

# Usage of grey literature in open archives



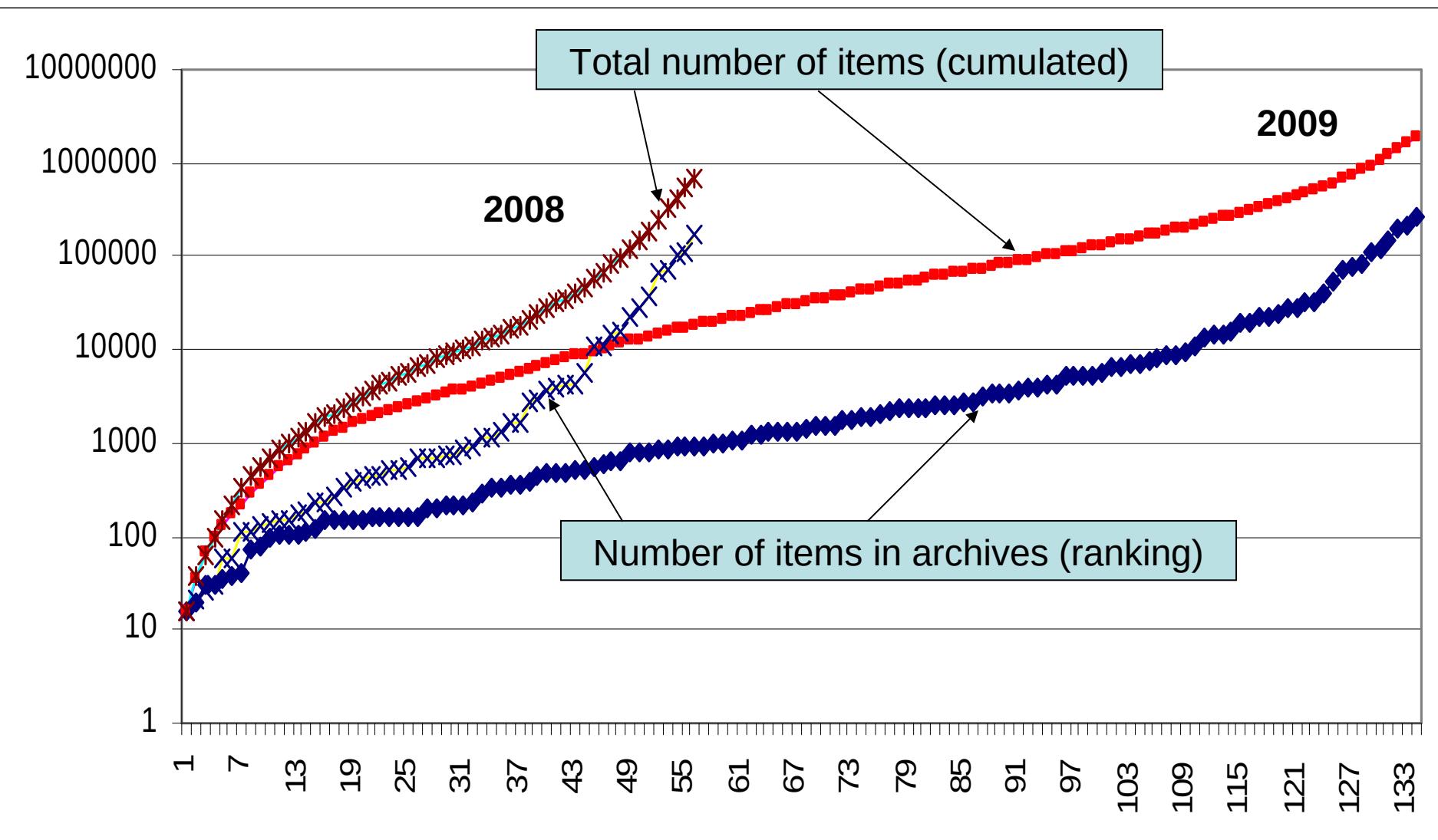
J. Schöpfel (University of Lille 3)



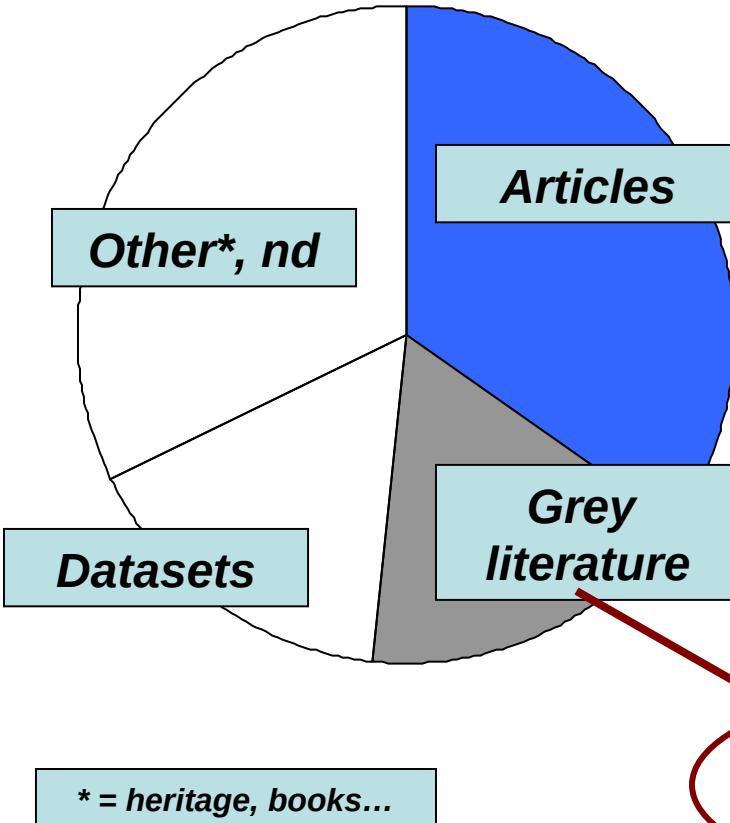
C. Boukacem-Zeghmouri (University of Lille 3)  
H. Prost (INIST-CNRS)



# Size of repositories and total number of items



# Content evolution



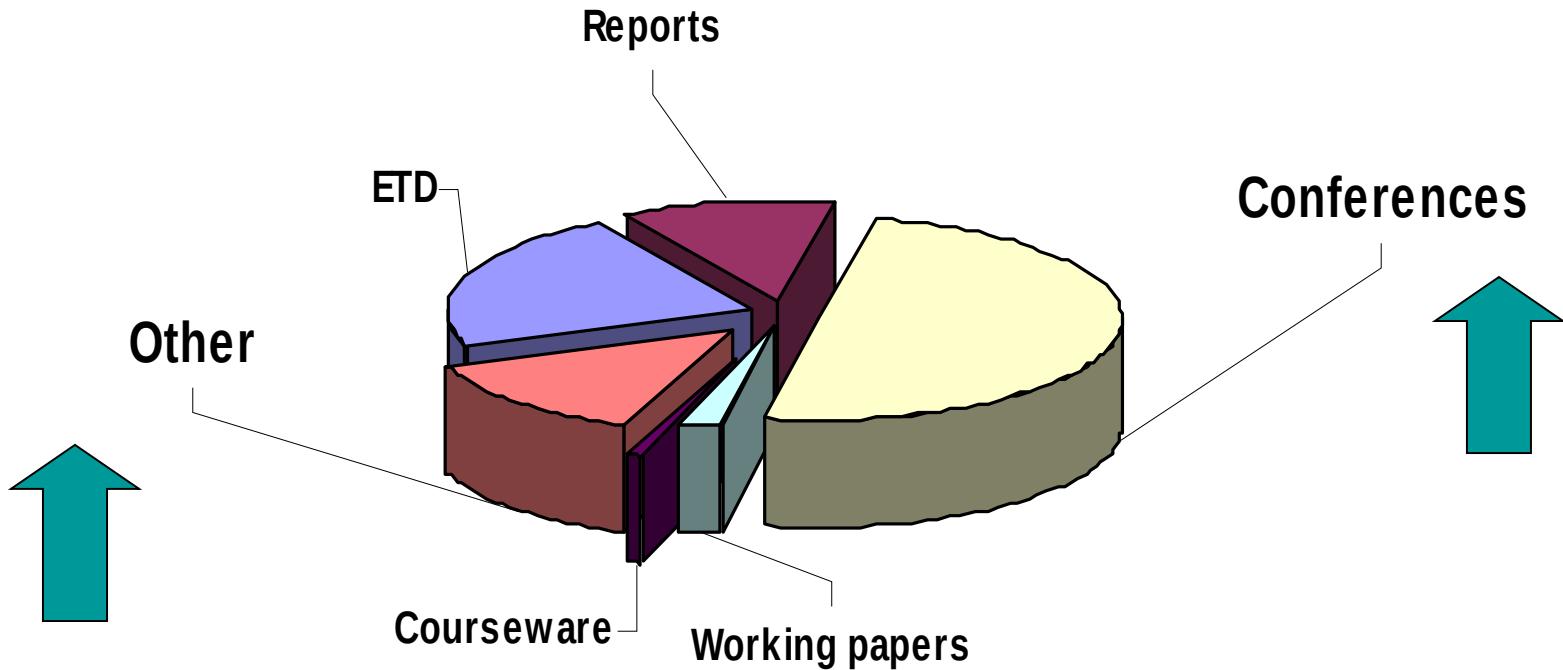
1.87m items

= x2,7 since 2008

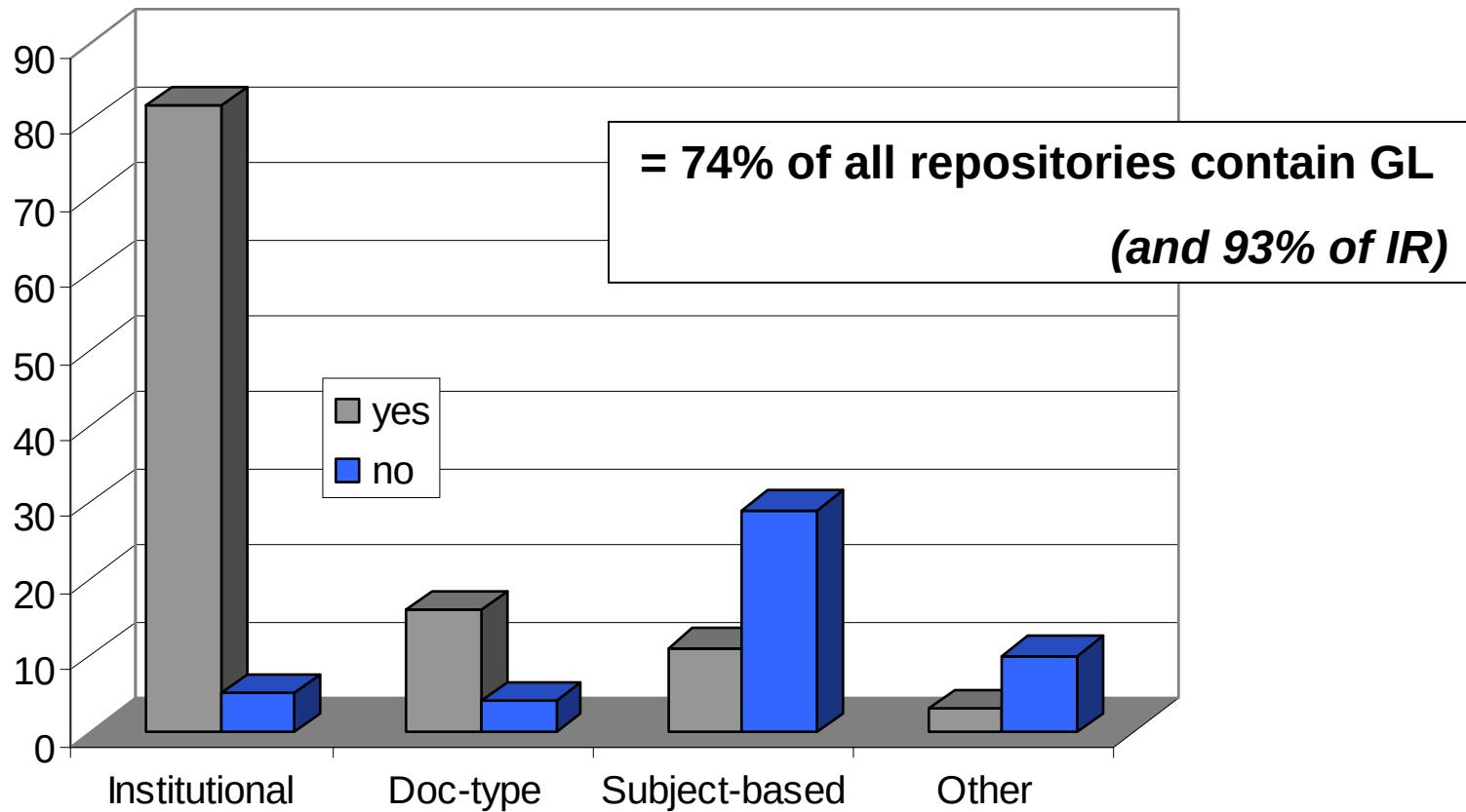
Representativity 10% (?)

Part of GL unchanged (17%)  
But: +200,000 new grey items

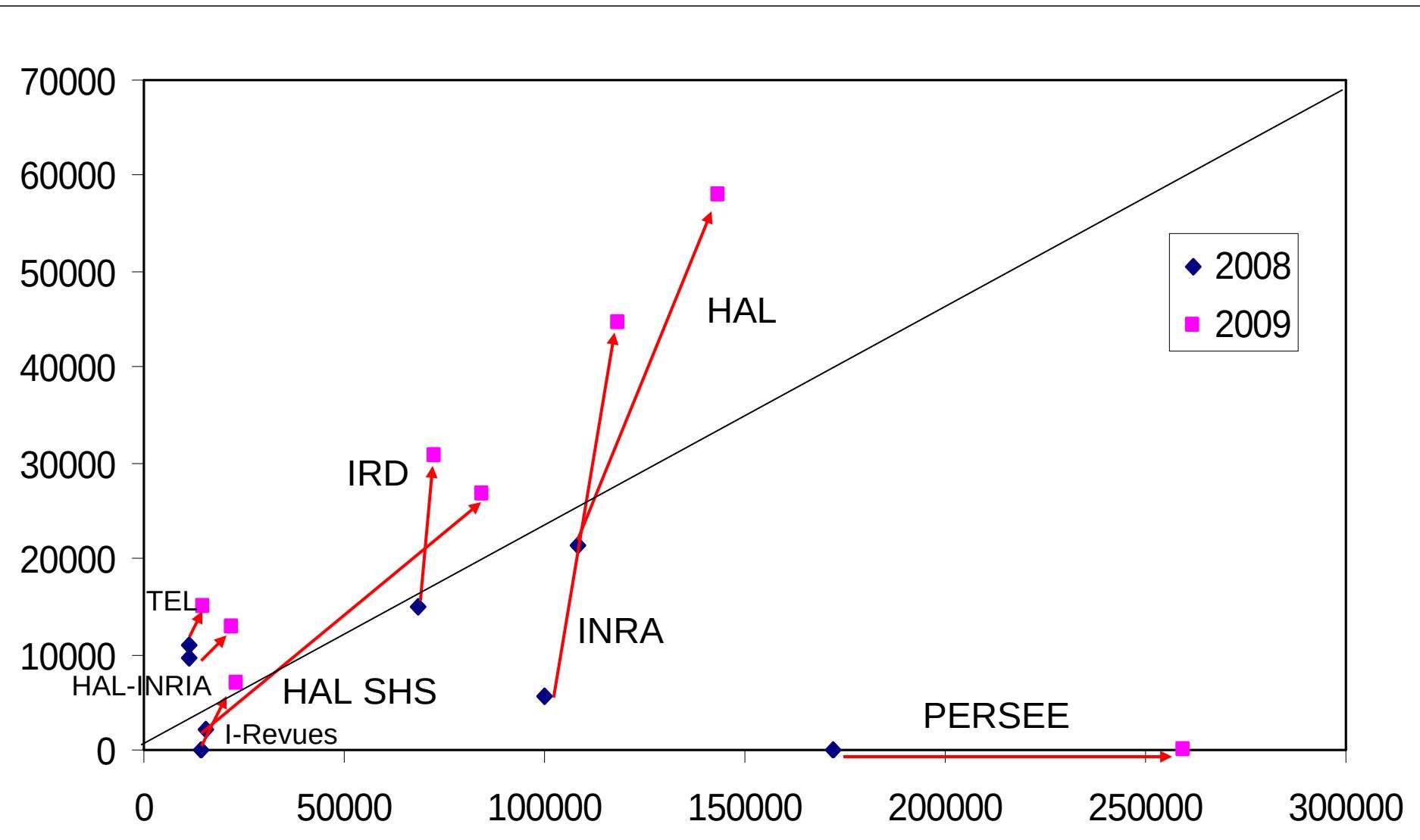
# GL document types



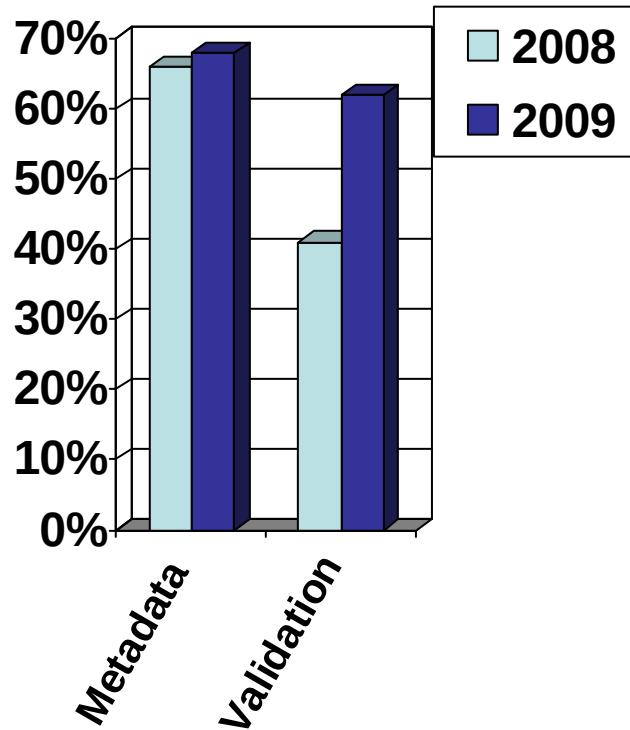
# Repository type and presence of grey literature



# Size of repository and number of grey items



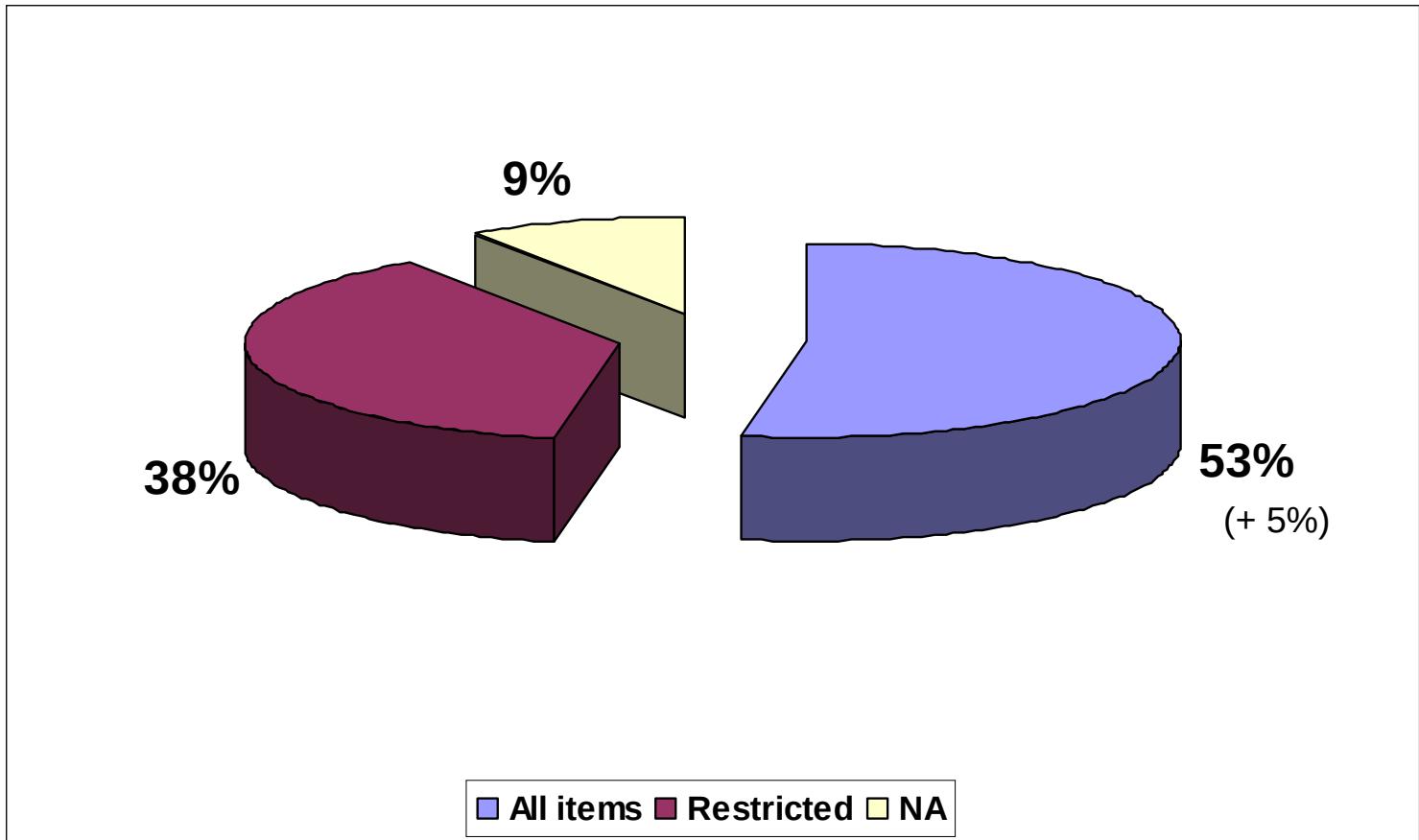
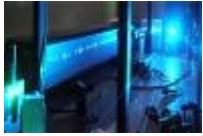
# Quality improvement



Slightly more archives with specific metadata for grey items

Significant more archives with some kind of content validation and/or quality control

# Access to full text...



# ... but items without fulltext



Half of all open archives contain bibliographic records that don't link to the document



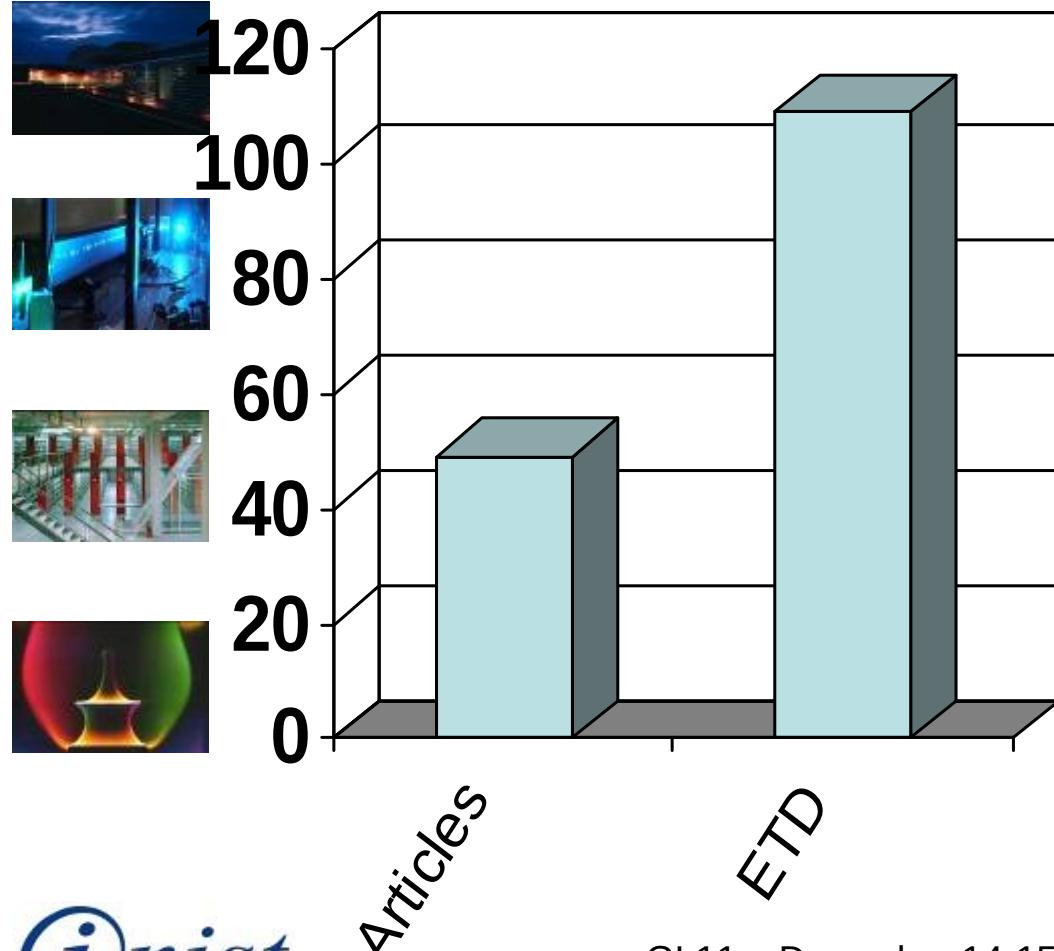
Part of these records varies from 5 to 90%



Overall part of records without fulltext: 16%



# Usage statistics of GL

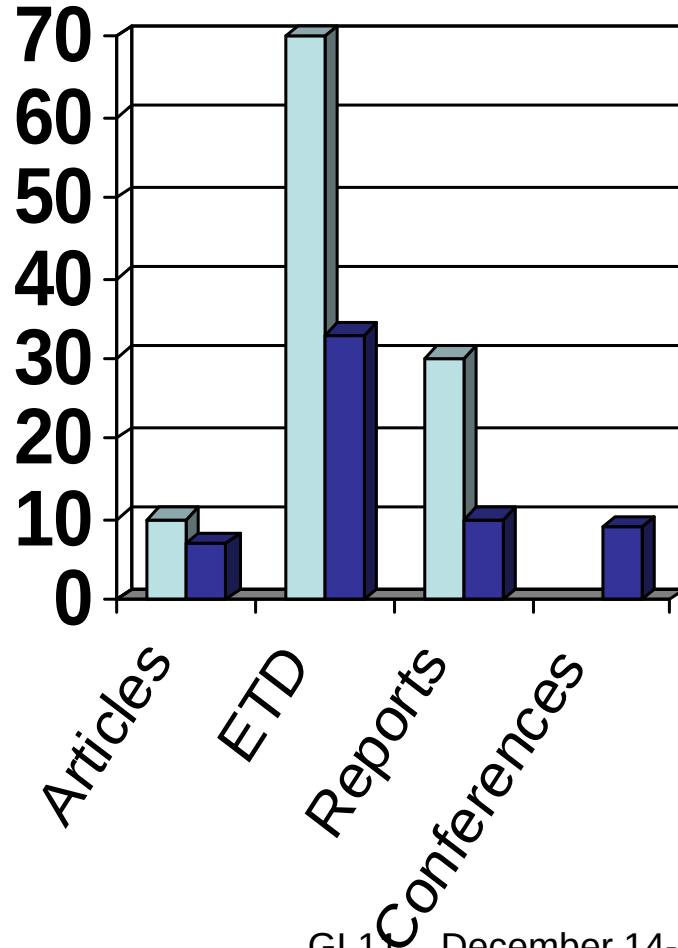


Average downloads per document type

Importance of grey literature: 2,2 (ETD)

University of Toulouse (OATAO)

# Usage statistics of GL



Average downloads per document type

Importance of grey literature:

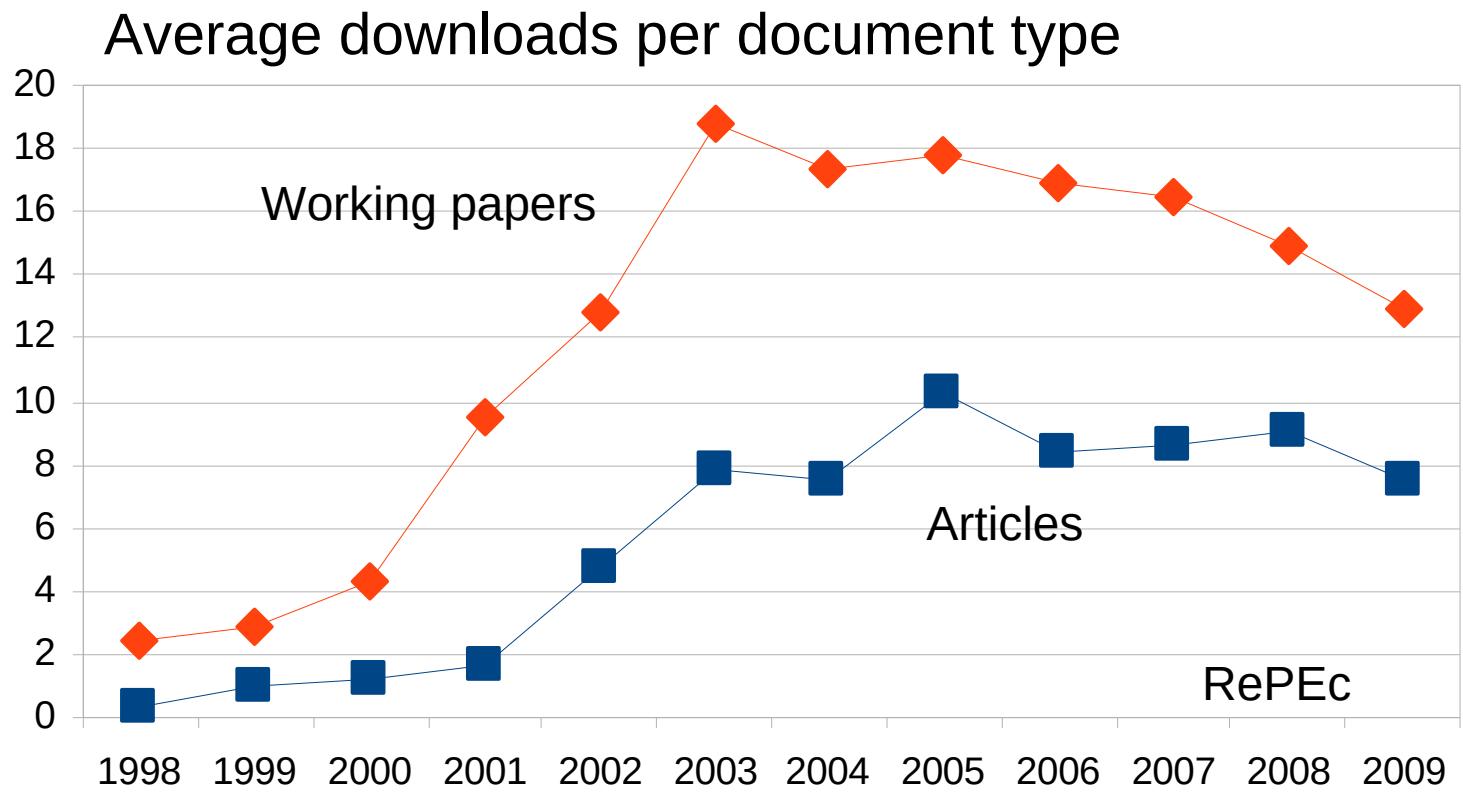
4,7 - 7 (ETD)

1,4 - 3 (reports)

1,3 (conferences)

IFREMER (Archimer)

# Usage statistics of GL



Importance of grey literature: 1,7 - 5 (working papers)

# Problems



Cumulative statistics



No history

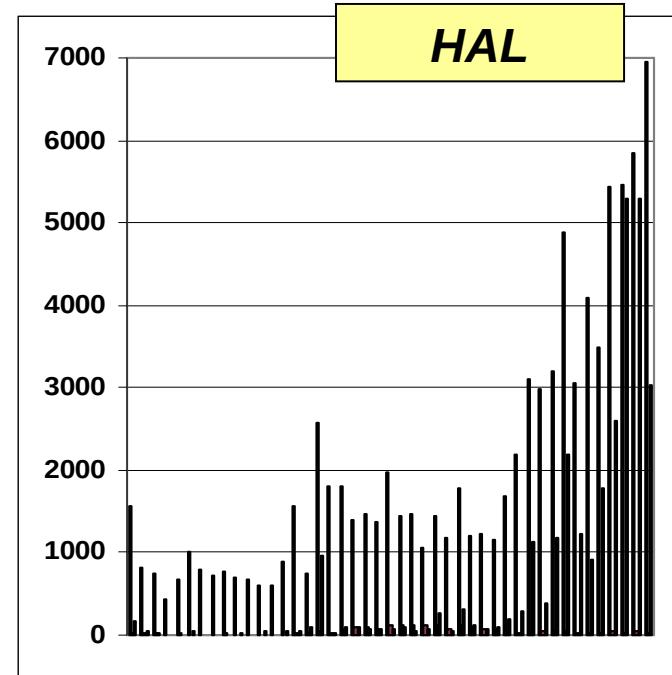


No details (formats, ...)



No specific information on GL

GL11 – December 14-15, 2009



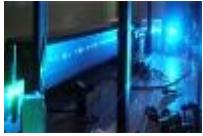
Without metadata, no statistics



# Metadata



On the one hand:



- Difficulties in identifying the types of documents.
- Only « published or unpublished » document.
- No count of results.



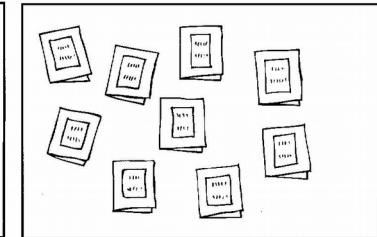
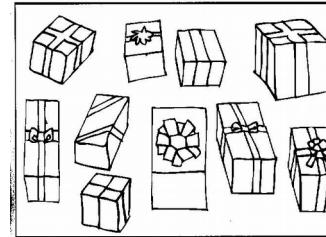
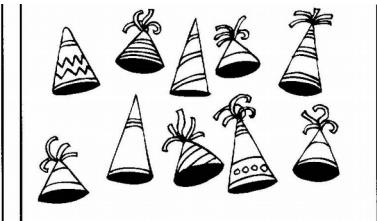
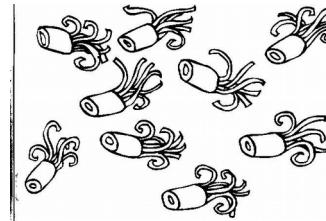


# Metadata



On the other hand:

- Query by:
  - author's affiliation
  - scientific department
  - research theme
  - document type
  - keywords
- Choice with date
- Choice with full text or not
- Ranking of results



# Log analysis (1)



(Figures in parentheses refer to the 7-day period ending 16-Nov-2009 00:00).



**Successful requests:** 132,810 (6,975)

**INP Toulouse**



**Average successful requests per day:** 914 (996)

**Successful requests for pages:** 132,810 (6,975)

**Average successful requests for pages per day:** 914 (996)



**Failed requests:** 84 (0)

**Distinct files requested:** 530 (526)

**Distinct hosts served:** 40,015 (3,609)

**Corrupt logfile lines:** 55

**Unwanted logfile entries:** 2,743,109

**Data transferred:** 172.81 gigabytes (9.18 gigabytes)

**Average data transferred per day:** 1.19 gigabytes (1.31 gigabytes)



Diversity of tools (Google Analytics / Sitemap, Webalizer Xtended, AWStats, HAL, PhpMyVisite, Analog ...)

# Log analysis (2)

reqs: %bytes: last time: file

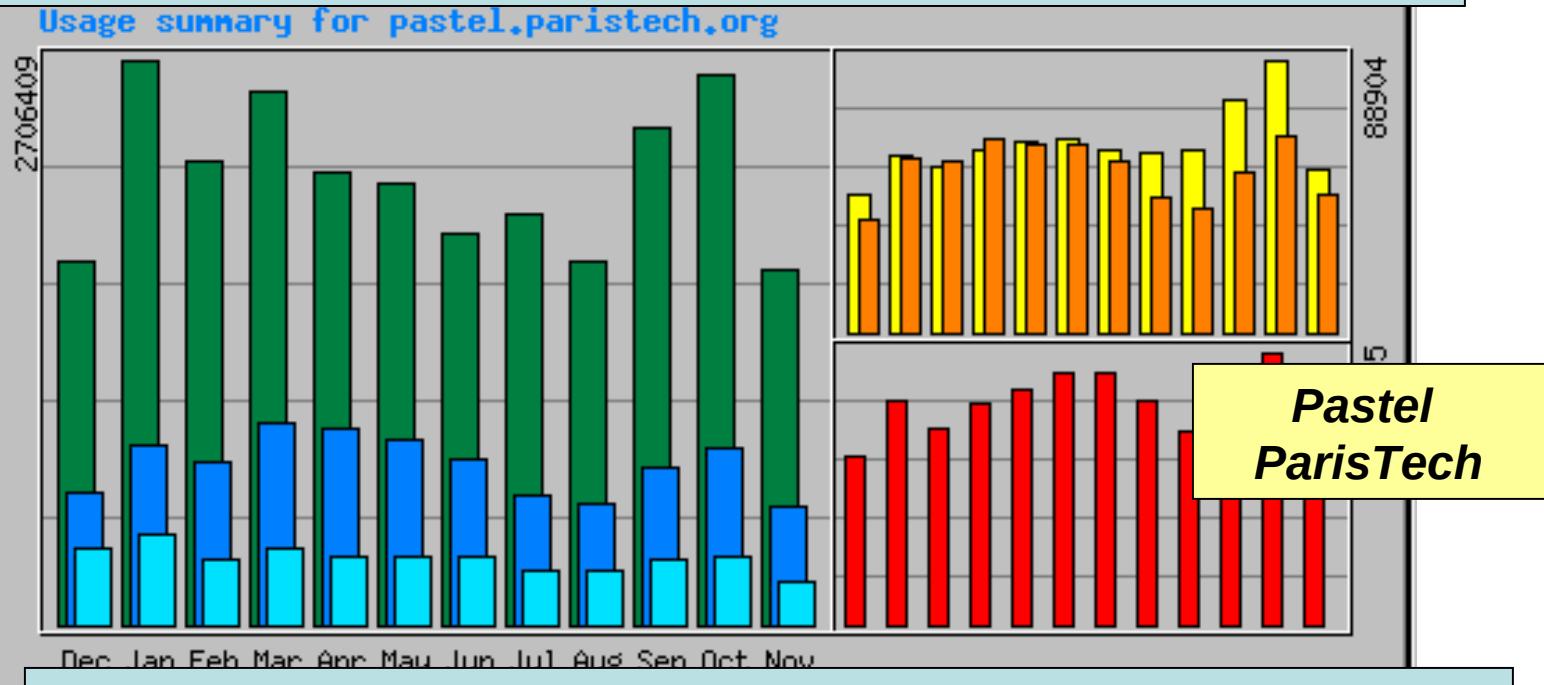
INP Toulouse



- 2044: 0.40%: 29/Nov/09 22:53: García Martinez (2009) Development and validation of the Euler-Lagrange formulation on a parallel a...
- 1283: 0.26%: 29/Nov/09 22:53: Delgado Zambrano (2009) Bioréacteur à membrane externe pour le traitement d'effluents contenant de...
- 1115: 0.30%: 29/Nov/09 22:53: Sepret (2009) Application de la PIV sur traceurs fluorescents à l'étude de l'entraînement...
- 1063: 0.21%: 29/Nov/09 22:53: Nerisson (2009) Modélisation du transfert des aérosols dans un local ventilé.
- 1057: 0.95%: 29/Nov/09 17:34: Delabrouille (2004) Caractérisation par MET de fissures de corrosion sous contrainte d'alliages...
- 1029: 0.14%: 29/Nov/09 22:53: Rajsiri (2009) Knowledge-based system for collaborative process specification.
- 1014: 0.79%: 29/Nov/09 22:10: Delay (2005) Analyse des écoulements transitoires dans les systèmes d'injection directe...
- 984: 0.88%: 29/Nov/09 22:55: Goncalo (2006) Procédé

# Log analysis (3)

Access to website: search engins,  
geographical origin, strategies, etc.



On site behaviour: bouncing, downloading,  
duration, domains, etc.

# Towards standardization: PIRUS (JISC)



*Publisher and Institutional Repository Usage Statistics*



For authors and institutions



Article level (DOI)

COUNTER compliant



Article Report 1: <title>Number of Successful Full-Text Article Requests by Month and DOI</title>



# Towards standardization: PIRUS 2 (JISC)



COUNTER standards & PIRUS results



Different « Article Reports » (*core set of standard usage statistics reports*)



Open Source software for production and sharing  
of usage statistics on article (item) level for OA



Cost analysis

Final report in December 2010

# Towards standardization: OA-Statistik (DINI)



For authors (usage follow-up), readers-scientists (relevance, alert), institutions (impact)



Article level (= document)



Tools for transfer/sharing (network)



Added-value services

# Towards standardization: other websites, projects



LogEc <http://logec.repec.org/> Usage statistics of the RePEc repository



IFABC <http://www.ifabc.org/> Definition of usage metrics (user, visit...)



SURF <http://www.surffoundation.nl/nl/projecten/Pages/SURE.aspx> Aggregation of log files



JISC Usage statistics review <http://ie-repository.jisc.ac.uk/250/> Proposal of standard

# Recommendations (1)



Recipient: authors, users, institutions



COUNTER principle: different levels, with a basic minimum level (*AR1*)



Selection of minimum elements for a basic log analysis

(*who, what, request type, when, identifier*)



# Recommendations (2)



Definition of elements and terminology  
(access, downloading, visit, request, hit...)



Periodicity (monthly) and delay (30 days)



Distinction full text / records



# Recommendations (3)



Added-value services\* :

Modular statistics (collections, document types, time period etc.)



Summary tables



Assistance-help / FAQ

Link with other tools measuring the impact of deposited items (citations, tagging etc.)



(...)

\* see PLoS <http://article-level-metrics.plos.org/>

# Forthcoming



2010 IRIS case study (Lille 1)



2010 Final report of DUAO-F project



??? Partnership with JISC/COUNTER and DINI



??? Project with CCSD and/or COUPERIN



# Thank you!



Joachim Schöpfel  
University Charles de Gaulle Lille 3  
[joachim.schopfel@univ-lille3.fr](mailto:joachim.schopfel@univ-lille3.fr)  
++ (0) 33 688 35 01 47

Chérifa Boukacem-Zeghmouri  
University Charles de Gaulle Lille 3  
[boukacemc@yahoo.fr](mailto:boukacemc@yahoo.fr)  
++ (0) 33 620 62 18 12

Hélène Prost  
INIST-CNRS  
[helene.prost@inist.fr](mailto:helene.prost@inist.fr)  
++33 (0) 383 50 47 12

