



Les statistiques d'utilisation d'archives ouvertes État de l'art



J. Schöpfel (Université de Lille 3)
H. Prost (INIST-CNRS)



Les archives ouvertes en France : un projet significatif



Volonté politique

Création du CCSD en 2000

Déclaration de Berlin 2003

Protocole 2005

Investissement public

Ressources humaines

Équipement etc.

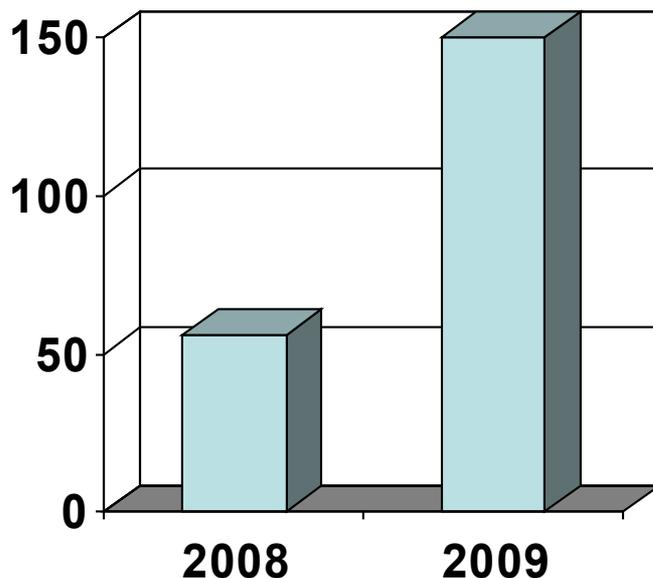
Résultats?

Usage défini en termes d'offre
(dépôts), pas en termes d'accès

= comme au début de « big deals »



©CNRS/IN2P3



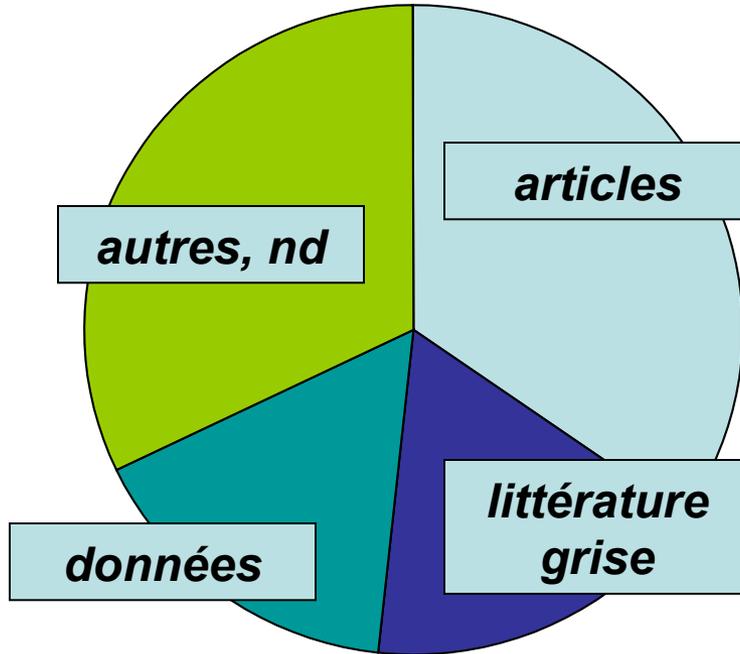
150 sites référencés
+ 94 depuis 2008

+ 45 nouveaux sites HAL*

** plus 3 autres depuis juin 2009*

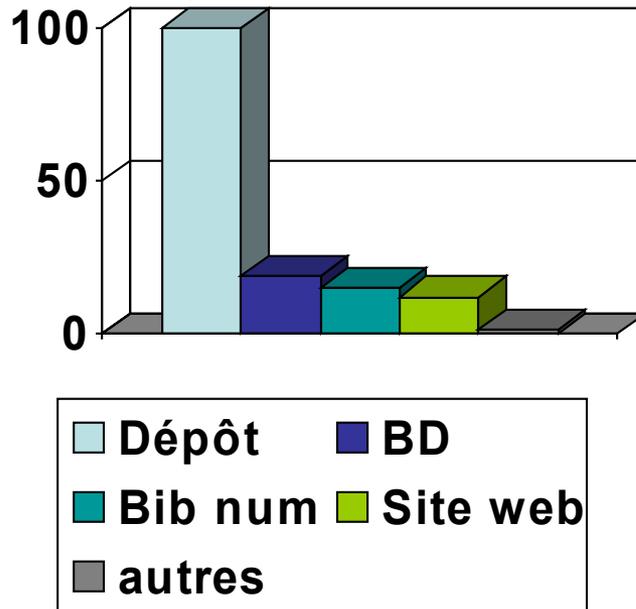
58% archives institutionnelles
20% archives thématiques

19 répertoires/listes dont
OpenDOAR, ROAR, BASE,
OAlster, Scientific Commons



1,87 millions de dépôts
= x2,7 depuis 2008

Représentativité 10% (?)



37 sites avec >50% de notices sans document

➤ Priorité à l'évaluation ?

30 sites à caractère patrimonial (= 20% des sites, 46% des docs)

➤ Priorité à la valorisation des archives ?

Statistiques à destination des auteurs et institutions



Accès aux notices



Accès aux documents



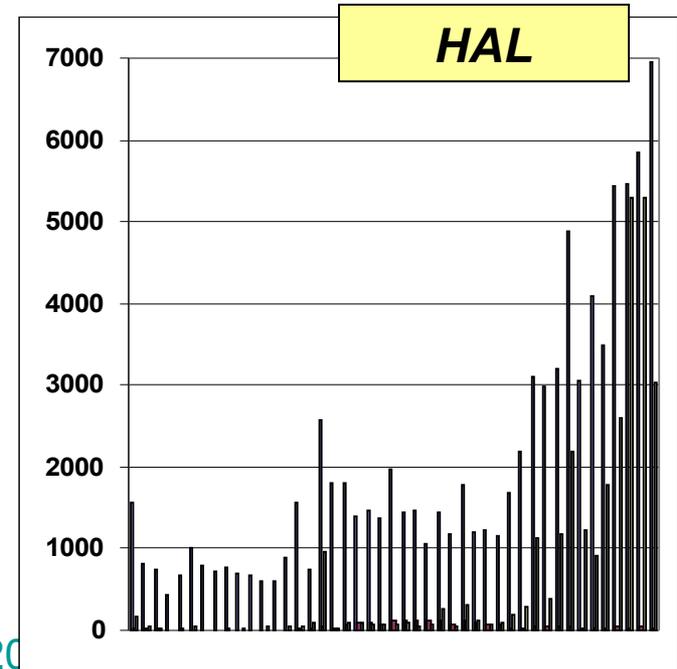
Statistiques cumulatives

Pas d'historique

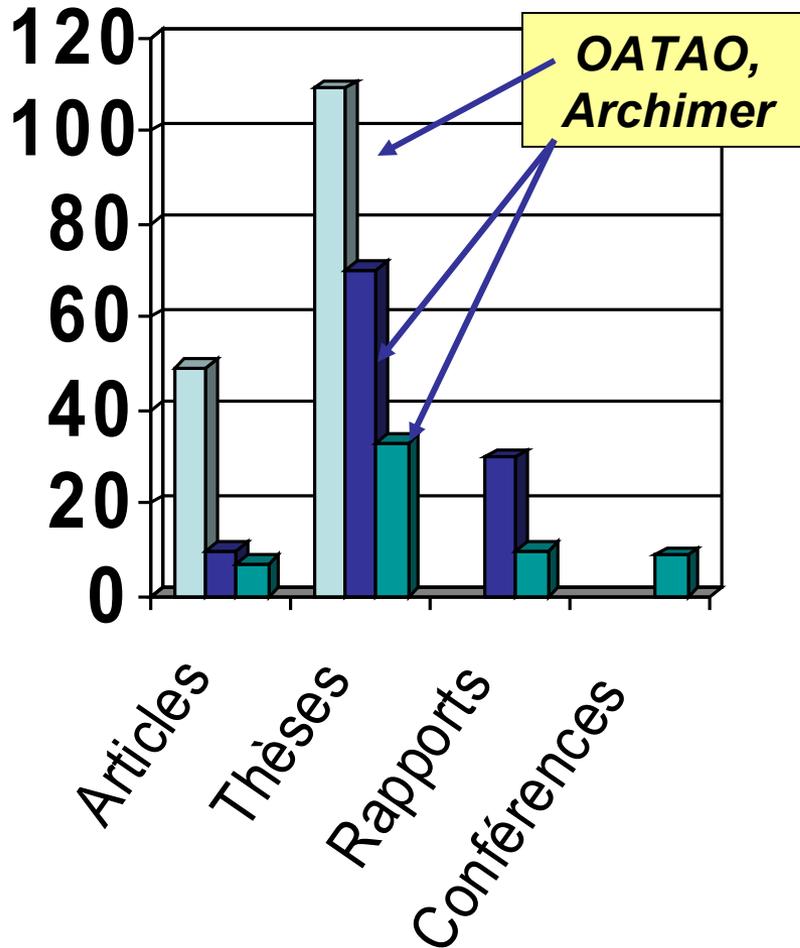
Pas de détail (formats, ...)

Cf. Centrale de Lyon

<http://bibli.ec-lyon.fr/documents/CS100720>



L'usage par document



Taux de consultation par type de document

Intérêt de la littérature grise

Période et nombre de dépôts hétérogènes

Analyse des logs (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

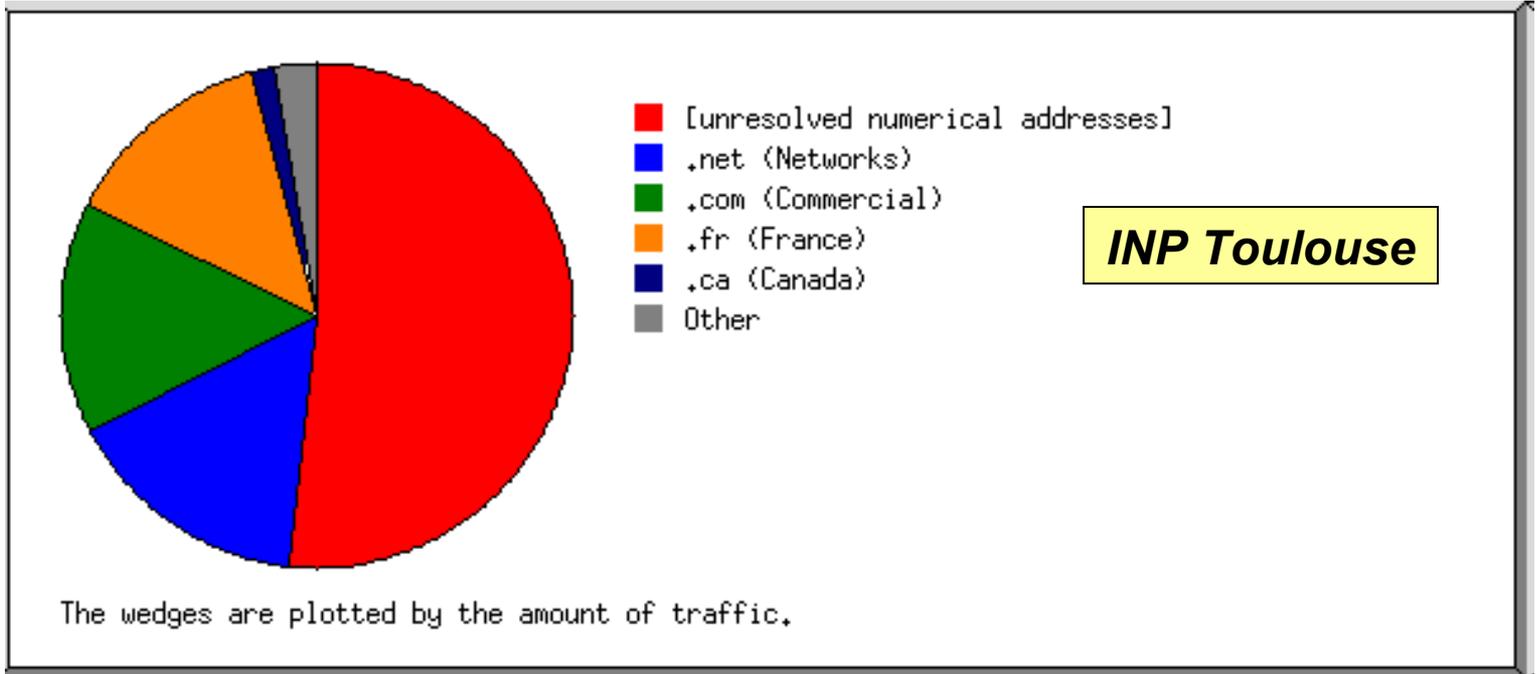
Unwanted logfile entries: 2,743,109

Data transferred: 172.81 gigabytes (9.18 gigabytes)

Average data transferred per day: 1.19 qigabytes (1.31 qigabytes)

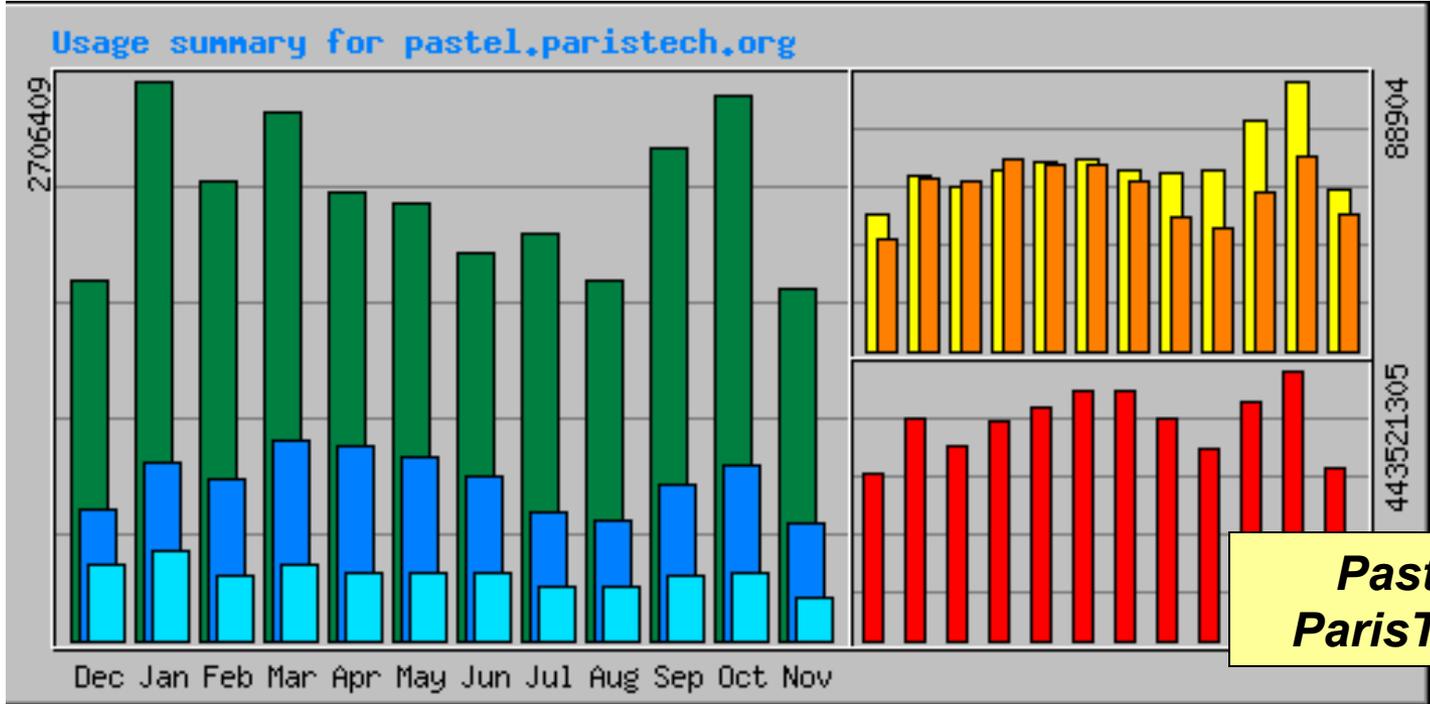
Diversité des outils (Google Analytics / Sitemap, Webalizer Xtended, AWStats, HAL, PhpMyVisite, Analog ...)

Analyse des logs (2)



Chemin d'accès : moteurs de recherche, origine géographique, stratégies, etc

Analyse des logs (3)



Comportement « sur site » : rebond, accès au texte intégral, durée, domaines, etc.



Vers une normalisation : PIRUS (JISC)



Publisher and Institutional Repository Usage Statistics

Pour auteurs et institutions



Niveau article (DOI)

COUNTER compatible



XML prototype

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



```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE reports PUBLIC "-//ProjectCounter//DTD reports//EN"
"http://www.projectcounter.org/dtd/2004/reports.dtd">
<reports xmlns="http://www.projectcounter.org/ns/2004/reports">
<article_report id="cranfield.ac.uk_12345" cop_version="1" cop_report="1"> ← 1
<header>
<title>Number of Successful Full-Text Article Requests by Month and DOI</title> ← 2
<timestamp>2008-12-01-T10:03:47Z</timestamp>
<vendor>
<vend_name>Cranfield University</vend_name>
<vend_imprint>Cranfield CERES</vend_imprint>
<vend_site>https://dspace.lib.cranfield.ac.uk/</vend_site> ← 3
<vend_contact>cranfieldceres@cranfield.ac.uk</vend_contact>
</vendor>
<customer>
<cust_name>Central Agency</cust_name>
<cust_ref>123-4567</cust_ref>
<cust_ip type="cidr">123.456.78.9</cust_ip>
<cust_username>centralagency</cust_username>
<cust_criteria>institution</cust_criteria>
</customer>
</header>
<article_data>
<article_doi="10.1016/j.jairtraman.2005.01.007" format="application/pdf"> ← 4
<requests start="2008-09-01" end="2008-09-30">108</requests>
<requests start="2008-10-01" end="2008-10-31">201</requests> ← 5
<requests start="2008-09-01" end="2008-10-31">309</requests>
</article>
```



Vers une normalisation : PIRUS 2 (JISC)



Normes COUNTER & résultats PIRUS



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



Open Source Logiciels pour génération et partage des statistiques d'utilisation au niveau article (*item*) dans des AO

Analyse financière des coûts d'un tel service



Rapport final en décembre 2010



Vers une normalisation : PEER (EC/STM)



Publishing and the Ecology of European Research
Étude de l'impact des archives ouvertes sur l'édition
Analyse des fichiers log (*Common Logfile Format*)

```
66.249.66.5 - - [12/jan/2009:20:31:53 +0100]
"GET/pdf_frontpage.Php?
Source_opus=87&startfile=egelhaaf_et_al_uniforsch
2002.Pdf HTTP/1.1" 302 414 "-" "mozilla/5.0
(compatible; Googlebot/2.1;
+http://www.google.com/bot.html)"
```





Vers une normalisation : OA-Statistik (DINI)



Pour auteurs (suivi d'usage), lecteurs-chercheurs (relevance, alerte), institutions (impact)

Niveau article (= document)

Outils de transfert/diffusion (réseau)

Services à valeur ajoutée



Aufsatz zugänglich unter
 URN: urn:nbn:de:bsz:93-opus-32590
 URL: <http://elib.uni-stuttgart.de/opus/vollt>

Protein Kinase D regulates several aspect
 Maier, Dieter ; Nagel, Anja C. ; Gloc, Hele

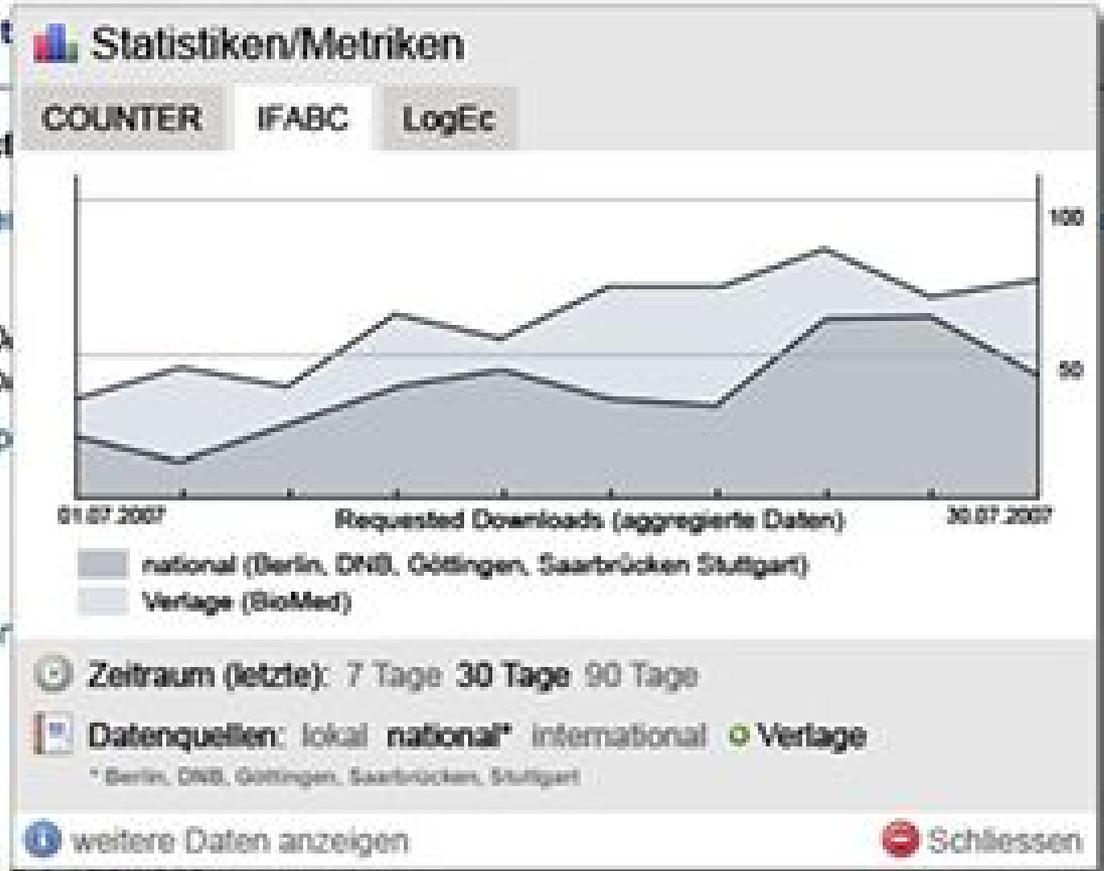
Originalveröffentlichung: (2007) BMC D
<http://www.bio>
 pdf-Format: Dokument 1.p

Statistiken:
 Gedruckte Ausgabe: Print-or

Dokument empfehlen:

Dokument in Google Scholar suchen:

SWD-Schlagwörter: Proteinkinase
 Institut: Institut für Zellbiologie und Immunologie
 DDC-Sachgruppe: Biowissenschaften, Biologie



s, Anette



Vers une normalisation : IFREMER



Conseils pour optimiser la visibilité des dépôts
(référencement)



Un seul fichier avec toutes les (méta)données
Renseigner les propriétés PDF
Réduire la taille des fichiers
(...)





Vers une normalisation : autres sites, projets



LogEc <http://logec.repec.org/> Statistiques d'utilisation
de l'archive RePEc



IFABC <http://www.ifabc.org/> Définition d'indicateurs
d'usage (utilisateur, visite...)



SURF <http://www.surffoundation.nl/nl/projecten/Pages/SURE.aspx>
Agrégation des fichiers log





Recommandations (1)



Destinataires : auteurs, utilisateurs, institutions



Principe COUNTER : plusieurs niveaux, avec un niveau minimum simple (*AR1*)



Choix des éléments minimum pour une exploitation des fichiers log

(qui, quoi, type de requête, quand, identifiant)





Recommandations (2)



Définition précise des éléments et termes
(consultation, déchargement, accès, *visit*,
request, *hit*...)

Périodicité (mensuel) et délai (30 jrs)

Séparation texte intégral / notices





Services à valeur ajoutée* :



Statistiques modulables (collections, type de documents, périodes etc.)

Tableaux synthétiques



Assistance-explication / FAQ

Lien vers d'autres outils pour mesurer l'impact d'un dépôt (citations, tagging etc.)

(...)



2010 Étude de cas IRIS (Lille 1)



2010 Rapport final du projet DUAO-F

2010 Étude des moteurs de recherche



??? Collaboration avec JISC/COUNTER et DINI

??? Projet avec CCSD et/ou COUPERIN



Merci



Joachim Schöpfel
Université Charles de Gaulle Lille 3
joachim.schopfel@univ-lille3.fr

++ (0) 33 320 41 69 38

++ (0) 33 688 35 01 47

Hélène Prost
INIST-CNRS

helene.prost@inist.fr

++33 (0) 383 50 47 12

