Výzkum a vývoj v oblasti výzkumných dat a disertací

Schöpfel, Joachim; Prost, Hélène; Malleret, Cécile
2017

Research and Development in the Field of Research Data and Dissertations
The *D4Humanities* Project at the University of Lille

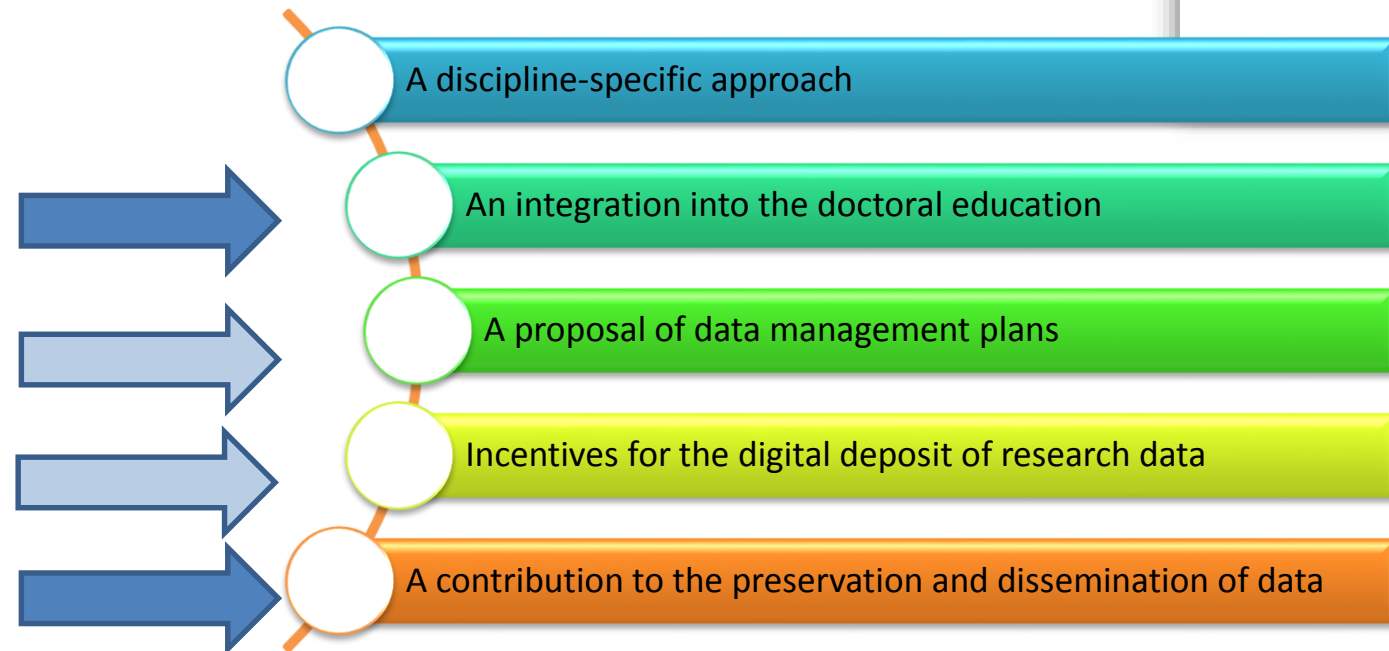## Joachim Schöpfel, Hélène Prost, Cécile Malleret

Université de Lille 3 SCIENCES HUMAINES ET SOCIALES

Gériico

MESHS Lille Nord de France

Région Hauts-de-France

Our project

- White paper on data in dissertations
- 2015-2018
- On the agenda:
  - Education
  - Workflow

A discipline-specific approach

An integration into the doctoral education

A proposal of data management plans

Incentives for the digital deposit of research data

A contribution to the preservation and dissemination of data
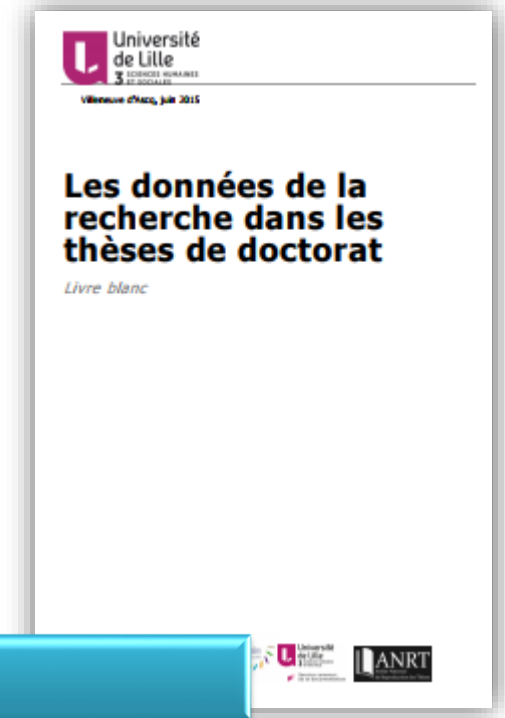
Data & ETD

National context

2006 : Centralized system STAR
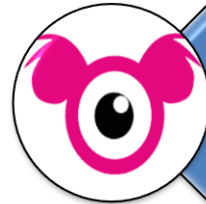
Platform NAKALA

Platform DoRANum

Service DMP OPIDoR

Data & ETD

Local context

Mandatory ETD policy	New institutional repository is under development (Dspace)	Strorage limited in the cloud

Motivation of PhD students	**Needs**
	• Storage and long term preservation solutions
	• Advice and assistance for RDM
	  • Deposit
	  • Description
	  • Legal aspects
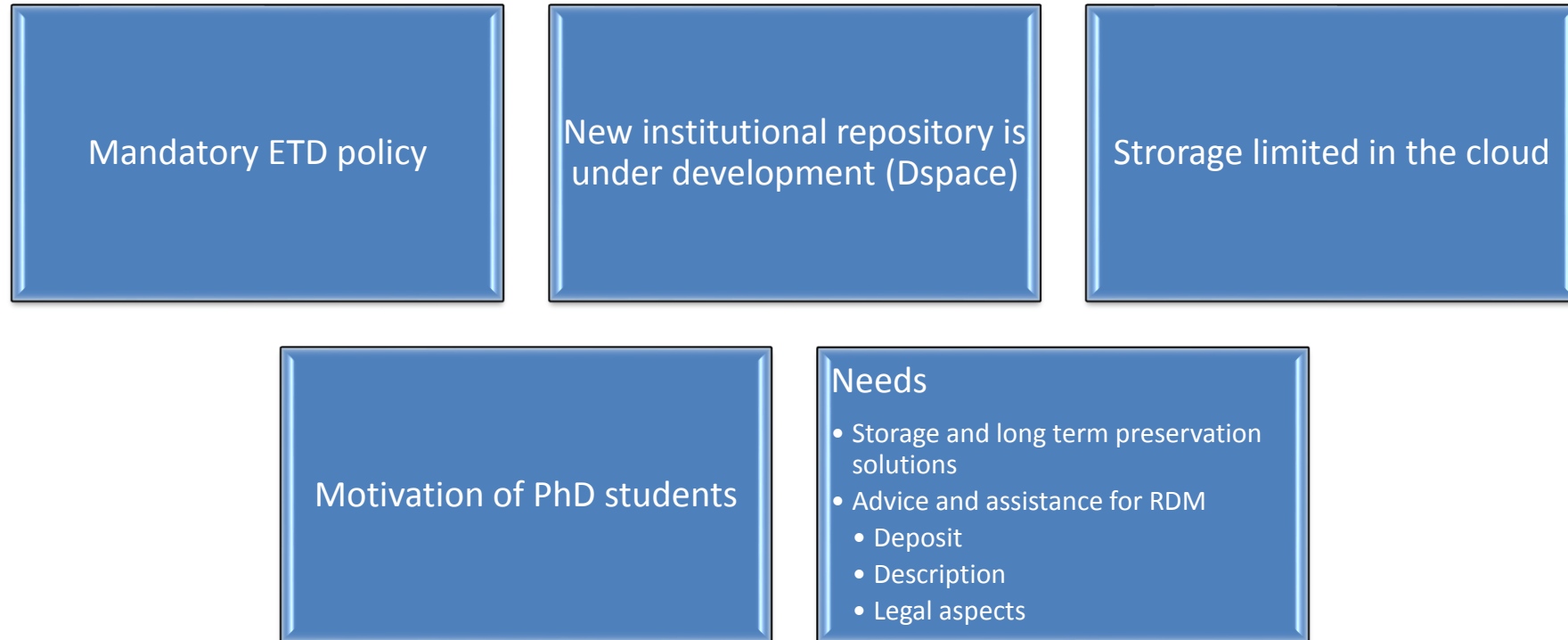
Local ETD data workflow
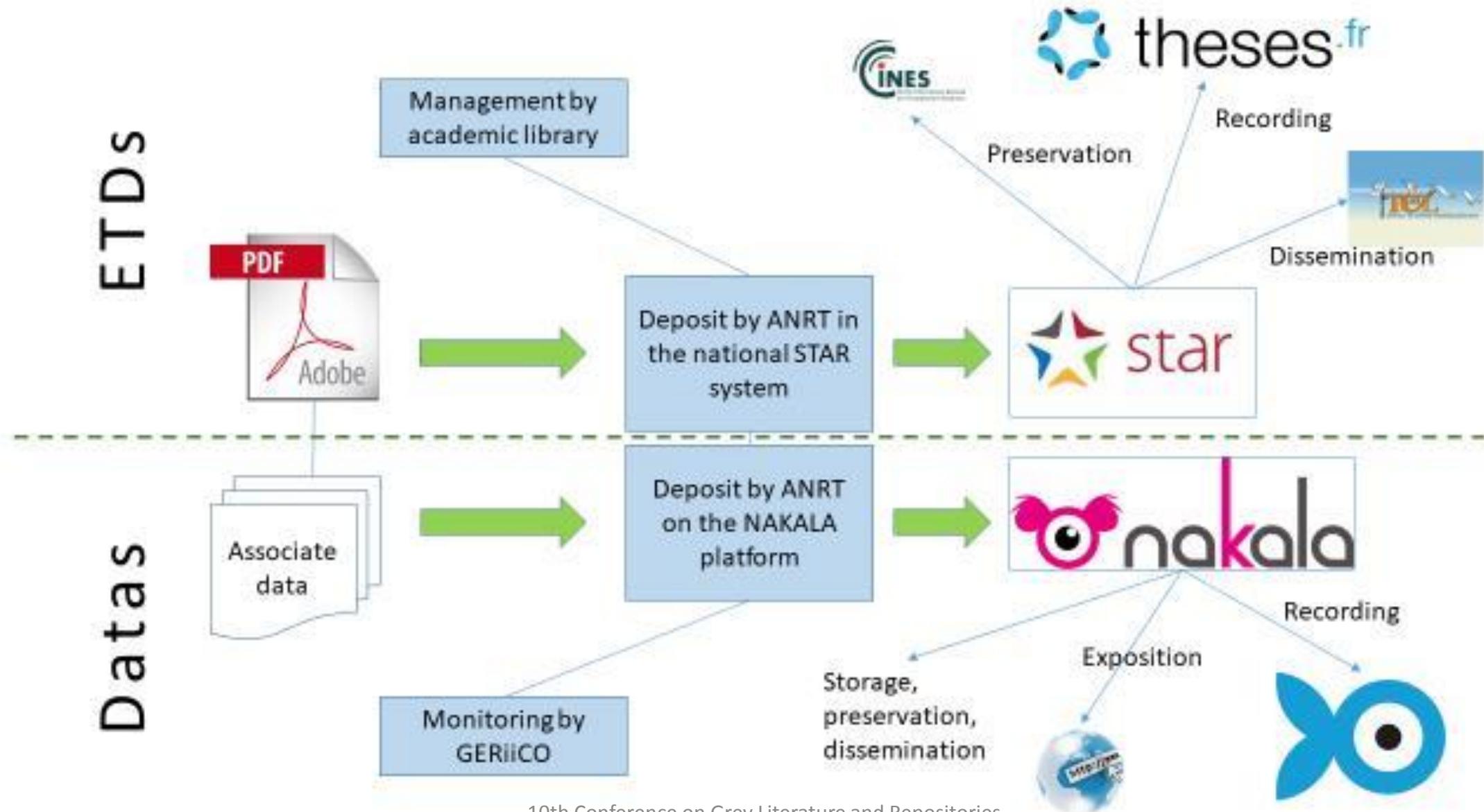
Not a development of a new system
but
the interconnection of two existing systems

- STAR & NAKALA

- Interface of deposit (?)

- Procedure

- Issues of compliance and interoperability

Main issues of discussion

- **Content and coverage**
  - Granularity
    - Reuse
  - Data format
    - Checklist
  - Data base

- **Metadata**
  - Indexing
    - ETD metadata
  - Data structure
    - METS
  - Referentials
    - 5 DC elements
  - Identifier
    - Handle
  - Source code

- **Other issues**
  - Legal aspects
  - Deposit
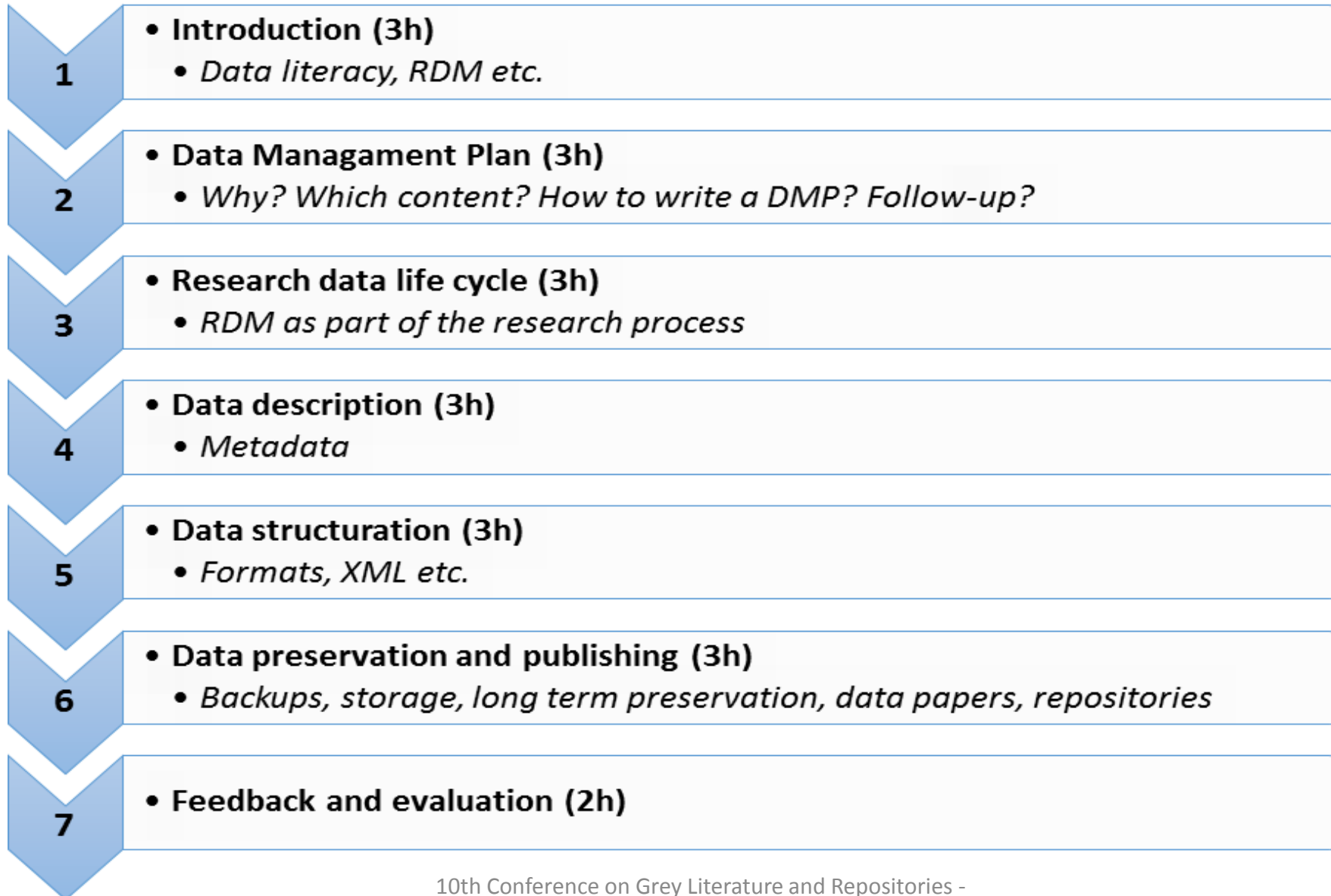   - Who has access?
  - Data size

Other issues

- Long term preservation
  - National infrastructure

- Quality
  - Validation?
  - Filter?

- Promotion
  - *But no « data sharing ideology »*
- Technical documentation
  - *For students*
  - *For staff*

The PhD training program

1
- **Introduction (3h)**
  - *Data literacy, RDM etc.*

2
- **Data Managament Plan (3h)**
  - *Why? Which content? How to write a DMP? Follow-up?*

3
- **Research data life cycle (3h)**
  - *RDM as part of the research process*

4
- **Data description (3h)**
  - *Metadata*

5
- **Data structuration (3h)**
  - *Formats, XML etc.*

6
- **Data preservation and publishing (3h)**
  - *Backups, storage, long term preservation, data papers, repositories*

7
- **Feedback and evaluation (2h)**

Key elements of training program

- Mixed team (scientists, librarian)
- Multidisciplinarity (but limited to SSH)
- 20 hours, six months
- Different levels of PhD projects
  - *May be discontinued*
- Mix of (some) theory and (much) practice
- PhD DMP as the red line of the training program
  - *With DMP platform*
- Individual follow-up and time for discussion
- Evaluation

LESSONS LEARNED

Not a technical problem

- RDM not a technical problem
- But often a « people problem » (Ward et al. 2011)
- "Improving tools are not the only steps necessary to overcome barriers. The next steps will likely involve training for scientists (...)" (Tenopir et al. 2015)
- Our focus is not on data but on people
- The most important decision of our project was perhaps the choice of a specific target group

Technical solution and education

- Data literacy (DMP, RDM) as good scientific practice
- It is not enough to develop RDM tools if you don't teach scientists how (and why) to use them
- Infrastructure will shape the training program
- But the training program contributes to further development
- An organizational learning process on the campus.

Training   Workflow
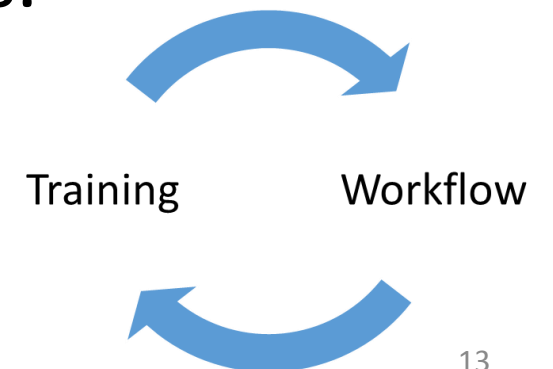
NOT a library project

- The library staff is part of the project, their contribution is essential for the success of the project
- Yet, designed as a research project with a doctoral training program, under the responsibility of the graduate school, with a scientific project management, and under political and strategic leadership
- Why:
  - legitimacy (in the sense that scientists have everyday experience with RDM);
  - scientists usually don't consider RDM as a "library affair" but as part of their daily reseach work with other scientists and technical staff.

Not limited to education and workflow

- Its main character is research, on two different levels
- Follow-up
  - monitoring and assessment of the training program
  - including feedback and continuous adjustment
  - evaluation of uptake and usage of the new data workflow
- Insight
  - type, format and content of datasets
  - impact of data on format and content of ETDs
  - the text and data mining of dissertations and data

Research — Follow-up / Insight
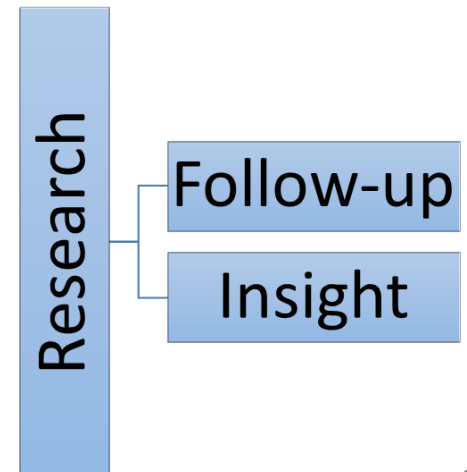
THANK YOU !

Contact
joachim.schopfel@univ-lille3.fr