



# **RDM Training for PhD Students**

Bernard Jacquemin, Joachim Schöpfel, Eric Kergosien, Cécile Malleret University of Lille, GERiiCO laboratory

"Improving tools are not the only steps necessary to overcome barriers. The next steps will likely involve training for scientists (...)" (Tenopir et al. 2015)

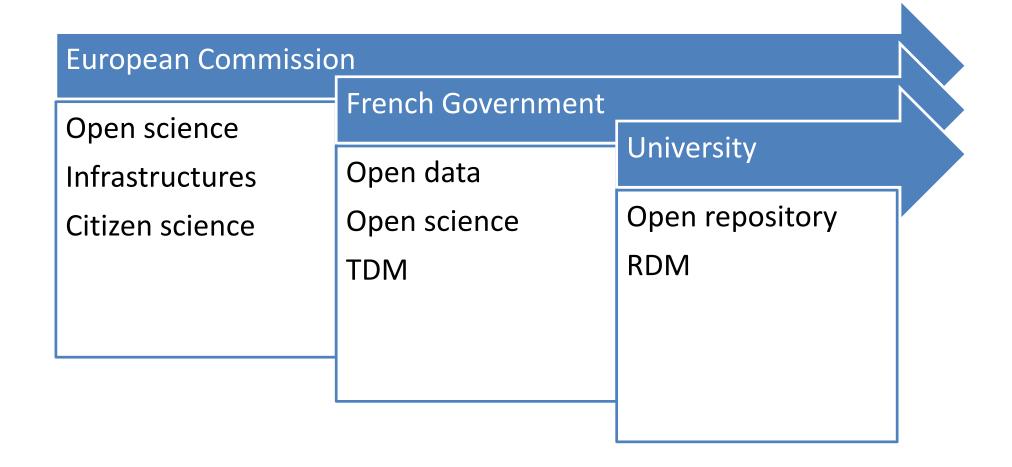








## The context

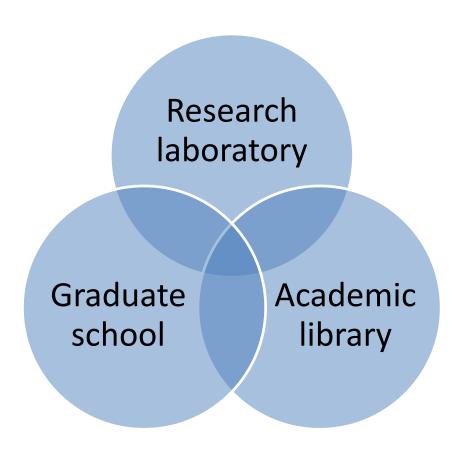


## The issue

- How to develop research data management as good scientific practice?
  - Also, to increase the number of successful project proposals
- How to improve data literacy?
  - Especially with early career scientists
- How to contribute to a « data culture » in social science and humanities?

## The framework

- Development of data literacy as part of a research project
  - Surveys on data behavior and related needs, studies on the concept of research data, on ethics, standards, policy, data mining, data repositories
  - Development of a local data workflow with national data infrastructure



# Launch of the program





- 2015: Three conferences at Lille on research data in ETDs
  - With PhD students
  - With scientists and data librarians from different countries
  - On RDM, legal issues and TDM
- 2015: White paper on RDM related to ETDs (nearly 4,000 downloads)
- 2015-2016: Launch of the training program with funding from the Graduate School, as part of PhD education in research methodology in social sciences and humanities
- 2016: International conference on ETDs at Lille



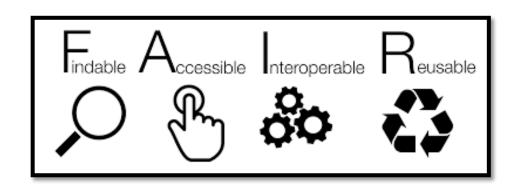


# The Lille training program

• Introduction (3h) • Data literacy, RDM etc. Data Managament Plan (3h) • Why? Which content? How to write a DMP? Follow-up? Research data life cycle (3h) • RDM as part of the research process • Data description (3h) Metadata • Data structuration (3h) • Formats, XML etc. Data preservation and publishing (3h) • Backups, storage, long term preservation, data papers, repositories • Feedback and evaluation (2h)

# Key elements of the Lille model

- Multidisciplinarity (but limited to SSH)
- 20 hours, six months
- Different levels of PhD projects
- Mix of (some) theory and (much) practice, with homework
- Interactive teaching and individual follow-up
- PhD DMP as the red line of the training program
  - With DMP platform (in the academic cloud)
  - H2020 FAIR template
- Evaluation of participation and DMP
- Mixed team (scientists, librarian)





### Which kind of DMP?

- H2020 FAIR DMP template
- Detailed DMP templates are better than simple DMP
- Problems are independent from motivation and attitudes towards data sharing
- Main problems are with
  - Standards (generic or disciplinary metadata, identifiers...)
  - Storage and preservation (choice of platform)
  - Estimation of required resources and related costs

## Students' feedback

#### Not good

- « Too much theory, not enough practice »
- Too heterogeneous
- Too long
- Too much focus on survey data
- Too much ethics and privacy

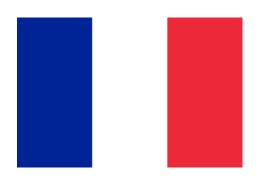




#### Good

- Very useful for PhD
- Clear and precise information
- Good discussion of intellectual property issues
- Customized approach
- Sharing of experience
- « Good atmosphere »

# Comparison with other French initiatives



What is specific to the Lille program:

Scientific governance and follow-up

Partnership with graduate school

Focus on PhD students as target group

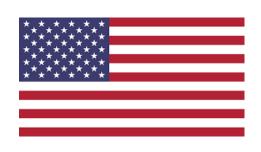
Schedule (nb of hours and sessions)

**Evaluation** 

# Comparison with initiatives from German, Dutch, Finnish, UK and US universities



What is specific to the Lille program:



Scientific governance and follow-up

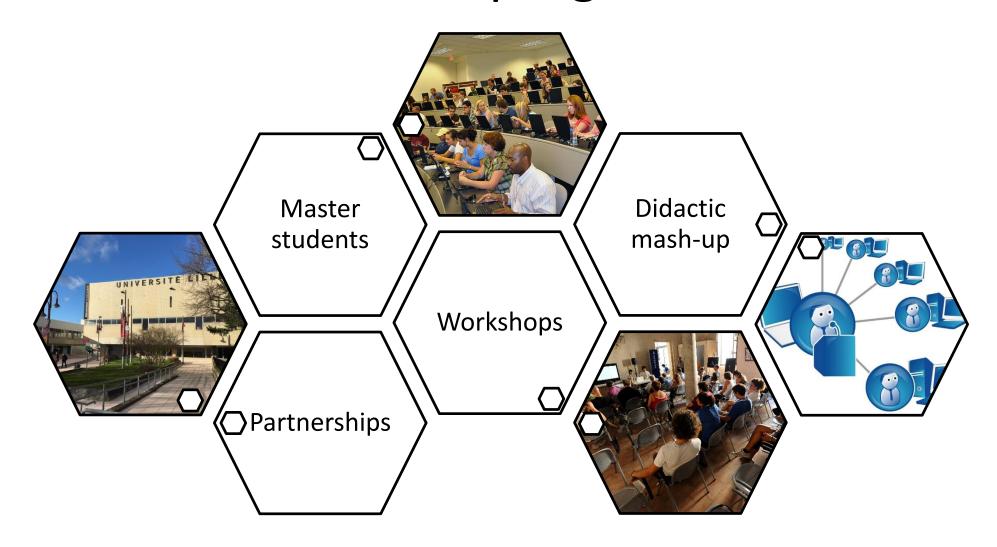
Partnership with graduate school

Completion of DMP with students' data

Schedule (overall length, nb of sessions)

**Evaluation** 

# Good ideas from other programs



## Lille model v2.0

- Including early career scientists (postdoc) and/or Master students?
- Modular workshops for specific needs, domains, laboratories?
- Tailored (customized) training programs?
- Written guidelines (tutorial) on metadata etc.?
- More case studies?
- Long-term follow-up, including RDM and deposit of datasets?

## Future research

- Follow-up
  - monitoring and assessment of the training program
  - evaluation of uptake and usage of the new data workflow
- Applied research (xDiss proposal)
  - The impact of open science on the writing and publishing of ETDs
  - TDM of ETDs and related data



## Thank you!

Corresponding author:

joachim.schopfel@univ-lille.fr



One part of the research was funded by the European Institute for Social Sciences and Humanities at Lille (MESHS) and by the regional authority Conseil Régional Hauts-de-France.

