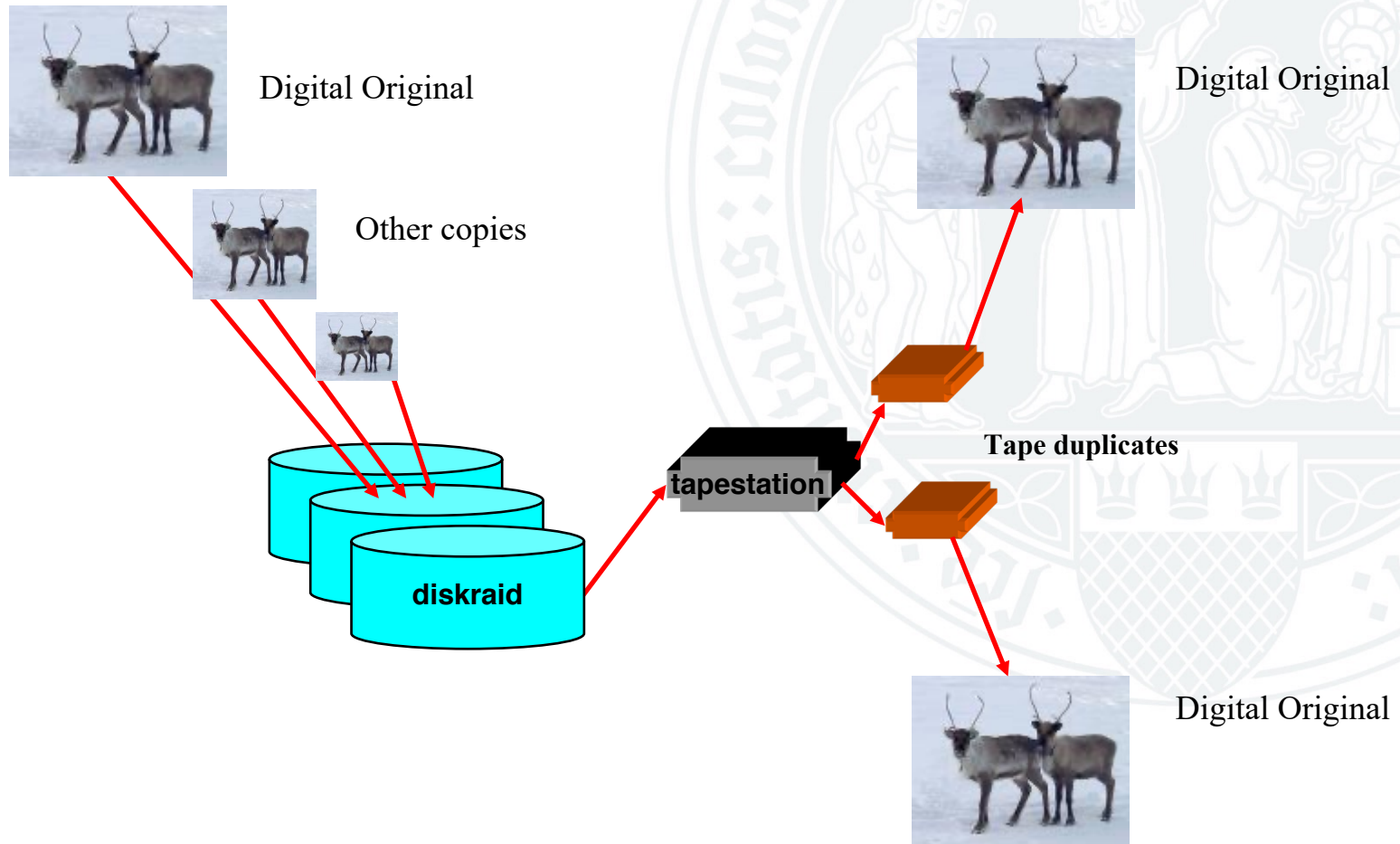
- A GUI user application is a frontend for:
 - cataloguing pictures (metadata)
 - importing pictures
 - changes and updates
- A command line application is a frontend for:
 - running import scripts
 - file list as parameter
 - meant for expert users
 - meant for large volumes
 - metadata as XML files
 - can link to pre-existing metadata
- Always connected to one discipline schema



Storage keeper



Long term preservation



Database

Discipline schema 1

- image_card
 - media_group_id
 - concent_metadata
- *subject specific tables*

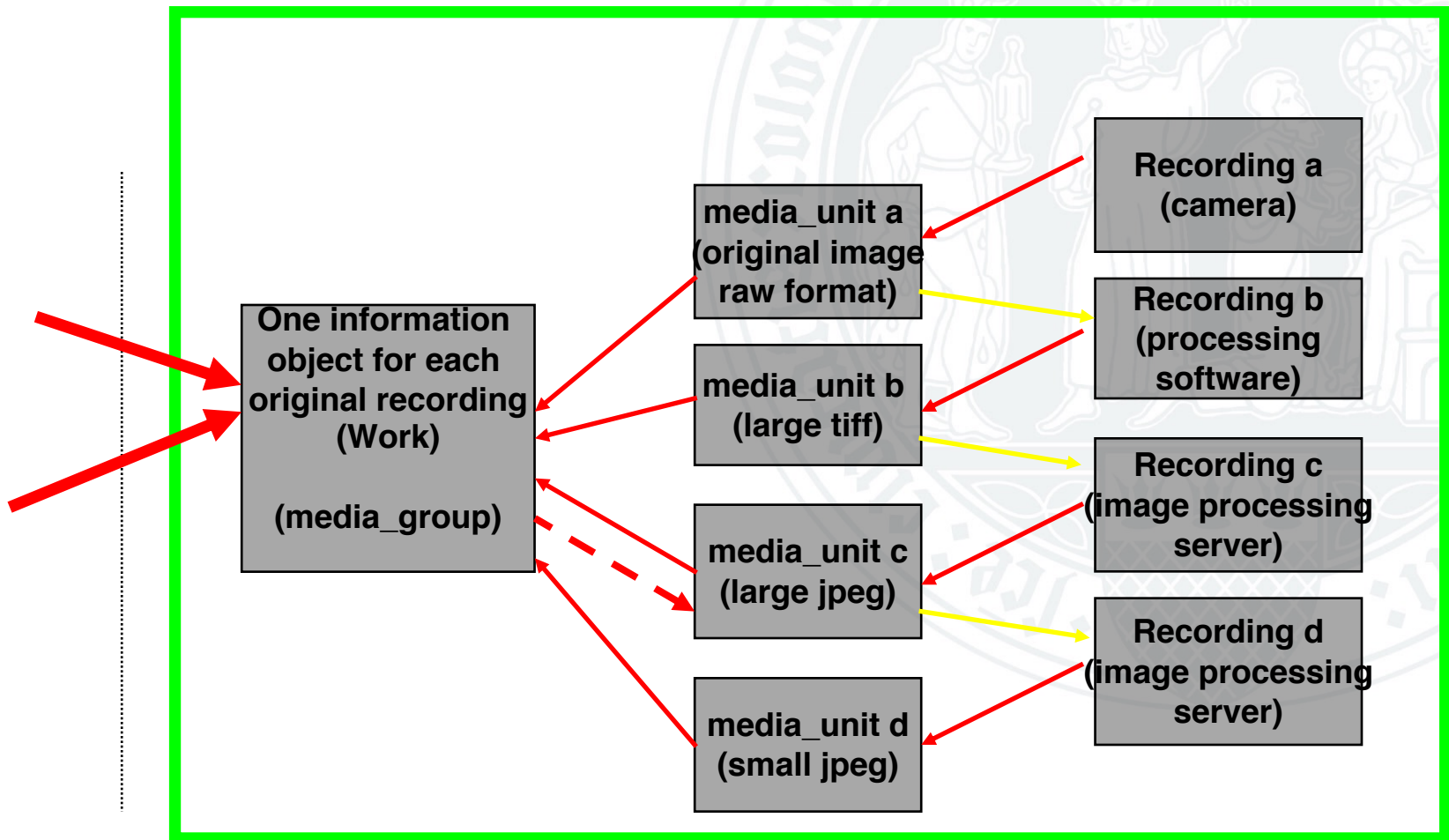
Common schema

- media_group
 - process_xml
 - tech_metadata
- media_unit
 - tech_metadata
 - default {0, 1}
- schema_setup
 - process_spec
 - delete_limit
- process
 - process_spec
 - status {0, 1, 2, 9}

Discipline schema 2

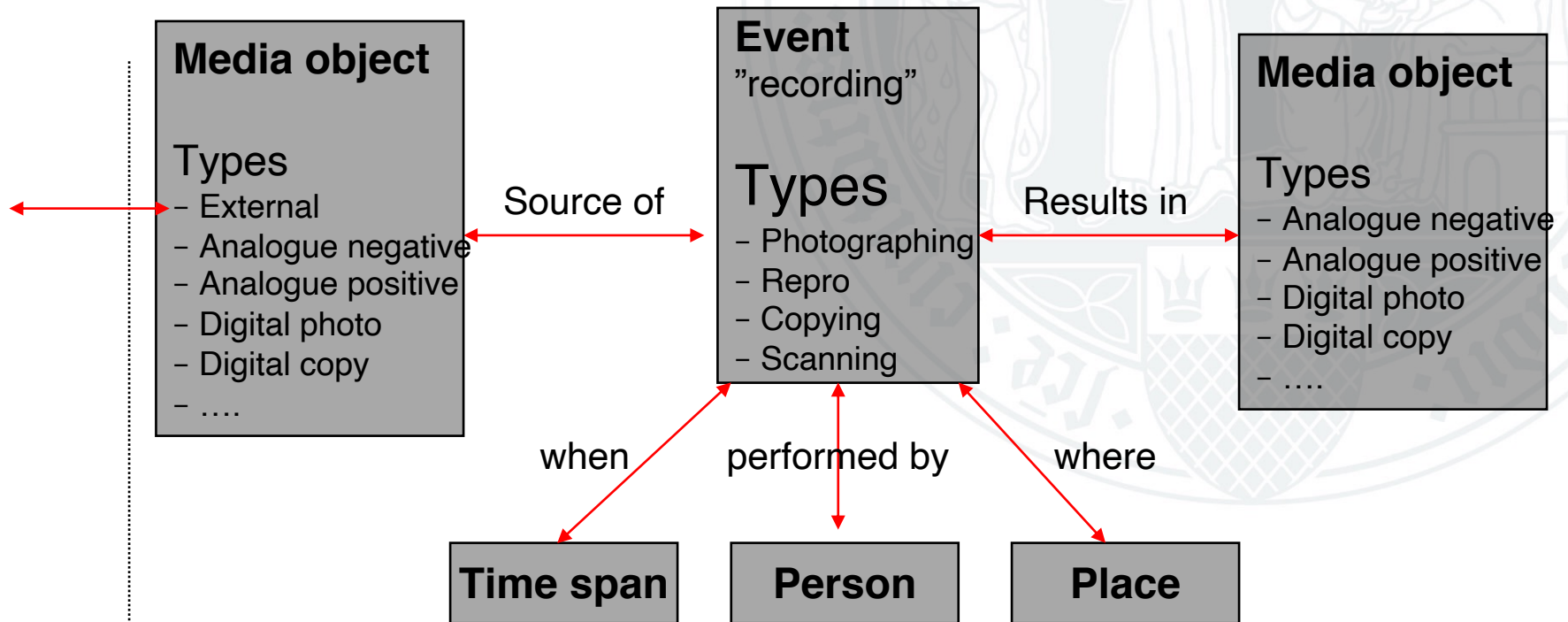
- image_card
 - media_group_id
 - concent_metadata
- *subject specific tables*

Example work flow (digital image)



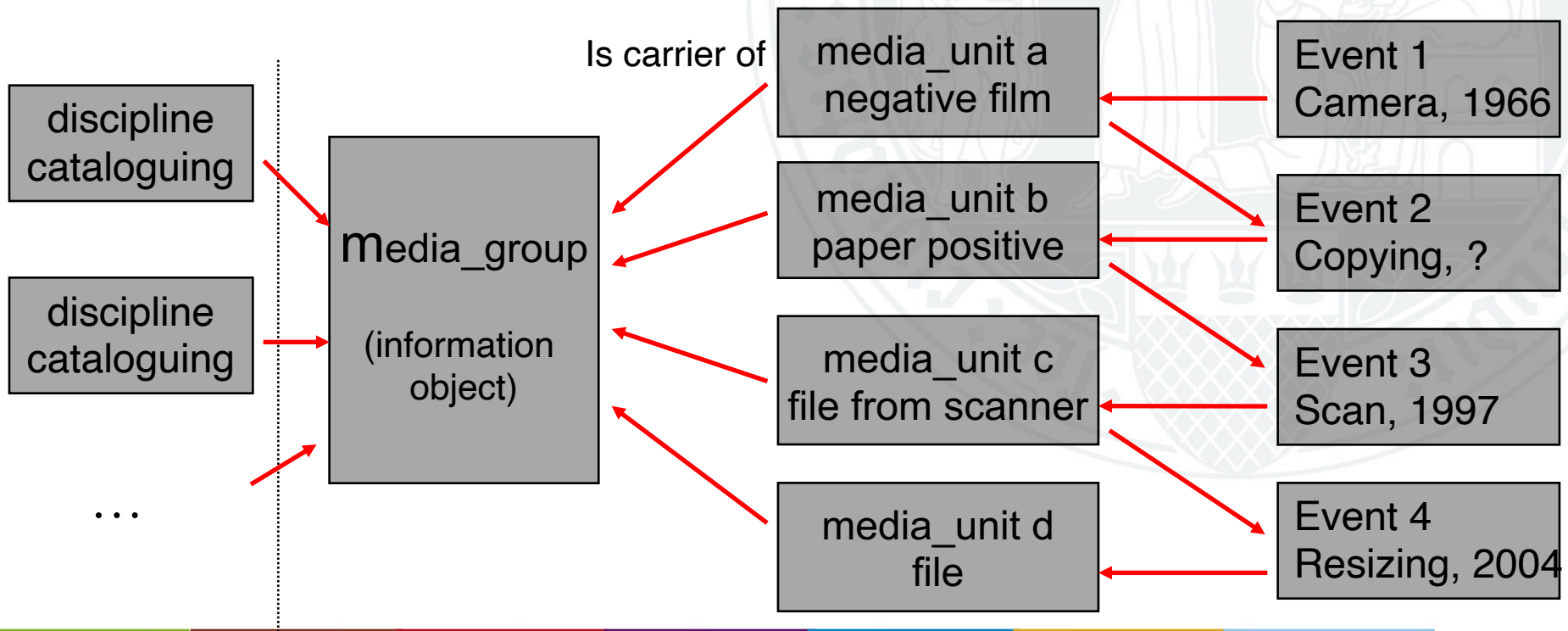
Event centric data model

source → recording → result



Data model example: digitised image

Separation between “information object” and “information carriers”



Remember: Database

Discipline schema 1

- image_card
 - media_group_id
 - concent_metadata
- *subject specific tables*

Common schema

- media_group
 - process_xml
 - tech_metadata
- media_unit
 - tech_metadata
 - default {0, 1}
- schema_setup
 - process_spec
 - delete_limit
- process
 - process_spec
 - status {0, 1, 2, 9}

Discipline schema 2

- image_card
 - media_group_id
 - concent_metadata
- *subject specific tables*

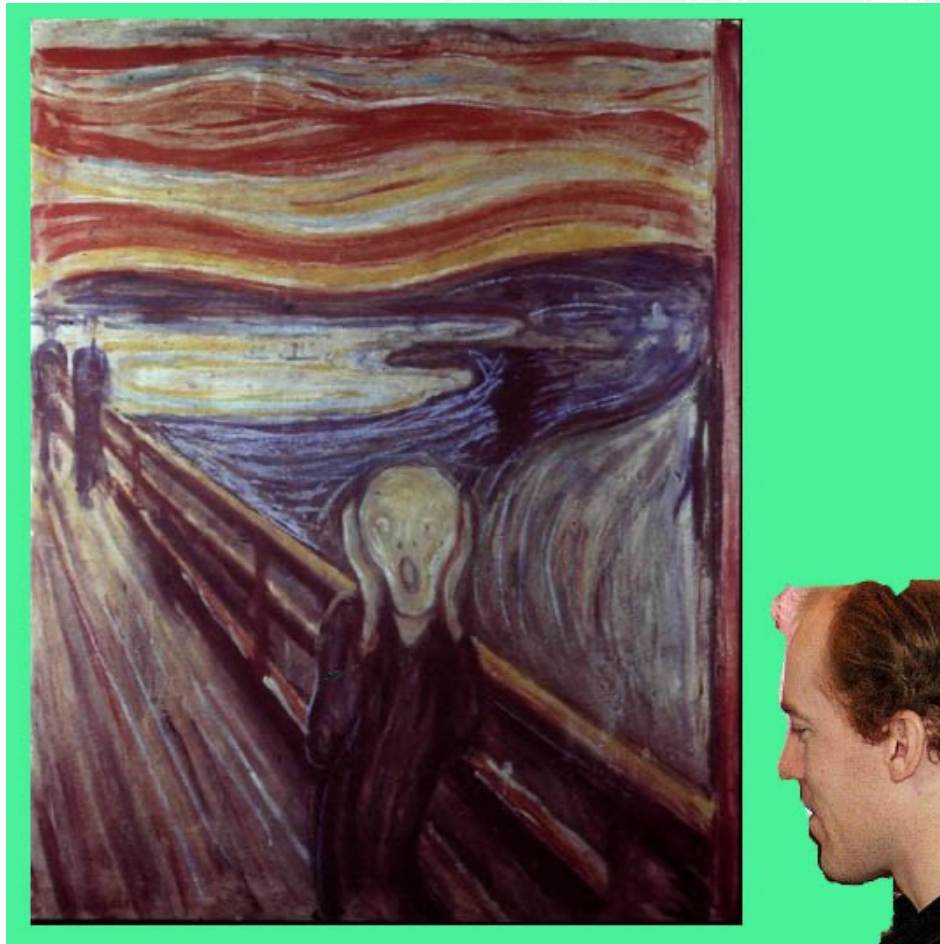
Which discipline?



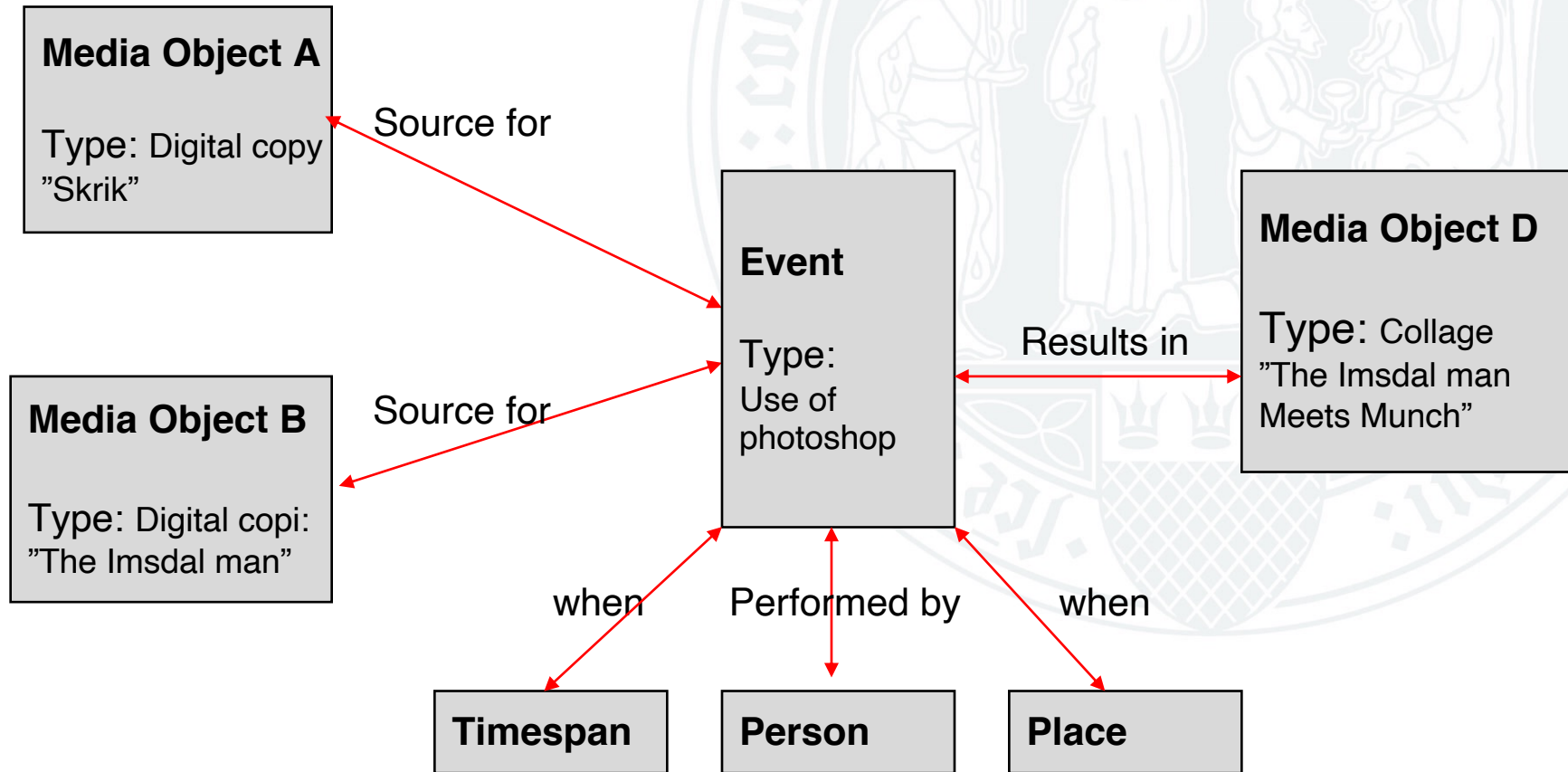
Which discipline?



Which discipline?



Event centric data model



Date model, concepts and things

