



národní
úložiště
šedé
literatury

Dealing with Research Data and Dissertations

Schöpfel, Joachim; Prost, Hélène
2017

Dostupný z <http://www.nusl.cz/ntk/nusl-369340>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte původ-Zachovejte licenci 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 20.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .



conference on
grey literature
and
repositories

NTK

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

Dealing with Research Data and Dissertations Workshop

Joachim Schöpfel & Hélène Prost

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-369340>



Time schedule

- 13:00 First part
 - What you should know about data
 - What you should know about data literacy, attitudes and needs
 - What you should know about data related to PhD dissertation
 - What you should know about service development
- 15:00 Pause
- 15:20 Second part
 - Presentation of Lille project
 - Discussion
- 16:50 Short feedback
- 17:00 End

Our objectives

Our intention is that each participant leaves the workshop with a better understanding of

- *A realistic model of RDM with PhD students on the campus*
- *Critical issues (anticipation of problems, risk analysis)*
- *Key success factors (governance, education, cooperation)*

Your expectations?

Preliminary questions

Do you know

- *the international directory of research data repositories [re3data](#) ?*
- *[DMPonline](#) for the creation of data management plans ?*
- *the two data repositories [Zenodo](#) or [figshare](#) ?*
- *the Educopia ETD+ [Toolkit](#) for the management of the students' research output ?*

What you should know...

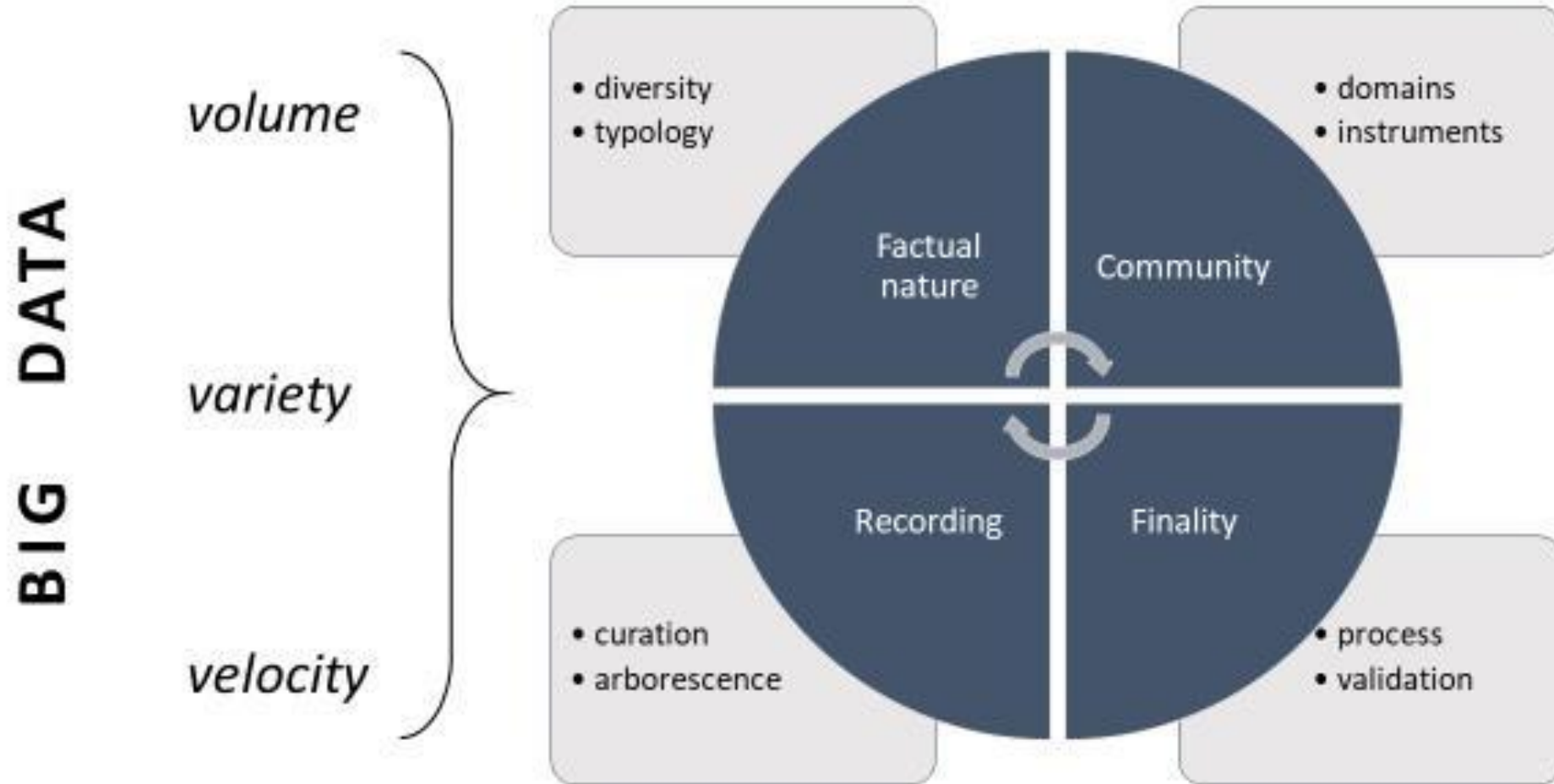
FIRST PART

What you should know about data

A popular but uncertain concept:

- « Big Data is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value » (De Mauro et al. 2016)
- « What constitutes data is determined by a given community of interest that produces the data. However, an investigator may be part of multiple, overlapping communities of interest, each of which may have different notions of what are data » (Koltay 2016)
- « The recorded factual material commonly accepted in the scientific community as necessary to validate research findings » (OMB Circular 110)

A conceptual approach



Typology of data

« Data are most often defined by example, such as facts, numbers, letters and symbols » (Borgman et al. 2015)

- Observation
- Experimentation
- Simulation
- etc.

Often more conditioned by instruments and methods than by disciplines and communities.

Primary and secondary data

- Data as material (resource) for research
- Data as research results
- Long tail of data
- Unequal categories
- Domain-specific profiles

	re3data	Prost & Schöpfel 2015, sources	Prost & Schöpfel 2015, résultats
Scientific and statistical data formats	63%	26%	49%
Standard office documents	59%		
Plain text	49%	64%	76%
Images	49%	25%	21%
Raw data	44%		
Structured graphics	38%		32%
Structured text	32%		
Archived data	23%	34%	
Audiovisual data	18%	6%	44%
Software applications	18%		9%
Databases	17%		37%
Networkbased data	6%		
Source code	4%		
Configuration data	2%		
Enquêtes et entretiens		47%	
Observations		41%	
Expériences		36%	
Cartes et plans			10%
Other	36%	7%	3%
Total	100%	100%	100%

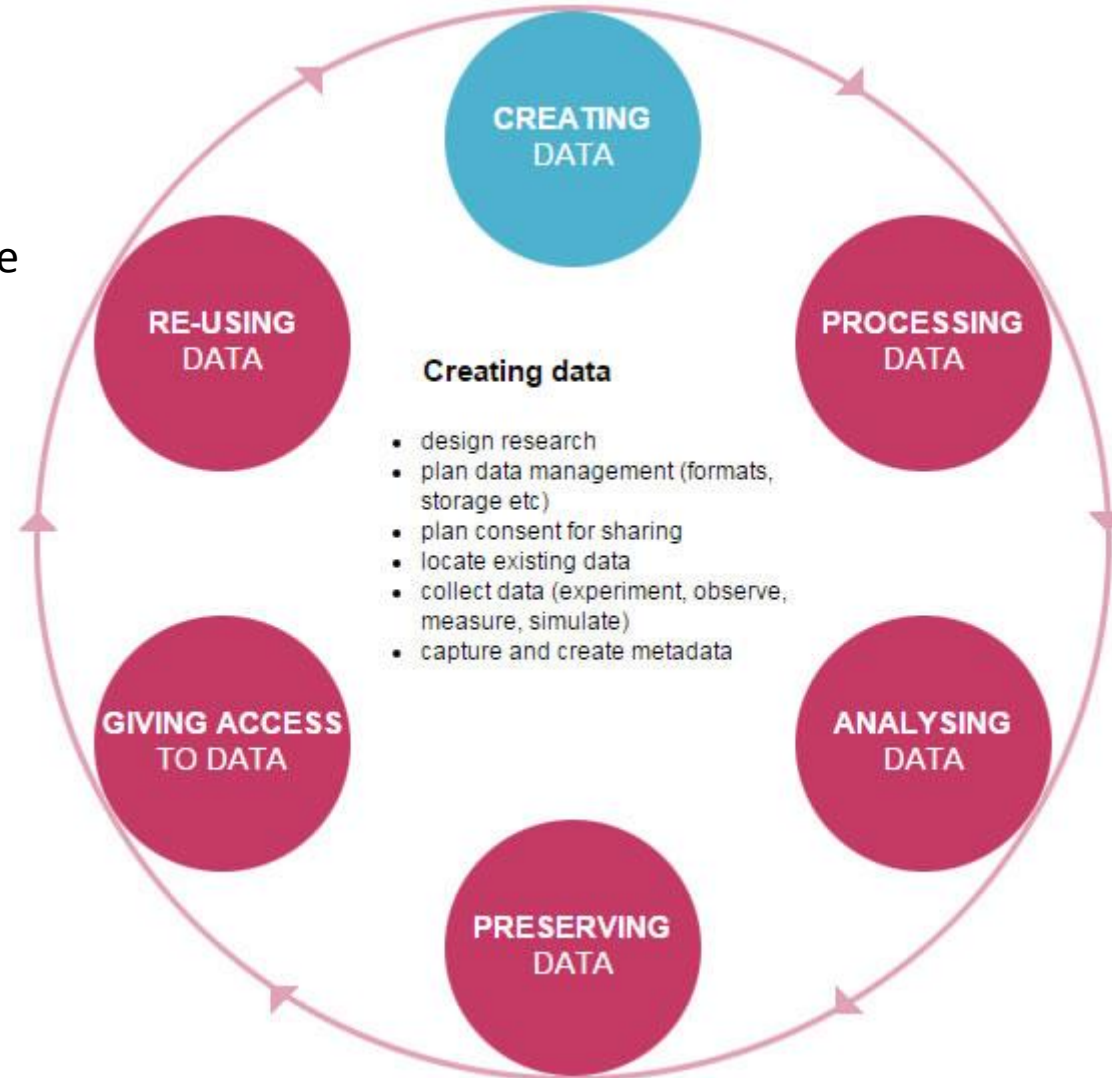
A functional approach

- Politics
 - Increase transparency
 - Increase efficiency of public action
 - Provide fuel for economy
- Economics
 - Optimize (valorize) public research
 - Accelerate innovation (health, environment)
- Science
 - Explore (reuse)
 - Visualize results (also: data journalism)
 - Compare and/or control results
 - Validate hypotheses
 - Also: citizen science



Data and research process

Data are dynamic
They have their own life cycle



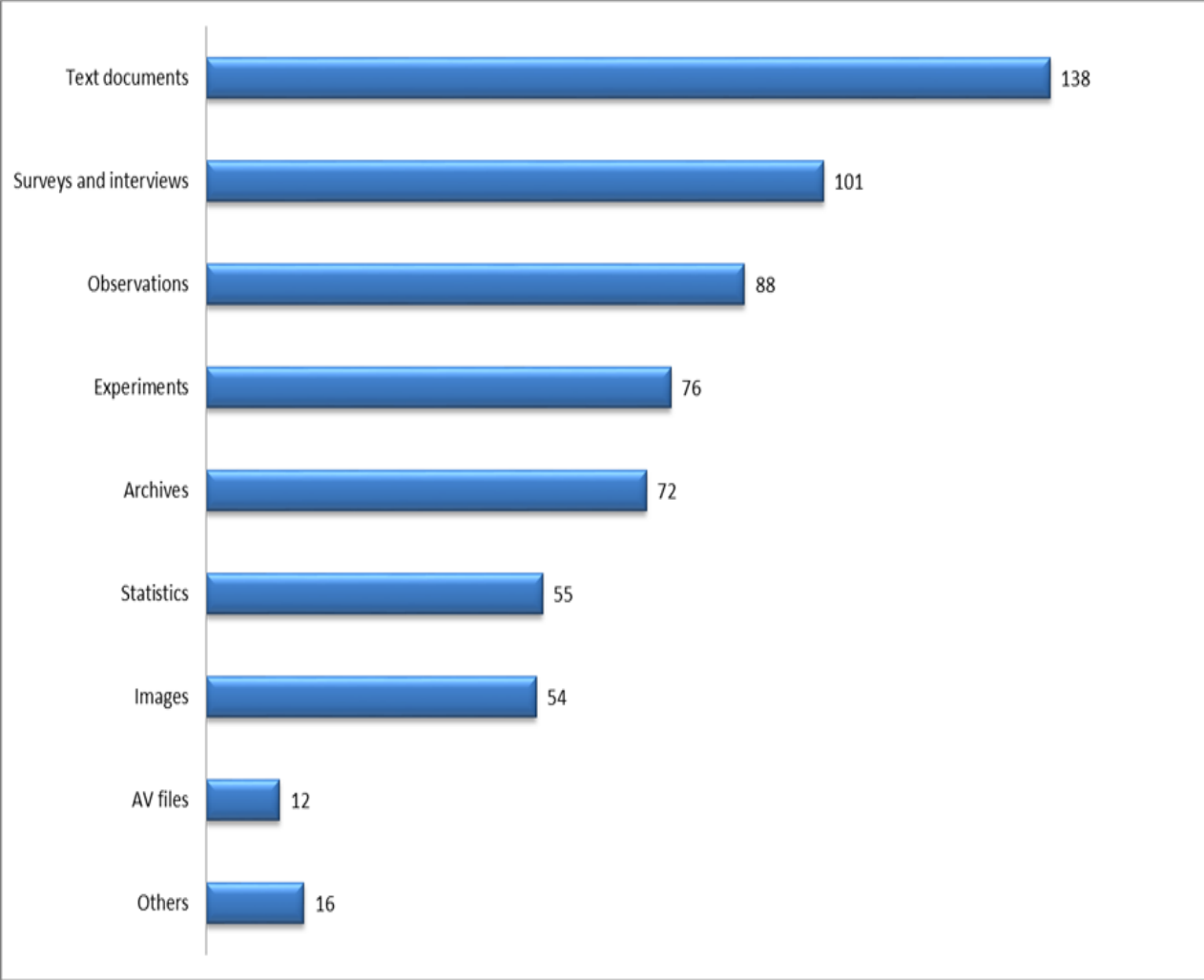
What you should know about data literacy, attitudes and needs

- An increasing number of surveys
- Institutional and disciplinary differences
- However, some common characteristics

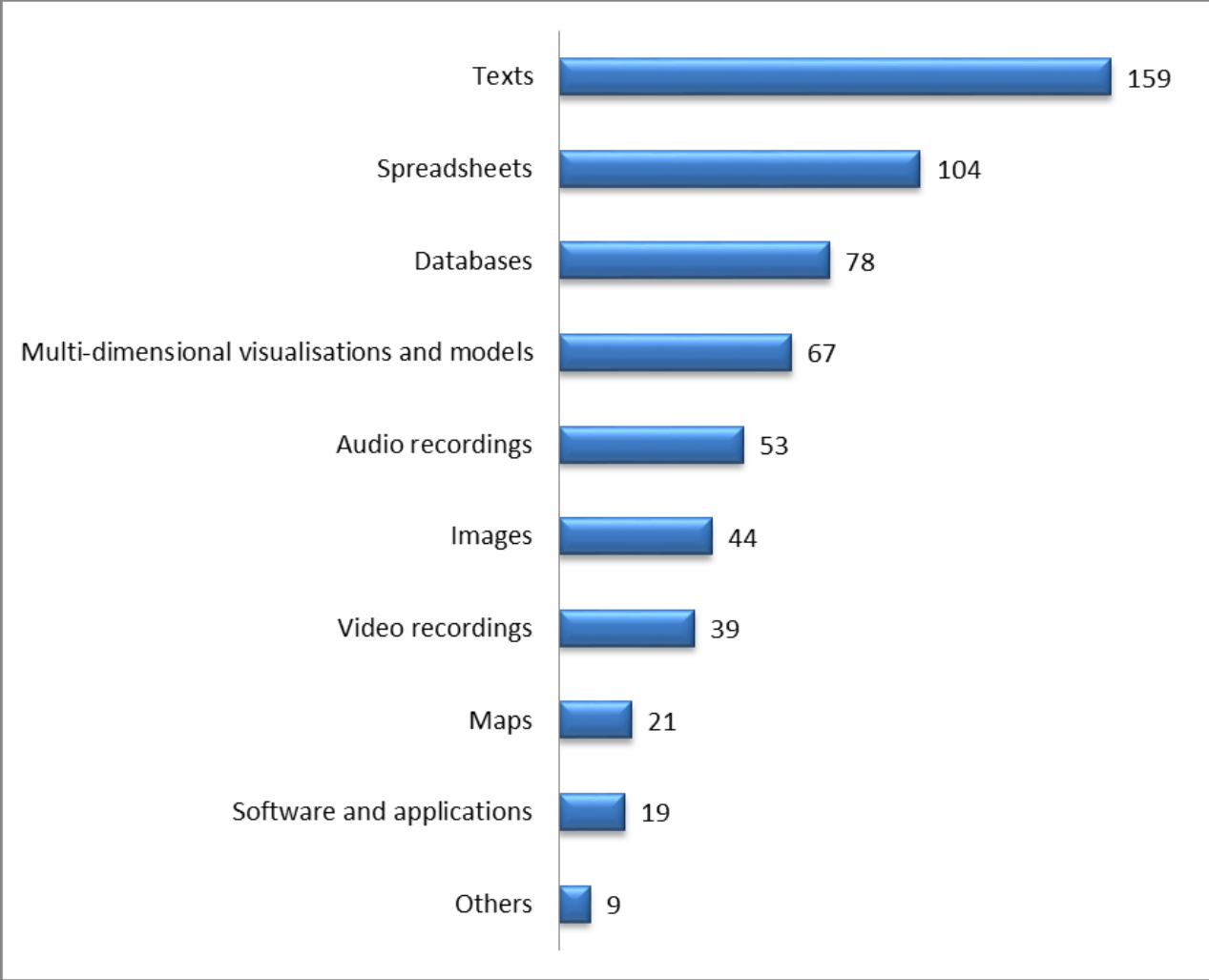
- Schöpfel, J., Prost, H., 2016. Research data management in social sciences and humanities: A survey at the university of Lille 3 (France). *LIBREAS. Library Ideas* 29, 98-112.
<http://hal.univ-lille3.fr/hal-01395816>

Data literacy

Diversity of data



Research data sources (n=214)



Research data results (n=211)

Data literacy

Storage and sharing

9/10

store their data on local



83% on private computer
49% on professional computer

97%

declare themselves responsible for data backup

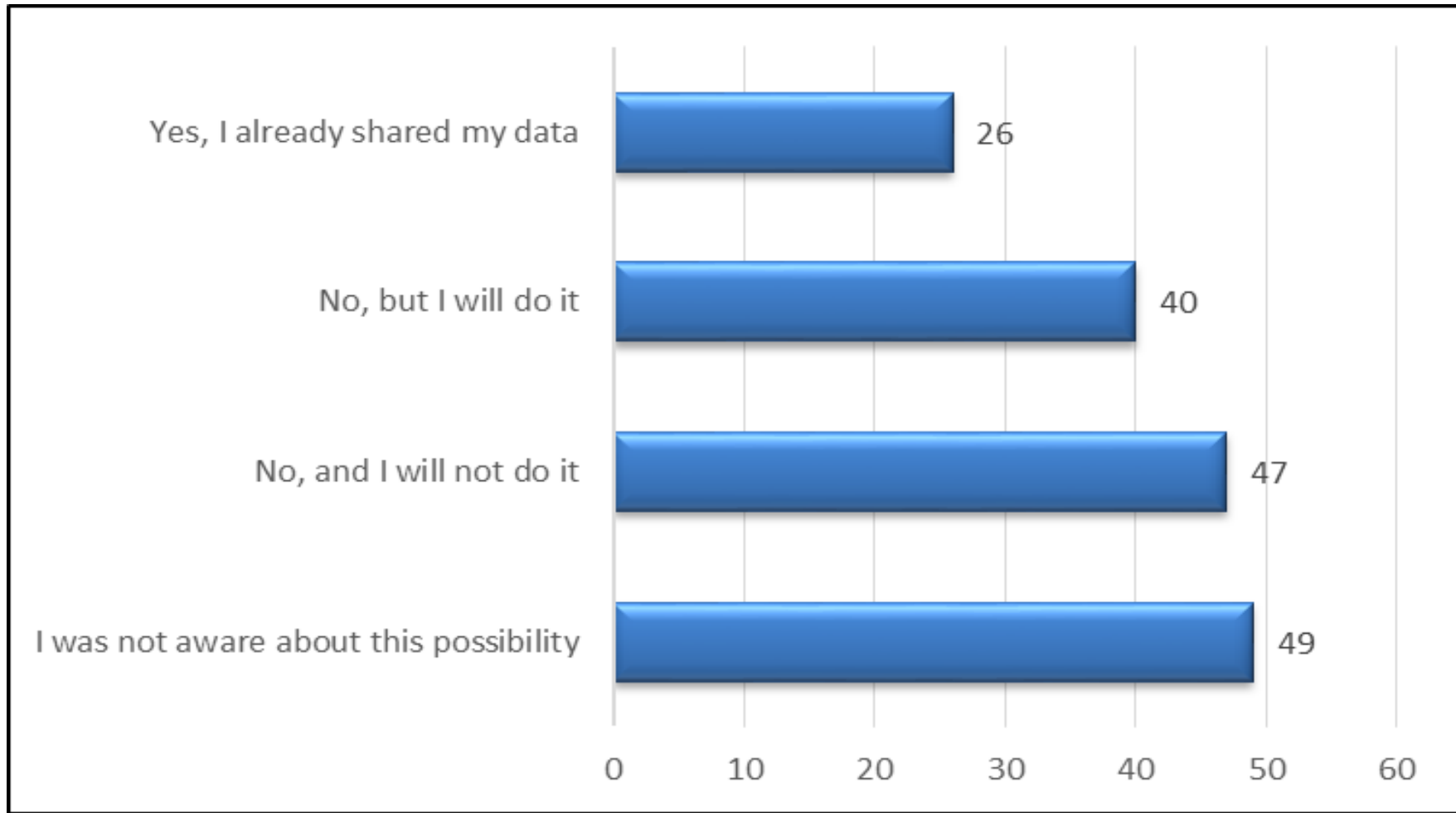
64%

do not share their research data with colleagues or other people
Nobody else has access to their data

Data sharing limited

Attitudes toward sharing

Experience, motivation

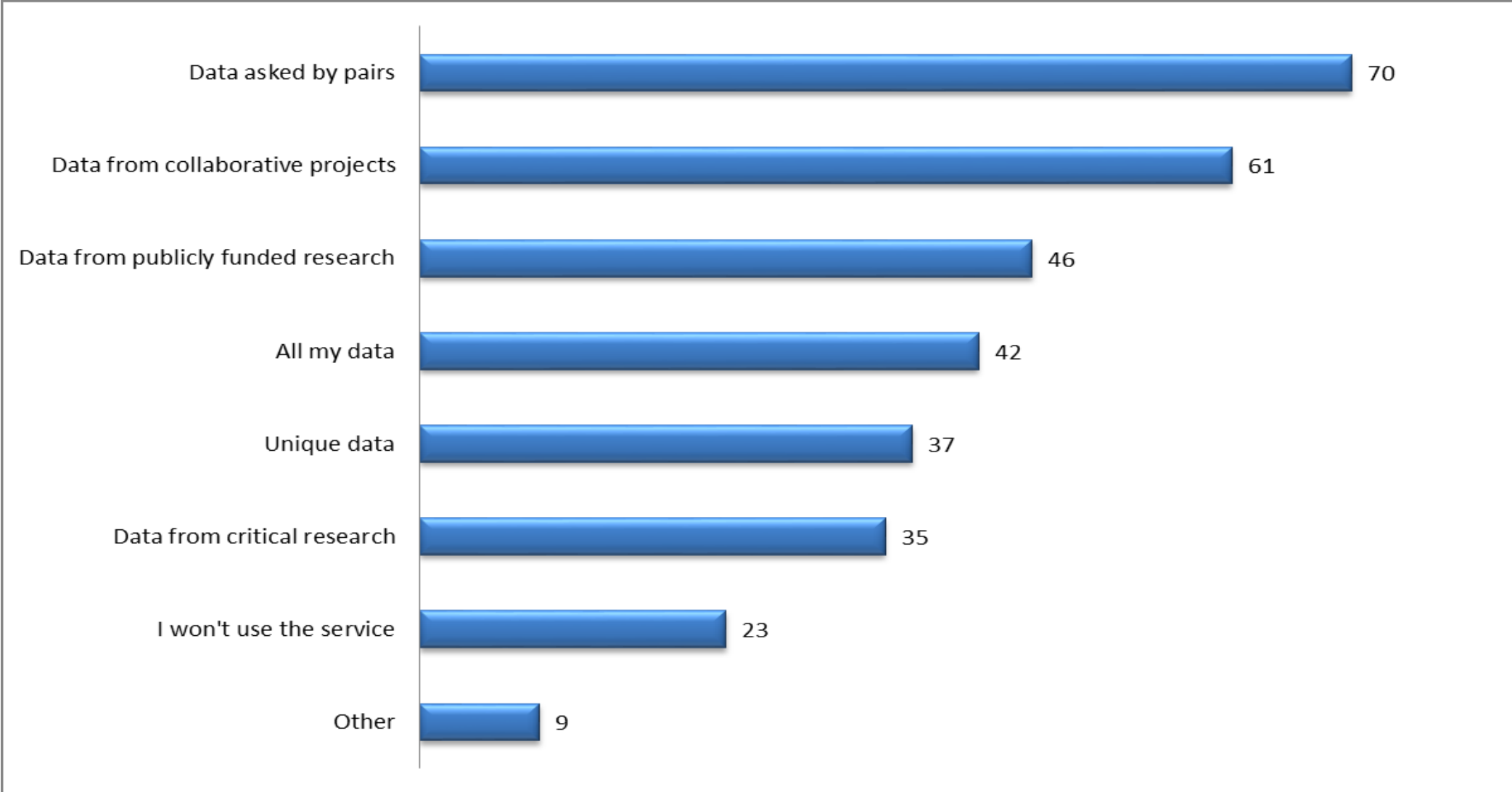


Deposit of research data in a data repository (n=162)

Workshop Data and Dissertations - October 18, 2017 - NTK

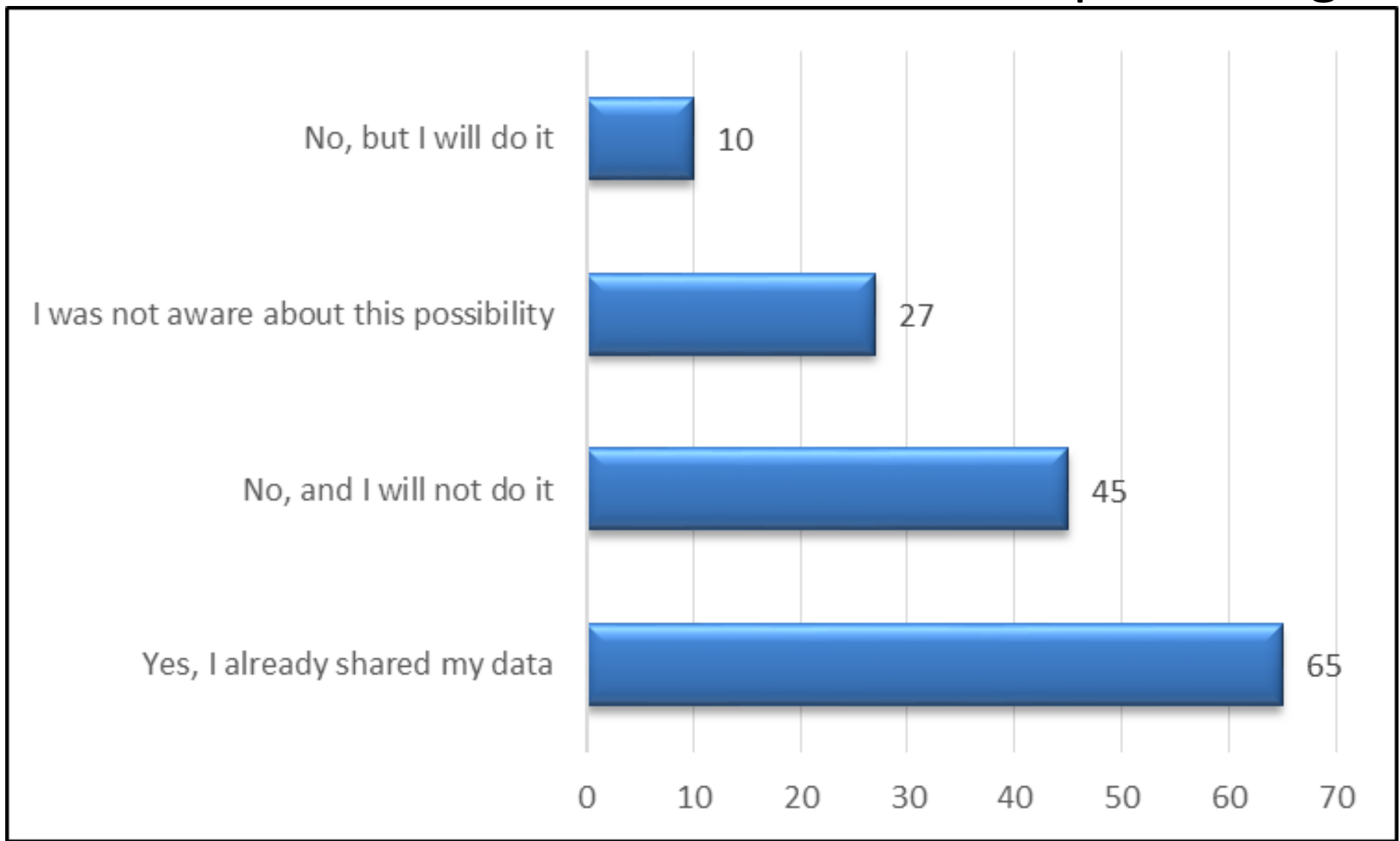
Attitudes toward sharing

Which kind of data?



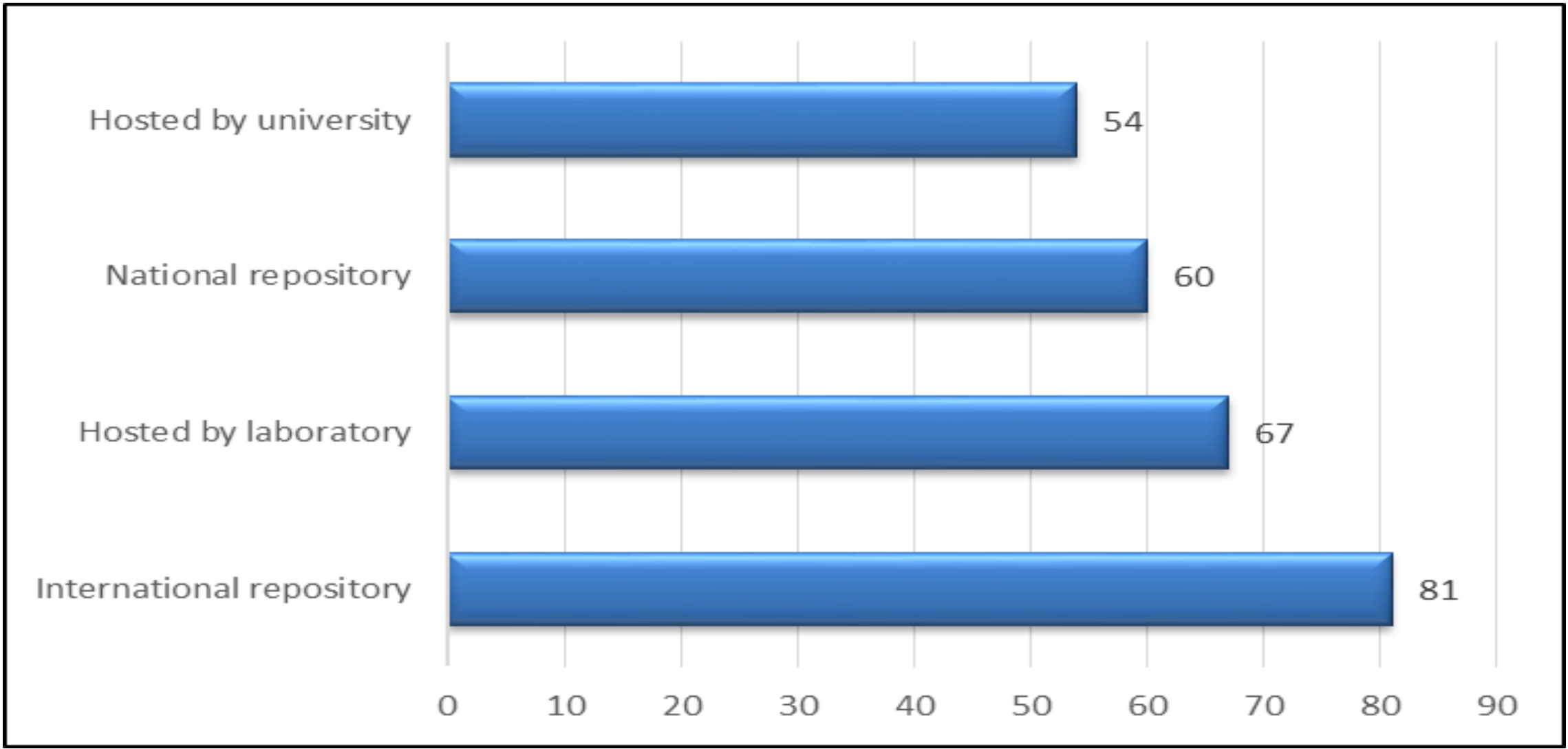
Attitudes toward sharing

→ Data publishing



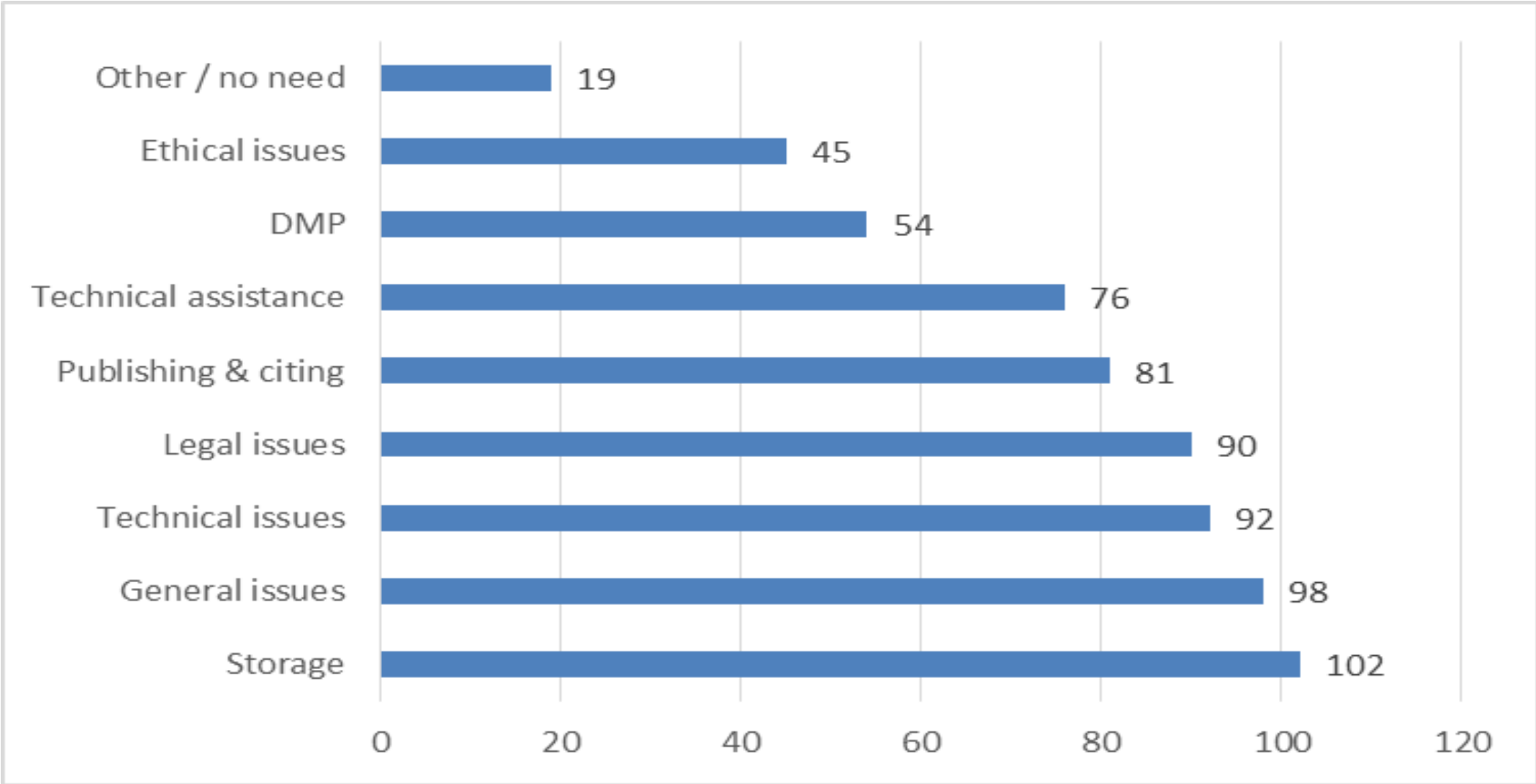
Attitudes toward sharing

Preferred data repository



RDM related needs

Above all, storage



Support and services needed (n=188)

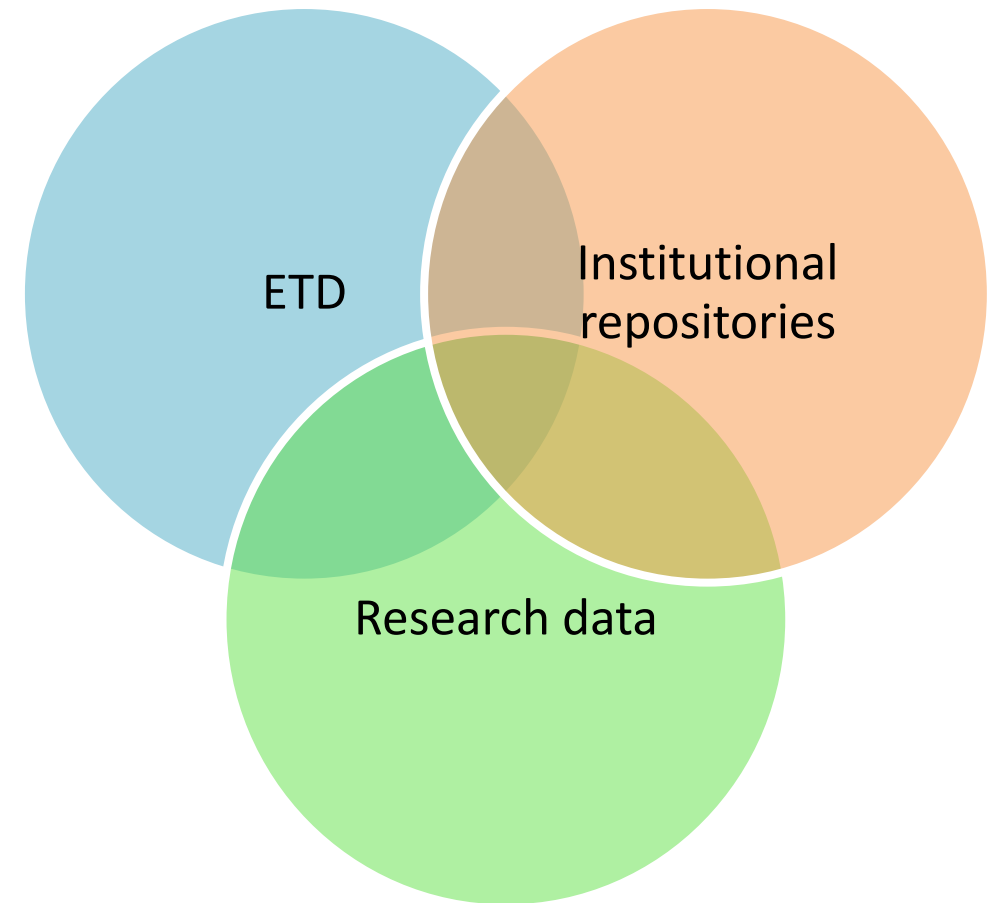
Workshop Data and Dissertations - October 18, 2017 - NTK

What you should know about data related to PhD dissertation

Originality

Linked to a scientific program

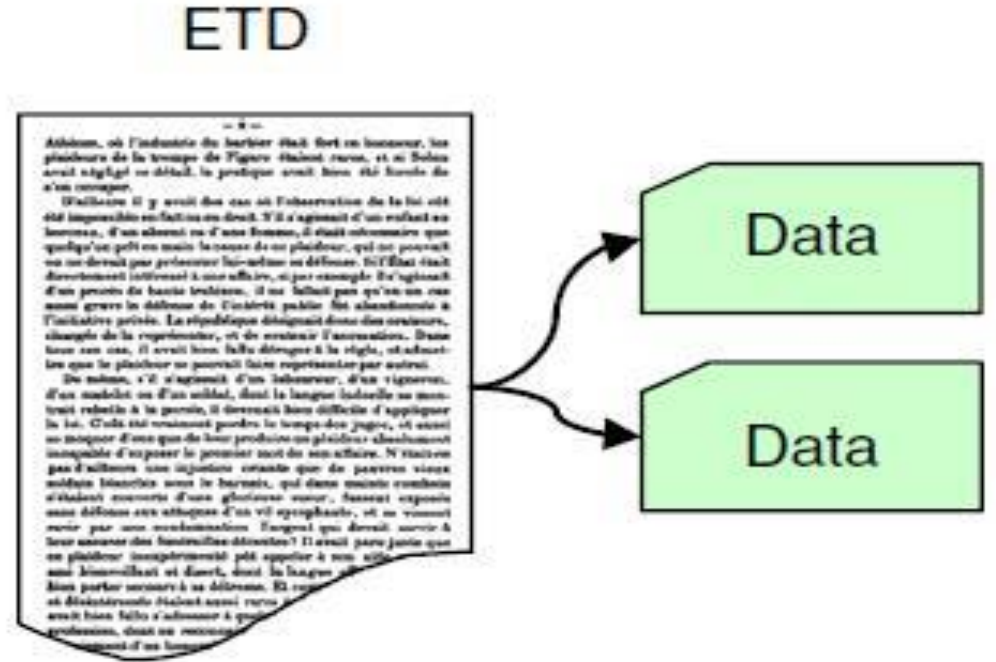
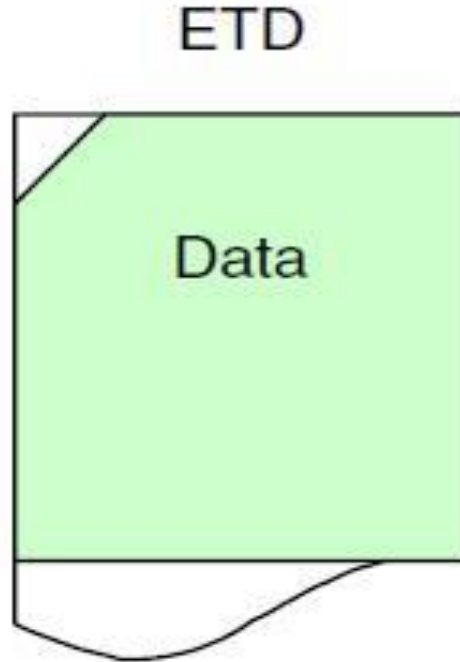
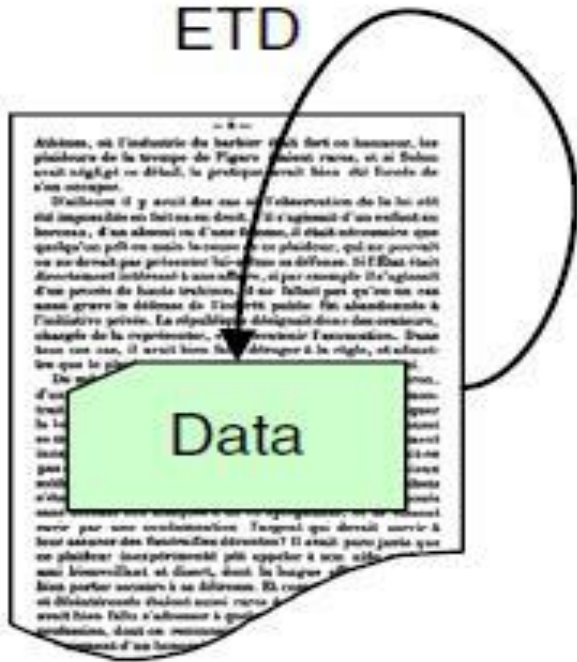
No commercial and public character



The potential of ETDs

- Contain the results of at least three years of scientific work
- Variety and richness of appendices
- Availability in open access
- Contribute to eScience

ETDs and data

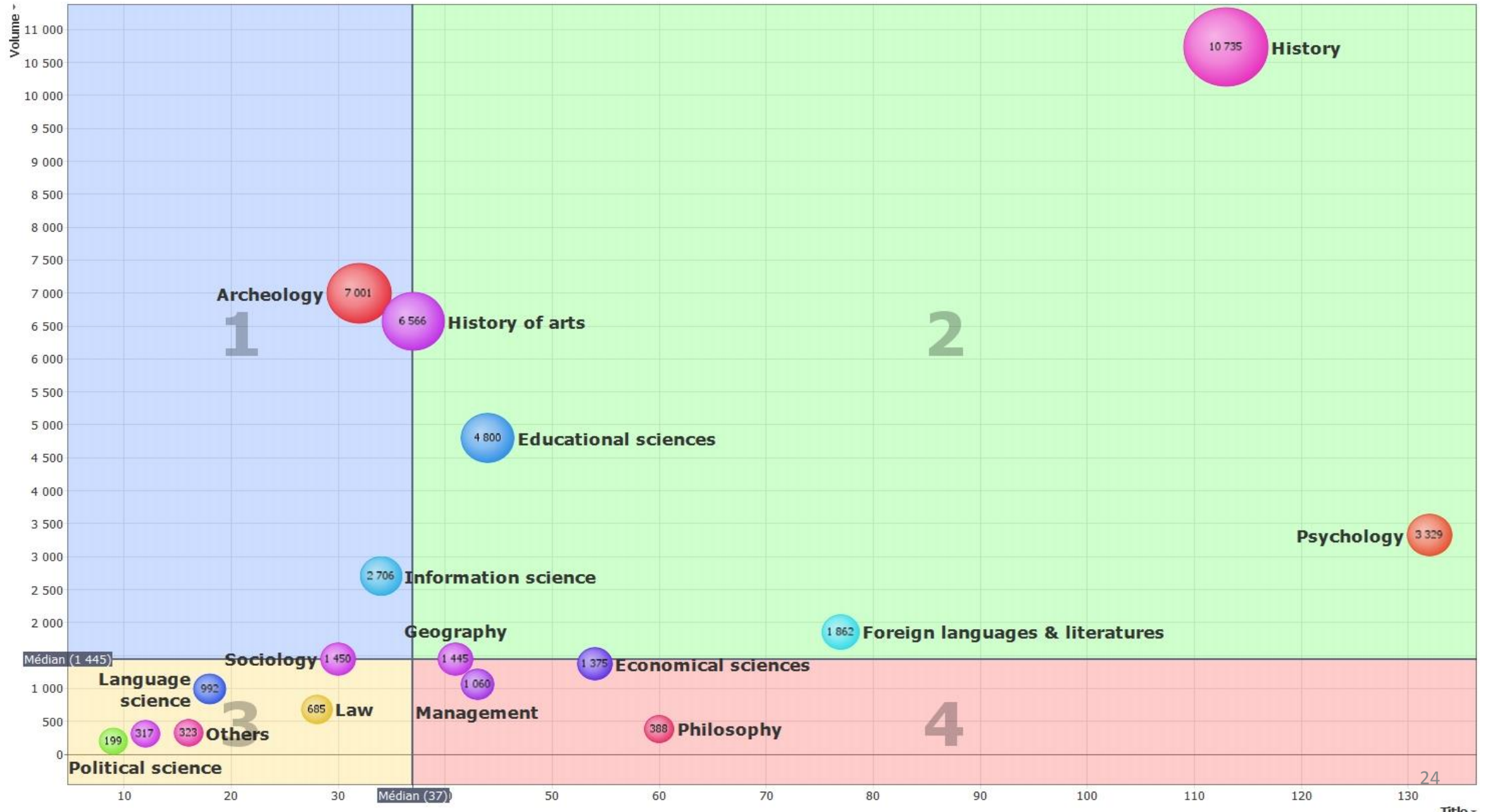


ETD as data vehicle

ETD as data

ETD as gateway to data

The size of data appendices



Data type and discipline

Y: Domain	Total										Tous
	Databases	Graphs - figures	Images - drawings	Maps	Others	Photographs	Statistics	Tables	Texts	ines	
Archeology	4	2	22	18		11	1	16	15	1	30
Economical sciences		16	1	5			2	31	36		43
Educational sciences		8	14	1			5	25	29	1	38
Foreign languages & literatures	1	1	20		1	1	6	21	36	1	46
French language & literature		1					1		5	1	6
Geography		13	7	13		5	3	27	23		33
History	16	22	39	27		26	14	44	65	12	88
History of arts	6		17	8	1	8		4	20	1	28
Information science	2	7	7	3	4	2	5	12	20	1	28
Language science	1	1	1				1	1	7		7
Law		1	3	2				4	5		7
Management	2	12	10	1		1	7	26	22	2	30
Others	1	2						2	4	1	6
Philosophy		2	2		1	1		1	11		11
Political science	1	1	4				1	6	2		6
Psychology	2	15	20	1		4	55	65	48		91
Sociology	2	7	8	4		6		21	28	2	28
Tous	Total	38	111	175	83	65	101	306	376	23	526

What you should know about service development

- Five basic questions of strategic service marketing
 - *What is our business?*
 - *Who are our customers?*
 - *What is our value for them?*
 - *Where is the business going?*
 - *Where should we go?*
- Collective choice
- Compliance with institutional policy

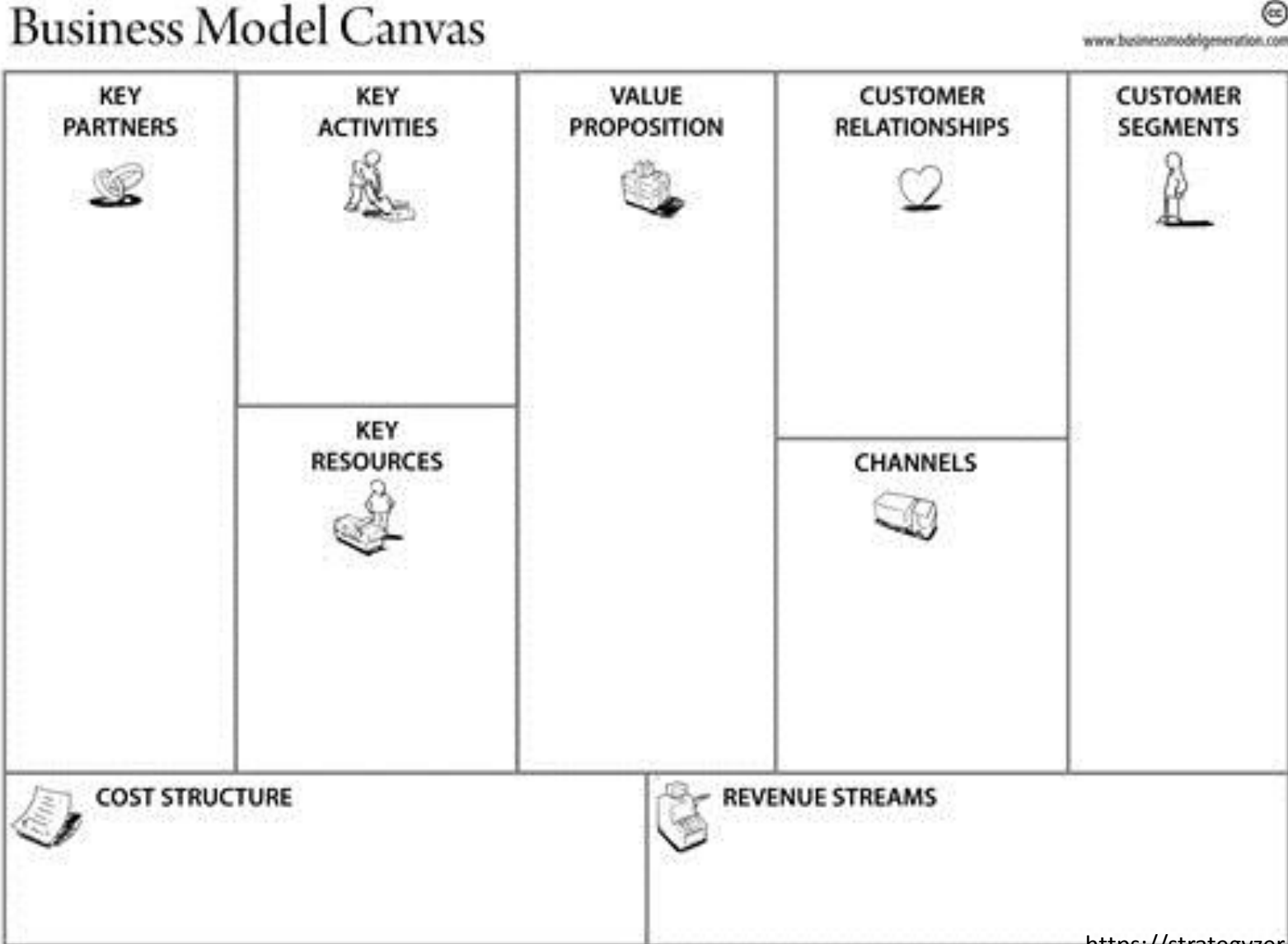
Strategical analysis with SWOT

Internal vs external factors

Helpful vs harmful factors



Preparing change with the *Strategyzer* Canvas



Taking Murphy's law seriously:

"Anything that can go wrong will go wrong"

- Risk analysis: what can go wrong, and what should be done?
 - Nature of risk
 - Probability
 - Impact (severity)
 - Overall assessment (probability * impact)
 - Prevention (what can be done to avoid risk)
 - Action (what can be done if it goes wrong)

		Impact		
		Low	Medium	High
Probability	High	low	medium	high
	Medium	low	medium	medium
	Low	low	low	low

Questions?

20 minutes

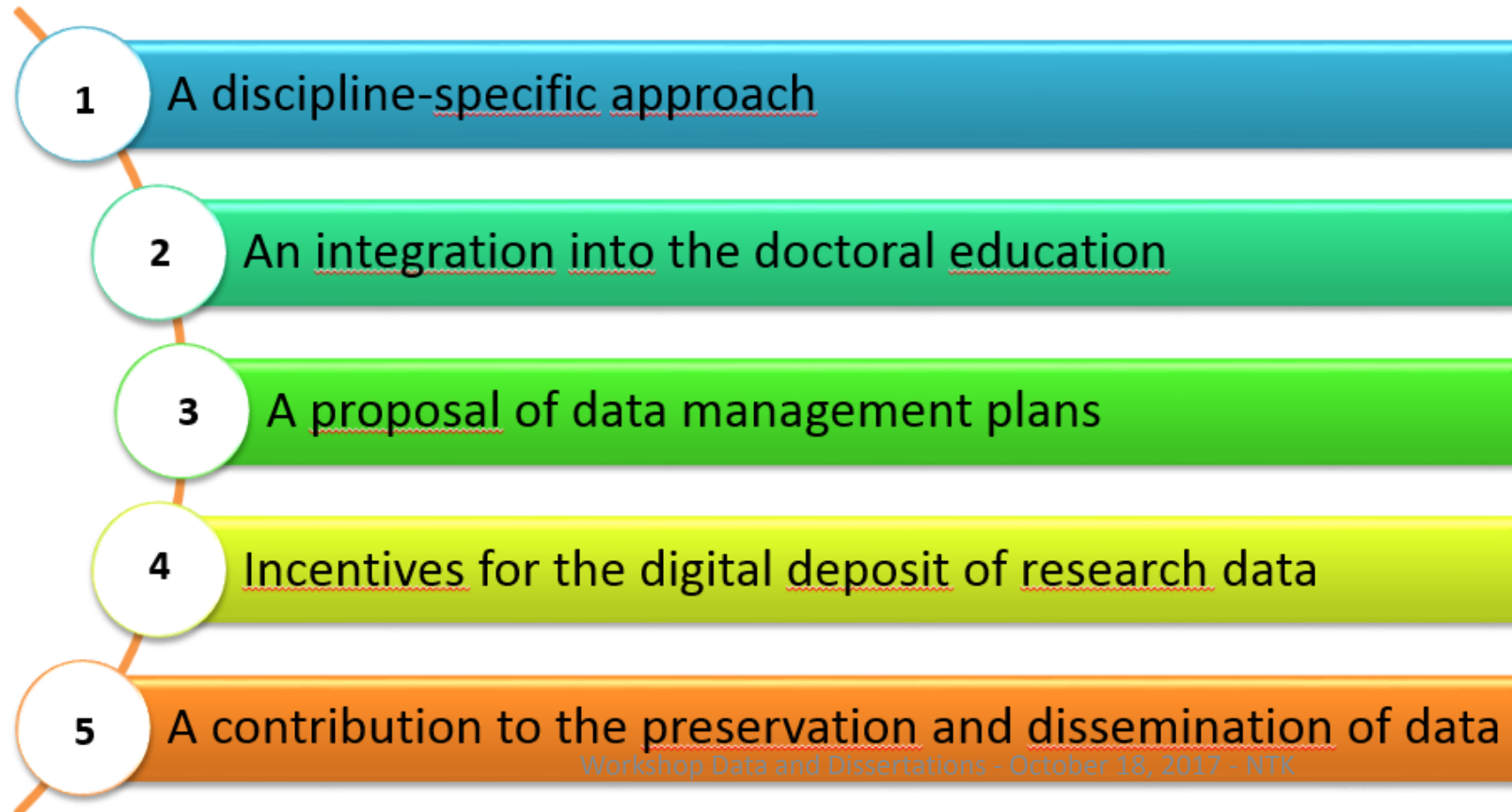
PAUSE

Our project, your initiatives...

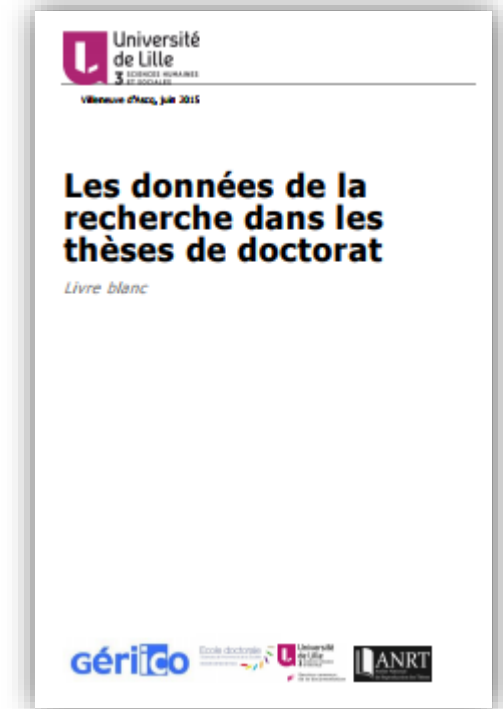
SECOND PART

The Lille project

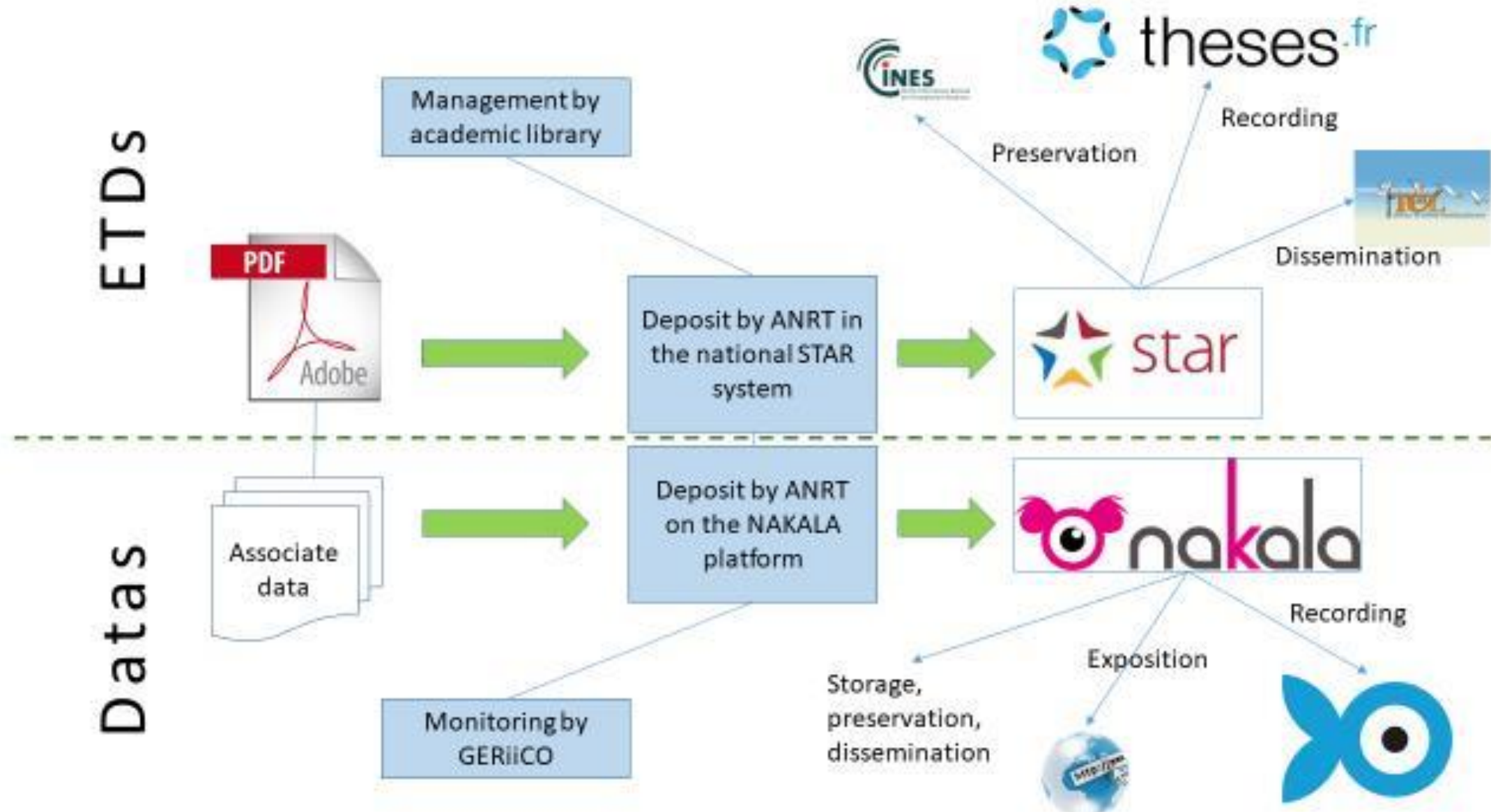
- White paper on data in dissertations
- 2015-2018



Workshop Data and Dissertations - October 18, 2017 - NTK



Local ETD data workflow



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

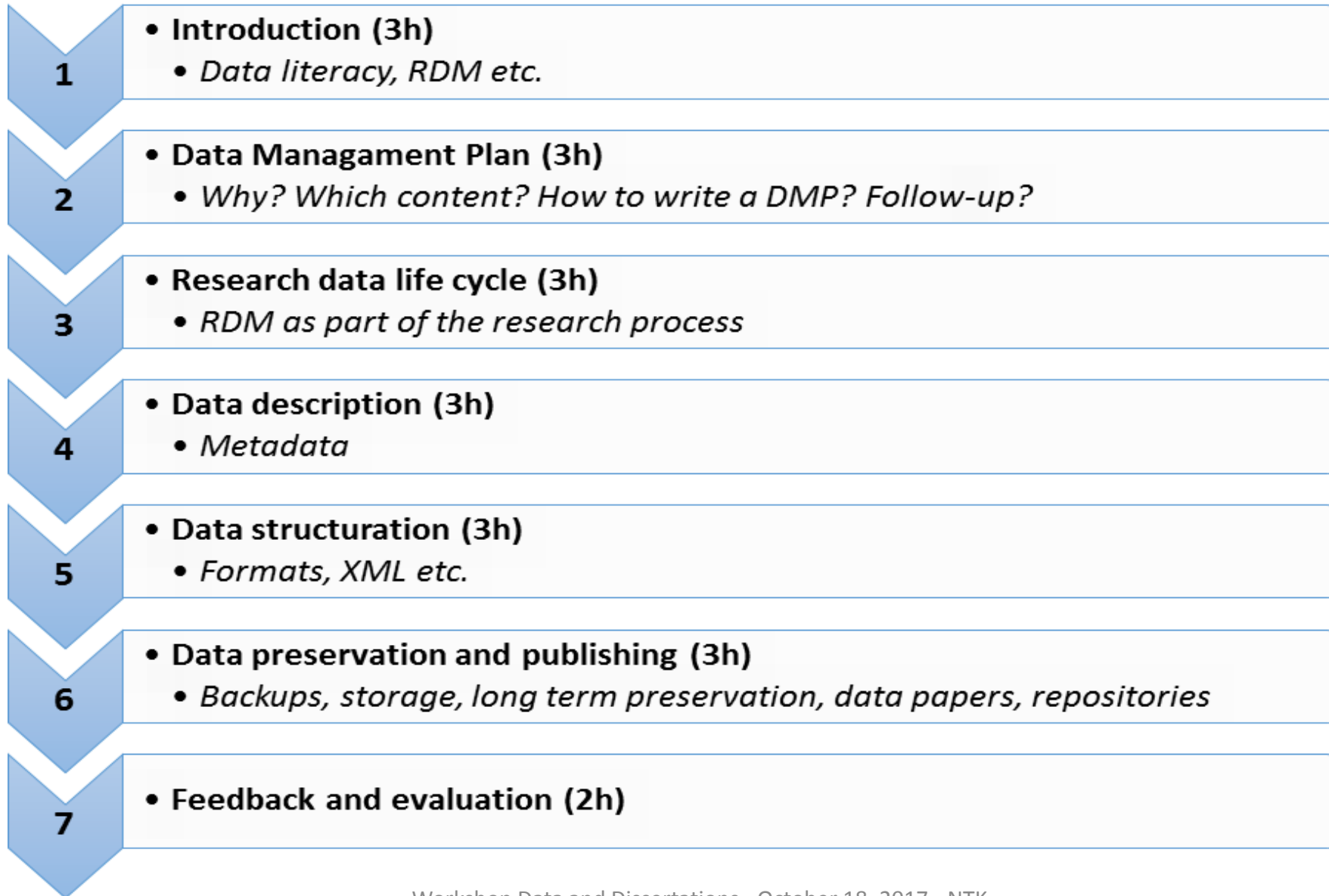
Other issues

- 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

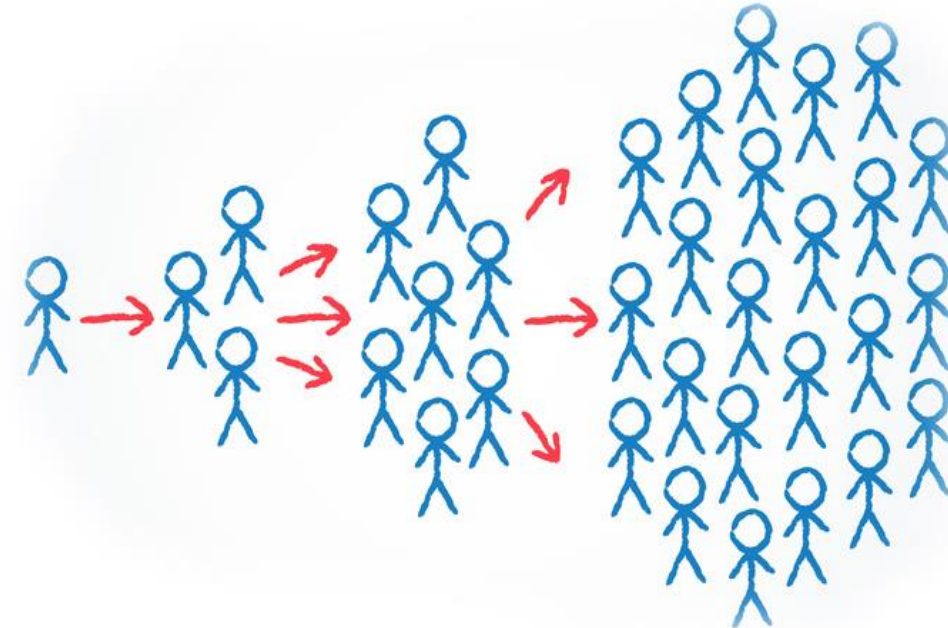
Examples of students' DMP

- On French CNRS platform [DMP OPIDoR](#)
- Dynamic character of DMP (initial, mid-term, final)
- Different data literacy level of students, depending on
 - Progress of dissertation project
 - Discipline (e.g. ethical and privacy issues in psychology, sociology...)

Your initiatives?

Questions?

- Main characteristics of a RDM program with (for) PhD students
- Key factors of success
 - Governance, education, *viral marketing*, partnerships...
- Major risks
 - Leadership, ideology (evangelism), focus on tools not people



Feedback?

THANK YOU !

References

http://www.citeulike.org/user/Schopfel/tag/data_management

Contact

joachim.schopfel@univ-lille3.fr