MAKING DATA IN PHD DISSERTATIONS REUSABLE FOR RESEARCH

Joachim Schöpfel
Hélène Prost
Cécile Malleret

8th Conference on Grey Literature and Repositories
Prague, October 21st , 2015

Research data and PhD theses

Context

Massive production

Heterogeneous formats and content

Big infrastructures

Big data

Small data

Individual projects

Open access

Public data

Dark data

Restricted open access
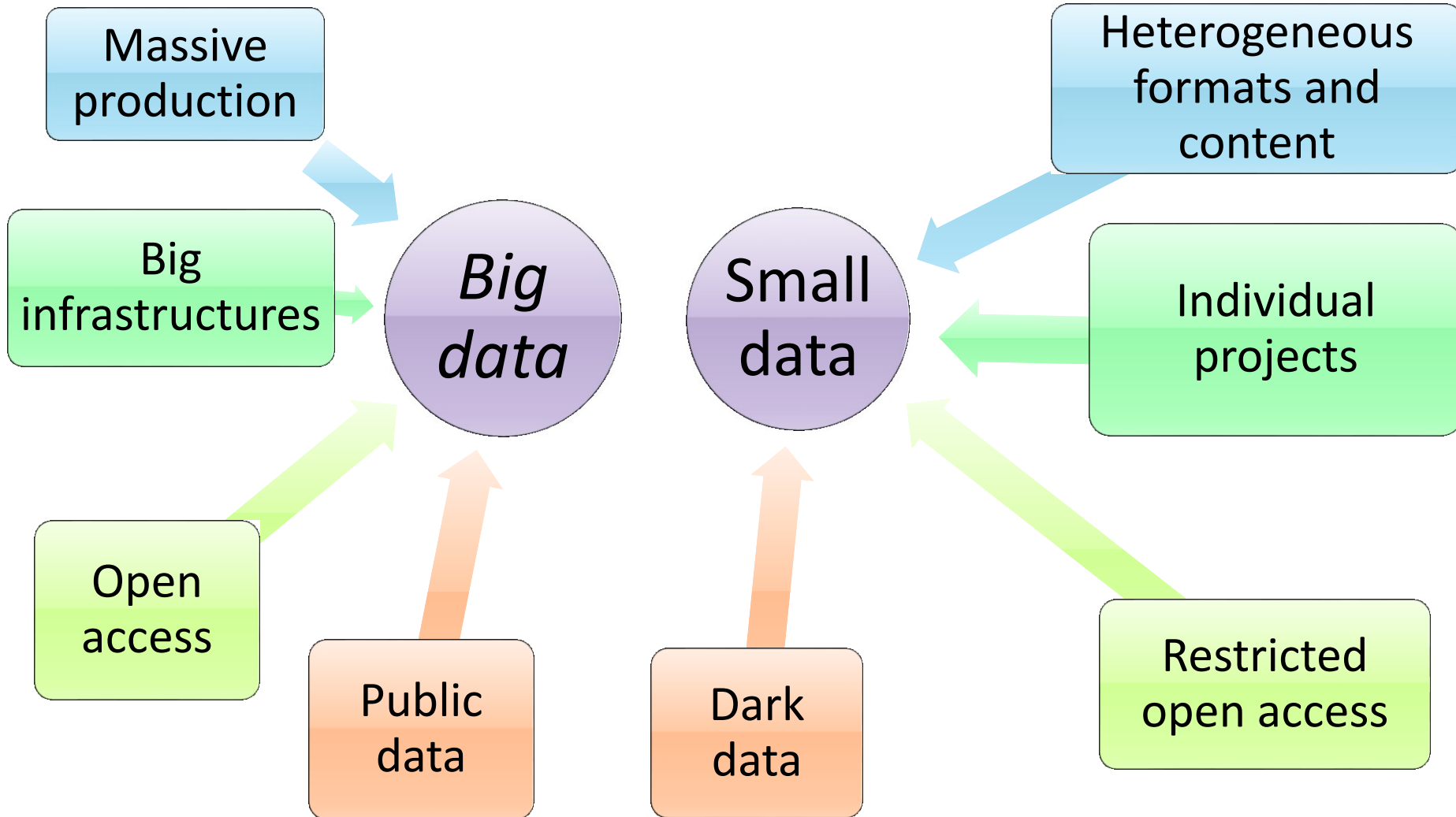
How to valorise listed data in the appendices ?
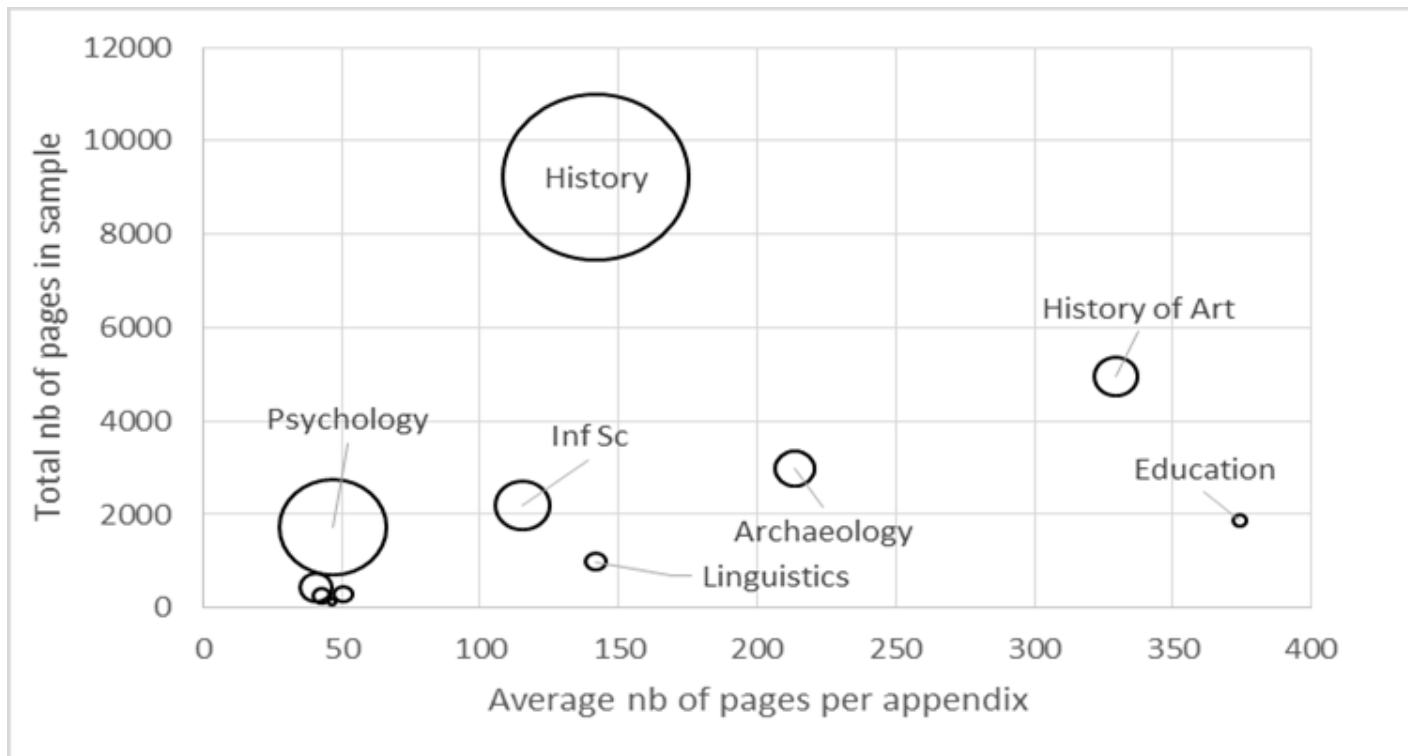
Data in PhD theses at Lille university

Volumetry

Corpus 283 theses

- 31% print
- 69% digital
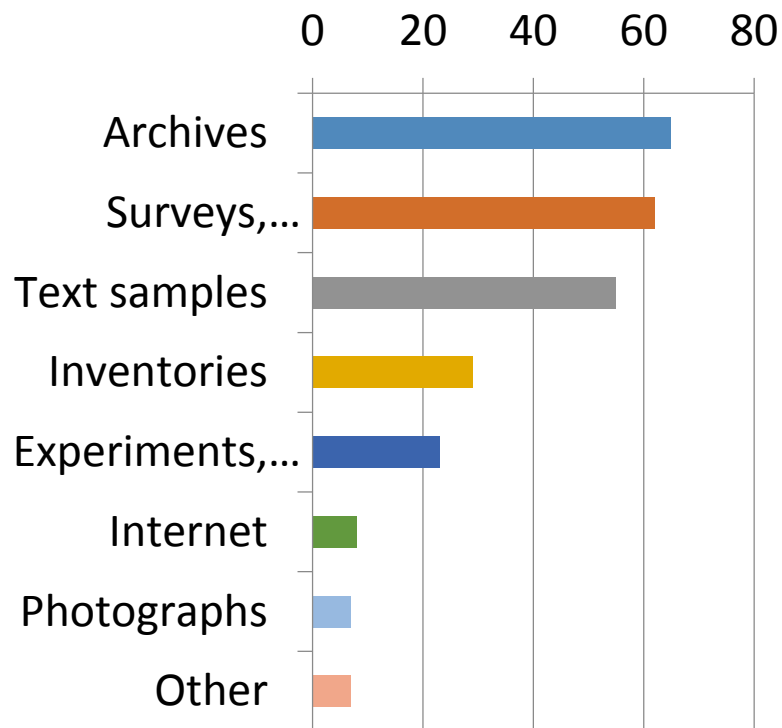- 66% with data in appendices

Appendices

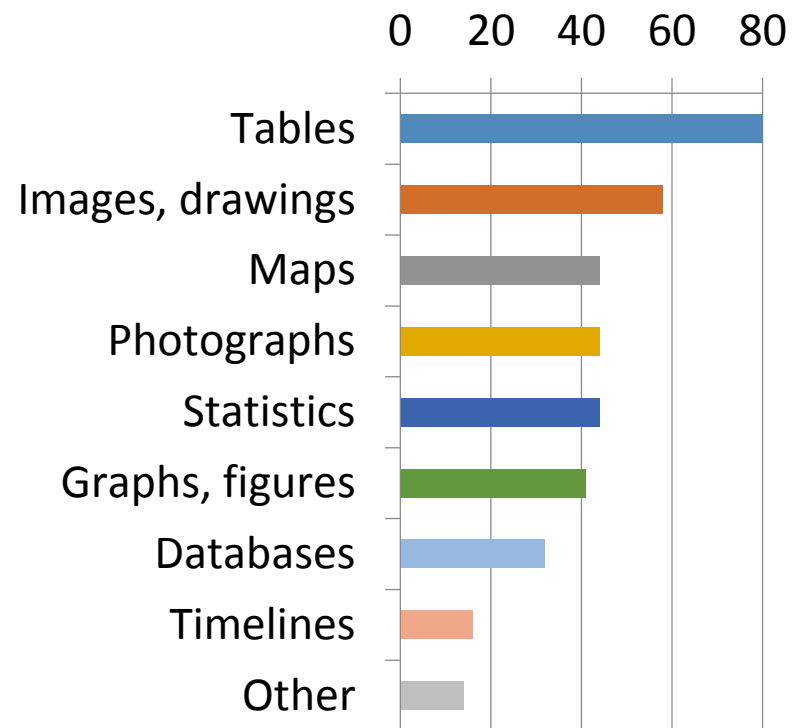- 25 000 pages
- from 5 to 829 pages
- a median of 81 pages

Diversity of research data in PhD

Sources of research data

Typology of research data

Number of relevant theses

The small data

Exemples in PhD theses

Databases

Experimental results

Surveys

Codes

Maps

Archives

Photos videos sounds

Graphs

Corpus exploration

Structure & presentation (1)

- Exemple : « The œiconography of Saint Anne in Spain at the end of the Middle Ages » from Alfonso de Salas (2010)
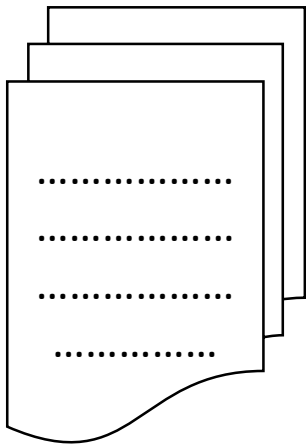
A main volume with text and classification plan
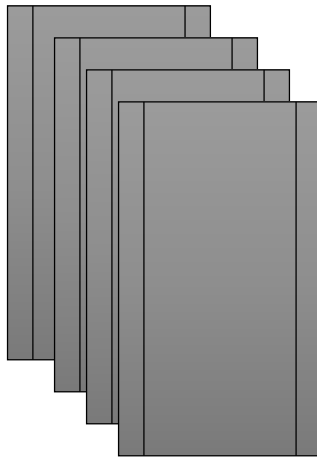
A volume of appendices including catalog with 179 notices

A volume of illustrations with 1581 images

Corpus exploration

- Exemple : « Coping strategies by athletes » from Annie Carton-Caron (2004)

Tables

Figures

A main volume with text

A volume with appendices with follows the plan of text

Numbered data non listed in appendices
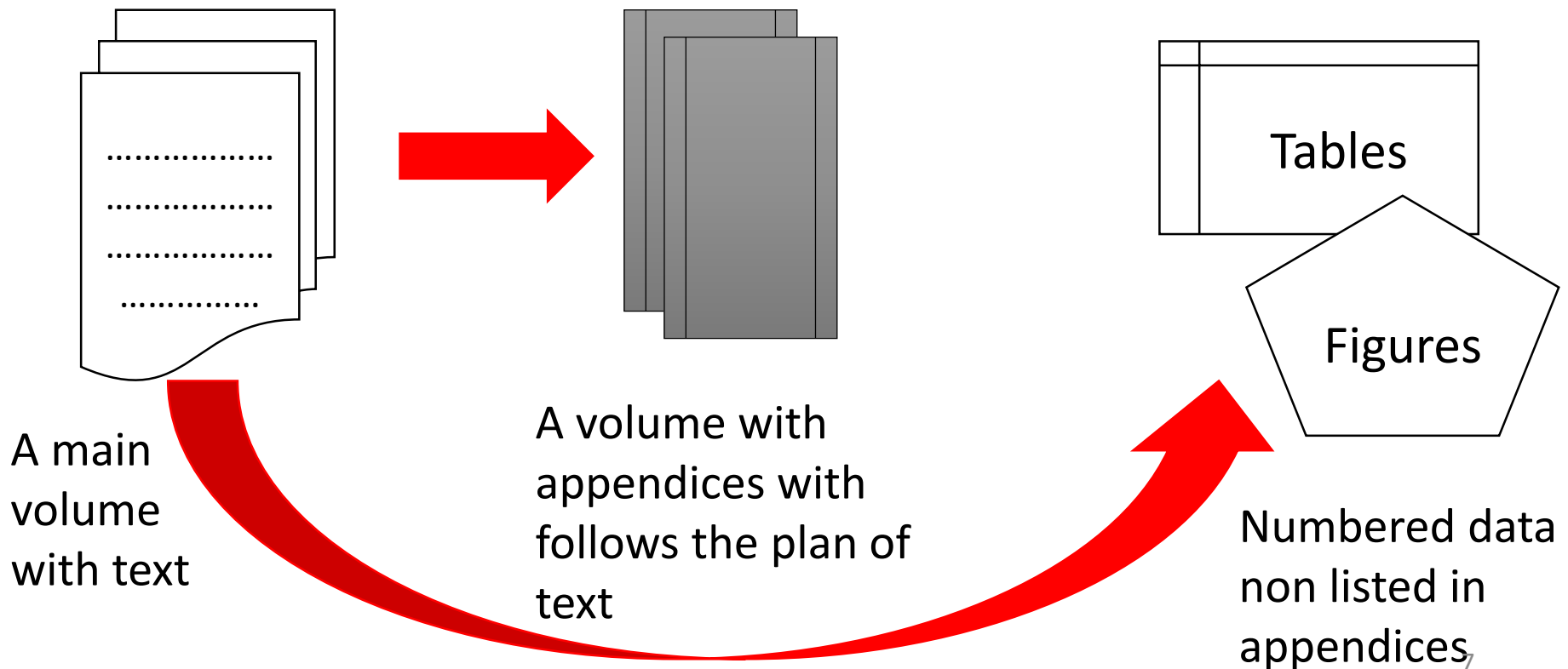
Barriers to open data
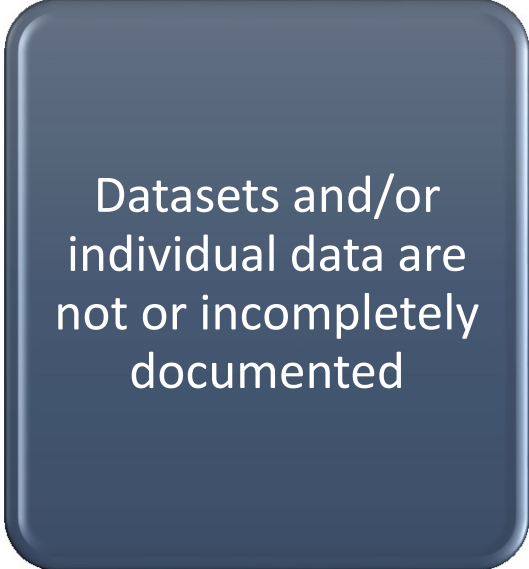
Incomplete, inadequate or missing description
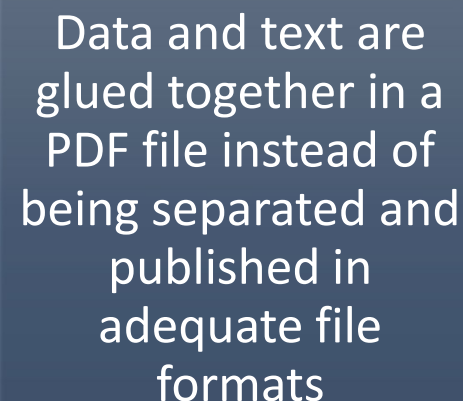
Datasets and/or individual data are not or incompletely documented

Missing organisation

Research data are presented without any structuration or organisation, often together with other, not reusable material in a kind of information mash-up not suitable for further research

Inadequate format

Data and text are glued together in a PDF file instead of being separated and published in adequate file formats

Potential use & valorisation of data

- Databases

- Lexical analysis, data mining

- Prosopographies

Images

Texts

Historic data
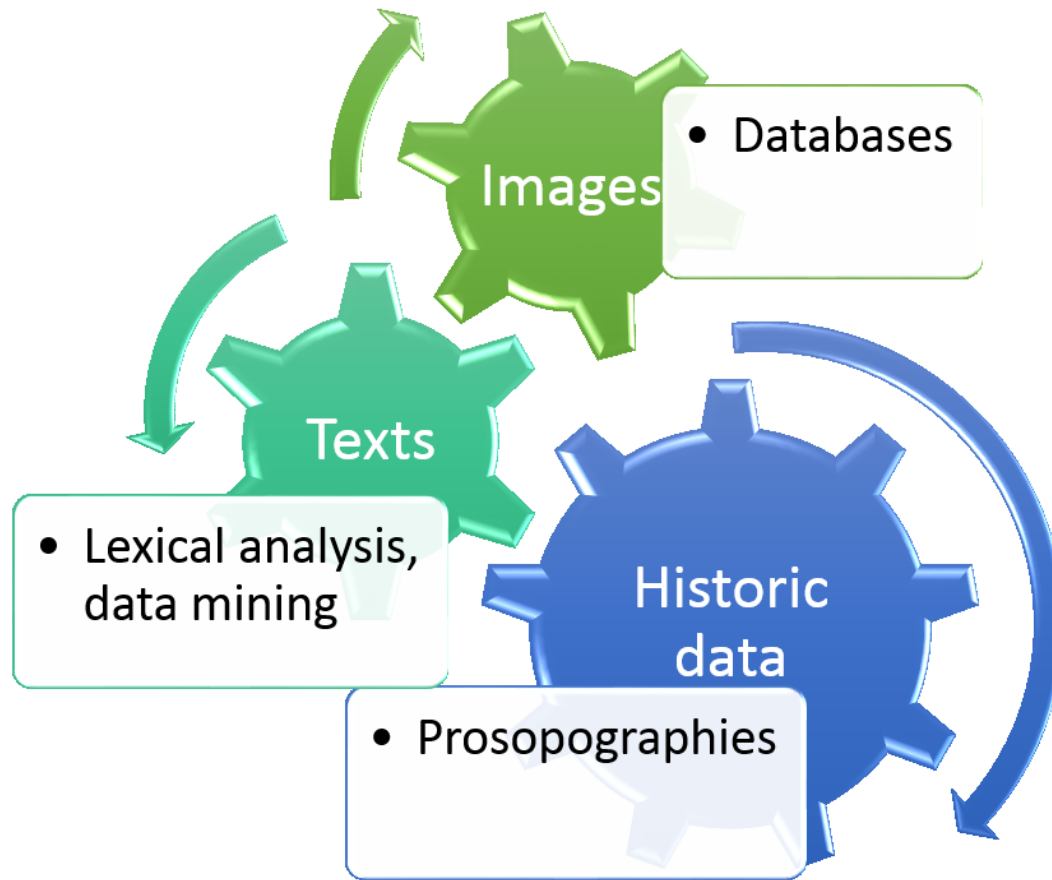
Data management : practices & needs

Survey at Lille 3

9/10 store their data on local

83% on private computer
49% on professional computer

97% declare themselves responsible for data backup

Expressed needs

Need	Percentage
An area of storage and archiving	54%
General advice	52%
Technical advice	49%
Legal advice	48%
Advice- citations publications	43%
Technical assistance	40%
Assistance in preparing DMP	29%
Advice - ethics	24%
Others - no need	10%

Data management : practices & needs

Preference for local or institutional repository

63% are motivated to submit data

Interested in the ethical and legal issues

The less experimented in the field of research data

Response rate **13**%

Seek advice for the publication of data
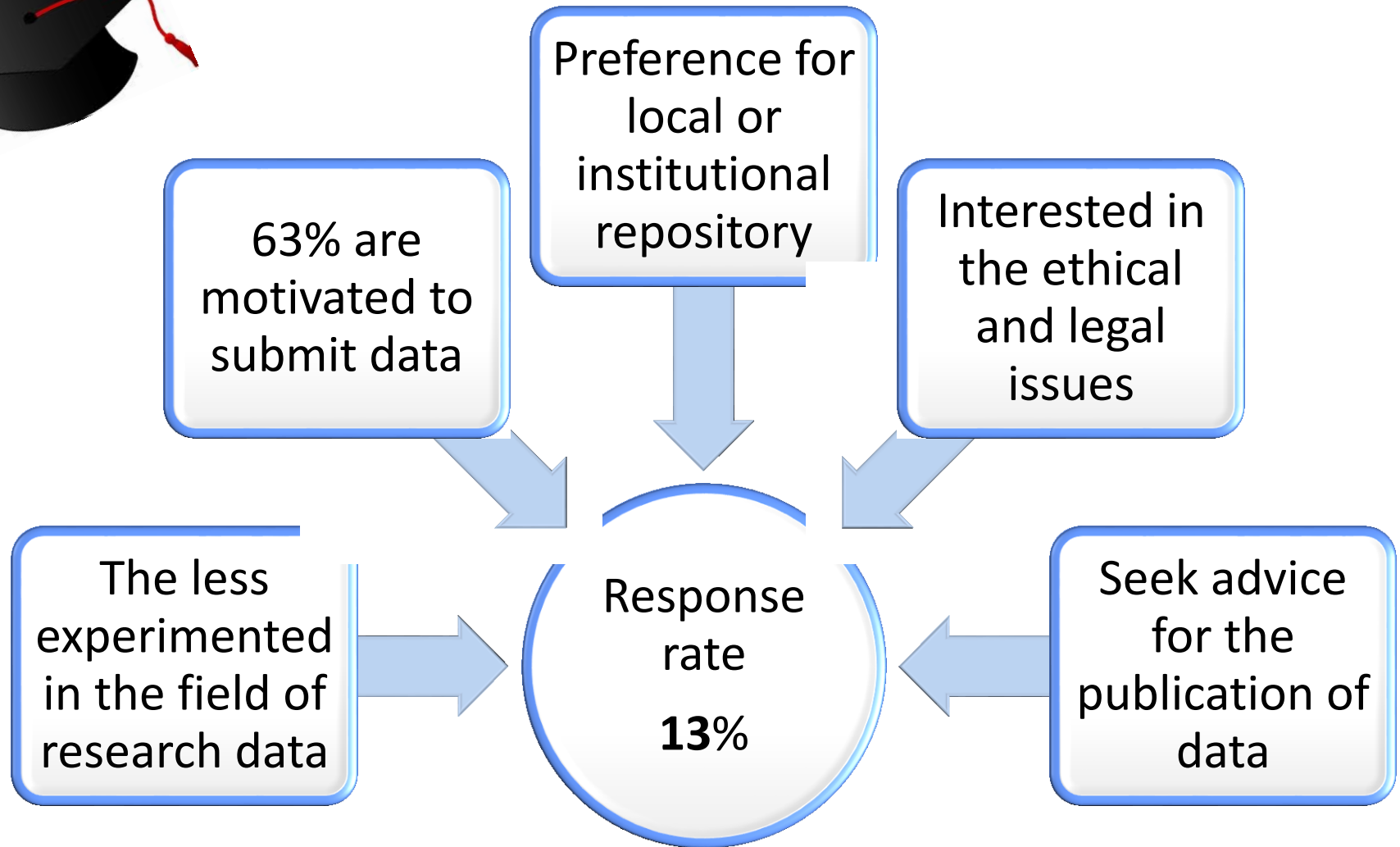
Recommandations

Clear separation of text and data. Digital research data must be submitted in different and separate files

Separation

Structuration of the research data, with a detailed and organized tagging (markup) of the datasets

Structuration

The data must be described in a standard language and format, with sufficient detail for retrieval and data mining

Metadata

Deposit in open and original format to facilitate long-term preservation and reuse

Format

Clearing of privacy and copyright issues

Copyright

Research data support at university

Services

Education	Advice, assistance	Infrastructures
• Seminars, conferences, training • Online resources (guidelines, FAQ…) • Alert service	• Legal and technical help (data management plan) • Technical assistance for deposit • Liaison with laboratories	• Mediation for deposit • Partnership with networks and repositories • Development of tools on the campus

Data management at university

Principles

1. A discipline-specific approach
2. An integration into the doctoral education
3. A proposal of data management plans
4. Incentives for the digital deposit of research data
5. A contribution to the preservation and dissemination of data

THANK YOU FOR YOUR ATTENTION

joachim.schopfel@univ-lille3.fr
helene.prost@inist.fr
cecile.malleret@univ-lille3.fr