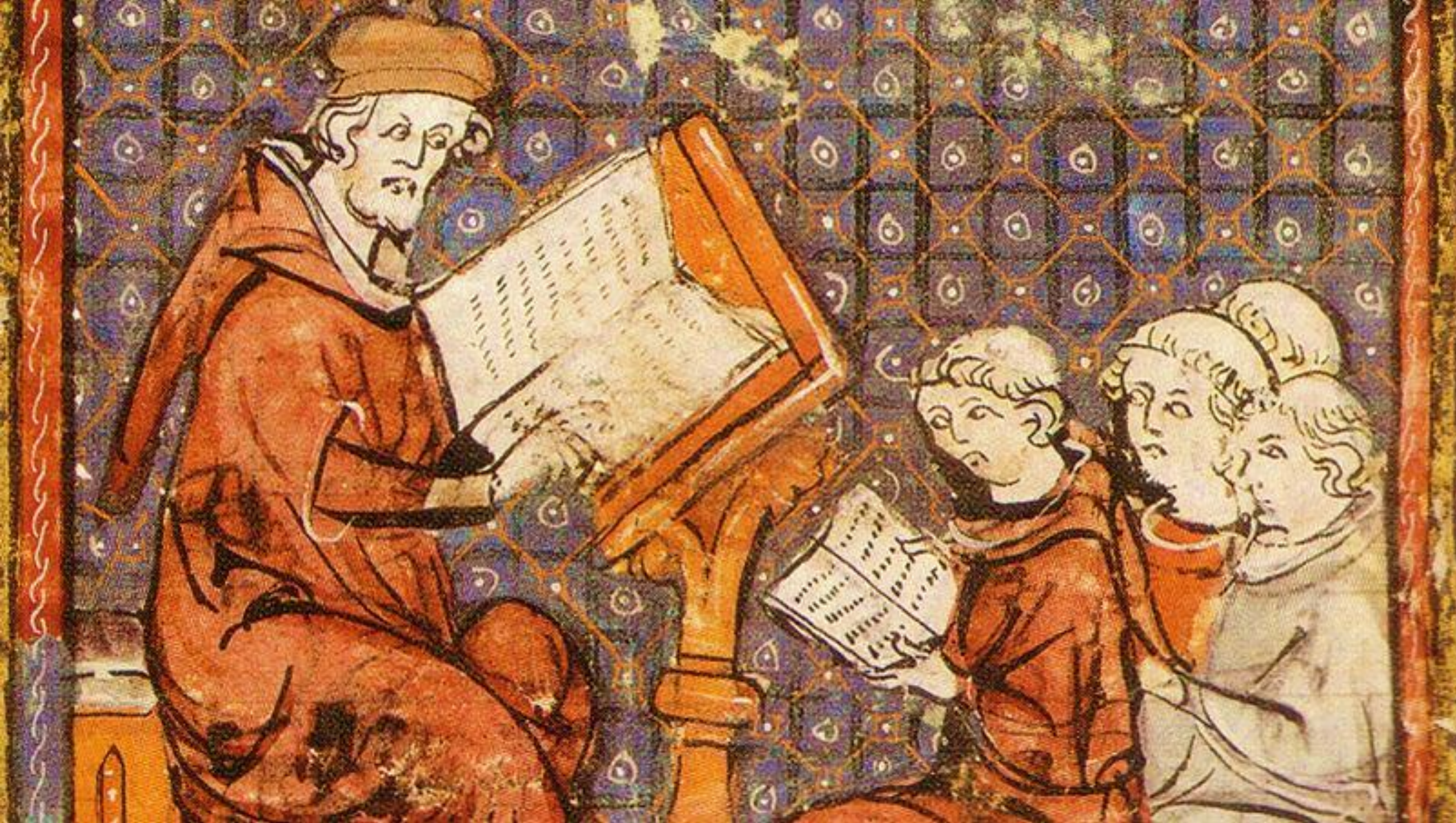


Dissertations as Data

Joachim Schöpfel, Eric Kergosien, Stéphane Chaudiron, Bernard Jacquemin
GERiICO laboratory, University of Lille (France)



11101111 00001000 01001011 11011110 00010000 10010111

00101111 01111000 01000010 01011110 11110000 01000110

10110111 00000011 01111110 10100001 00011100 01100100

11001111 11111110 11100011 11000001 00001011 11011011

11101111 10010010 10010101 11011111 00100100 00101001

00110000 01010100 01010011 11101100 01100111 10111010

11101111 00010111 11010111 01001011 10111010 00011111

10010110 00001010 00010000 10011000 00111011 00100000

01101101 00000011 11001100 11001100 11001110 01111001

10001101 01100001 01011010 01010010 10111011 10010101

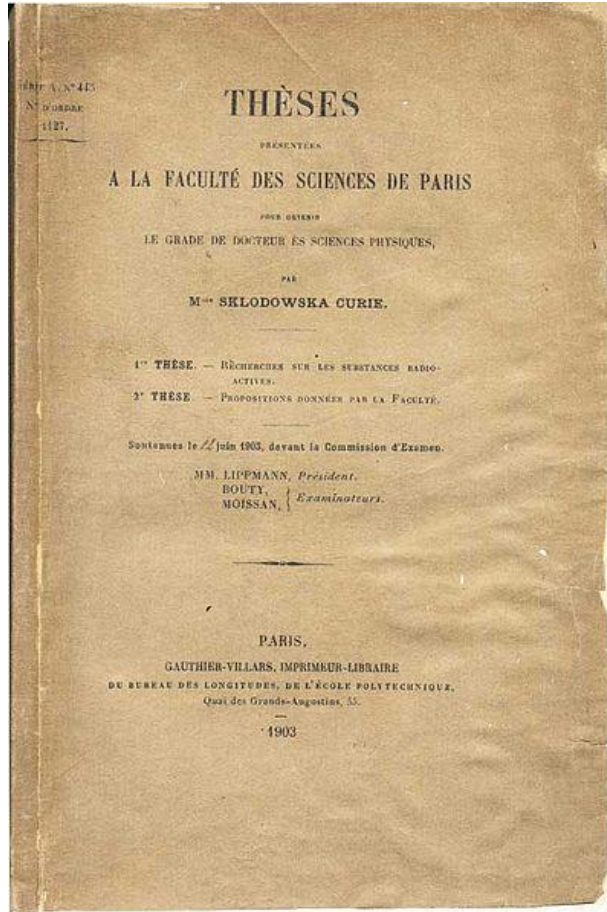
Overview

- Conceptual approach and empirical application
- About dissertations
- Dissertations and data
- Dissertations as data
 - Examples
 - Barriers
- Concluding remarks

About dissertations

Three ways to describe a dissertation

Dissertation as a document



document

form

sign

medium

Source: RTD-DOC, CNRS 2003-2006

Dissertation as a form

= an object whose materiality allows the users to use it and to manipulate it

- Format
- Media
- Structure
- Figures
- References
- Annexes, other material



Dissertation as a

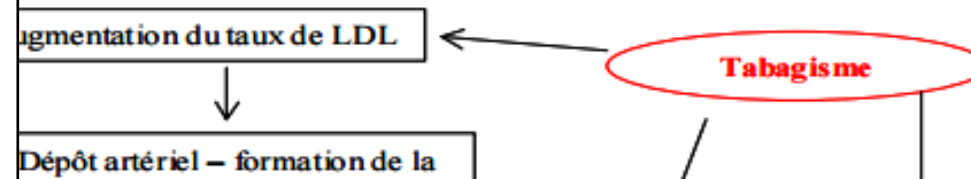
sign

= *content, a semantic network of information, part of a collection*

- Terminology
- Semantics
- Relevance
- Context of interpretation
- Knowledge

tabagisme, la dyslipidémie et l'hypertension artérielle.

Parmi les facteurs de risque réversibles, le **tabagisme** est le plus fortement corrélé avec le risque d'IDM [Yusuf *et al.* 2004]. Cette corrélation est d'autant plus importante que le sujet est jeune puisque 58% des IDM chez les hommes de moins de 40 ans sont imputables au tabagisme [Cambou 2008]. Le tabagisme passif augmente également le risque cardiovasculaire [Giannuzzi *et al.* 2005]. Sur le plan physiopathologique, la consommation de cigarettes va entraîner une augmentation du taux de LDL, ce qui induit une réponse inflammatoire qui augmente le risque de rupture de la plaque d'athérome, qui peut conduire à la thrombose [Ambrose *et al.* 2004]. Par ailleurs, le tabagisme induit une dysfonction endothéliale qui a pour conséquence une diminution du flux sanguin en augmentant la résistance des artères à ce flux mais également en diminuant la vasodilatation induite par l'NO [Ambrose *et al.* 2004]. L'accumulation de ces différents facteurs va augmenter le risque d'IDM (Figure 9).



Dissertation as a **medium**

= a trace of social activities, a tangible element of communication between human beings, a social phenomenon

- Institutional character (legitimacy)
 - Prescription (format, content...)
 - Legal and administrative status
 - Process (deposit, preservation, dissemination...)
- Research project (scientificity)
 - Evaluation (impact)
 - Network (altmetrics)
 - Open science (outreach)
- Creative work (authorship)
 - Intellectual property
 - Competitive strategy (career)
 - Financial interests (publication, patent...)

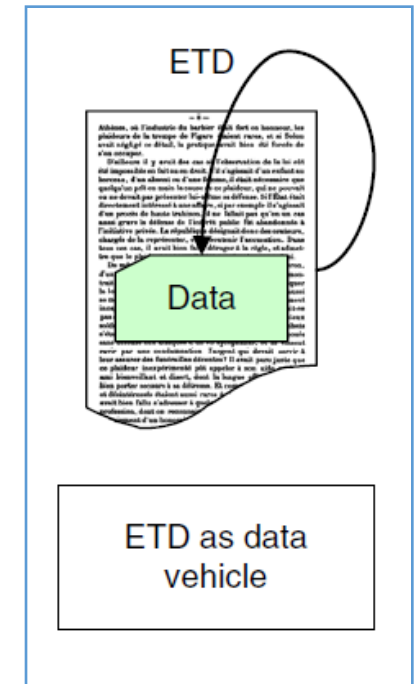


Dissertations and data

A reminder

Data vehicle

- Dissertations contain data
 - Primary data (sources)
 - Secondary data (results)
- In the text
 - Figures
 - Tables
- In appendix
 - Various formats, often non compliant with content mining
 - More or less structured
 - Sometimes on different medium
 - No identifiers
 - Metadata?



Data in PhD dissertations (survey 2015, N=780)

Annexe III : approche épigraphique

Figure 71 : Tableau des phylactères compilés par cette étude.

Texte du phylactère	Traduction	Référence biblique/Burlique	Référence date funéraire en bourgoise	Nombre d'occurrences
miserere mei Deus secundum magnam misericordiam tuam	Priez moi de moi, Seigneur, dans votre grande miséricorde	Psaume 51 (90) verset 3,4	Arcollet Mualard, 1491 ; Yabeau Molson, 1502 ; Jean Berthout, Vive sainte ; Renaud Lenoire, 1482 ; Don Calixte Maris, 1518 ; Jehan Legrand, s.d. ; Guillaume Millet, 1486 ; Thomas Gillet ; Pierre Tisserand, 1544 ; Charles de Maynot, 1515 ; Claude Mescaiger, 1476 ; Guyot Gaudin, 1471 ; Jehu de Sully ; Jean Vunry, 1480 ; Nicolas Desnos, 1554	8
miserere mei Deus / Domine	Priez moi de moi, Seigneur	Psaume 51 (90) verset 3	Pierre Tisserand, 1544 ; Catheline Fevre, 1476 ; Etienne Guartier, 1529 ; Gantier Maloube, 1502 ; Jeanne des Potots, 1477 ; Jean de Belgigny, 1501	7
O miter Dei memento mei	O mère de Dieu, souviens-toi de moi	Cantique populaire Ave Maria, Virgo Serena	Pierre Tisserand, 1544 ; Catheline Fevre, 1476 ; Etienne Guartier, 1529 ; Gantier Maloube, 1502 ; Jeanne des Potots, 1477 ; Jean de Belgigny, 1501	6
In manus tuas domine committit spiritus meus / (ibi soli peccavi)	Donne tes mains Seigneur, je remets mon âme / contre toi seul j'ai péché	Luz (100), 40	Humbert de Polley, 1468 ; Pichnot de Normand, 1502 ; Jehan Legrand, s.d. ; Jean de Beaune, 1439	4
Credo quod redemptor meus de deum	Je crois que mon rédempteur est dieu			

Databases

	DCO		NTK		NH4	
	kg	mg/L	kg	mg/L	kg	mg/L
	4660553	47	365217	26	203	
	4535603	31	347360	18	200	
	4814508	38	369005	23	219	
	14010664		1081581		6231	
	4670221	39	360527	22	207	
	139703	8	11560	4	104	
	4814508	47	369005	26	219	

Experimental results

si vous êtes retraité(e), celle du chef de famille si vous êtes étudiant(e), celle du (de la) conjoint(e) si vous êtes inactif (inactive).

Agriculteur Commerçant, artisan

Cadre, Profession intellectuelle supérieure

Profession intermédiaire Employé Ouvrier

2. La case correspondant à votre catégorie socio-professionnelle est :

Classe dite populaire Classe dite moyenne Classe dite supérieure

3. Quel est votre sexe ?

Masculin Féminin

4. Quel âge avez-vous ?

Moins de 25 ans De 25 à 45 ans De 46 à 60 ans Plus de 60 ans

5. Portez-vous du parfum ?

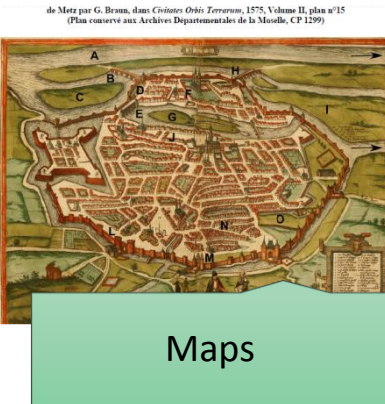
Non Oui

6. Si e

Surveys

```
0,2634.60,100.00,595.20,104215.60,111000.00,
ral load where vl[i] is the log viral load :
the log viral load for the second sample, ei
aset1_raw.csv",header=T)
set1_Boolean.csv",header=T)
rm weight is added so that the rest of ra
fu
```

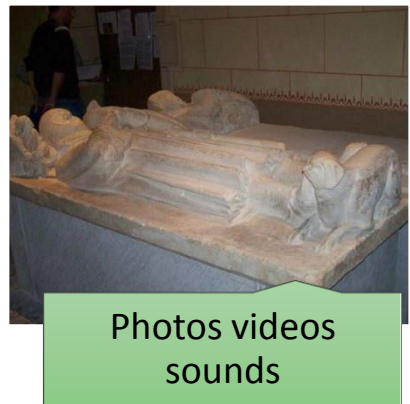
Codes



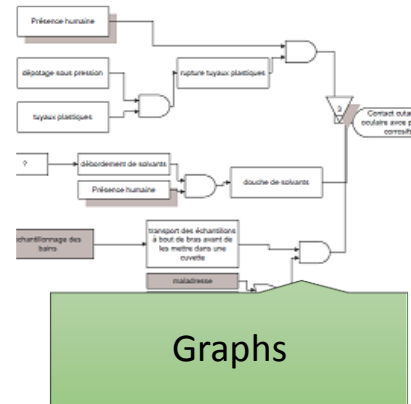
Maps



Archives

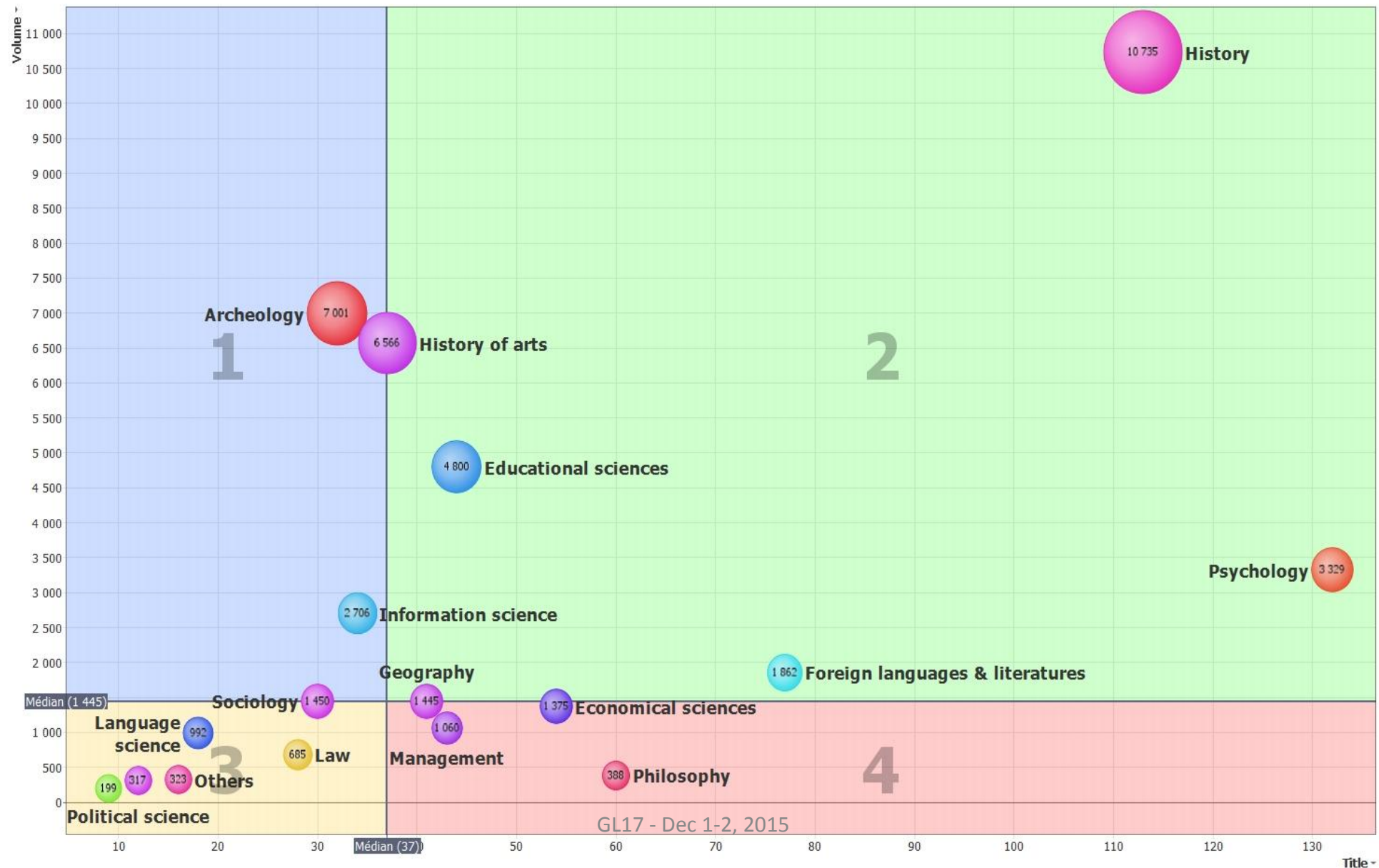


Photos videos sounds



Graphs

Data in PhD dissertations (survey 2015, N=780)



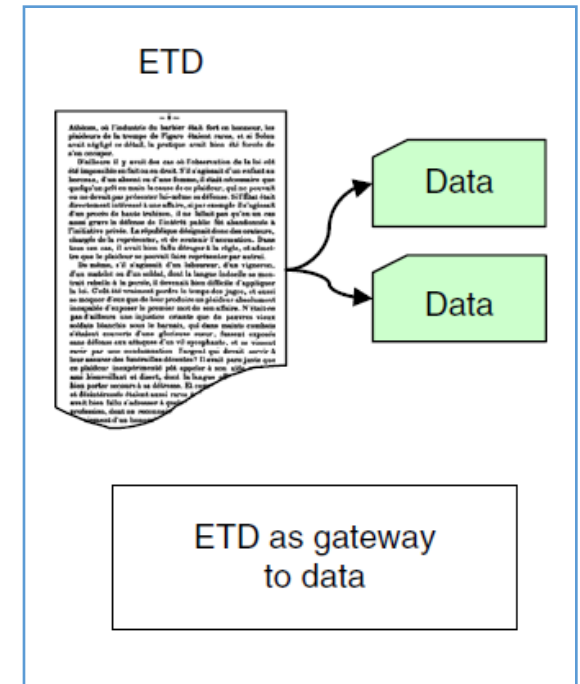
Data in PhD dissertations (survey 2015, N_{Total} 780)

Y: Domain	Databases	Graphs - figures	Images - drawings	Maps	Others	Photographs	Statistics	Tables	Texts	Timelines	Tous
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

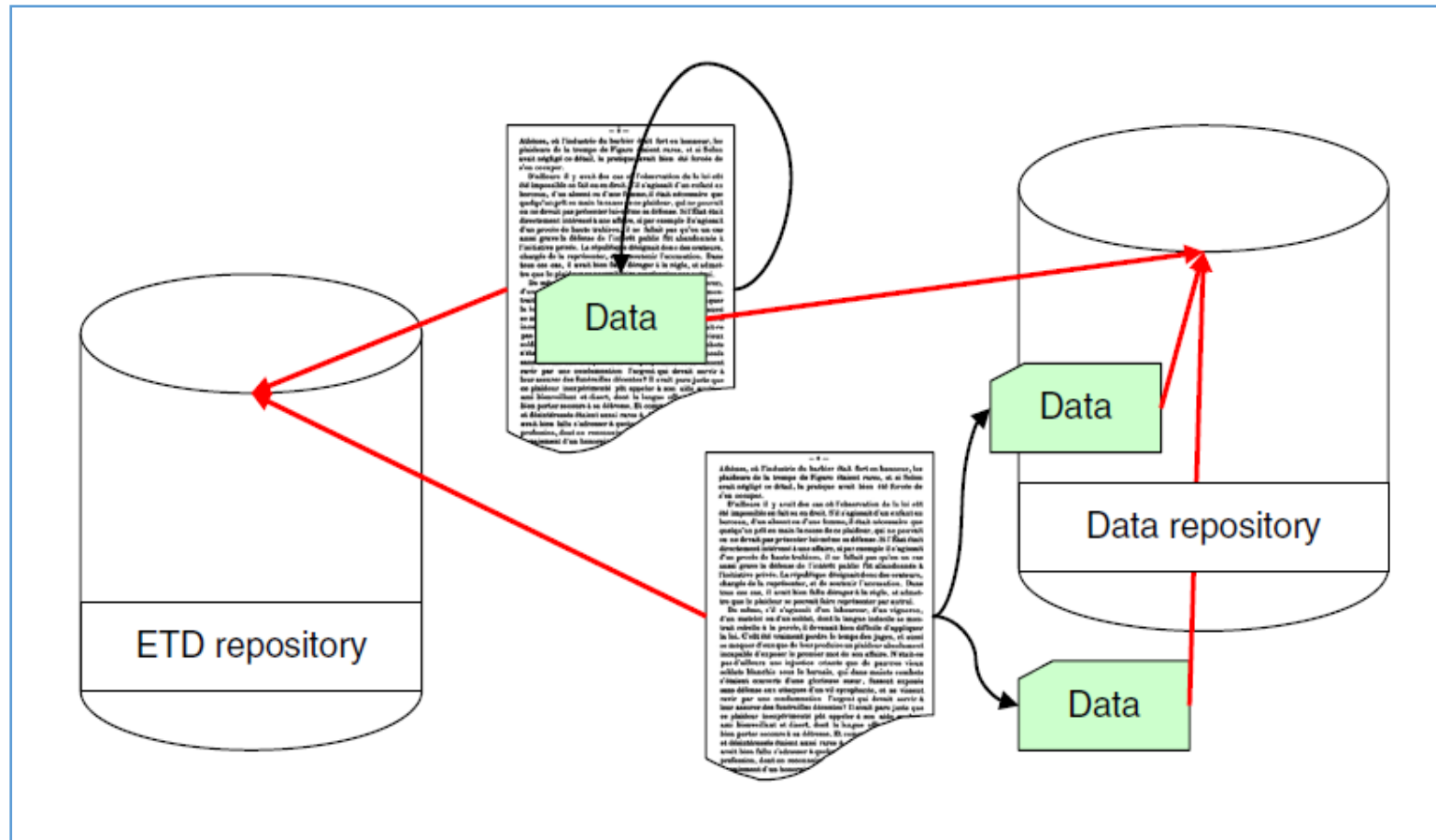
GL17 - Dec 1-2, 2015

Gateway to data

- Dissertations (can) link to data
 - Primary data (sources)
 - Secondary data (results)
- Where is the data?
 - In data repositories
 - In institutional repositories
 - In social networks, ...
- Format
 - Initial (production), preservation, dissemination
 - More or less structured
- Metadata
 - Identifiers (DOI)
 - Generic v disciplinary



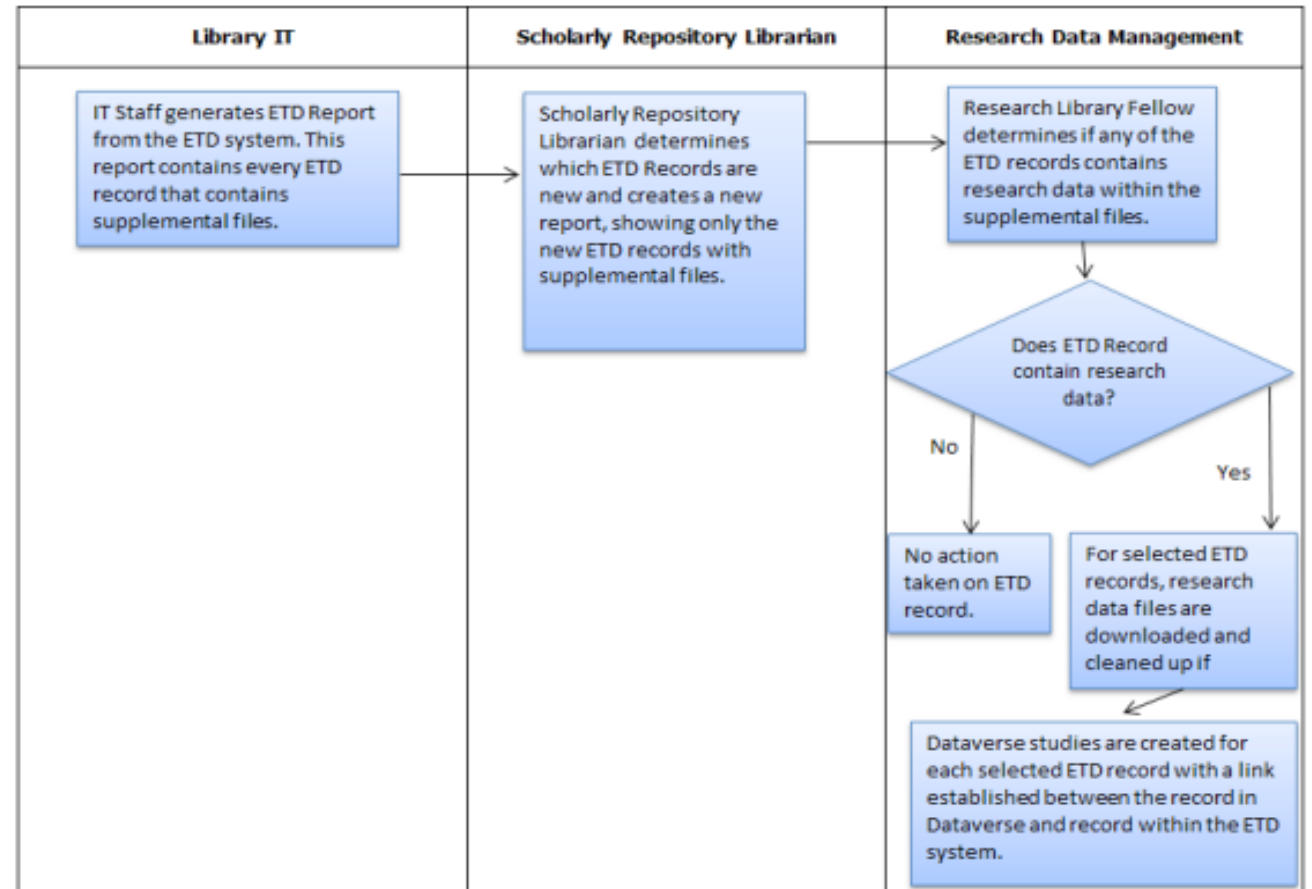
Separation of text and data



Workflow – Dataverse ETD Pilot Program

- Metadata
- Documentation
- Identifier
- Link

Source: Dataverse ETD Pilot Program at Emory (Doty et al. 2015)



Workflow – ETDplus project (Educopia Institute)

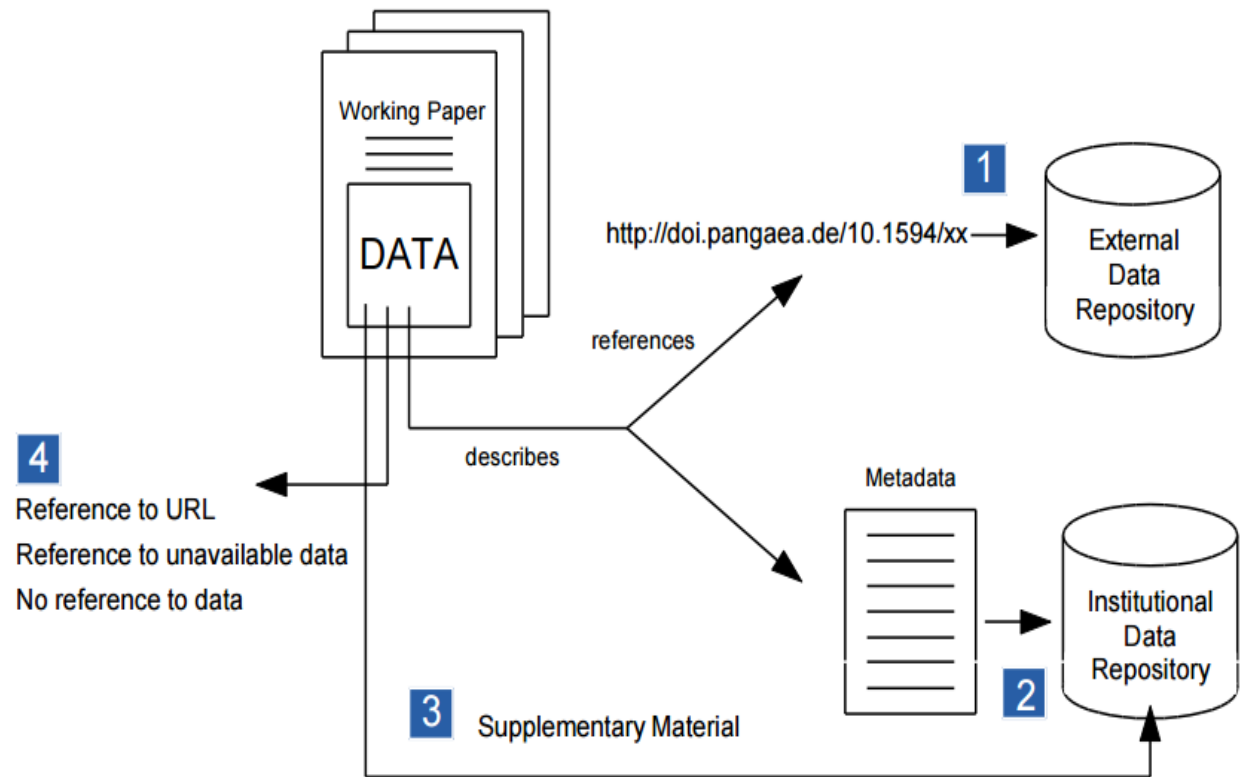
- *Separation of the dissertation text files and the related “complex content objects” whenever possible.*
- *“Embedding multimedia components within the full text might seem advantageous in that they would then be inseparable. However, when the time comes that it is necessary to migrate either the full text itself or one of the multimedia components, having separate files would greatly simplify matters” (p.5-9).*
- *Metadata and persistent identifiers like handle, PURL, ARK or DOI are supposed to provide the “glue” that binds together the text, multimedia and data files.*
- *Preservation-worthy research data (survey data, measurements, laboratory notebooks, measured spectra etc.) might be stored as part of an “ETD package” or, after transformation into archival files, in a separate data repository, and “links to the data repository from the ETD metadata would then enable researchers to access these research data in the future” (p5-11).*

Source: Guidance Documents for Lifecycle Management of ETDs (Schultz et al. 2014)

Workflow – Bielefeld PUB

- For working papers but also for ETDs
- Research data as a discrete resource
- Attribution of DOI (DataCite)
- Distinction between institutional and disciplinary data repositories

Source: Bielefeld PUB (Schirrwagen & Vompras 2015)

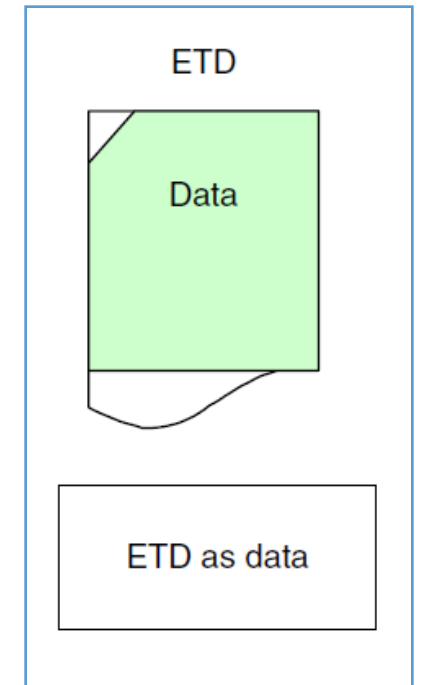


Dissertations as data

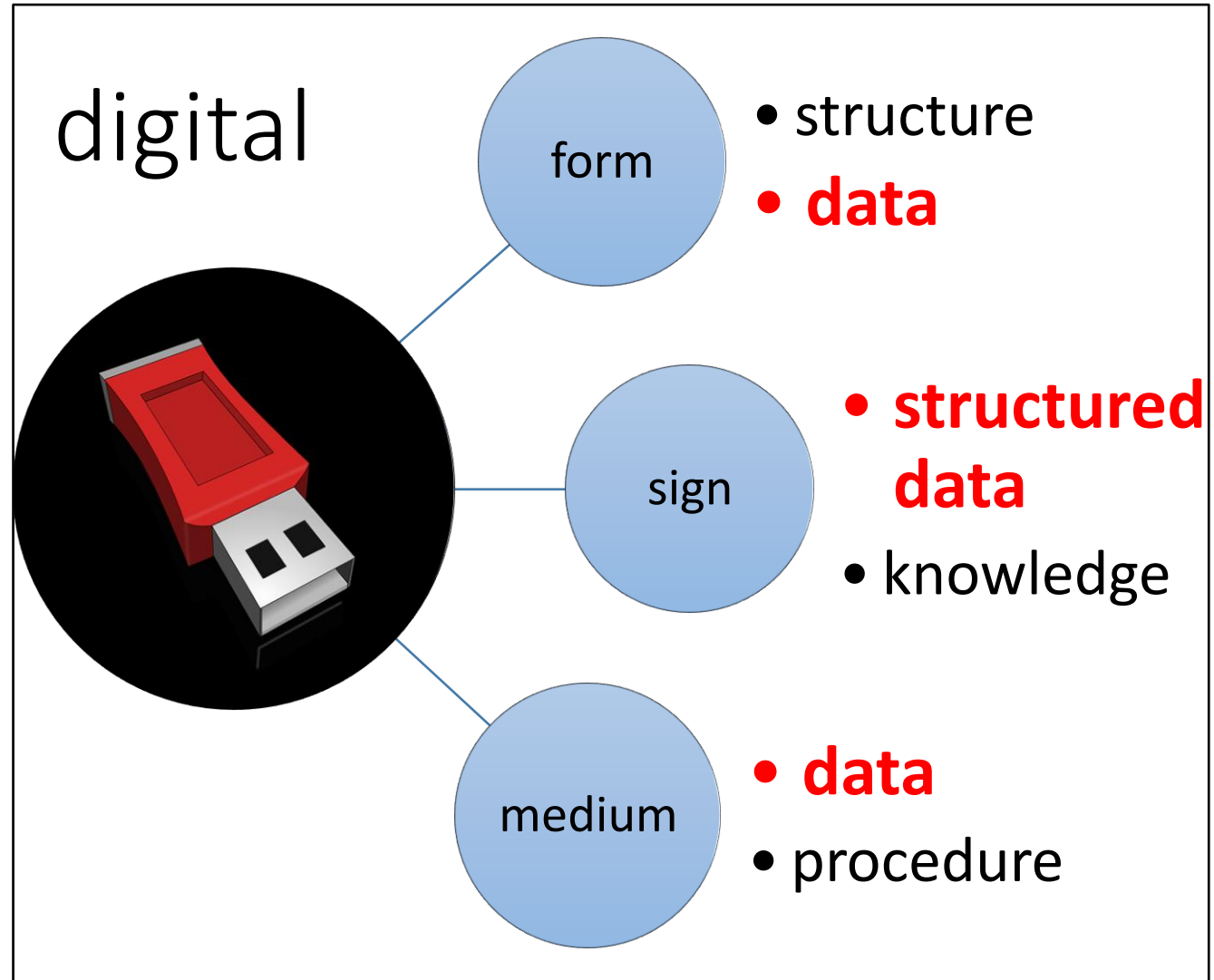
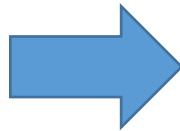
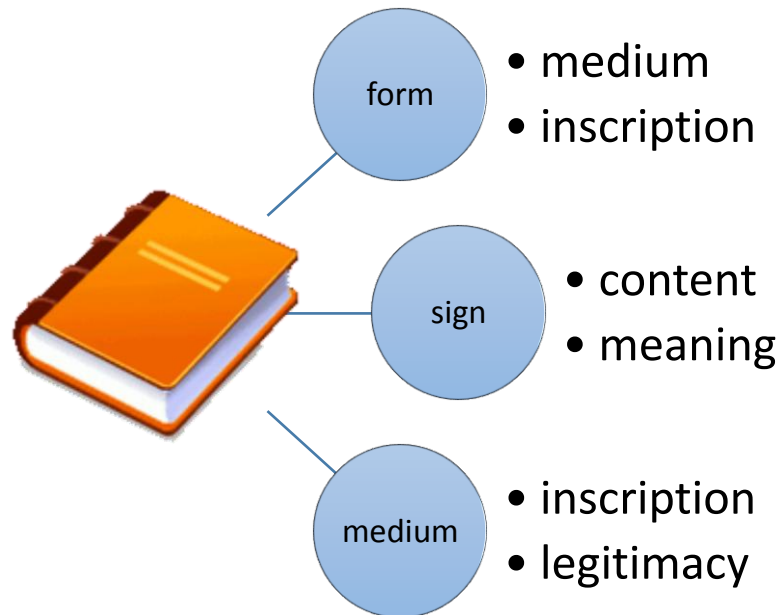
Another way to see and handle dissertations

Data

- No distinction between text and data
- The whole dissertation is data
 - set of values of qualitative or quantitative variables
 - some existing information or knowledge represented in some form suitable for better usage or processing
 - can be measured, collected and reported, and analyzed, to create information suitable for making decisions
- Distinction between data and metadata
 - Yet, metadata is data, too



From print to ...



ETD as a

form

= *structure + data; allows for information extraction and filtering*

- Format: XML (syntax)
- Media: Web
- Structure: DTD
- Text encoding: TEI (grammar)
- Resource Description Framework
- Metadata

```
<!ELEMENT submission - O
  ( #PCDATA | em | strong | tt | q | term | foreign |
    in | link | target | a | sup | sub | br )* >
```

3.4. Submission Type

The element type **submission type** (*submission*) indicates that the contained text is the type of submission.

```
<!ELEMENT submission - O
  ( #PCDATA | em | strong | tt | q | term | foreign |
    in | link | target | a | sup | sub | br )* >
```

3.5. School

When it appears in the front matter, the element type **school** (*school*) indicates that the contained text is the name of the school.

This element type may also appear in a bibliographic citation. In that instance, the contained text is the name of the school.

```
<!ELEMENT school - O
  ( #PCDATA | em | strong | tt | q | term | foreign |
    in | link | target | a | sup | sub | br )* >
```

3.6. Degree

The element type **degree** (*degree*) indicates that the contained text is the name of the degree the author received.

```
<!ELEMENT degree - O
  ( #PCDATA | em | strong | tt | q | term | foreign |
```

<http://etd.vt.edu/etd-ml/dtdetds.htm>

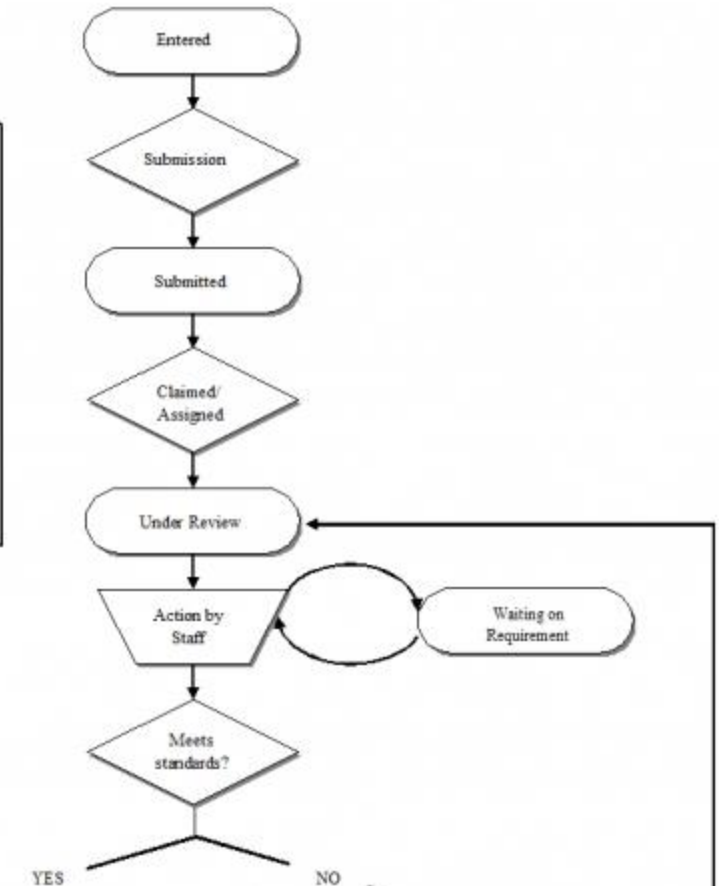
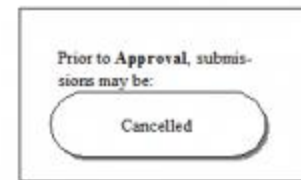
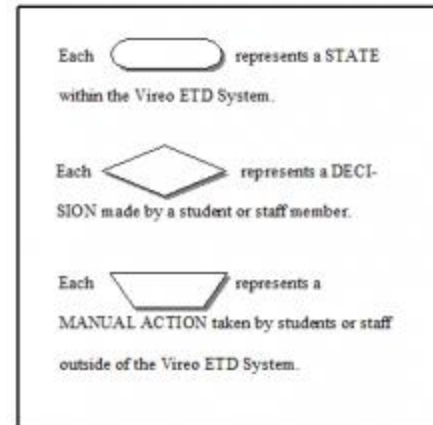
ETD as a

medium

= *data + procedure*

- Rules (prescription)
 - Format
 - Structure
- Legal status
 - IP exception
 - Licensing
 - Access restrictions

ETD Workflow/State Diagram



https://wikis.tdl.org/tdl/ETD_Workflow_Diagram

Examples

Re-use to enable new insights from text and data

- Text and data mining (TDM) or content mining of machine-readable text and data
- Extraction of information (knowledge) from large volume of content
- “The main barriers against the uptake of TDM are not technical, but primarily a lack of awareness among academics, and a skills gap. They relate to legal issues around copyright and database rights, and to some policy choices of restrictions being implemented by publishers on (for instance) access to APIs. **These problems are all soluble, but require non-technical solutions.**”

Brook, M., Murray-Rust, P., Oppenheim, C., Nov. 2014. The social, political and legal aspects of text and data mining (TDM). *D-Lib Magazine* 20 (11/12). <http://dx.doi.org/10.1045/november14-brook>

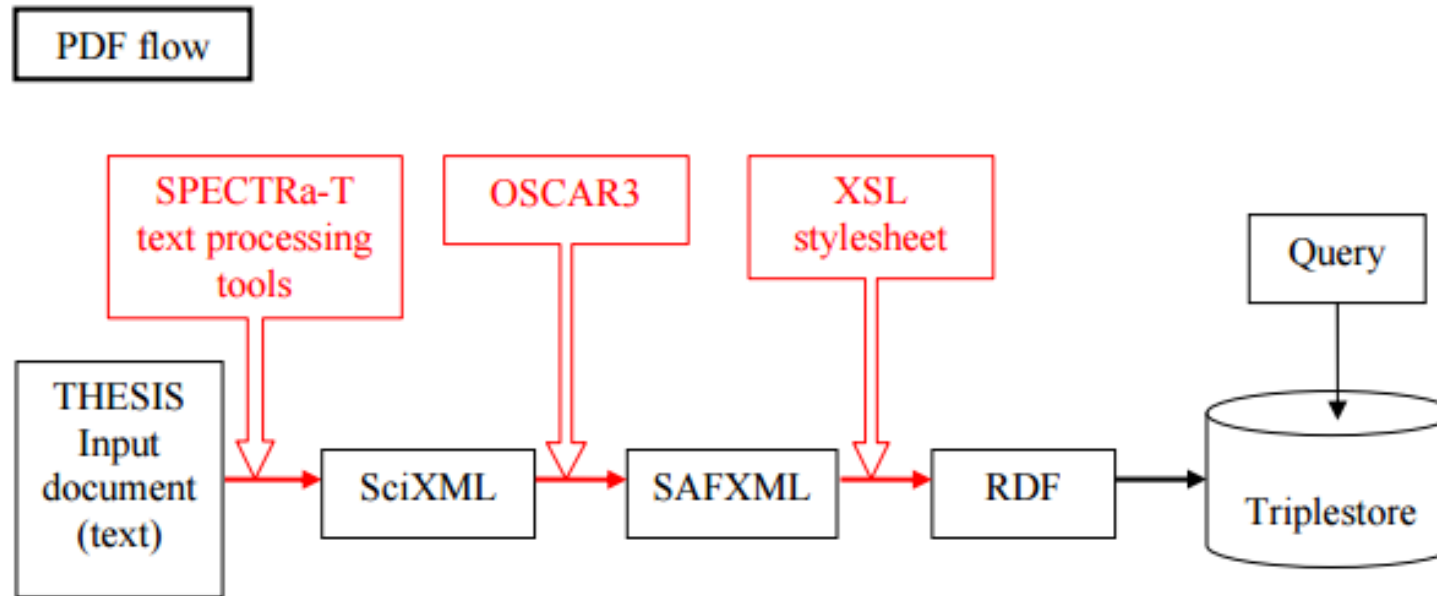
SPECTRa-T – a JISC project 2007-2008

- A proof-of-concept approach focusing on chemistry research data in molecular and related subjects
- To develop text-mining tools and processes for the automatic extraction of experimental research data (chemical objects and named chemical entities)
- To transform the extracted data into metadata and ingest them into data repositories and RDF triplestores, thus enabling RDF-based semantic querying of the contents

Morgan, P., Downing, J., Murray-Rust, P., Stewart, D., Tonge, A., Townsend, J., Rzepa, H., 2008. Extracting and re-using research data from chemistry e-theses: the SPECTRa-t project. In: *ETD 2008 11th International Symposium on Electronic Theses and Dissertations*, 4 - 7 June 2008, The Robert Gordon University, Aberdeen, UK.

SPECTRa-T workflow

Various ETD formats – LaTeX, Postscript, Adobe Portable Document Format (PDF), Microsoft Word (DOC), Microsoft Office Open XML (DOCX)



<http://www.dspace.cam.ac.uk/handle/1810/230116>

SPECTRa-T on ETDs as **form**

- “**PDF** (...) is significantly less well-suited to text-mining than is marked-up text.”
- “For data mining in chemistry, **a marked-up document file format** (such as Office Open XML) for deposited e-theses should be available in preference to PDF.”
- “Require deposit of the Word version of the thesis, and then to convert it to XML **as a standard process** (...)”

SPECTRa-T on ETDs as **sign**

- “Discipline-specific text-mining tools, together with the triplestore and data repositories into which text-mined outputs are to be deposited, will **best be managed by subject experts** operating within the same disciplinary environment.”
- “Much of the effort involved in developing text-mining tools and workflows is **unavoidably specific to individual disciplines.**”
- “Those who are responsible for designing and delivering institutional repository services **need to work closely with researchers in different subjects**, in order to understand their needs and co-ordinate departmental and institutional activities.”

SPECTRa-T on ETDs as **medium**

- “Institutions to determine **what purpose theses might serve** in each discipline beyond their immediate role in assessing a student's research (...)”
- “In the sciences theses should be regarded not simply as objects destined for preservation and archival status, but rather as **unique resources containing potentially valuable data** that must be made extractable and re-usable.”
- “Those authorities with responsibility for overseeing the writing, submission and subsequent management of theses (should) **implement appropriate strategies**, applying both institution-wide and discipline-specific policies, guidelines and practices designed to ensure optimal conditions for re-use. In other words, **institutional regulations** should ensure that each thesis is fit for purpose.”

Lille project

TERRE-ISTEX project for geographic indexing and analysis
from scientific corpora

Thematic project (funded by ISTEEX 2016 – 2017)

Gérico

STL
savoirs
langage
extes



ISTEX
L'excellence documentaire pour tous

Lille : ISTEEX project

- 1 objective: To offer an unique and exceptional concentration of scientific documentation archives
- 2 parts: acquisition of scientific digital archives, creation of the ISTEEX platform
- 3 level of services: basic services, value added services, thematic projects
- 5 partners (+): CNRS, ABES, Couperin, MESR, Université de Lorraine for engineering CPU (+ R&D)
- Budget: 60m € (« Investissements d'Avenir »)
- Duration: 3 years
- <http://www.istex.fr/>



Lille : Issues

- To study what happens on a territory on the thematic « climate change » from heterogeneous scientific data (scientific articles from CIRAD and ISTEEX, PHD Thesis)
- To give an help to Scientists to do their state-of-the-art (on a territory, on a thematic...)

Questions :

- What is a **territory** ? → A set of geographical information
 - Geographical information = spatial entity (where) + thematic entity (what) + temporal entity (when)
 - *Example : a study on climate changes realized in the south of Madagascar in 1981.*
- How to identify this kind of information? → **text mining methods in abstract and content, experts validation**

A use case for research information:

- What are the scientific documents related to Madagascar and Sénégal?
- In these documents, what are the studied thematics and the associated authors?
 - Development of a Geographical Research Tool based on Elastic Search
 - Spatial visualization
 - the results analysis is made by experts

Lille : Data

CIRAD data

Multilingual data Données coming from the Agrotrop database: CIRAD open archives

92 000 references and 25 000 complete documents: scientific publications and grey literature (reports, etc.)

3000 documents related to the thematic « climate change »

Metadata : title, author, abstract, indexed labels coming from lexicons (**AGROVOC thesaurus** and **Agris/Caris from the FAO**). Thèmes Agris : <https://agritrop.cirad.fr/view/subjects/>, geographical information (countries)

PHD Thesis data (ANRT)

200,000 PhD theses, metadata, SUDOC link (ABES).

70,000 digital documents

ABES: RAMEAU thesaurus from the BnF for label description

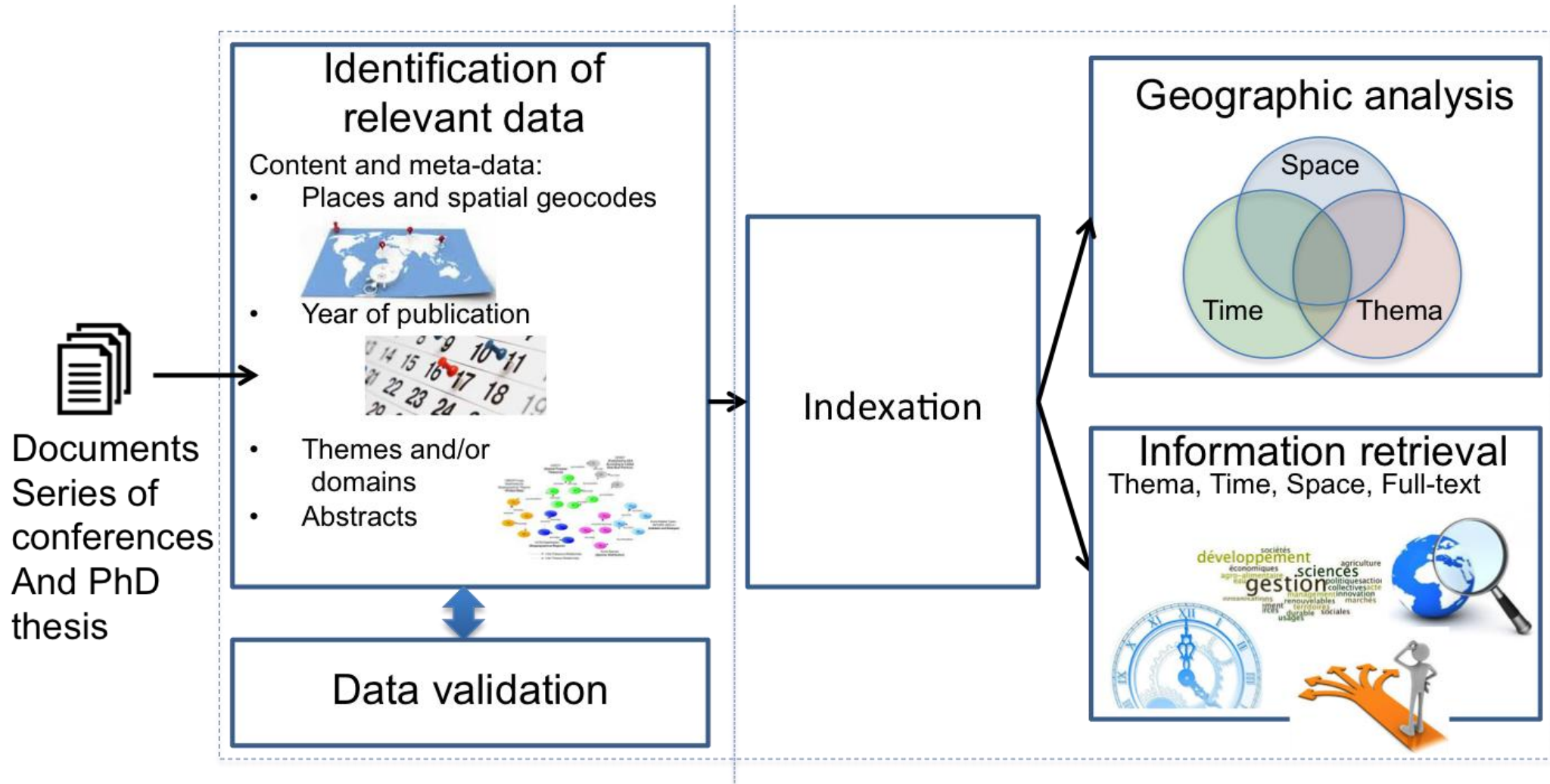
400 PhD theses related to the thematic « climate changes »

ISTEX Data

8,190,000 scientific articles (Emerald, Wiley, Elsevier, etc.) related to the thematic « climate change » (10,000 are in french)

Metadata into MODS (XML format)

Lille : General process



Lille : First work (proof of concepts)

- Geographical analysis on the EGC conference publications (1200 articles): Indexing, Research information, Analysis

- Build a first geographic index (MODS model for metadata)

Information thématique	Nom ville conférence Noms villes auteurs Noms auteurs Titre article Résumé article Session Domaine
Information spatiale	Coordonnées villes auteurs Coordonnées ville conférence
Information temporelle	Année conférence
Information plein texte	Termes titre article Termes résumé article

- Implementation of a Multicriteria Research Information tool based on Elastic Search

Lille



: First work (proof of concepts)

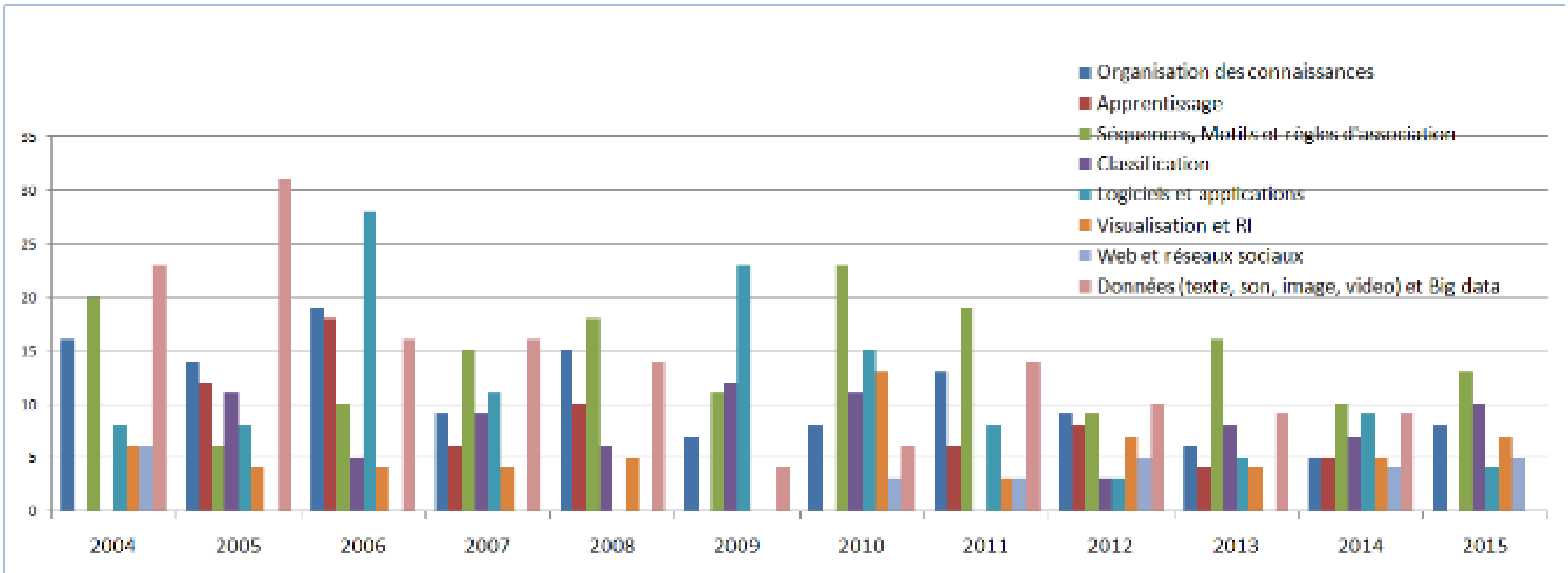
The sentiAnnotator viewer for the extraction process and experts validation (ISWC, october 2015)

- ✓ Upload of Big corpora (french and english)
- ✓ NLP process to extract features
- ✓ Experts can (un)validate marked data
- ✓ Possibility to download results

<http://siso.teledetection.fr/>

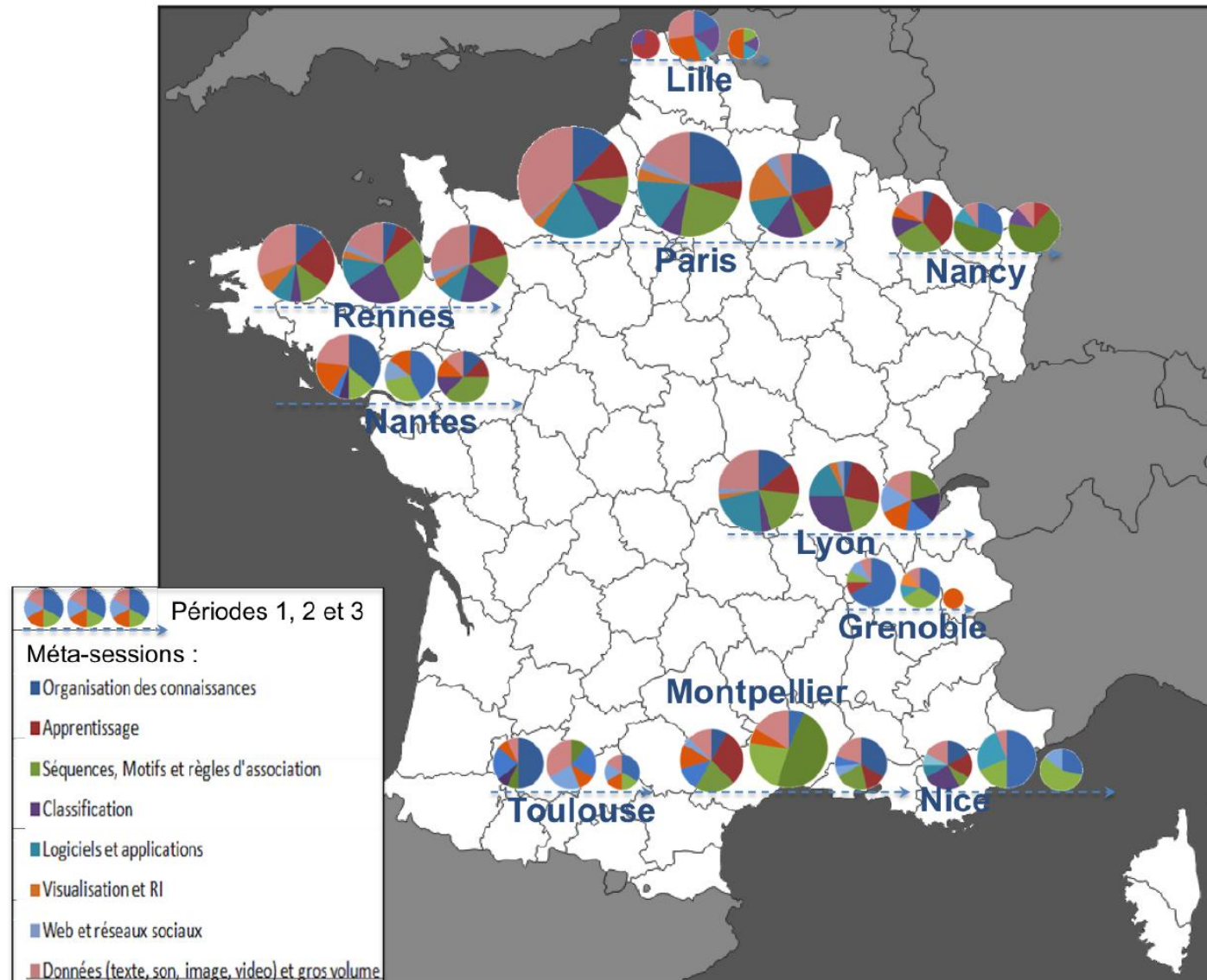
Lille : First work (proof of concepts)

- Geographical analysis on the EGC conference publications: Indexing, Research information, Analysis
 - A set of geographical analysis available (in french language) : Elastic Search and Kibana tool
<http://ekergosien.net/DefiEGC/index.html>



Lille : First work (proof of concepts)

- A set of geographical analysis available : Elastic Search and Kibana tool

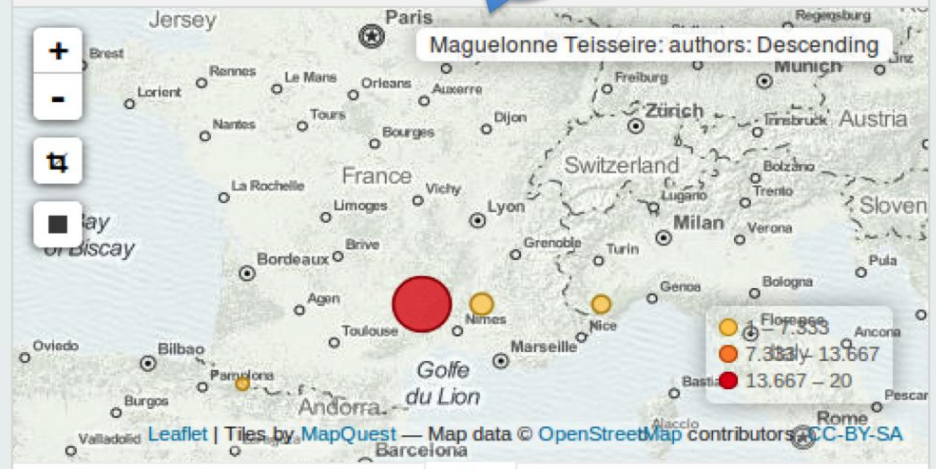


1auteur&Coaut&annee

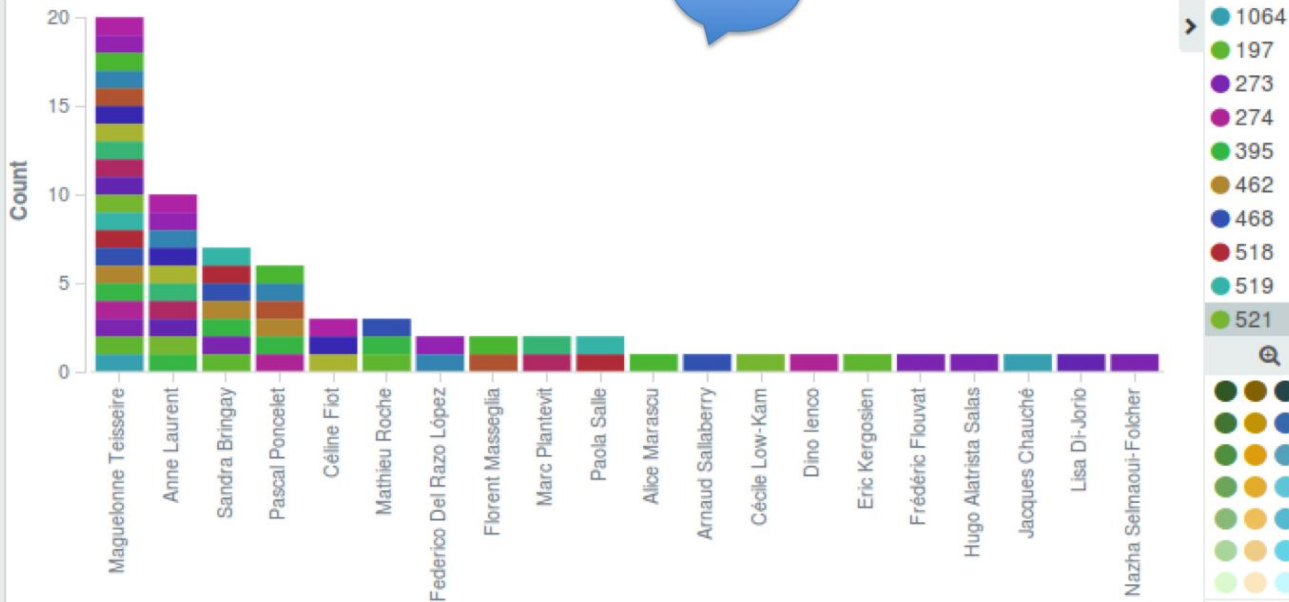


authors: Descending

GeoVeteran

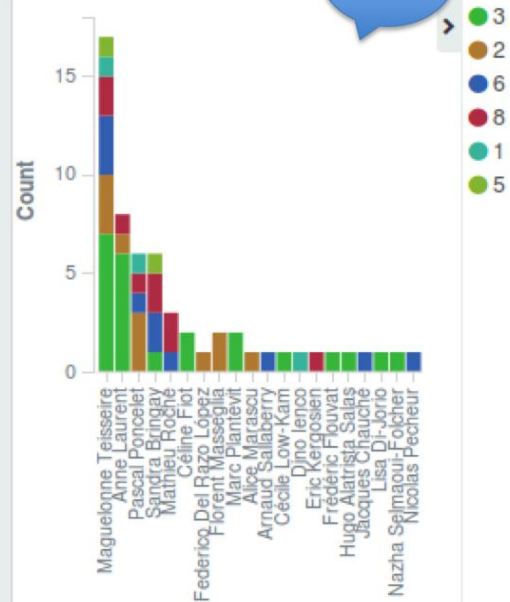


1auteur&Coaut&idart



authors: Descending

1auteur&Coaut&ms



authors: Descending

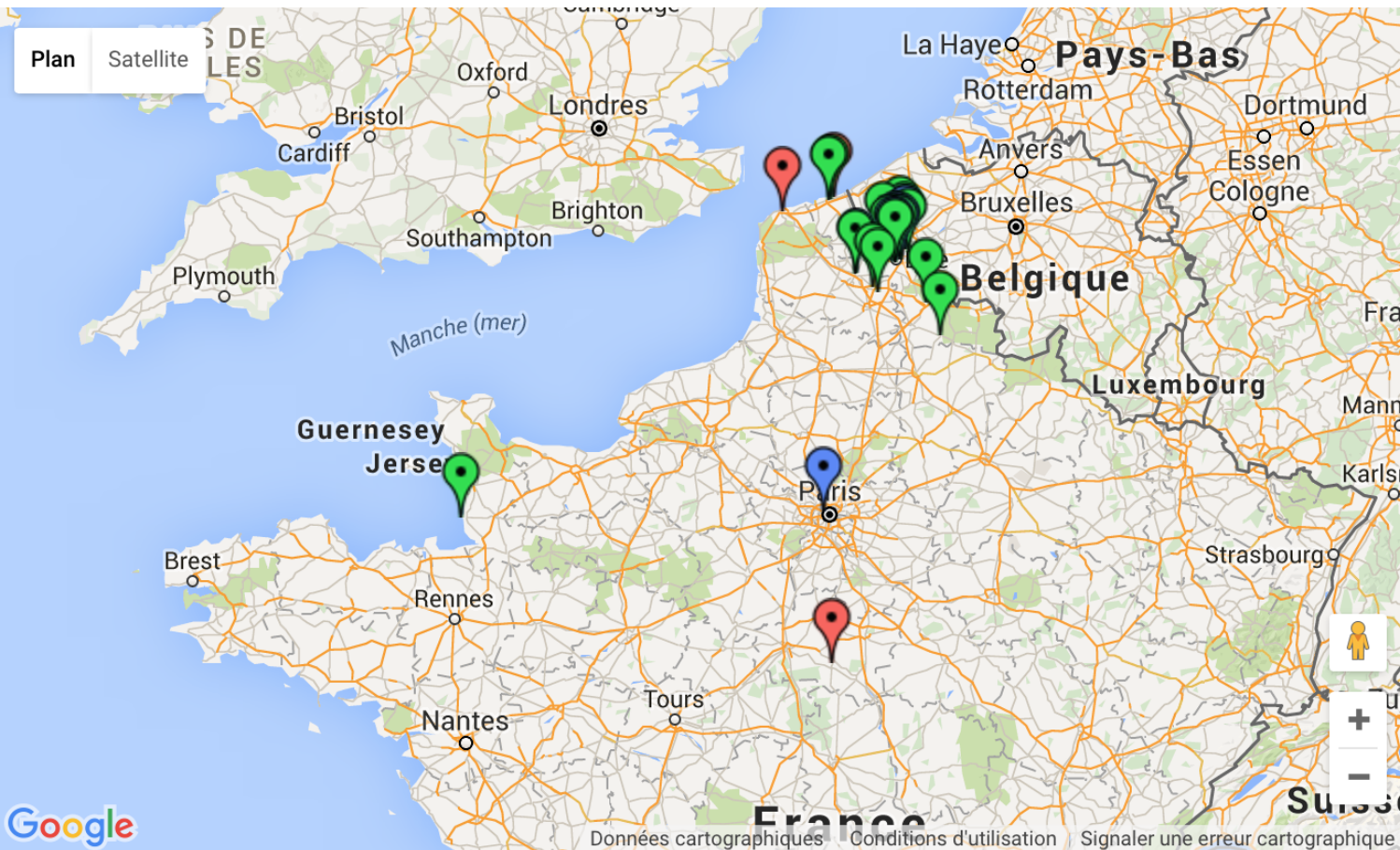









Prospects: « big data », Valorization

- « Big data »
 - Heterogeneous data from ISTEEX, PhD theses and Cirad corpus :
→ Normalization of metadata with the MODS model (XML)
 - Volume:
→ Test the scalability of our text mining application on bigger corpora
 - Tests on other thematics (history, technology, culture...)
 - Integrate other resources (web, bases de données, archives ouvertes, plateformes...)

Master and BA dissertations

- Indexing students dissertations (for Lille 3 students: thousands of dissertations)
- Implementation of a dynamical geographical information research tool

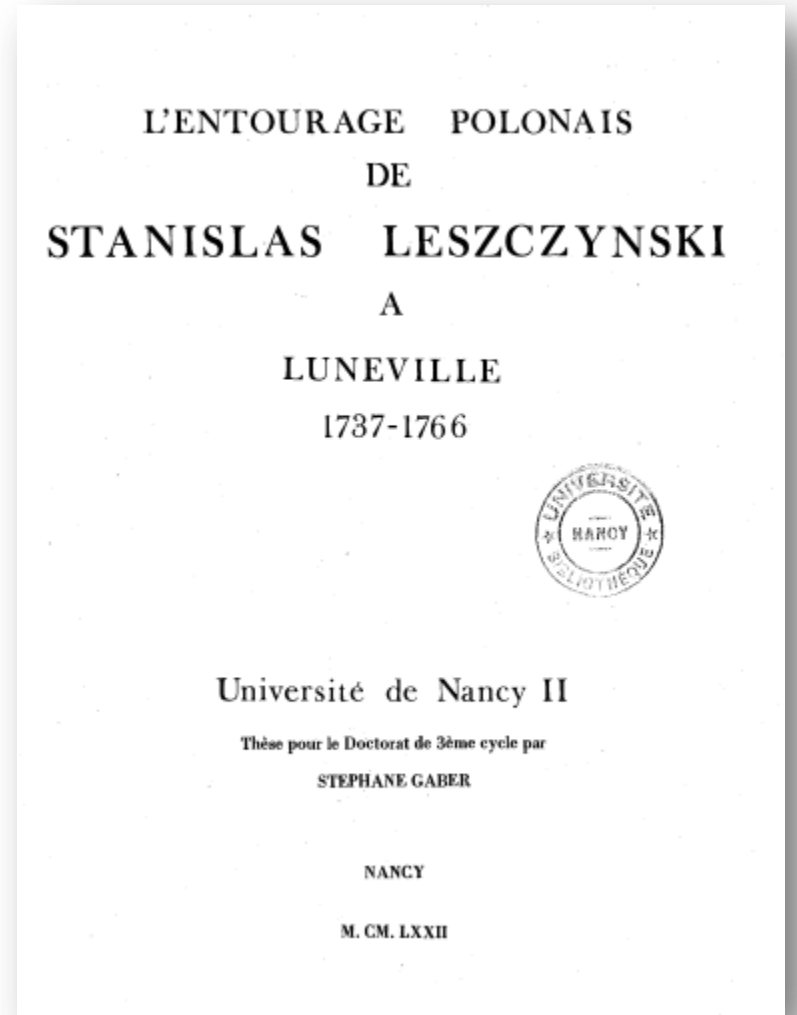


Type	Categorie	Nombre de stage	Marqueur
Médiathèque	public	27	
Bibliothèque	public	7	
Centre de ressources documentaires	autres	4	
Réseau de lecture publique	associatif	3	
Learning Center	autres	1	
CDI	autres	1	
BCD	autres	1	

Barriers

Availability

- One part of dissertations is not digital
- Especially older dissertations
- Digitization of print dissertations: problem with OCR (see SPECTRa-T)



http://docnum.univ-lorraine.fr/public/BUL_T_1972_0006_GABER.pdf

Accessibility

- One part of dissertations is not online
 - And/or not (easily) identifiable
- Some (few) are confidential
- More are embargoed (publishing)
- Others are restricted to on-campus access only
- Social networks, private websites...

Institution	Sample	Open Access	Restricted Access	On Campus	Embargo
ProQuest (US) ⁶	500,000+	(95%)	5%	<i>n/a</i>	<i>n/a</i>
Texas, (Austin)	11,539	92%	8%	8%	<i>n/a</i>
PUC Rio de Janeiro (Brazil)	2,787	89%	11%	<i>n/a</i>	11%
West Virginia	4,600+	85%	15%	15%	0.3%
Florida State	3,709	84%	16%	<i>n/a</i>	<i>n/a</i>
ABES (France)	3,709	84%	16%	<i>n/a</i>	<i>n/a</i>
Lille 1 (France)	833	79%	21%	15.5%	5.5%
Auckland (Australia)	3,088	72%	28%	28%	<i>n/a</i>
Lorraine (France)	52	71%	29%	29%	<i>n/a</i>
Maryland	2,050	68%	32%	<i>n/a</i>	32%
Valenciennes (France)	35	63%	37%	31%	6%
Liège (Belgium)	191	57%	43%	33%	10%
Virginia Tech	20,386	54%	46%	<i>n/a</i>	46%
Amherst	<i>n/a</i>	48%	52%	32%	20%
Lille 3 (France)	124	40%	60%	<i>n/a</i>	<i>n/a</i>

<http://www.dlib.org/dlib/march15/schopfel/03schopfel.html>

Legality

- Dissertations are online but TDM is not allowed
 - No law
 - No license
- Gratis v libre
 - Harnad v Murray-Rust
- Orphan dissertations?

18. ALLOW TEXT AND DATA MINING

#tdm

“Text & data mining” (TDM) or “content mining” refers to a variety of analytical procedures to derive additional information and knowledge from a large set of publications.

EXAMPLE: *A scientist may use a computer to automatically analyse a large number of research papers to discover a previously unknown association between two diseases.*

Some rightsholders insist that this requires two separate licences, one for reading the works and another one for TDM.



A specific and mandatory exception to remove text and data mining for scientific purposes from the reach of European copyright and database law should be considered.

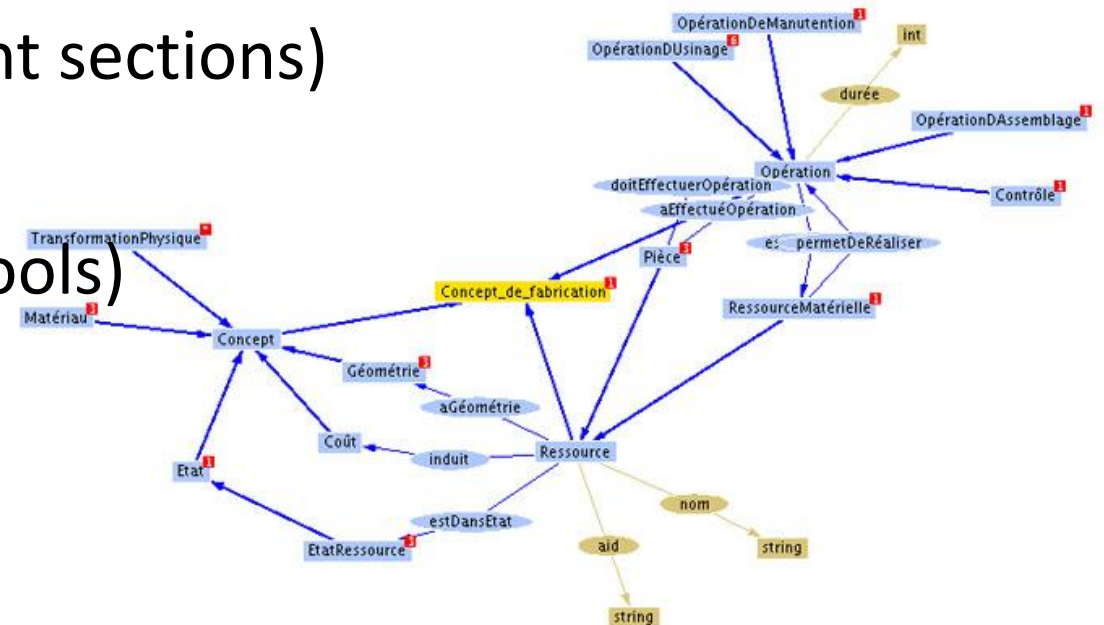
—EC TDM expert group recommendation

THE REPORT RECOMMENDS: Clarify that lawful access to data includes the right to mine it through automated analytical techniques.

<https://juliareda.eu/copyright-evaluation-report-explained/#tdm>

Feasibility

- Dissertations are online (ETDs) but TDM is not possible (optimal)
- Format (PDF processing not optimal for TDM)
- Structure (non recognition of relevant sections)
- Metadata (discipline, domain...)
- Discipline-specific knowledge (and tools)

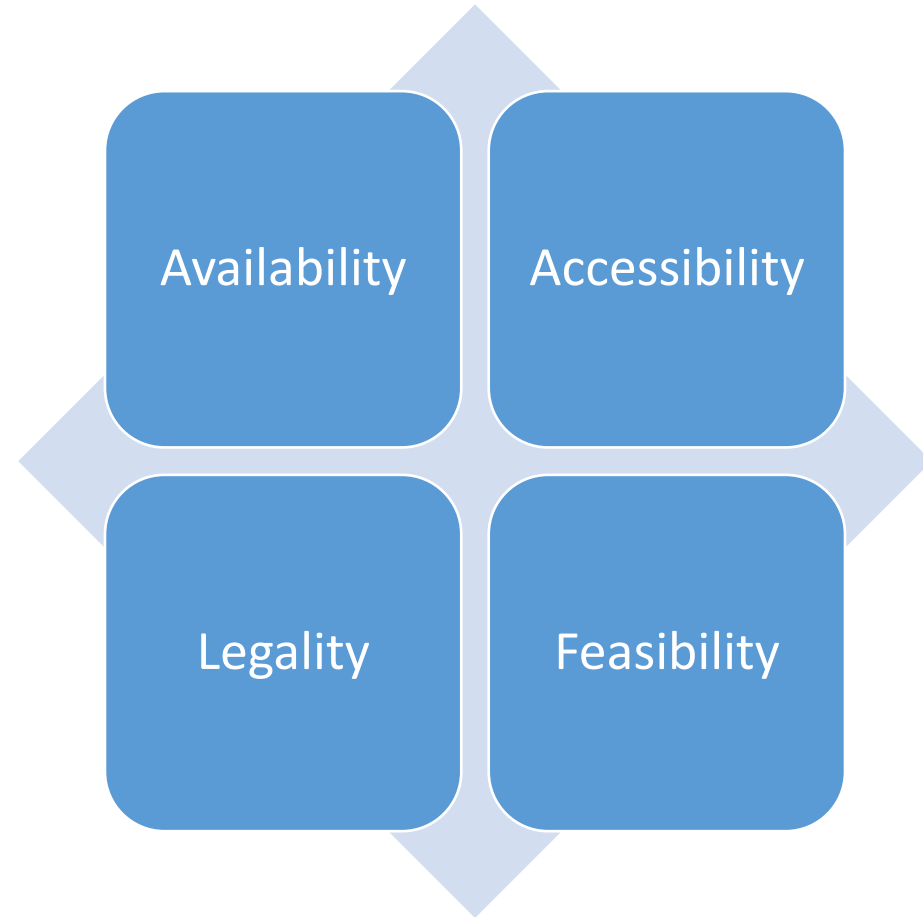


Concluding remarks

Open science

- Amsterdam Call for Action (2016)
 - EU copyright reform and national initiatives
 - Licensing and TDM-friendly formats by publishers
- “The Right to Read is the Right to Mine” LERU 2016 (League of European Research Universities)
- EU copyright reform (2017?), objective TDM:
 - “A mandatory exception for research and education purposes;
 - a mandatory exception that will enable users to text and data mine all content to which they have legal access”
- French new Law for a Digital Republic (2016):
 - TDM exception for scientific documents (if legally accessed)
 - For scientific purpose, no commercial use

Conditions – A2LF



Recommandations

Availability

- Retrodigitization
- Text and complementary material

Accessibility

- Libre open access
- Institutional repositories

Legality

- TDM exception (or fair use)
- Clearance of third party rights (prescription rules)

Feasibility

- Standards (prescription rules)
- Rich metadata and file structure

References

- Pedauque, R. T., 2003. *Document: Form, sign and medium, as reformulated for electronic documents*. Report, CNRS, Paris. http://archivesic.ccsd.cnrs.fr/sic_00000594/fr/
- Murray-Rust, P., 2014. Liberating ETDs to the world and making them useful. In: *ETD 2014 17th International Symposium on Electronic Theses and Dissertations*, July 23-25, 2014, Leicester, UK. <http://www2.le.ac.uk/library/etd2014/plenaries/liberating-etds>
- Schöpfel, J., Prost, H., Piotrowski, M., Hilf, E. R., Severiens, T., Grabbe, P., 2015. A French-German survey of electronic theses and dissertations: Access and restrictions. *D-Lib Magazine* 21 (3/4). <http://www.dlib.org/dlib/march15/schopfel/03schopfel.html>
- Schöpfel, J., Juznic, P., Prost, H., Malleret, C., Cesarek, A., Koler-Povh, T., 2015. Dissertations and data (keynote address). In: *GL17 International Conference on Grey Literature*, 1-2 December 2015, Amsterdam. <https://hal.archives-ouvertes.fr/hal-01285304>
- All references available at <http://www.citeulike.org/user/Schopfel/tag/etd2016>
- Corresponding author: joachim.schopfel@univ-lille3.fr



That's all Folks!