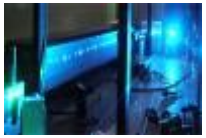
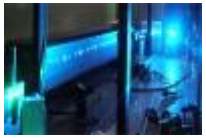


Usage Statistics of Open Archives State of the Art

J. Schöpfel (University of Lille 3)

H. Prost (INIST-CNRS)

Open archives in France: a significant project



A political will

Creation of the CCSD in 2000

Berlin declaration in 2003

Protocol in 2005

Public investment

Human resources

Equipment etc.

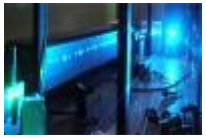
Results?

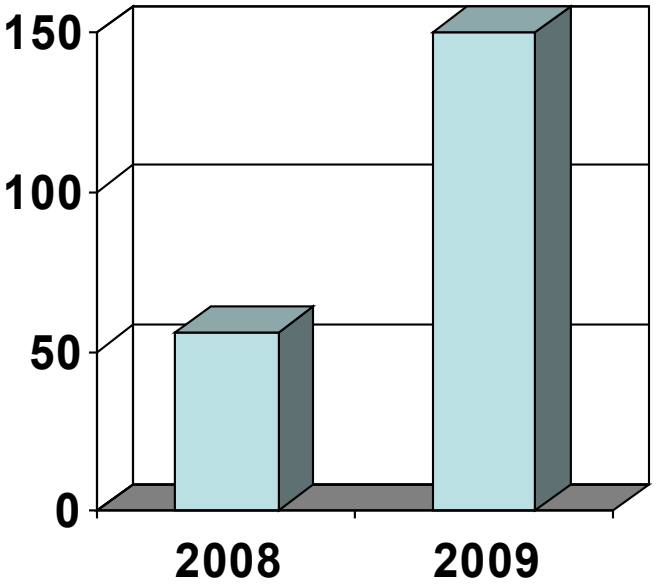
Usage defined as offer (deposits), not as access (downloads)

= as in early « big deals »

©CNRS/IN2P3

Explosion of OA sites





150 referenced websites
+ 94 since 2008

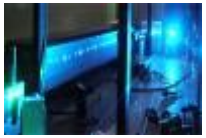
+ 45 new HAL sites*

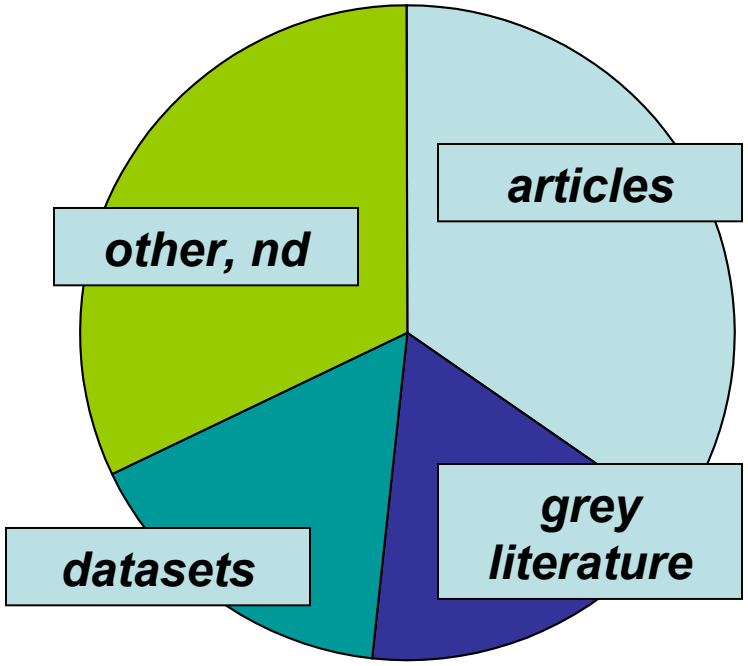
** +3 since June 2009*

58% institutional repositories
20% subject-based archives

19 directories/lists, such as OpenDOAR, ROAR, BASE, OAIster, Scientific Commons

Content evolution



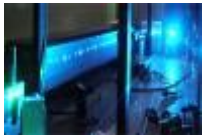


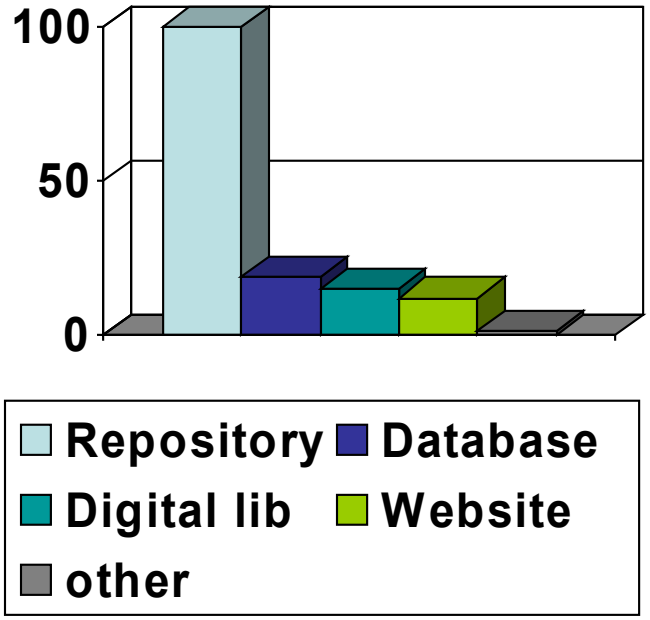
1.87m items

= x2,7 since 2008

Representativity 10% (?)

Limits of a concept





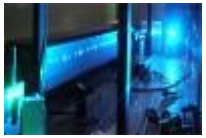
37 sites with >50% records without full text

➢Priority for evaluation ?

30 sites with heritage character (= 20% sites, 46% items)

➢Priority for valorization of archives ?

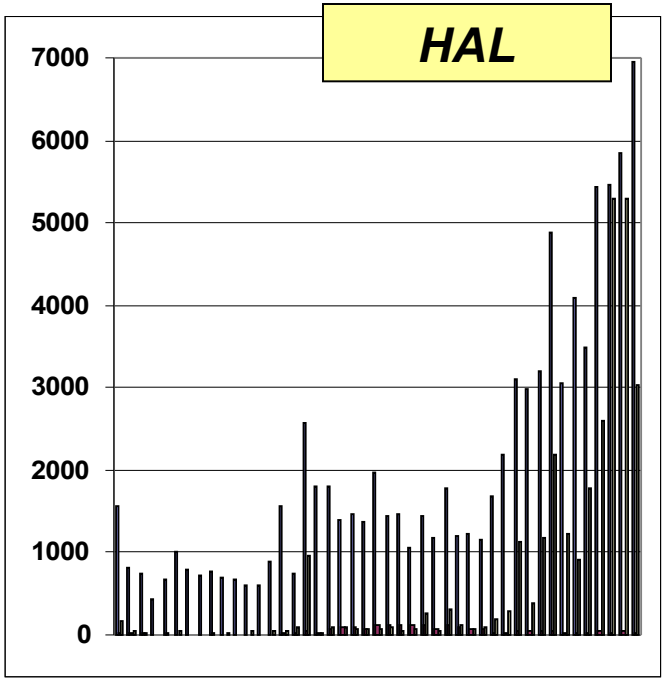
Statistics for authors and institutions



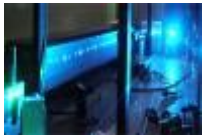
Access to records
Access to documents

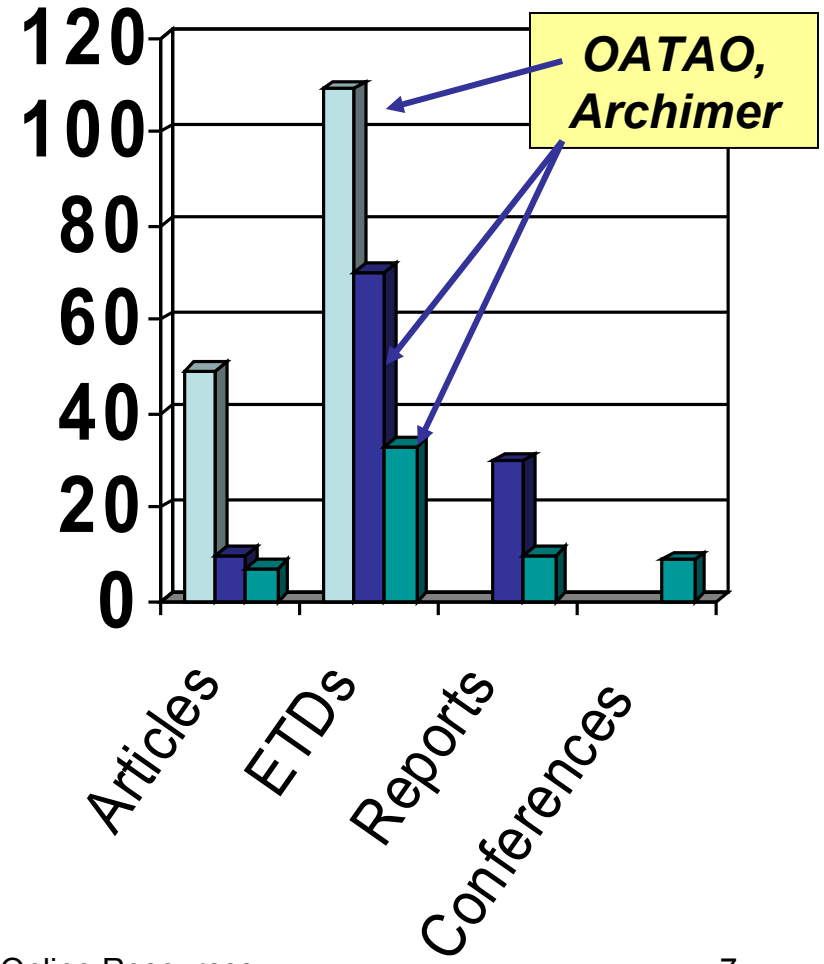
Cumulative statistics
No history
No details (formats, …)

See Centrale de Lyon
http://bibli.ec-lyon.fr/documents/CS10072008.pdf



Document-specific usage





Downloads per document type

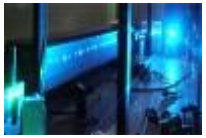
Importance of grey literature

Heterogenous time periods and item numbers

Log analysis (1)

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

INP Toulouse

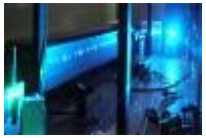


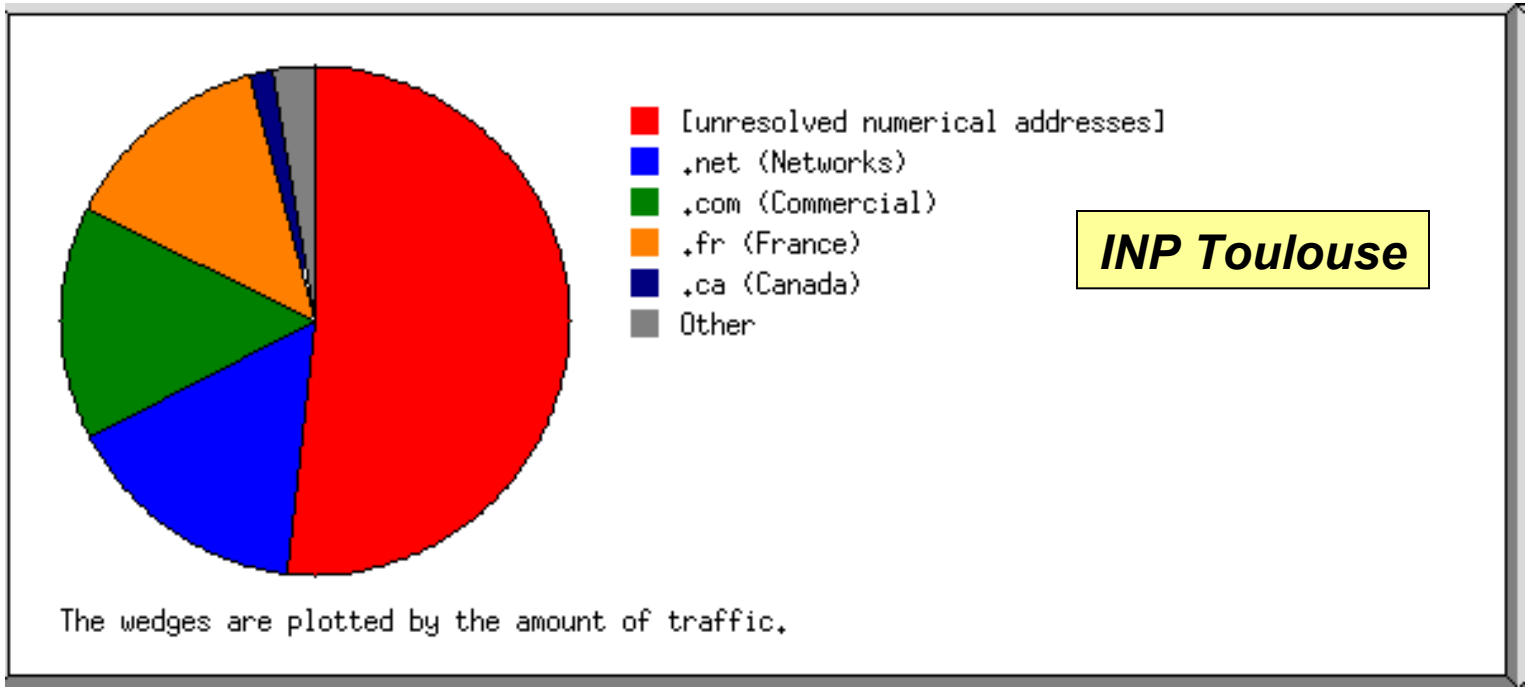
Successful requests: 132,810 (6,975)
Average successful requests per day: 914 (996)
Successful requests for pages: 132,810 (6,975)
Average successful requests for pages per day: 914 (996)
Failed requests: 84 (0)

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)

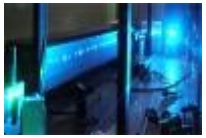


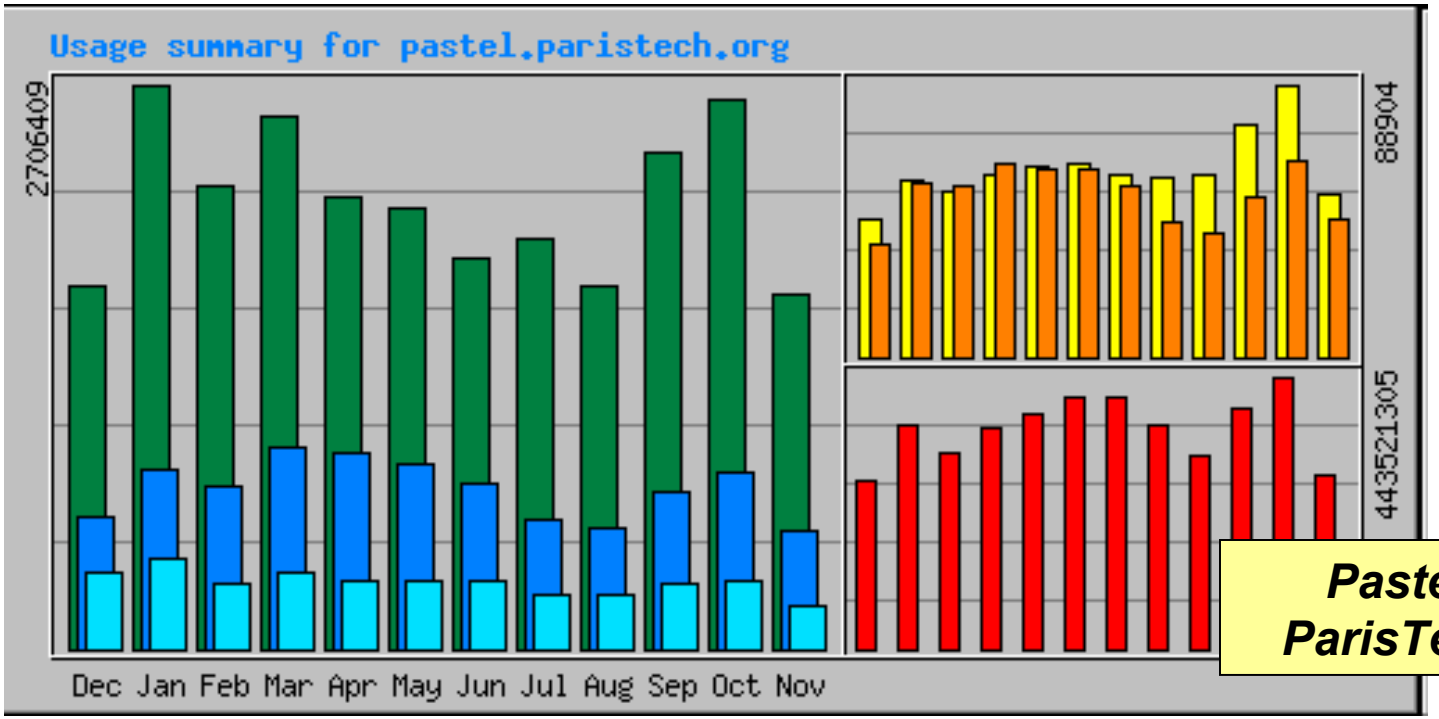


INP Toulouse

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

Log analysis (3)

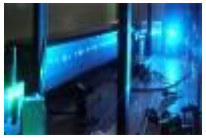




Pastel ParisTech

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

Towards standardization: PIRUS (JISC)



Publisher and Institutional Repository Usage Statistics

For authors and institutions

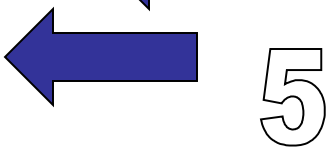
Article level (DOI)

COUNTER compliant

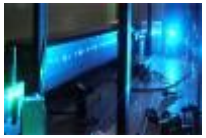
XML prototype

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">
<header>
<title>Number of Successful Full-Text Article Requests by Month and DOI</title>
<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>
<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">
<requests start="2008-09-01" end ="2008-09-30">108</requests>
<requests start="2008-10-01" end ="2008-10-31">201</requests>
<requests start="2008-09-01" end ="2008-10-31">309</requests>
</article>
```

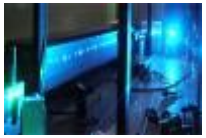


Towards standardization: PIRUS 2 (JISC)



- 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: PEER (EC/STM)



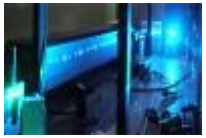
Publishing and the Ecology of European Research

Study on impact of open archives on publishing

Log analysis (*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)"

Towards standardization: OA-Statistik (DINI)



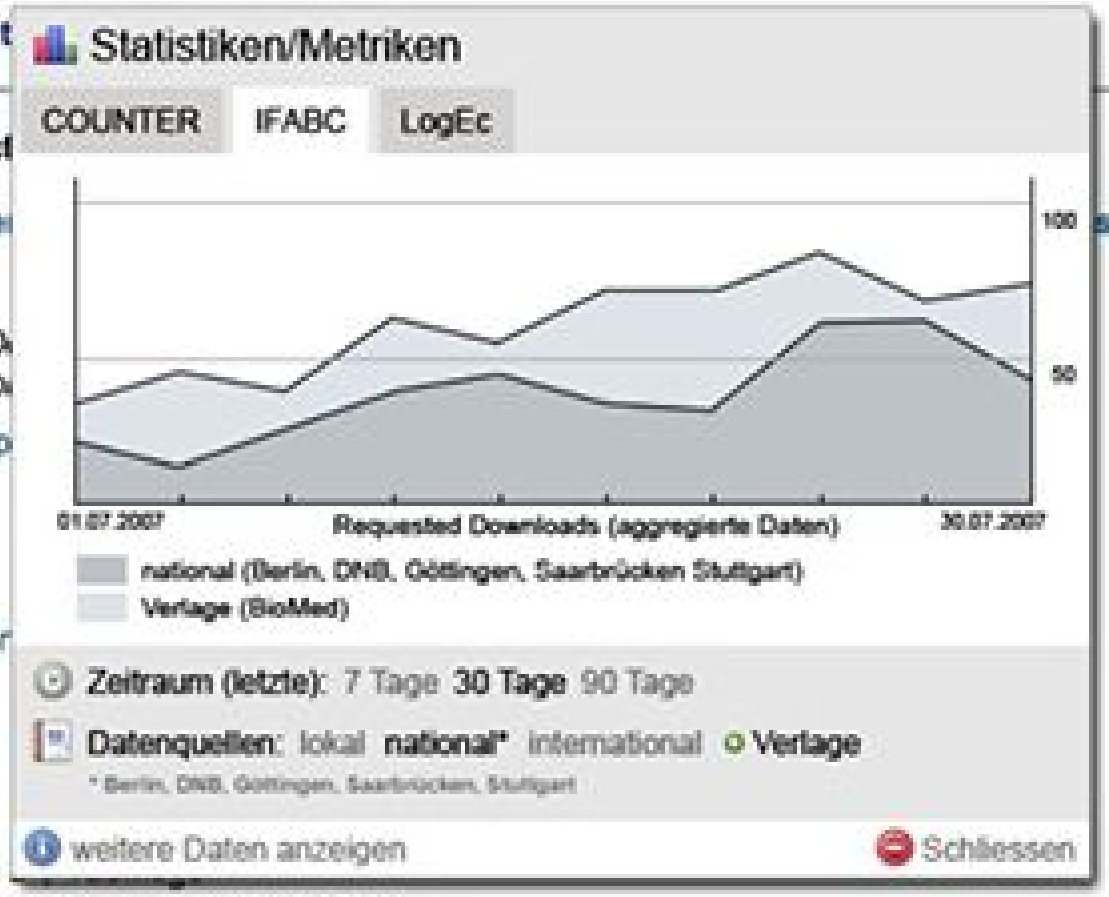
- For authors (usage follow-up), readers-scientists (relevance, alert), institutions (impact)
- Article level (= document)
- Tools for transfer/sharing (network)
- Added-value services

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

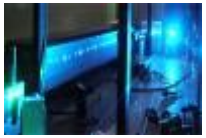
Protein Kinase D regulates several aspect

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



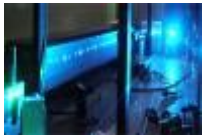
Towards standardization: IFREMER



Advice for optimization of deposit visibility (referencing)

One unique file with all (meta)data
Adding information to PDF properties
Downsizing the file
(…)

Towards standardization: other websites, projects

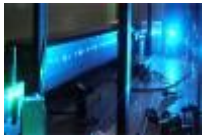


LogEc http://logec.repec.org/ Usage statistics of the RePEc repository

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

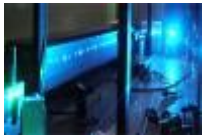
Recommendations (1)



- Recipient: authors, users, institutions
- COUNTER principle: different levels, with a basic minimum level (*AR1*)
- Selection of minimum elements for a basic log analysis

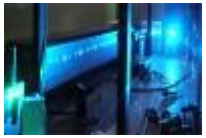
  (who, what, request type, when, identifier)

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

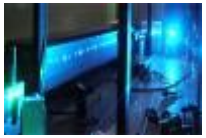
Link with other tools measuring the impact of deposited items (citations, tagging etc.)

(…)

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

Forthcoming

2010 IRIS case study (Lille 1)



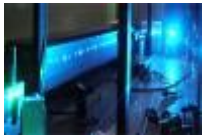
2010 Final report of DUAO-F project

2010 Study on search engins

??? Partership 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
++ (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