Université de Lille 3 SCIENCES HUMAINES ET SOCIALES

Univerza v Ljubljani

Data and dissertations

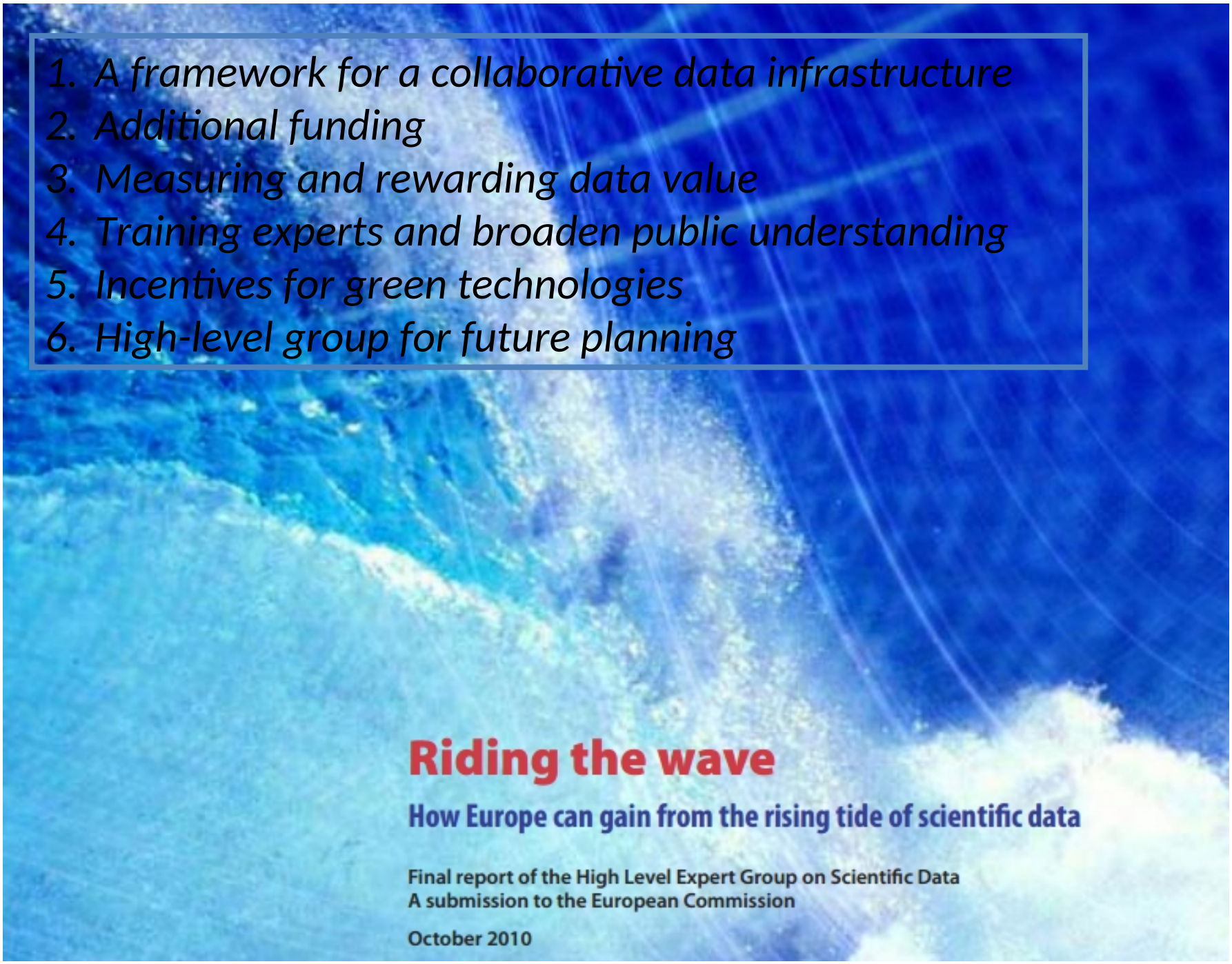
Joachim Schöpfel
Primož Južnič
Hélène Prost
Cécile Malleret
Anna Češarek
Teja Koler-Povh

cnrs advancing the frontiers

Data and dissertations

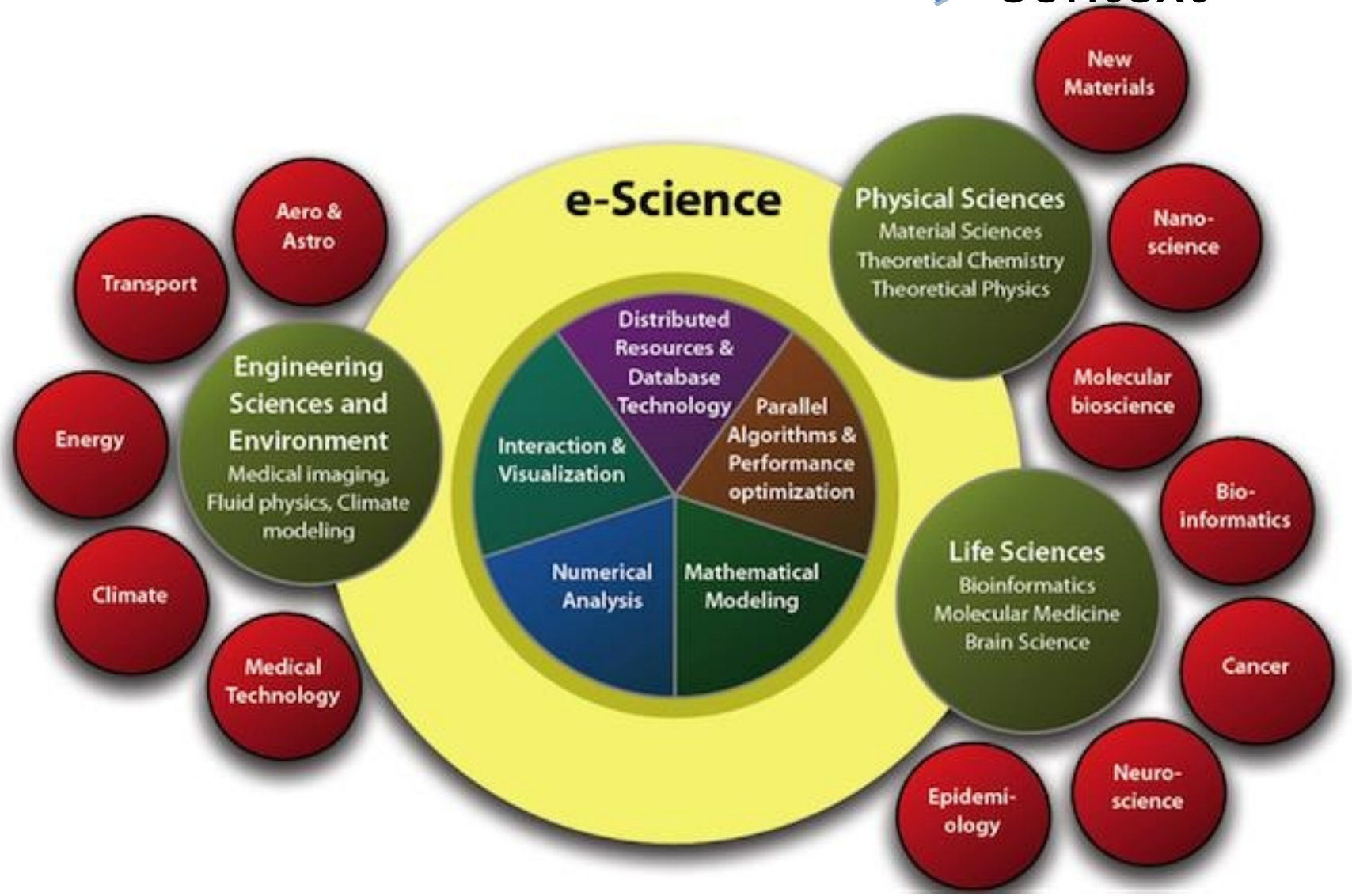
Context

1. *A framework for a collaborative data infrastructure*
2. *Additional funding*
3. *Measuring and rewarding data value*
4. *Training experts and broaden public understanding*
5. *Incentives for green technologies*
6. *High-level group for future planning*



E-science

Context



https://www.kth.se/en/forskning/forskningsplattformar/ict/forskning/e-vetenskap-1.323973

eScience

- Data-driven science
- Data-intensive scientific discovery
  - Data deluge
- *4th paradigm* (Hey et al. 2009)
  - Distributed computer and knowledge systems
  - Information and communication technologies
  - Large-scale and collaborative sciences and engineering
  - Mathematical modelling, numerical analysis, visualization techniques, data mining…
  - Integration of theories, simulations and experiments
- Scientific information has become a continuum between publication and data

Research data

- « Recorded factual material commonly accepted in the scientific community as necessary to validate research findings »

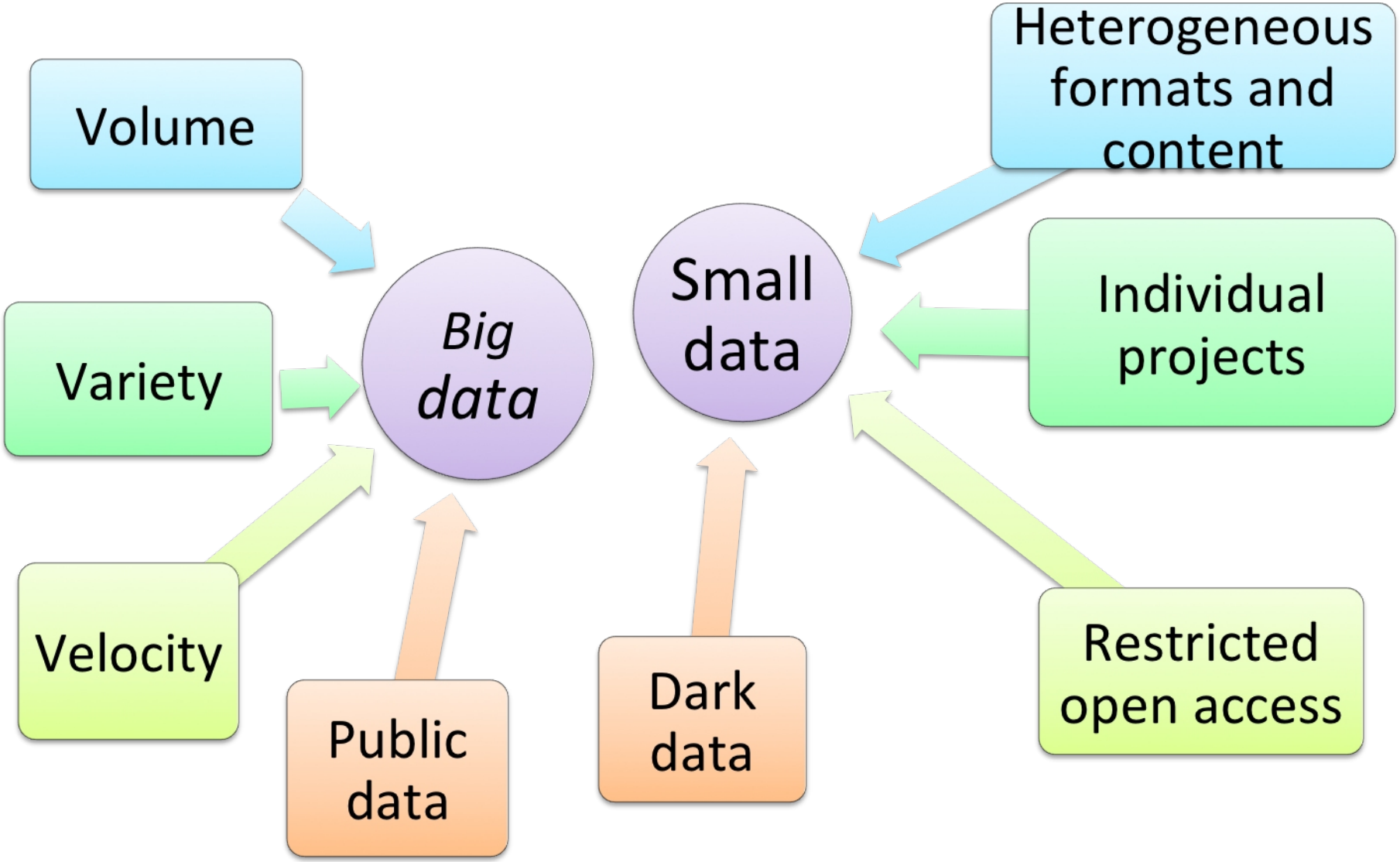
  US OMB Circular 110

- « Re-usable research results, collected, observed or created for purposes of analysis to produce original results »

  University of Edinburgh (cited by Burnham 2013)

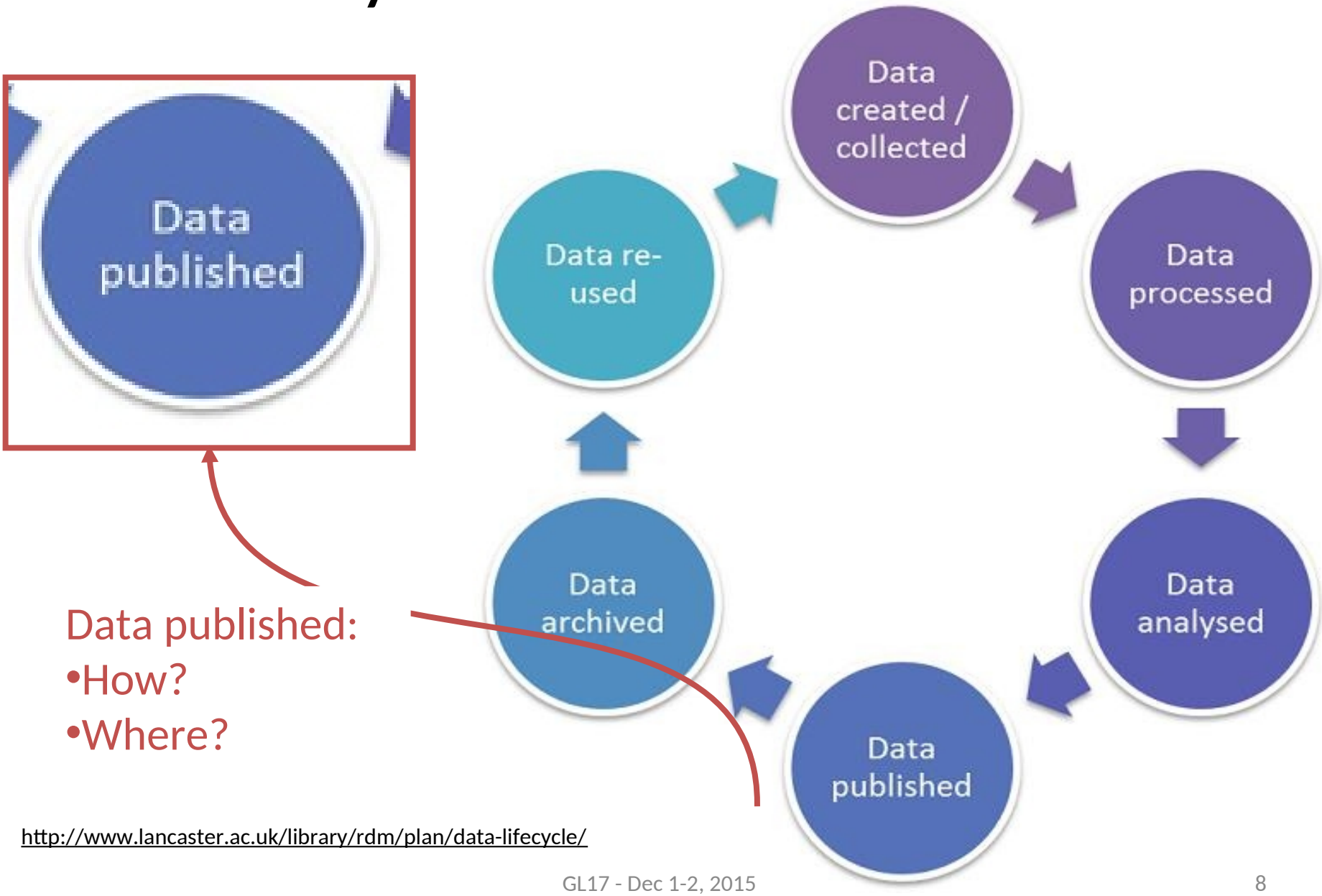
  – Large variety of formats, sources and types
  – Data as material (input) v. Data as results (output)

Big data vs small data



How to valorise listed data in the appendices ?

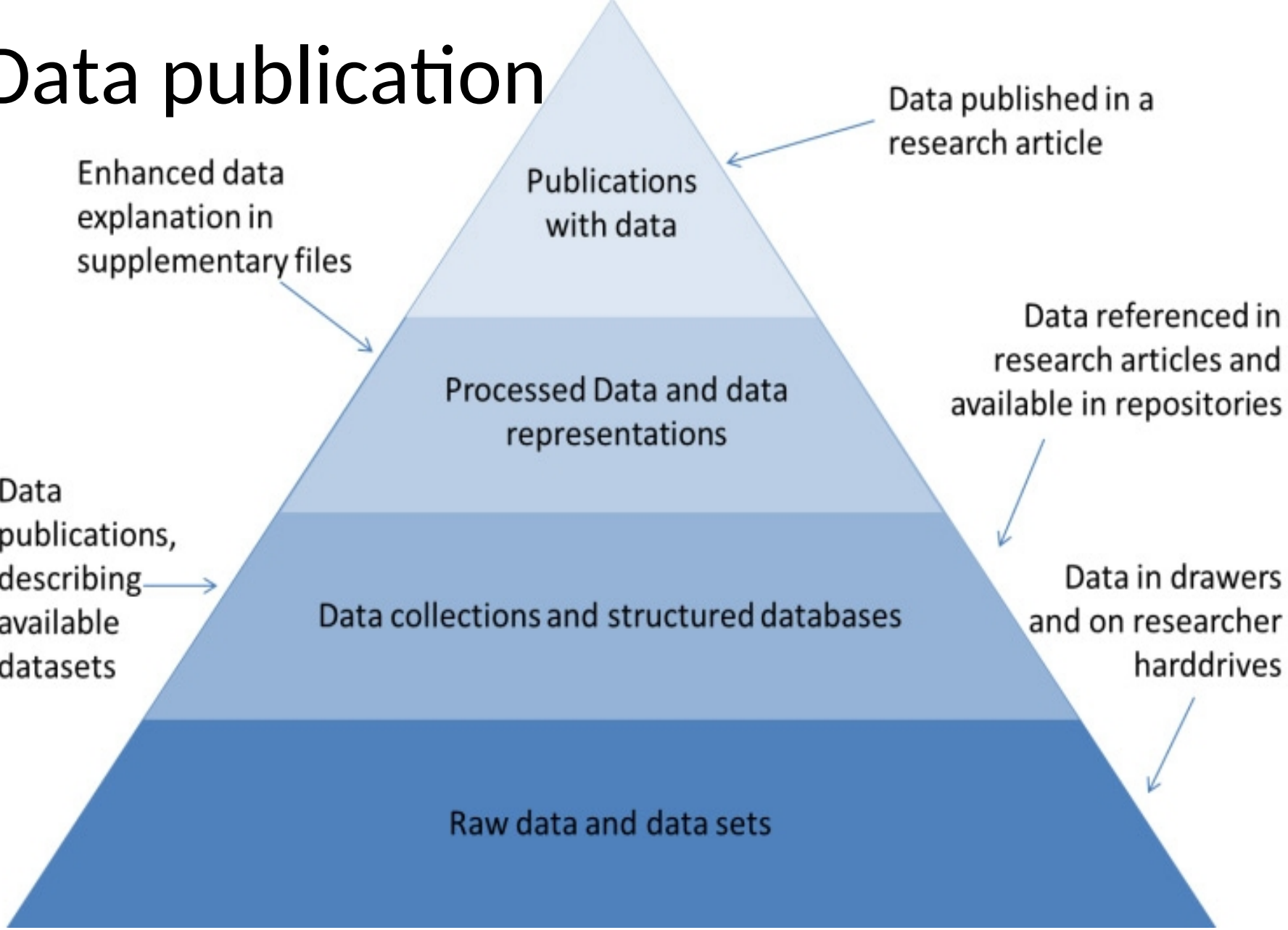
Data life-cycle



Data published:
- How?
- Where?

http://www.lancaster.ac.uk/library/rdm/plan/data-lifecycle/

Data publication

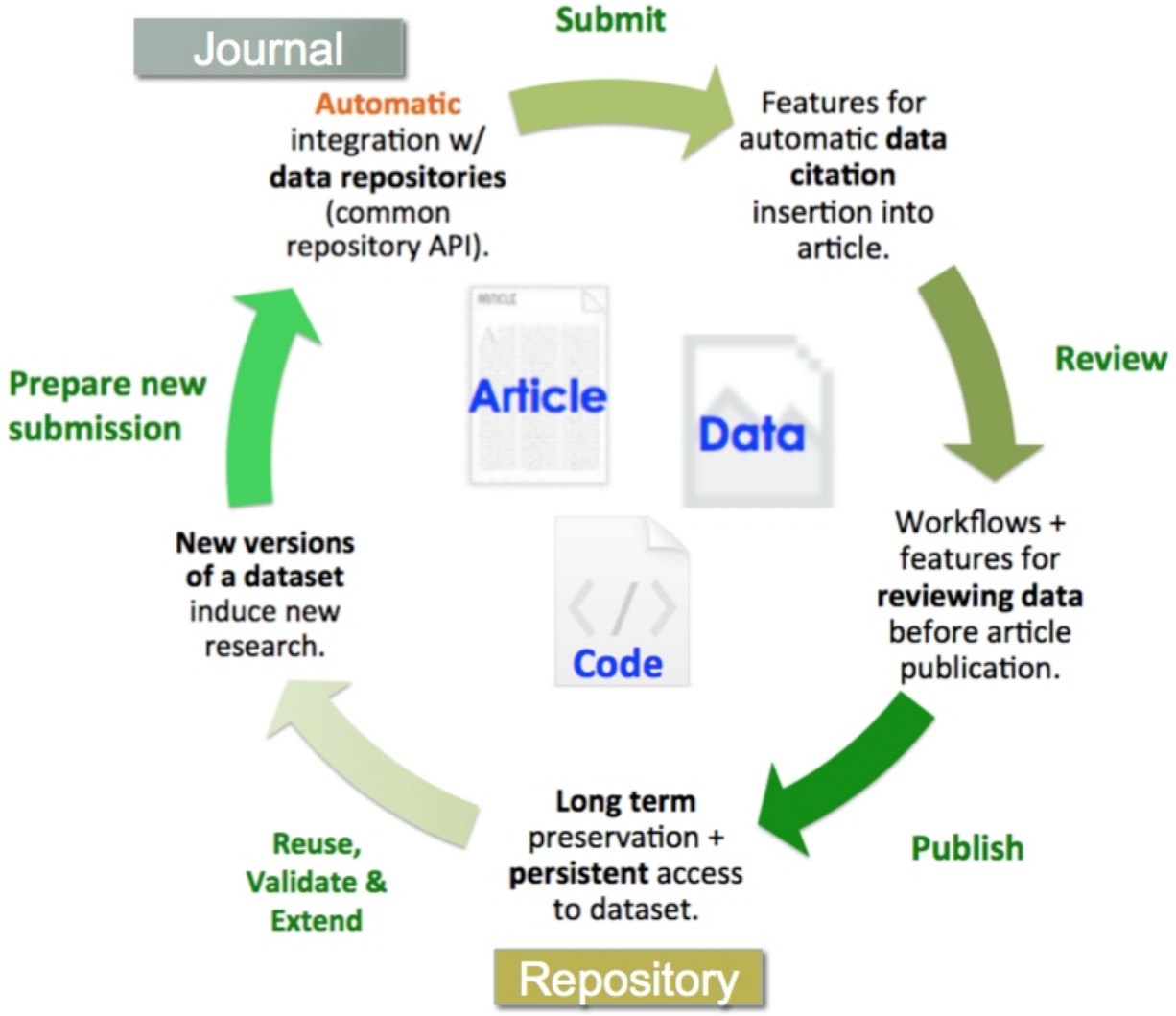


https://www.elsevier.com/connect/can-data-be-peer-reviewed

Publication and data

- Document as data
  - Exploited as primary data source for TDM
- Data vehicle
  - Supplementary materials of publication
- Gateway to data
  - Publication contains links to data, integrated or not in the text

Integration of DataVerse and OJS – An example of the gateway function



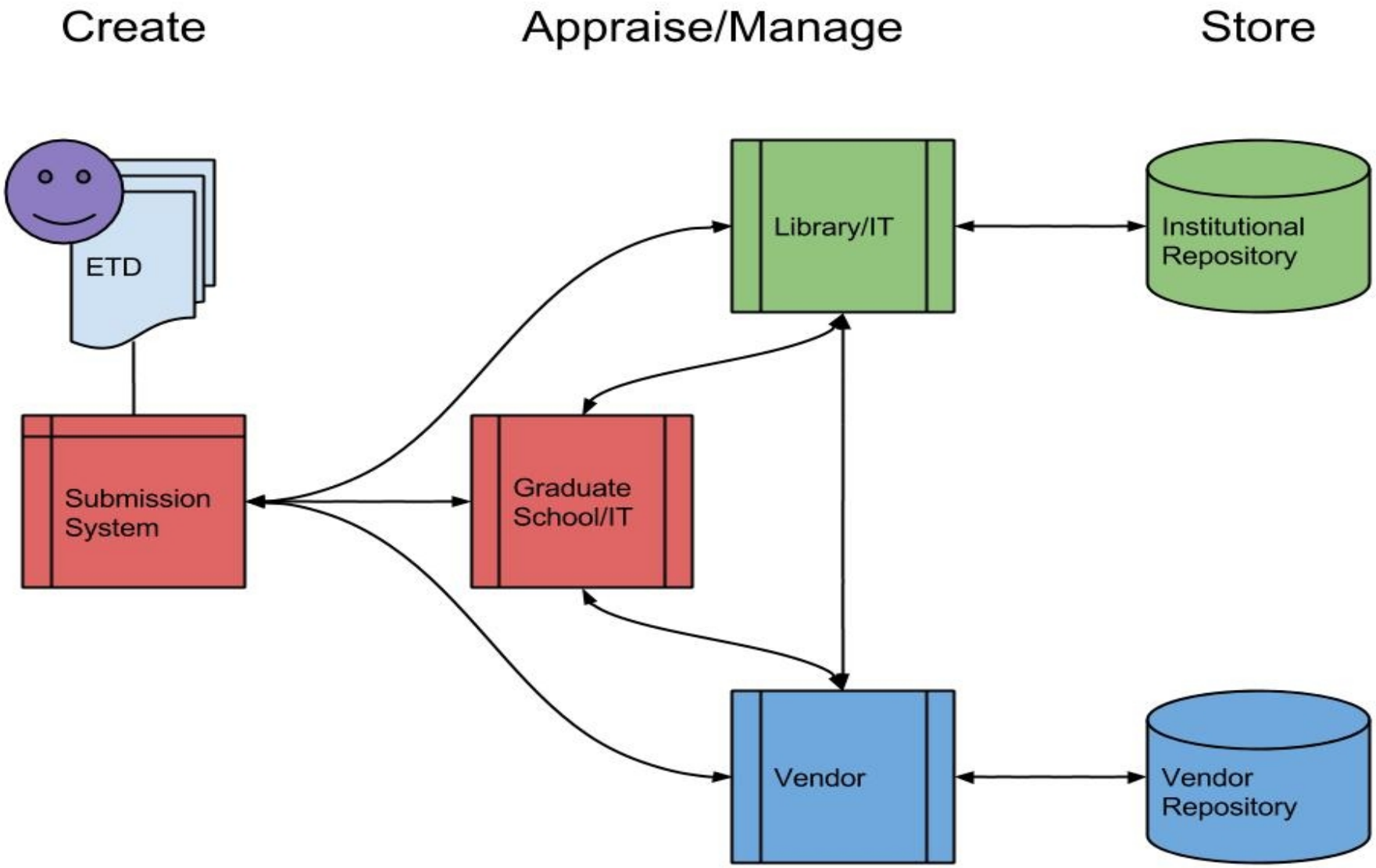
http://journal.code4lib.org/articles/10989

Data and dissertation

The case of ETDs

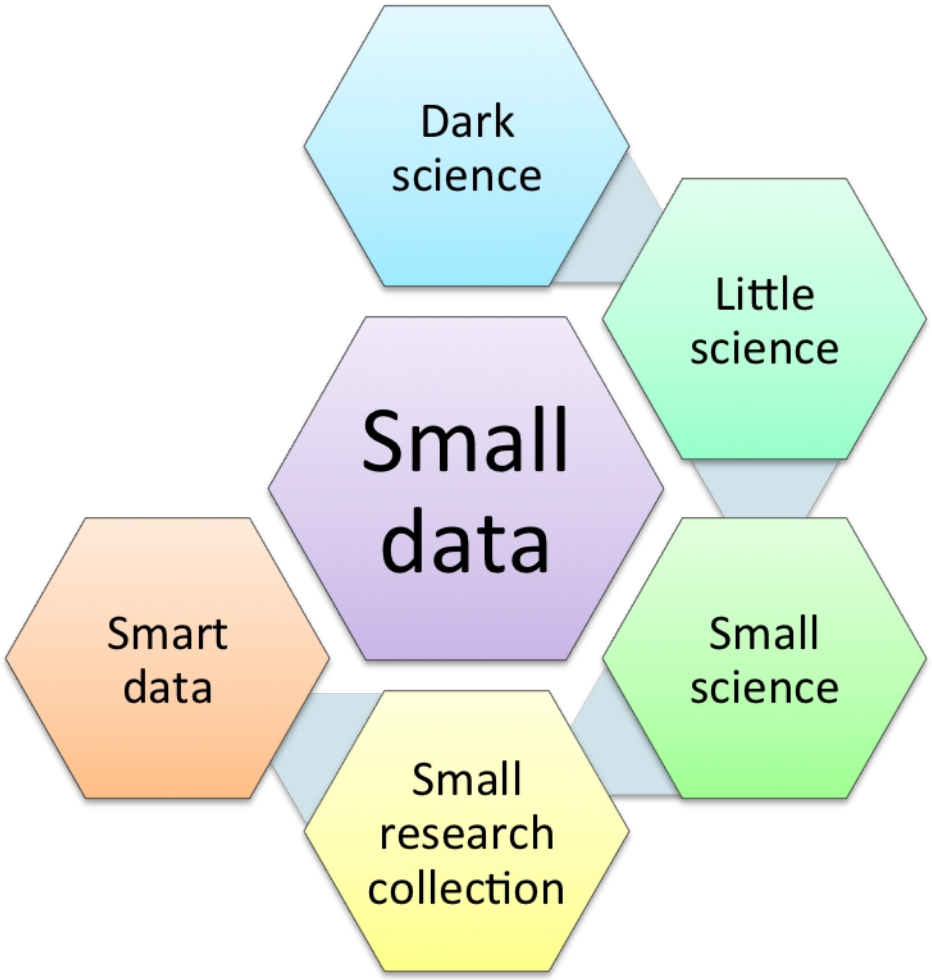
The challenge of ETDs

Life-cycle management

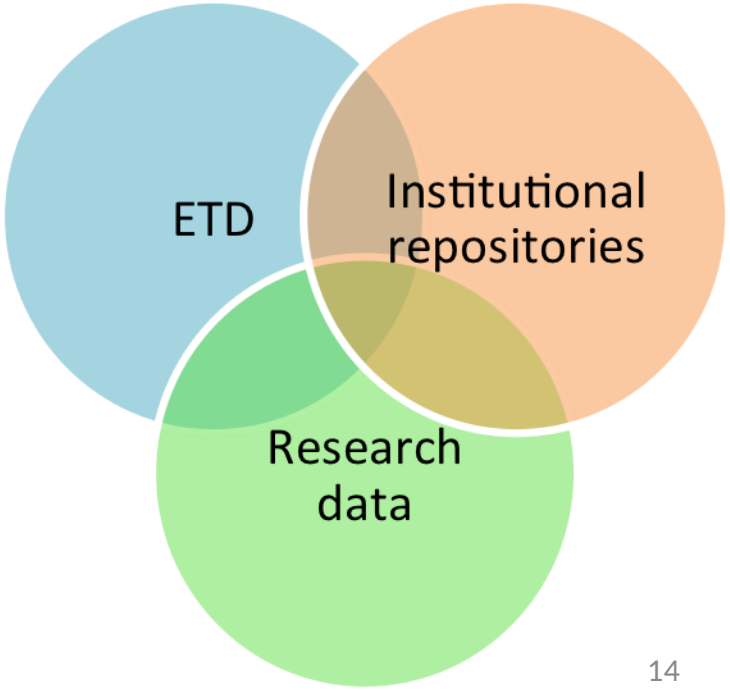


http://educopia.org/research/electronic-theses-and-dissertations

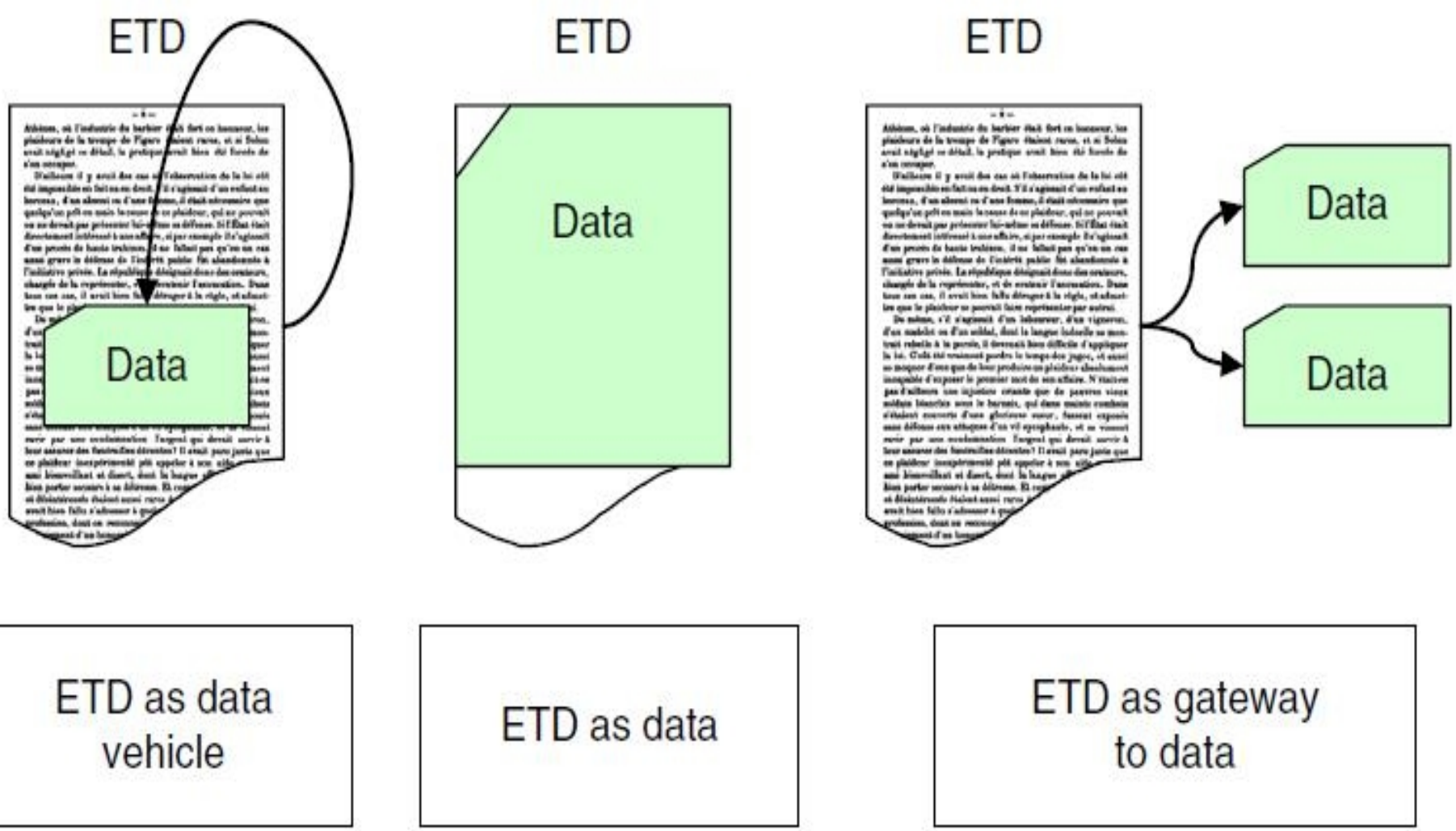
The small data of ETDs



- Originality
- Linked to a scientific program
- No commercial and public character



ETDs and data



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

Data and dissertations

Empirical results

Data management : practices & needs

Survey at Lille 3

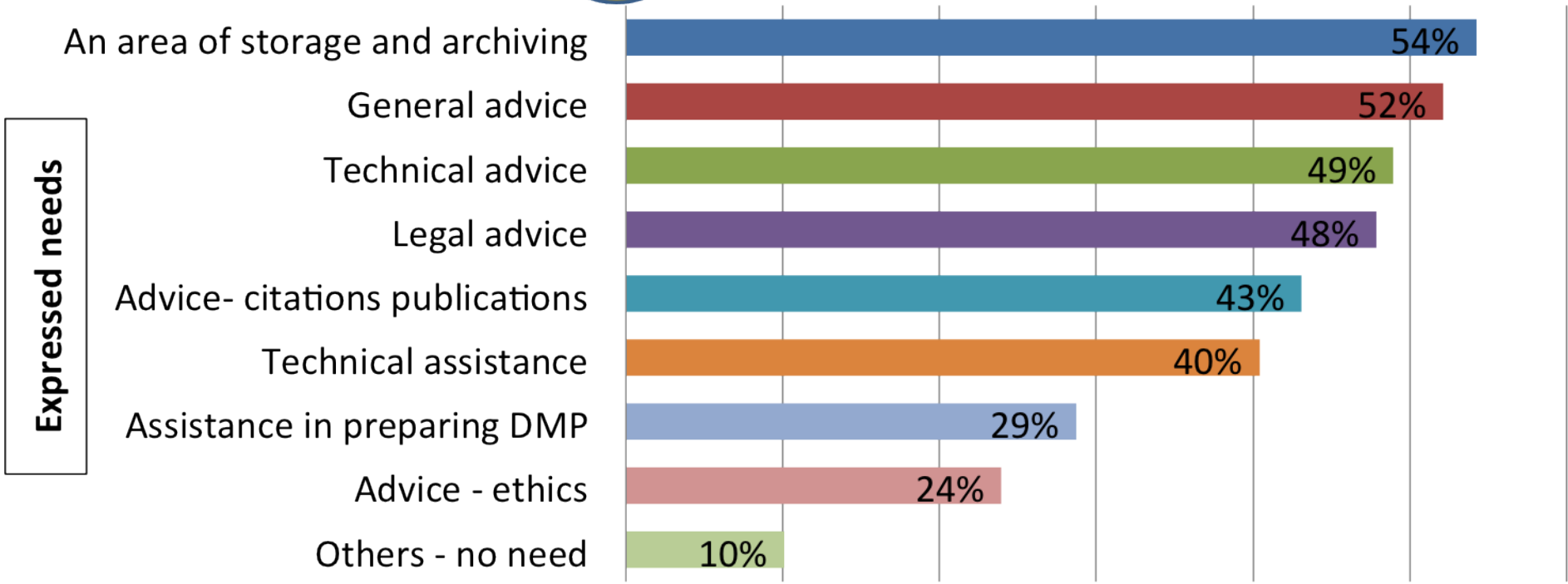
9/10

store their data on local

83% on private computer

49% on professional computer

97% declare themselves responsible for data backup

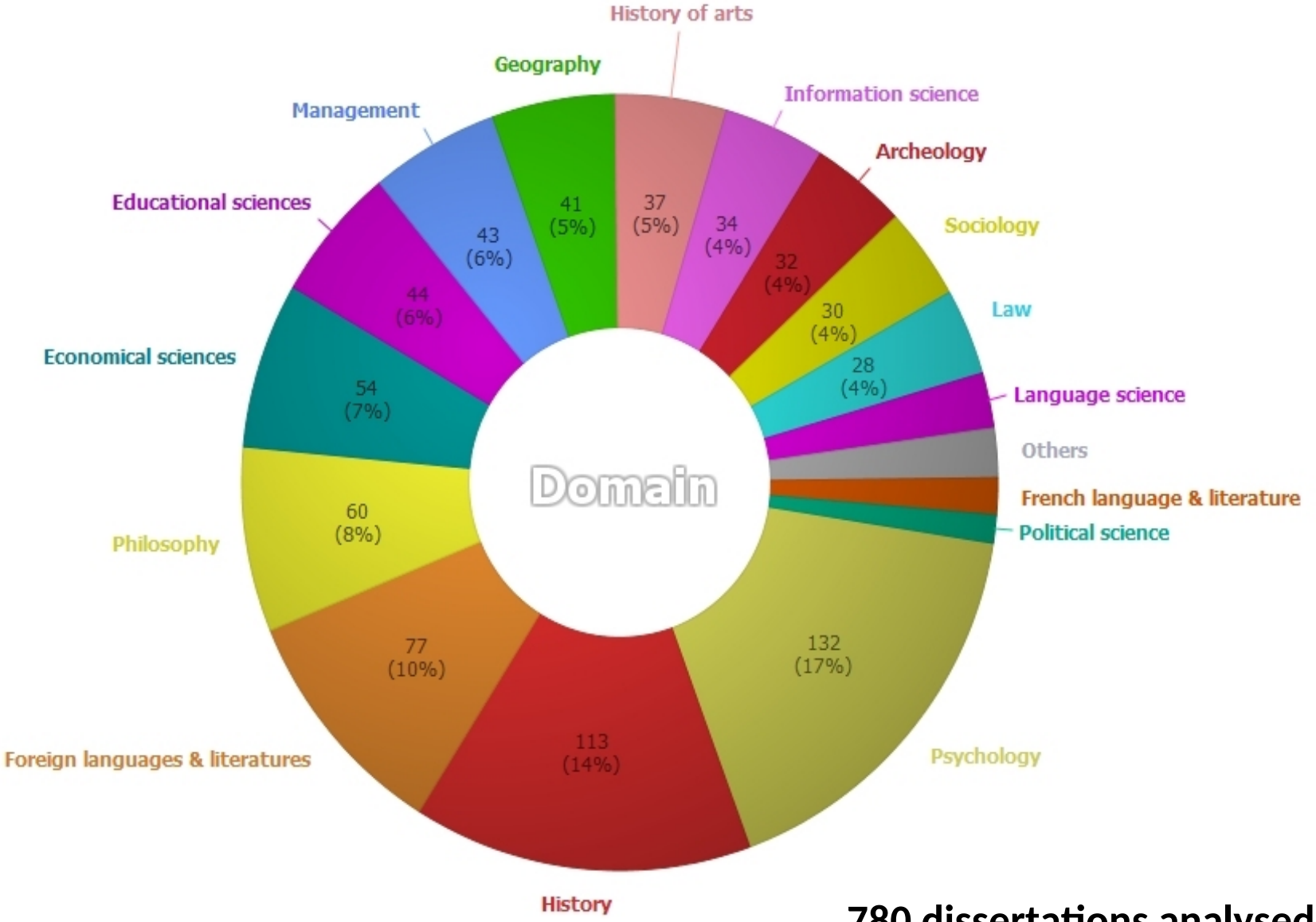


Data management : practices & needs

The PhD students

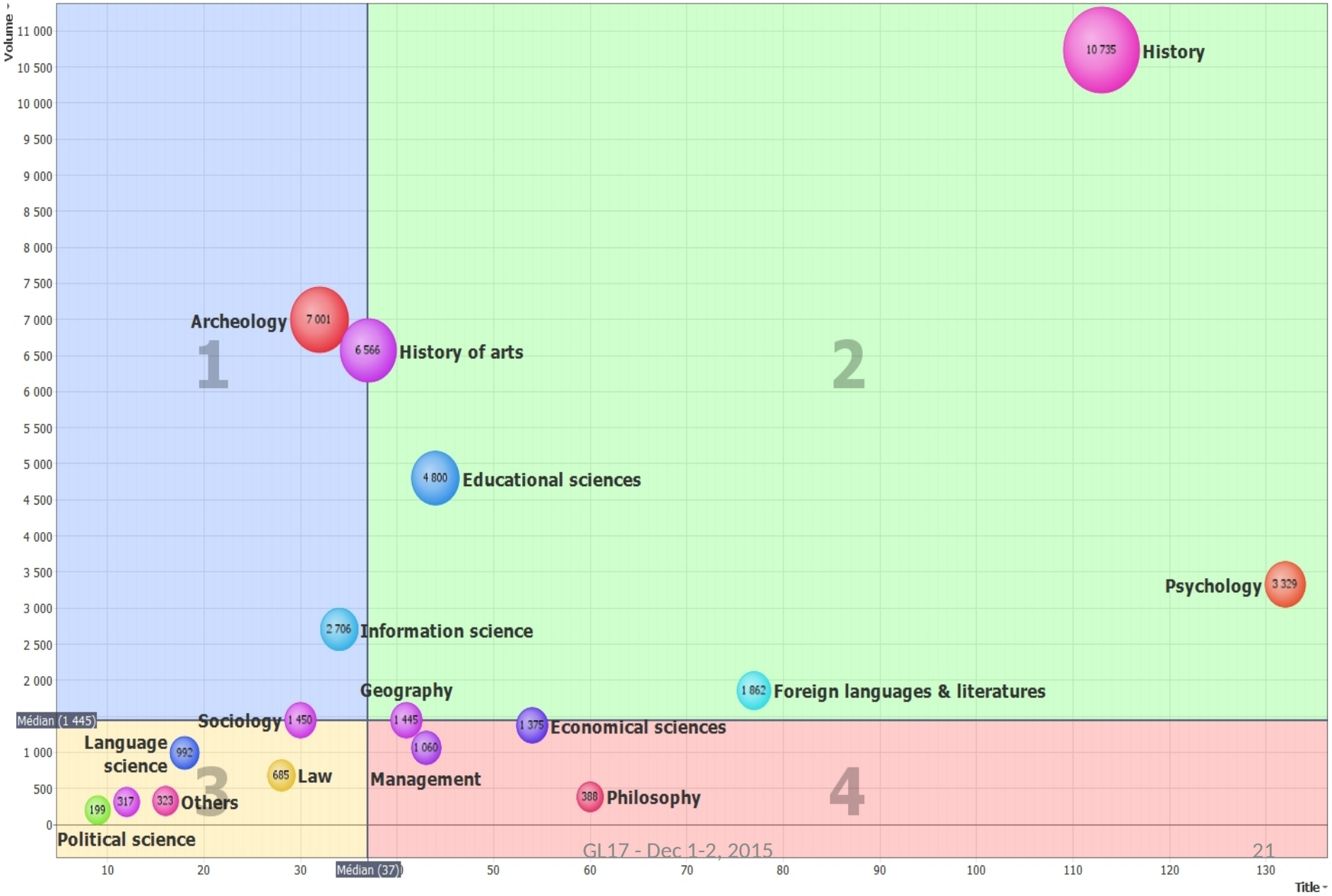
Research data in dissertations

A French-Slovenian survey

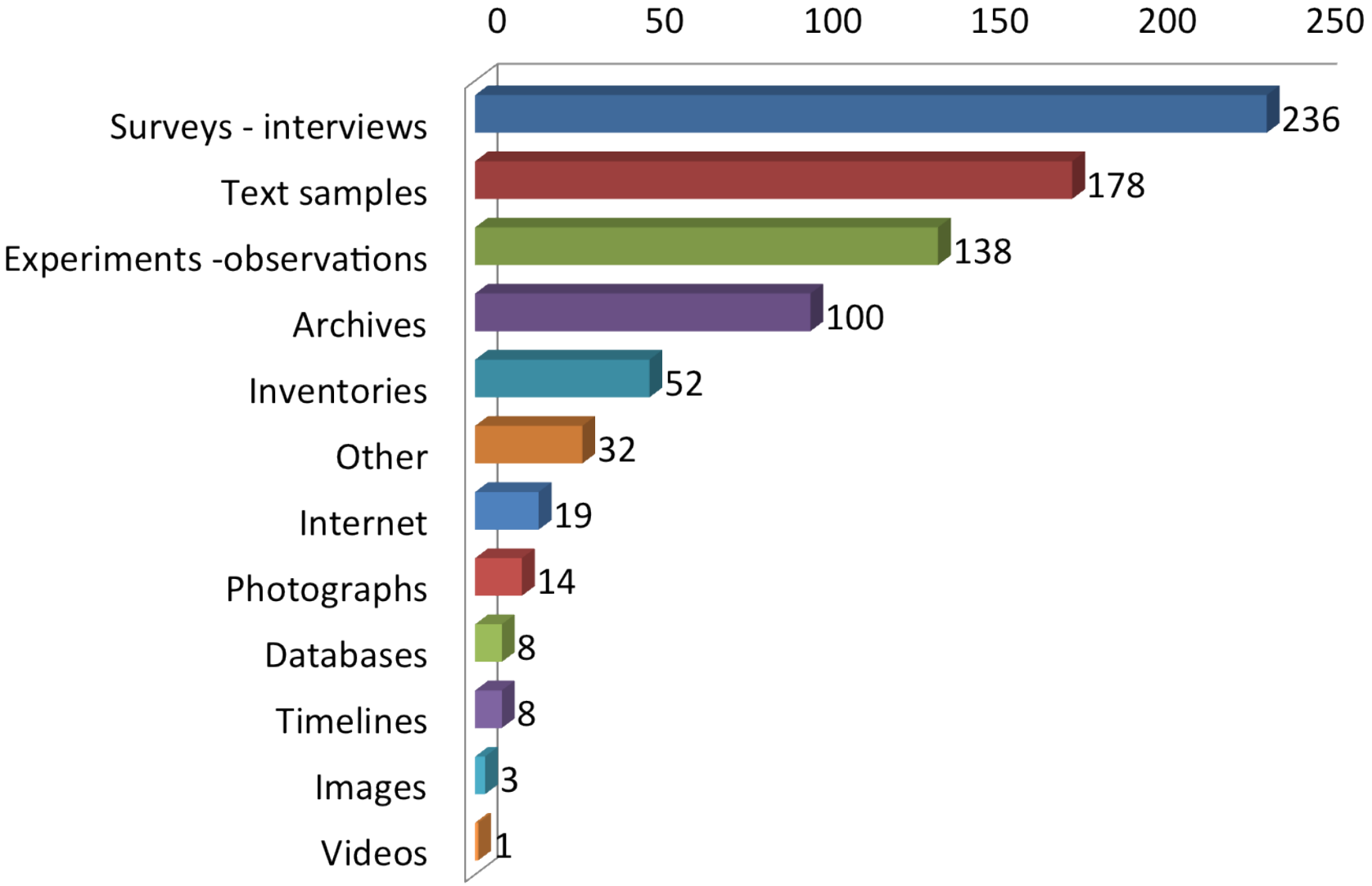


780 dissertations analysed

The size of appendices



Sources of appendices



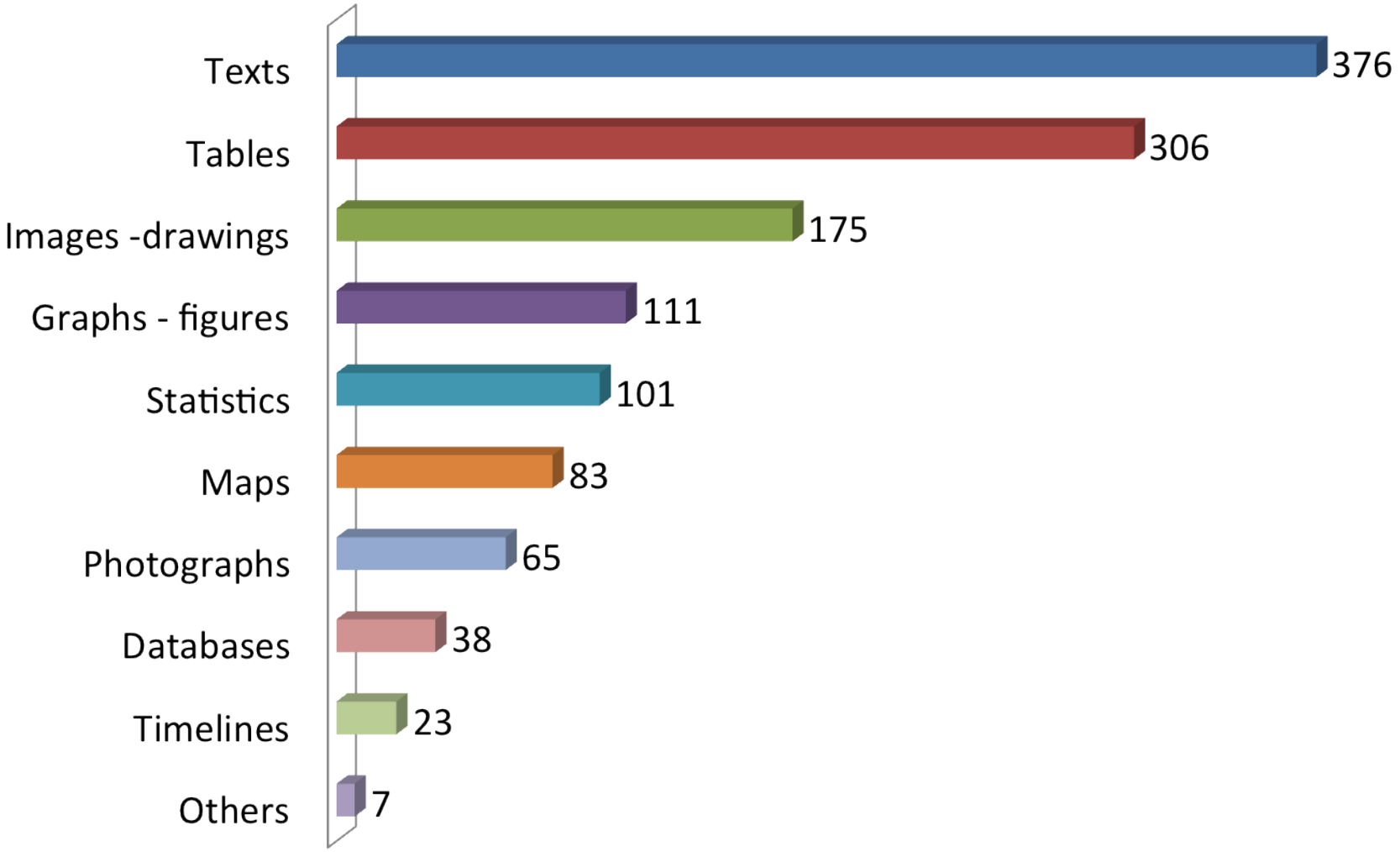
Sources of appendices

Domains

Y: Domain	Archives	Databases	Experiments - observati...	Images	Internet	Inventories	other	Photographs	Surveys - interviews	Text samples	Timelines	Vidéos	Total
Archeology	1		14			25		5		3			30
Economical sciences		1	27		3	1			14	18			43
Educational sciences			6		3			1	33	13	2		38
Foreign languages & literatures	4		5		1		11		14	35	1		46
French language & literature									2	3	1		6
Geography	3	2	27		1		5		15	8		1	33
History	76		4	1		7		2	9	21	2		88
History of arts	10		4			17	7	2	3	4	2		28
Information science	1	1	1		6		6		17	10			28
Language science			1						3	5			7
Law					2				2	5			7
Management	1	3	7		1	1			21	14			30
Others		1	4						1				6
Philosophy	1		2					1	1	9			11
Political science									6	3			6
Psychology			30				3	2	71	12			91
Sociology	3		6	2	2	1		1	24	15			28
Total	100	8	138	3	19	52	32	14	236	178	8	13	526

X: Sources

Typologies of appendices

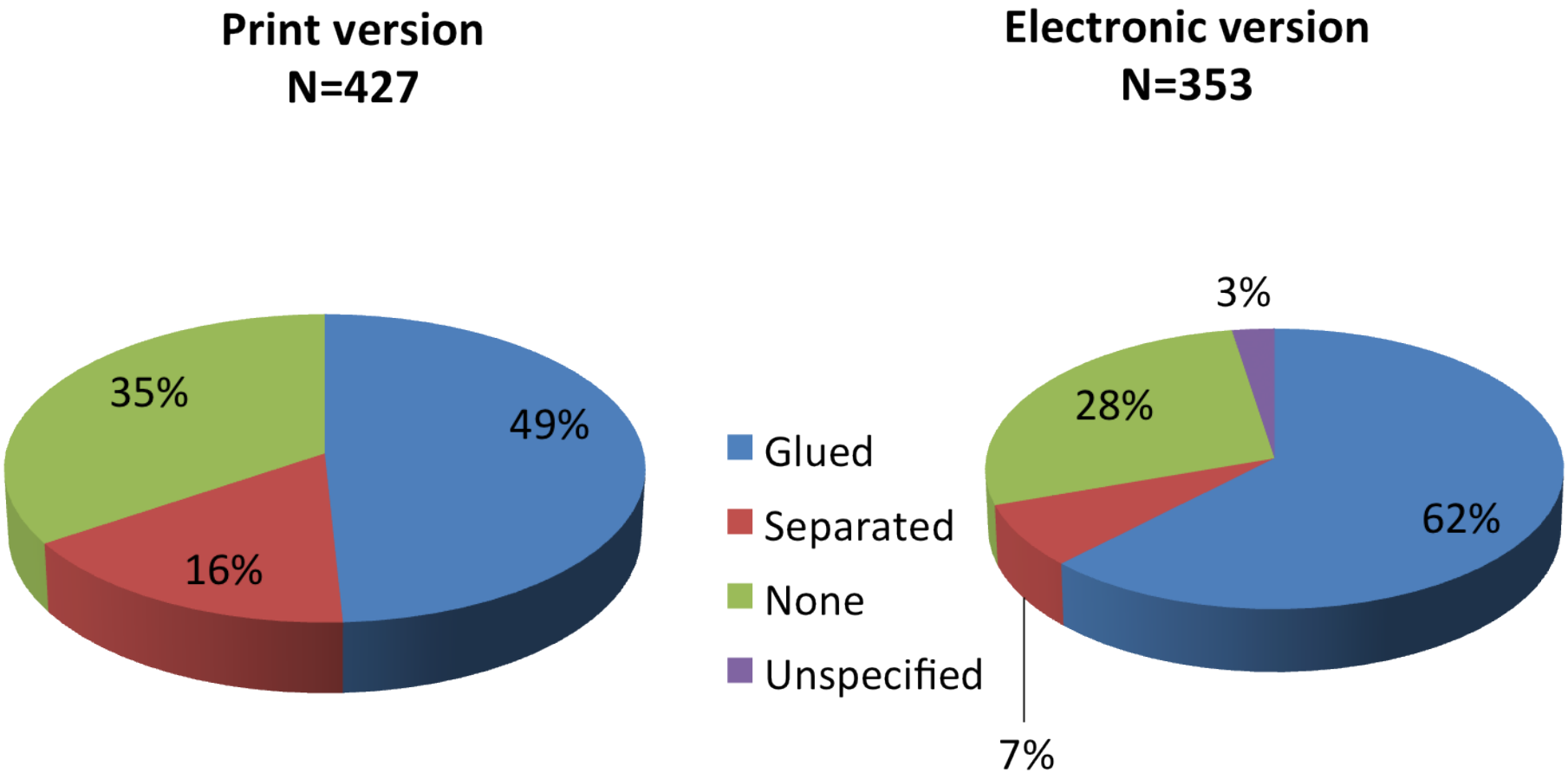


Typologies of appendices

Domains

Y: Domain	Databases	Graphs - figures	Images drawings	Maps	Others	Photographs	Statistics	Tables	Texts	Timelines	Total
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
Total	38	111	175	83	7	65	101	306	376	23	526

Link between text and data appendices



Data and dissertations

observations

Text and data

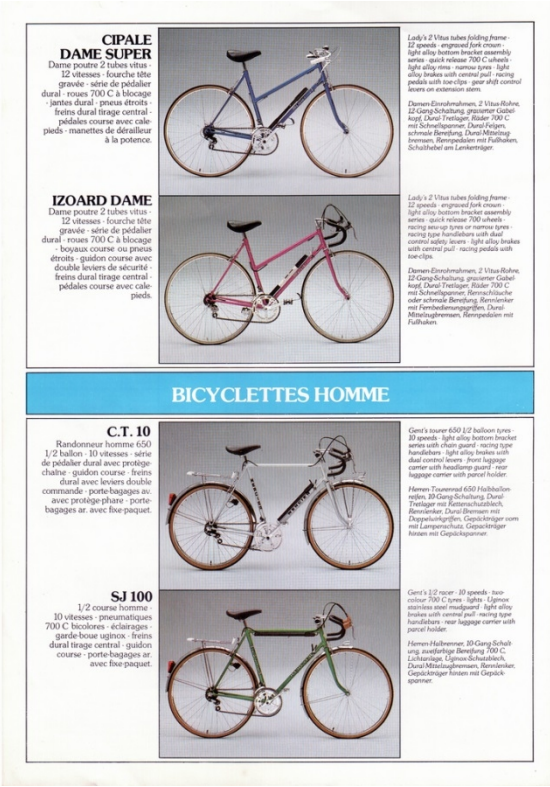
→ Structure & presentation

E-119 **TRI+**

E-119 **TRI**

E-117 **TRI+**

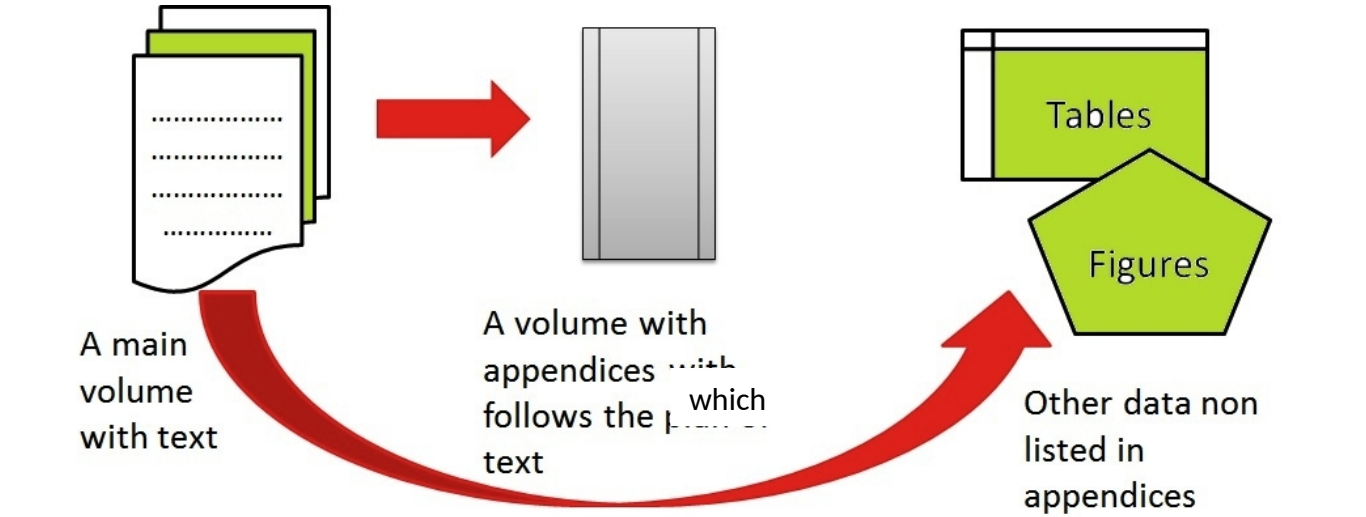
E-117 **TRI**

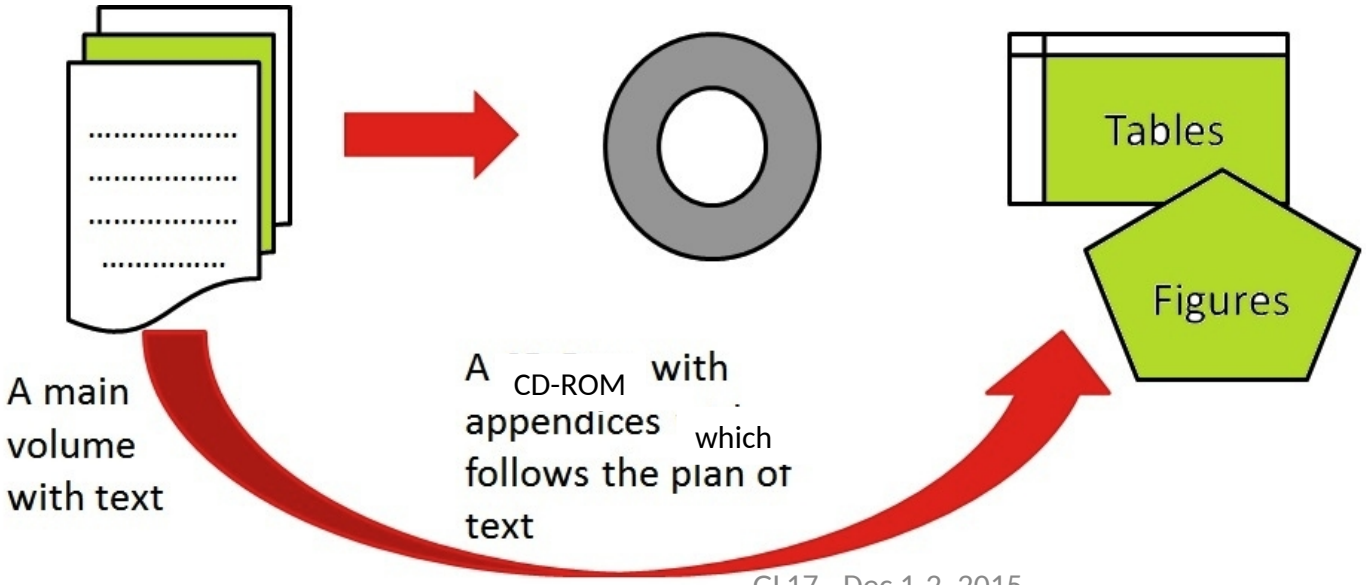


It’s like bicycles !

Text and data

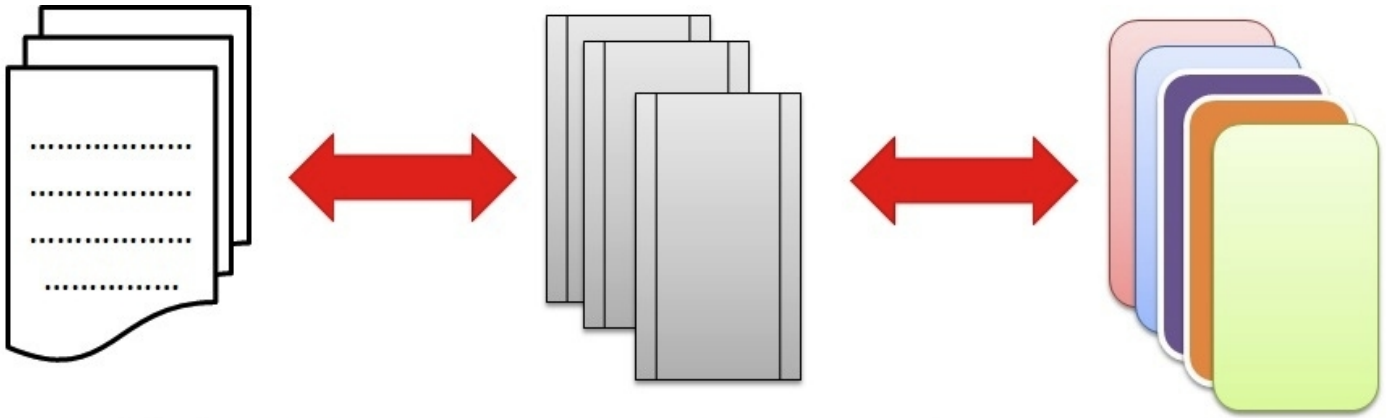
Text and appendices separated (1)





Text and data

Text and appendices separated (2)



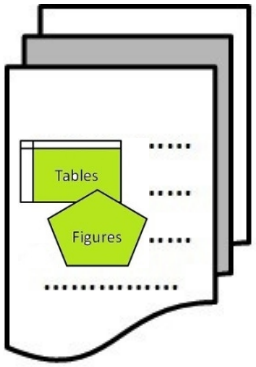
A main volume with text and classification plan

A volume of appendices including catalog with 179 notices

A volume of illustrations with 1581 images

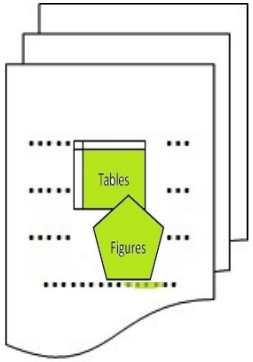
Text and data

Text and data attached



Appendices are inserted at end of each chapter

Other data are included in text and classified



No appendices

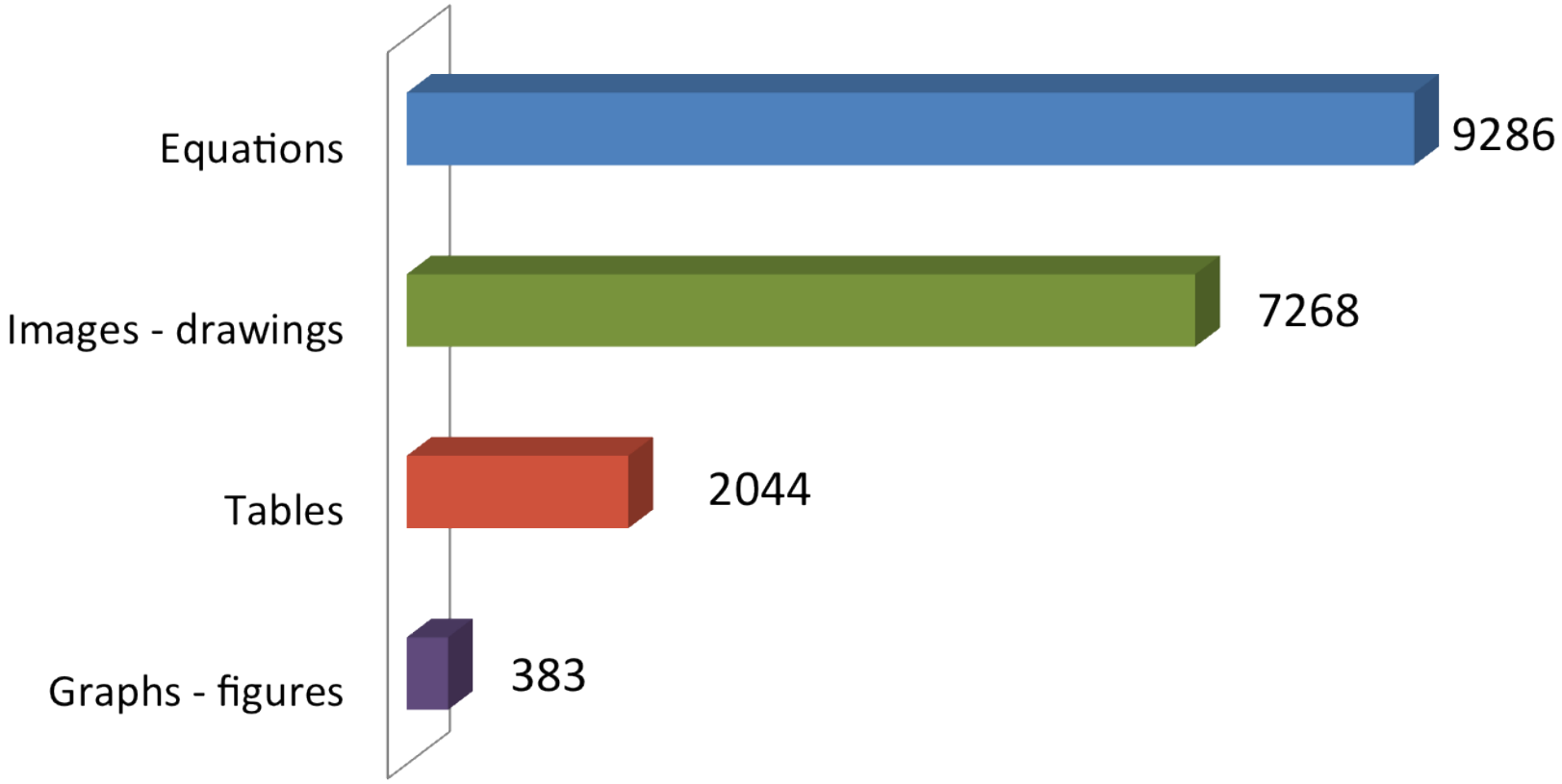
All data are included in text with classification or not

ETDs in engineering science
A case study from IR DRUGG

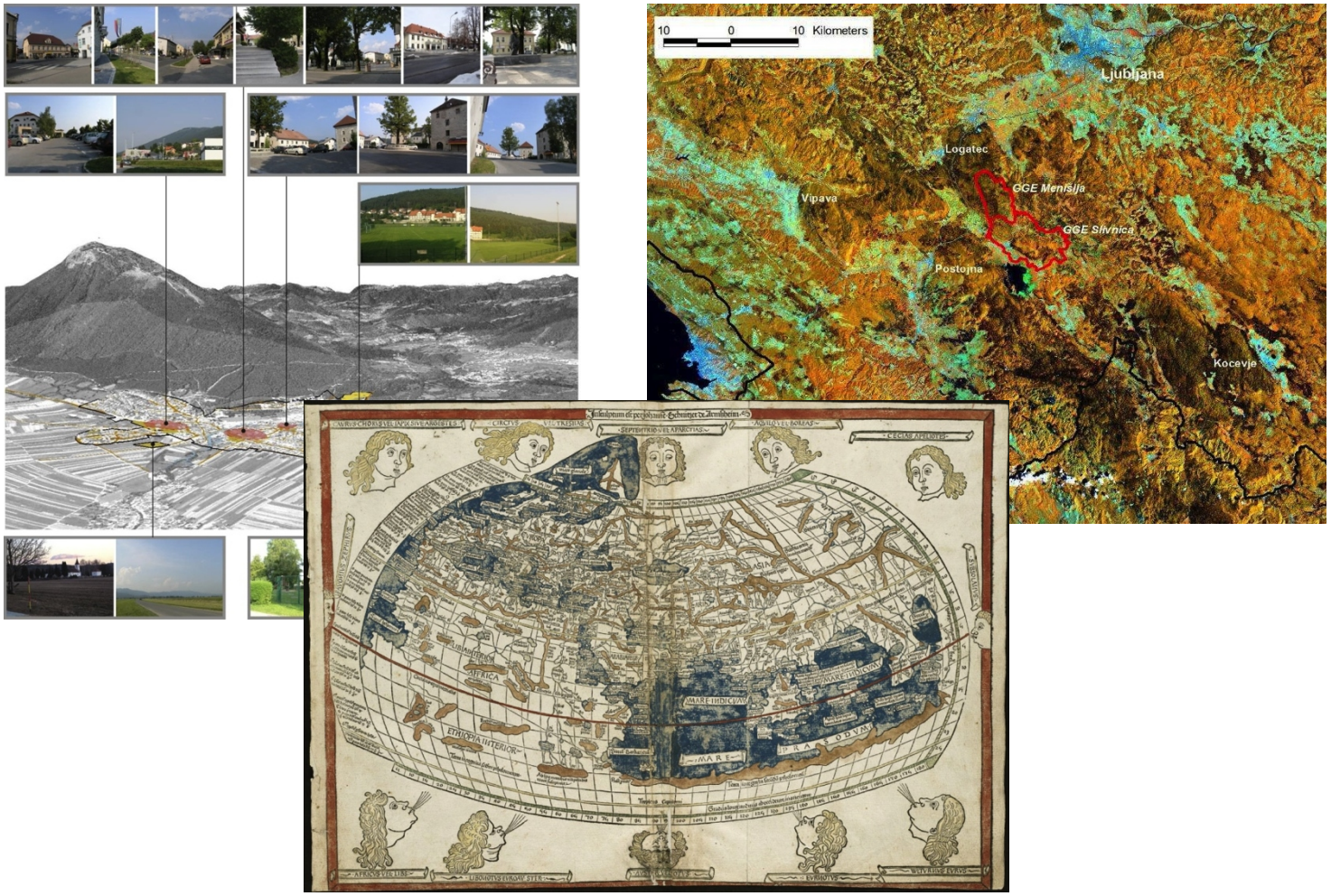
- All dissertations(100 %) are archived in the DRUGG.
- In period 2008-2014 = 86 dissertations.
- 18,981 appendices integrated in all 86 dissertations

  +
- also 237 **attached** appendices in 28 dissertations
- All content described by metadata
- Most data in format pdf

DRUGG - typologies of appendices



Photo? Map? Figure!



The high frequenced appendices in engineering

1. Equitations (total = 9,286).

   They are a part of 65 dissertations (from 86).

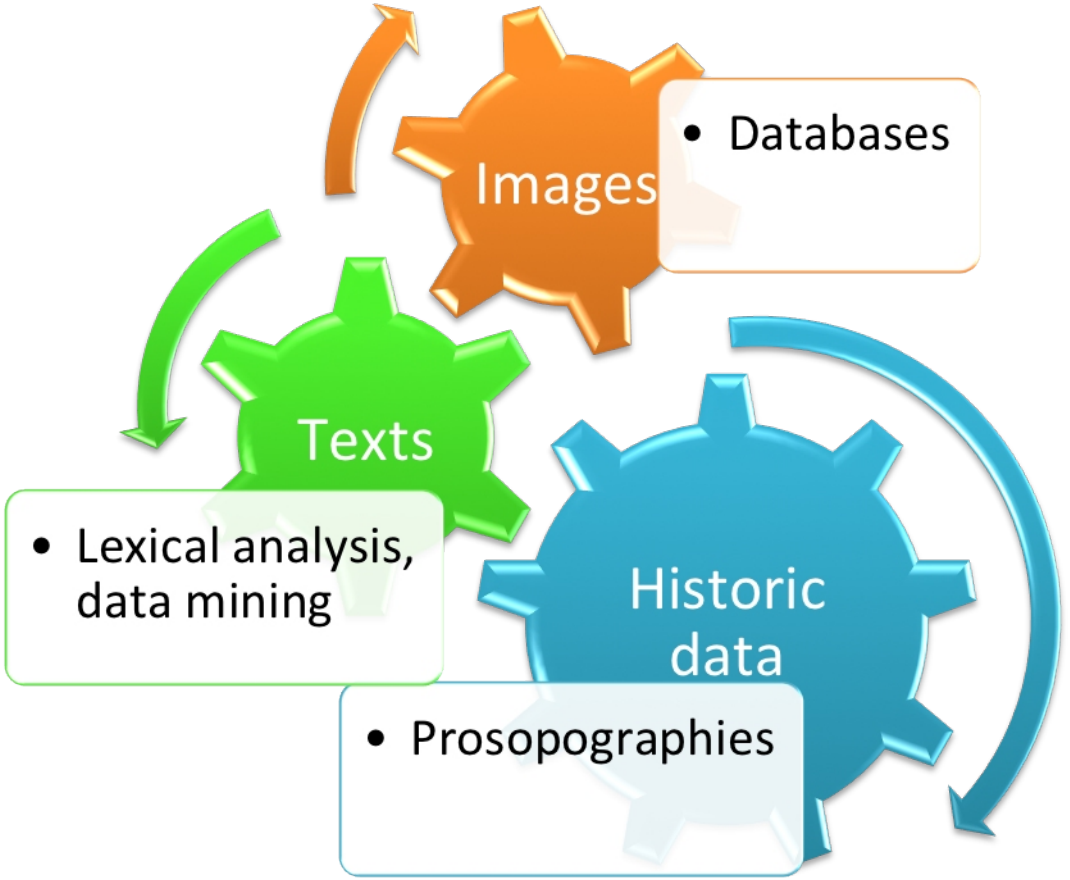
2. Figures (total = 7,591).

   They are presented in all of 86 dissertations.

Specialty: Maps and photos (they are many) are treated as figures!

Conclusion: The appendices are important source of information in engineering.

Potential use & valorisation of data



Barriers to open data

Datasets and/or individual data are not or incompletely documented

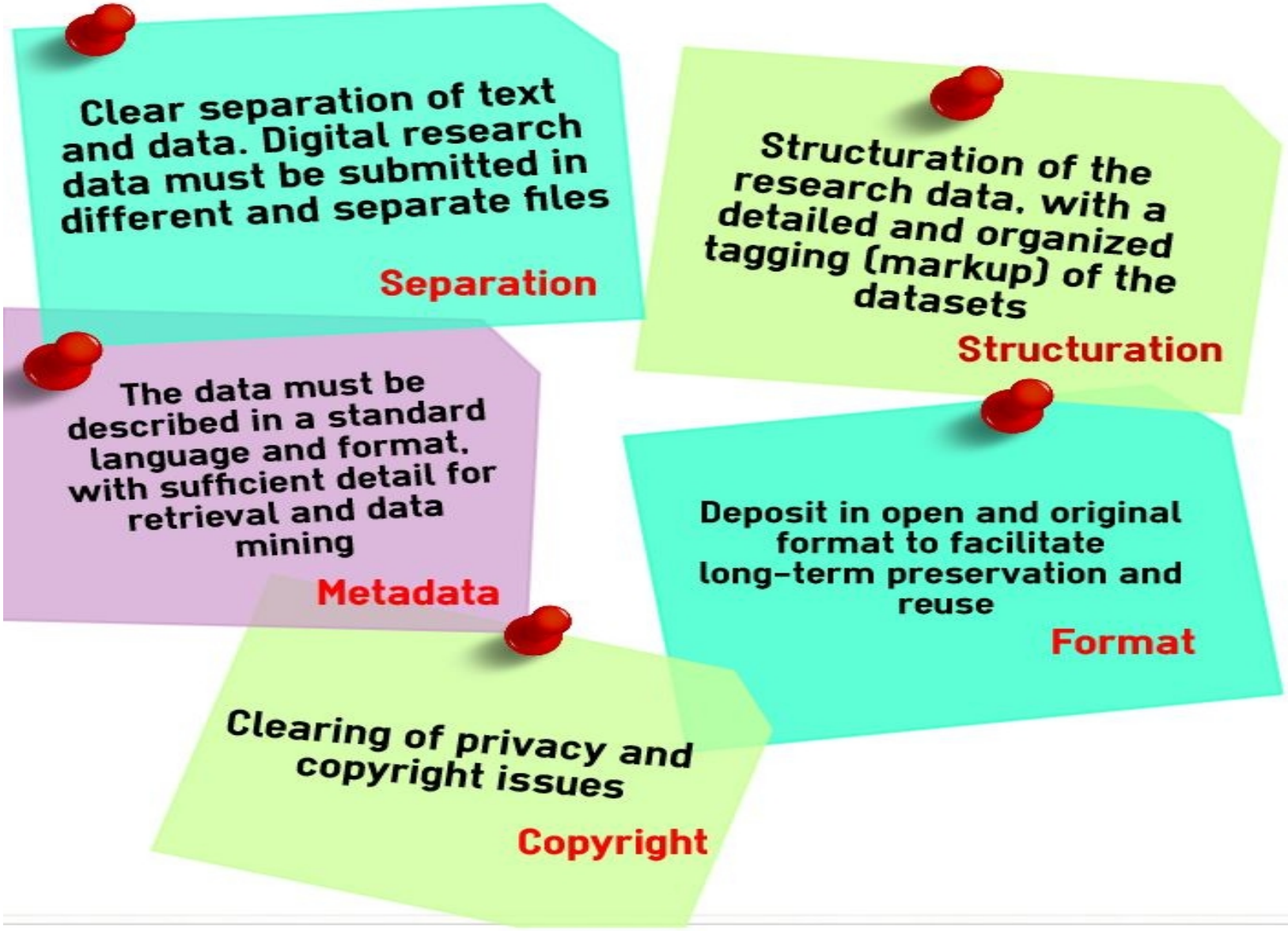
Research data are presented without any structuration or organisation, often together with other, not reusable material in a kind of information mash-up not suitable for further research

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

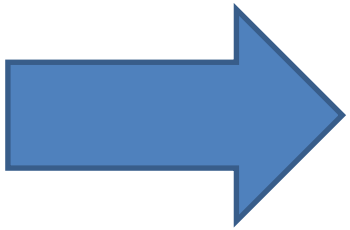
Data and dissertations

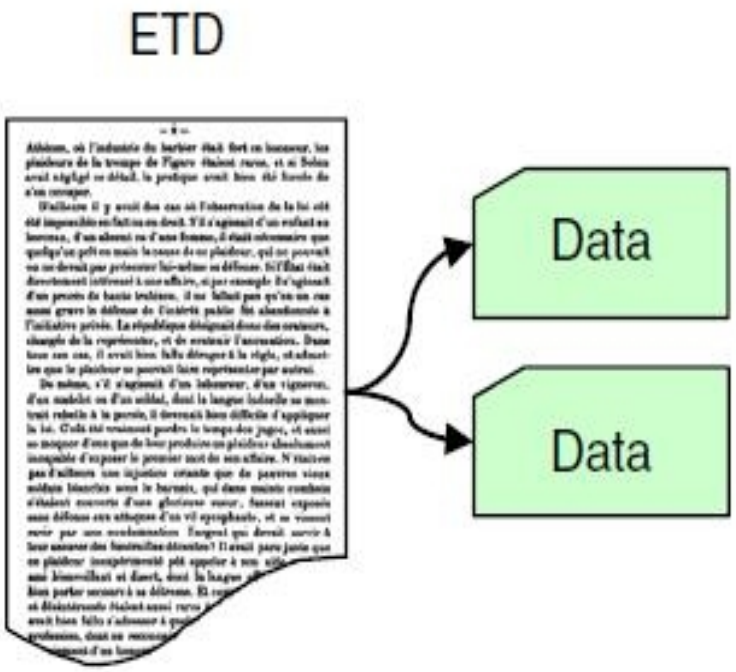
Recommendations

Recommandations



ETDs and data

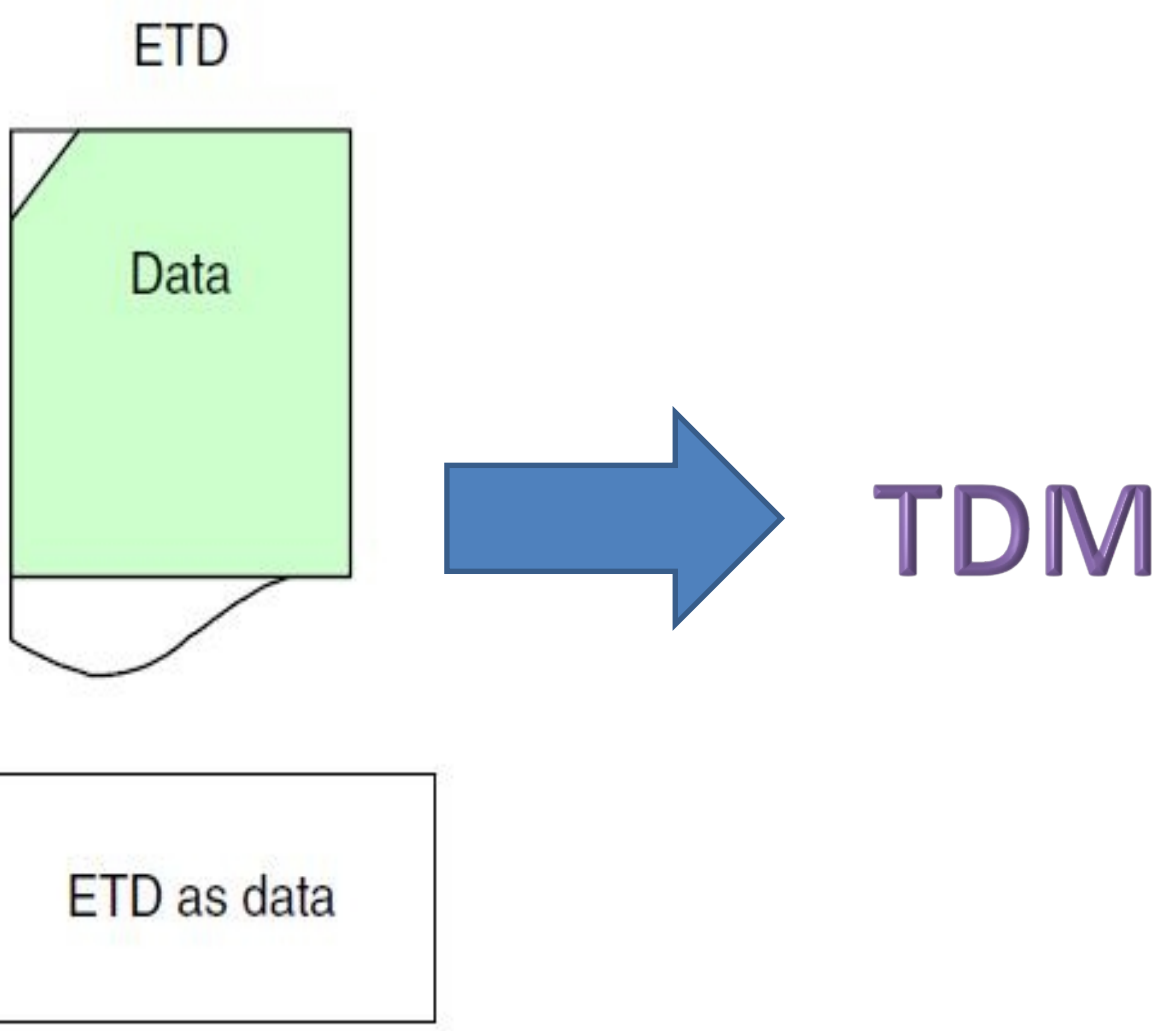




ETD as data vehicle

ETD as gateway to data

ETDs and data



Helping PhD students to manage data

Services

- Seminars, conferences, training
- Online resources (guidelines, FAQ...)
- Alert service

Education

- Legal and technical help (data management plan)
- Technical assistance for deposit
- Liaison with laboratories

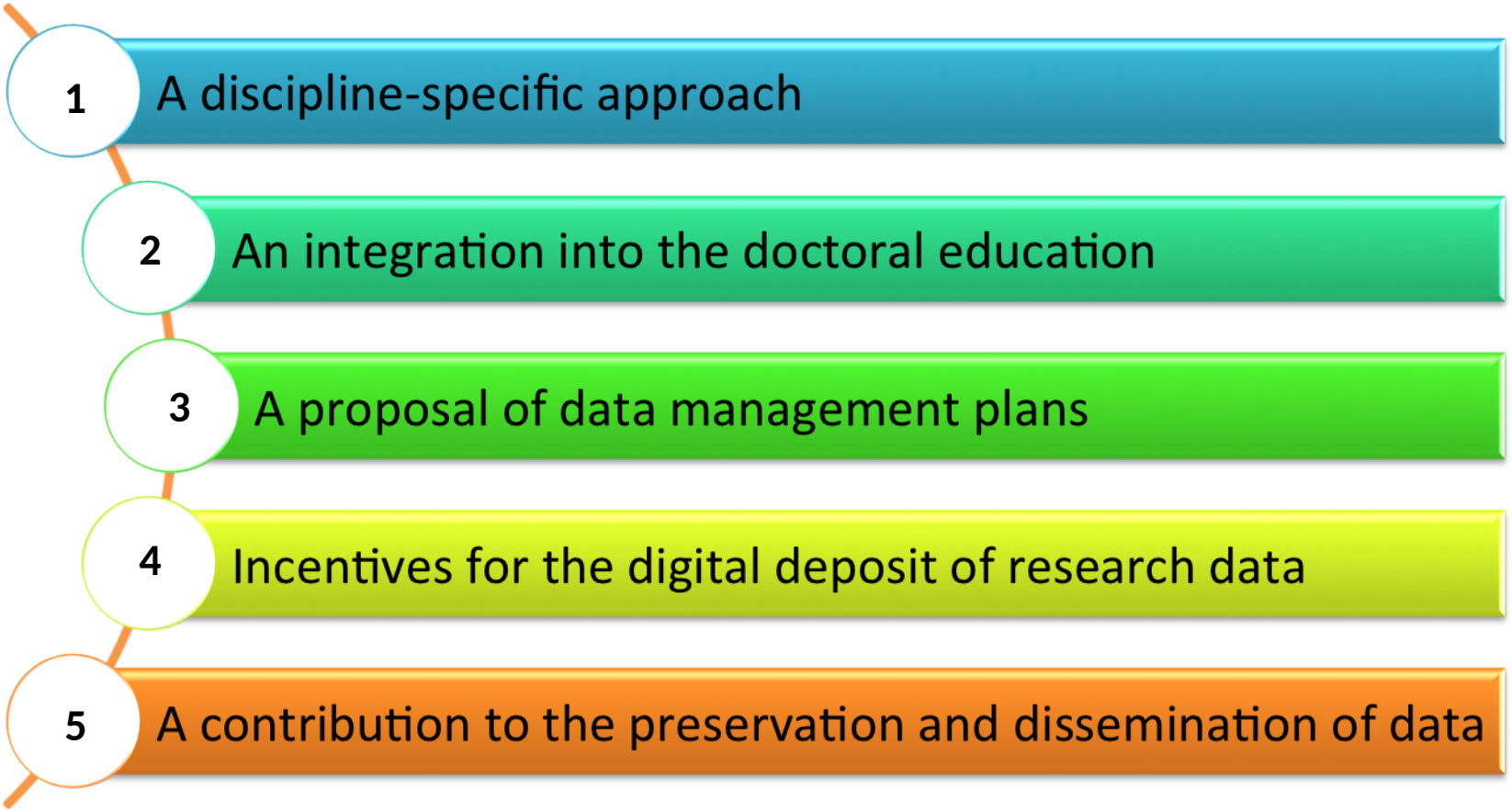
Advice, assistance

- Mediation for deposit
- Partnership with networks and repositories
- Development of tools on the campus

Infrastructures

Data management at university

Principles



Thank you !

Contact
joachim.schopfel@univ-lille3.fr
primoz.juznic@ff.uni-lj.si

References
http://www.citeulike.org/user/Schopfel/tag/gl17