



conference on  
grey literature  
and  
repositories

# NTK

50°6'14.083"N, 14°23'26.365"E  
Národní technická knihovna  
National Library of Technology

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

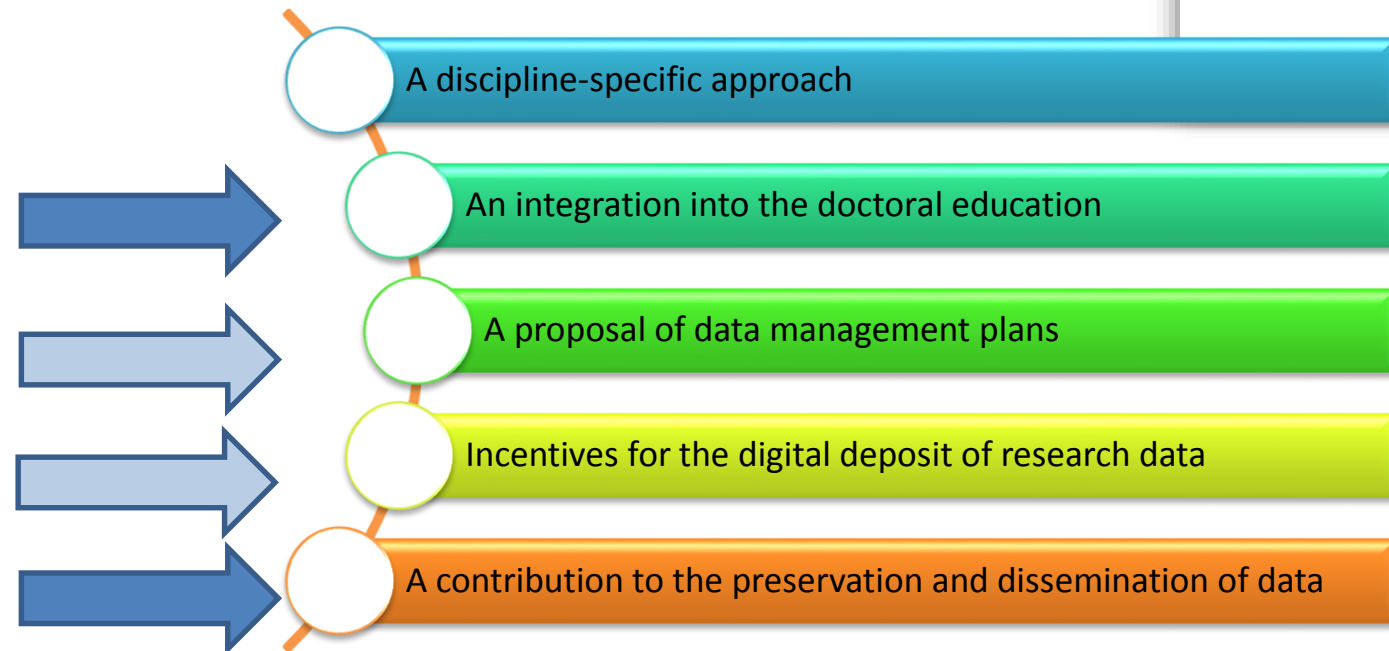
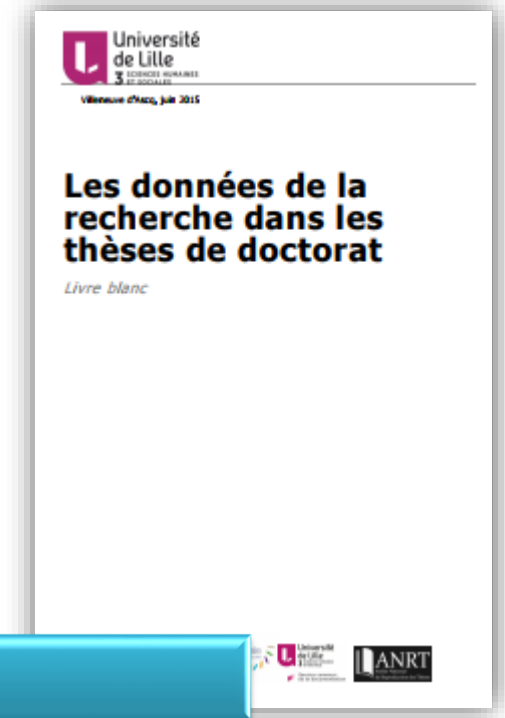
**Joachim Schöpfel, Hélène Prost, Cécile Malleret**

This presentation is licensed under the Creative Commons: [CC-BY-SA-4.0](https://creativecommons.org/licenses/by-sa/4.0/),  
via <http://www.nusl.cz/ntk/nusl-367310>



# Our project

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

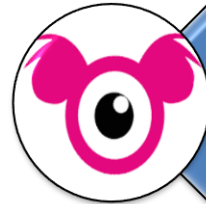


# Data & ETD

National context



2006 : Centralized system STAR



Platform NAKALA



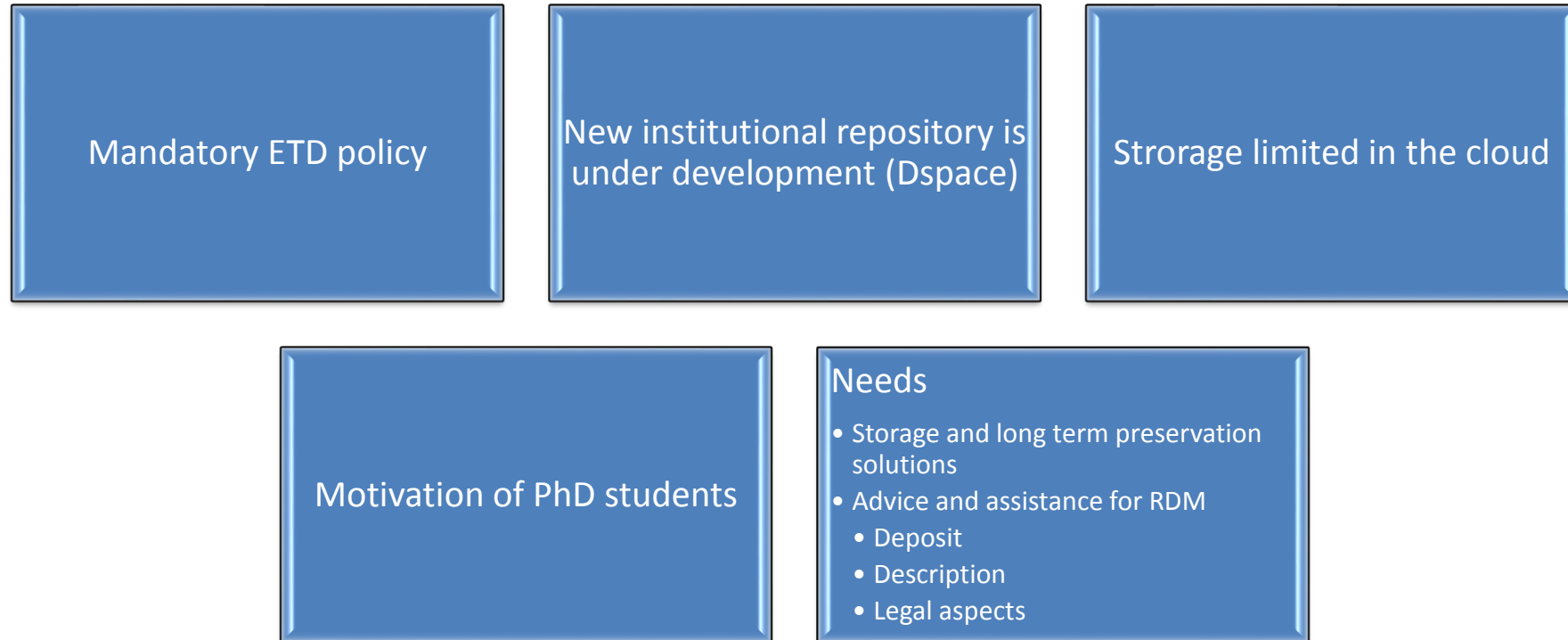
Platform DoRANum



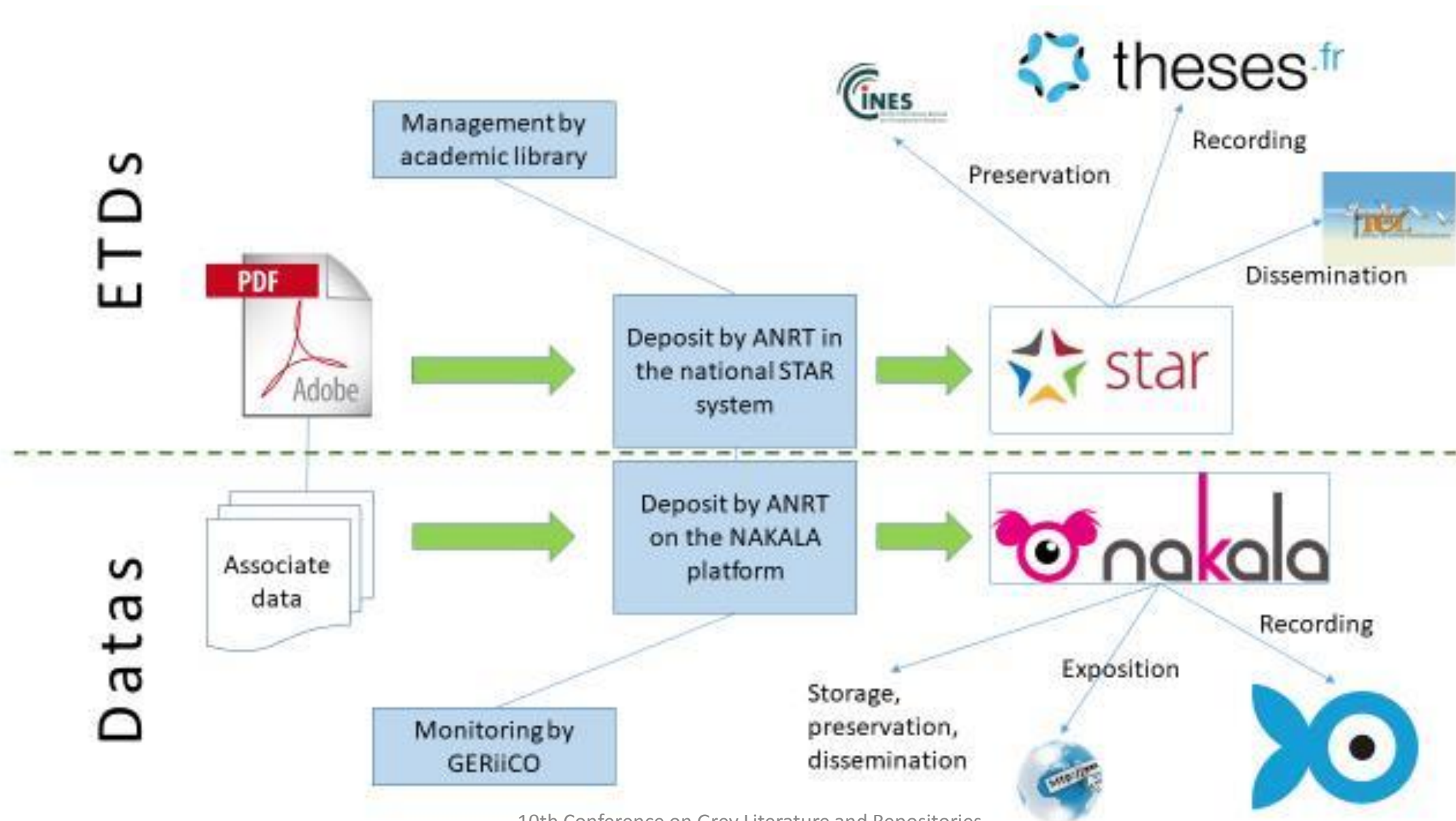
Service DMP OPIDoR

# Data & ETD

Local context



# 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

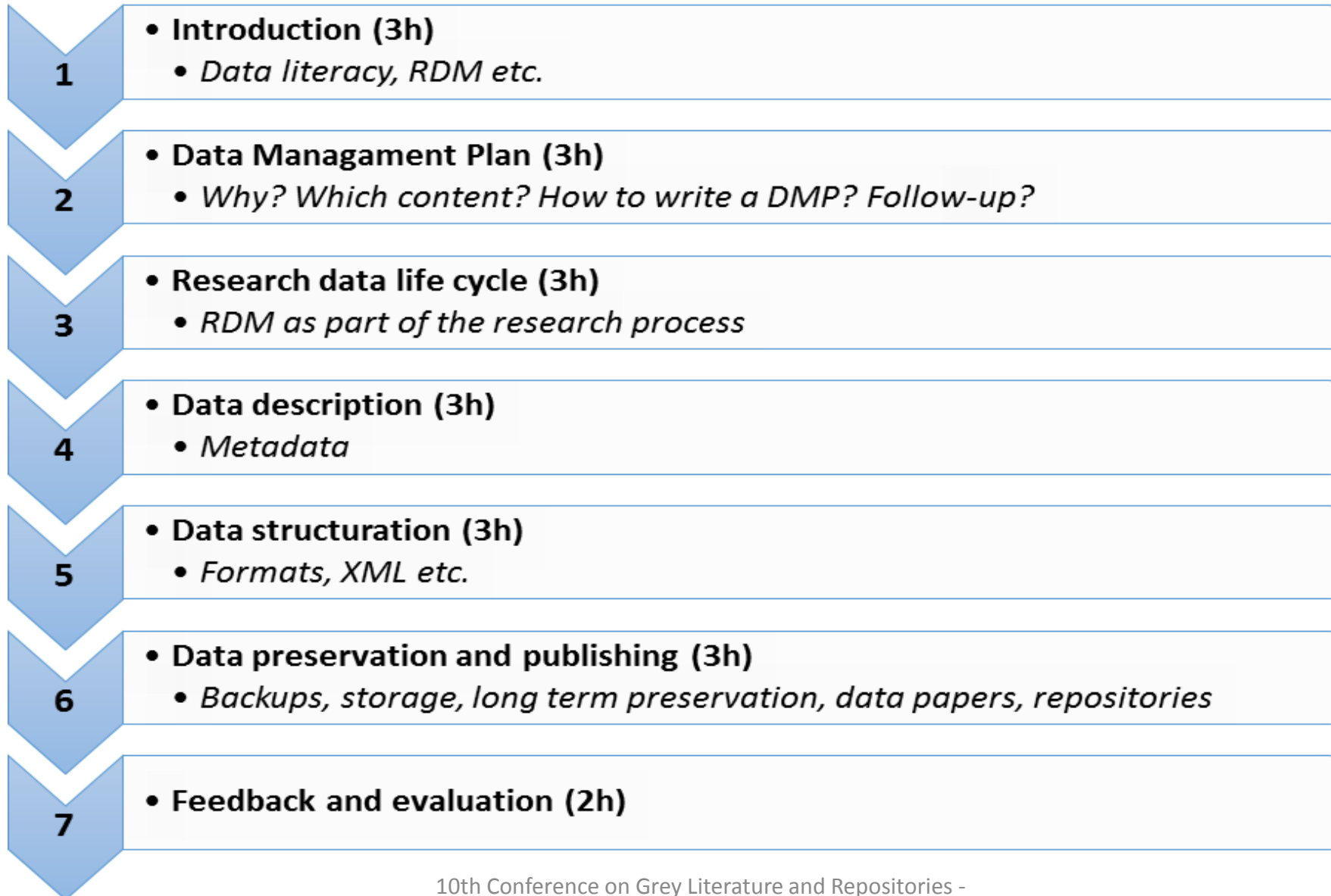
- Long term preservation
  - National infrastructure

- Quality
  - Validation?
  - Filter?

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



# The PhD training program



# 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

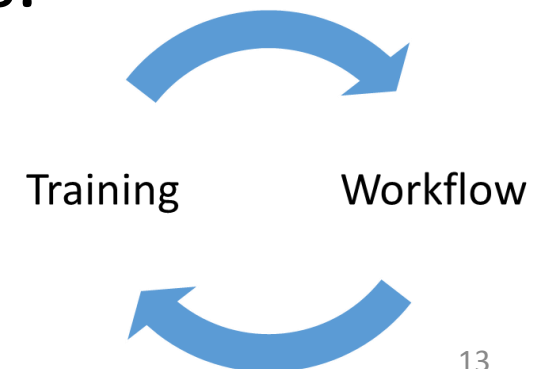
# LESSONS LEARNED

# 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.

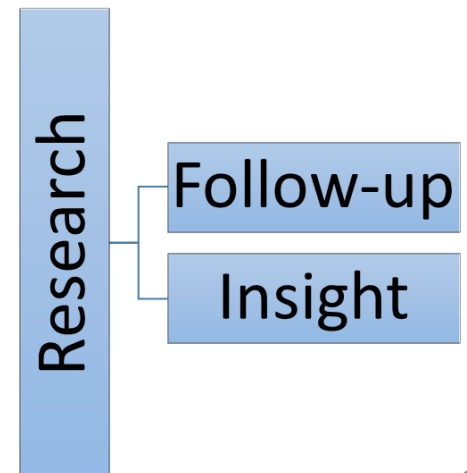


# NOT a library project

- 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 research 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



# THANK YOU !

Contact

[joachim.schopfel@univ-lille3.fr](mailto:joachim.schopfel@univ-lille3.fr)