



# Usage data: Workshop Objectives from the perspective of the DINI - DFG - JISC projects

Frank Scholze  
Nils Windisch

**JISC Usage Statistics Workshop  
Humboldt University Berlin,  
Erwin Schrödinger Center, 7.-8. July 2008**



## Cluster of projects

- open-access.net
- OA Repository Network
- OA Statistics
- OA Citations
- DRIVER
- JISC usage statistics review !





## DINI certificate on usage statistics

- Minimum standards
  - Every document repository must keep consistent access statistics (web server log files)
  - Web server log files must be anonymized for long-term archiving purposes
  - Criteria used to collect or filter the data must be documented
- Recommendations
  - Access to documents by automated agents, robots or similar is filtered out (and documented)
  - Web server log files are edited according to the Counter Code of Practice
  - Access statistic is attached to the document as dynamic metadata and visible to the end-user



## What do we count?

- Practical definition of meaningful items
  - Files?
  - Publications (journal articles etc.)?
- Identification of meaningful items
  - Checksums
  - Persistent identifiers
  - Distributed heterogeneous publication network





## How do we count?

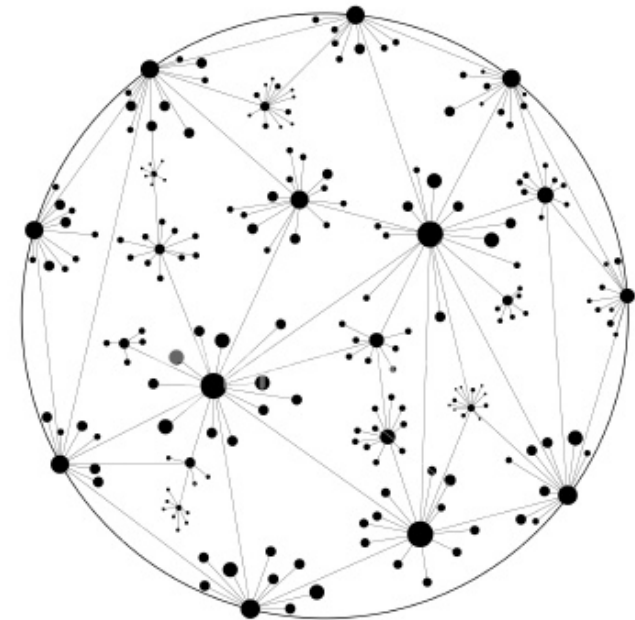
- Practical and pragmatic definition of usage
  - Access
  - Click spans
  - Definition of non-human access
  - Pseudonymization
  - Deleting or tagging
  - Sessions





## How do we aggregate?

- Technically
  - processing on which level
  - OpenUrl ContextObjects
  - SUSHI
- Organisationally
  - DRIVER
  - OA Repository Network
  - ...
- Co-operation

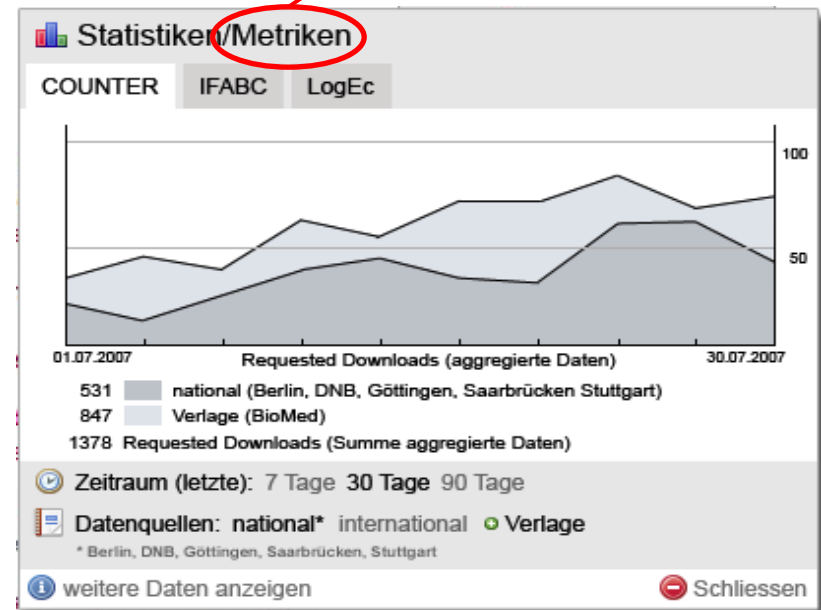




## What do we report?

- Access over time
- Sources of aggregation
- Standards for processing

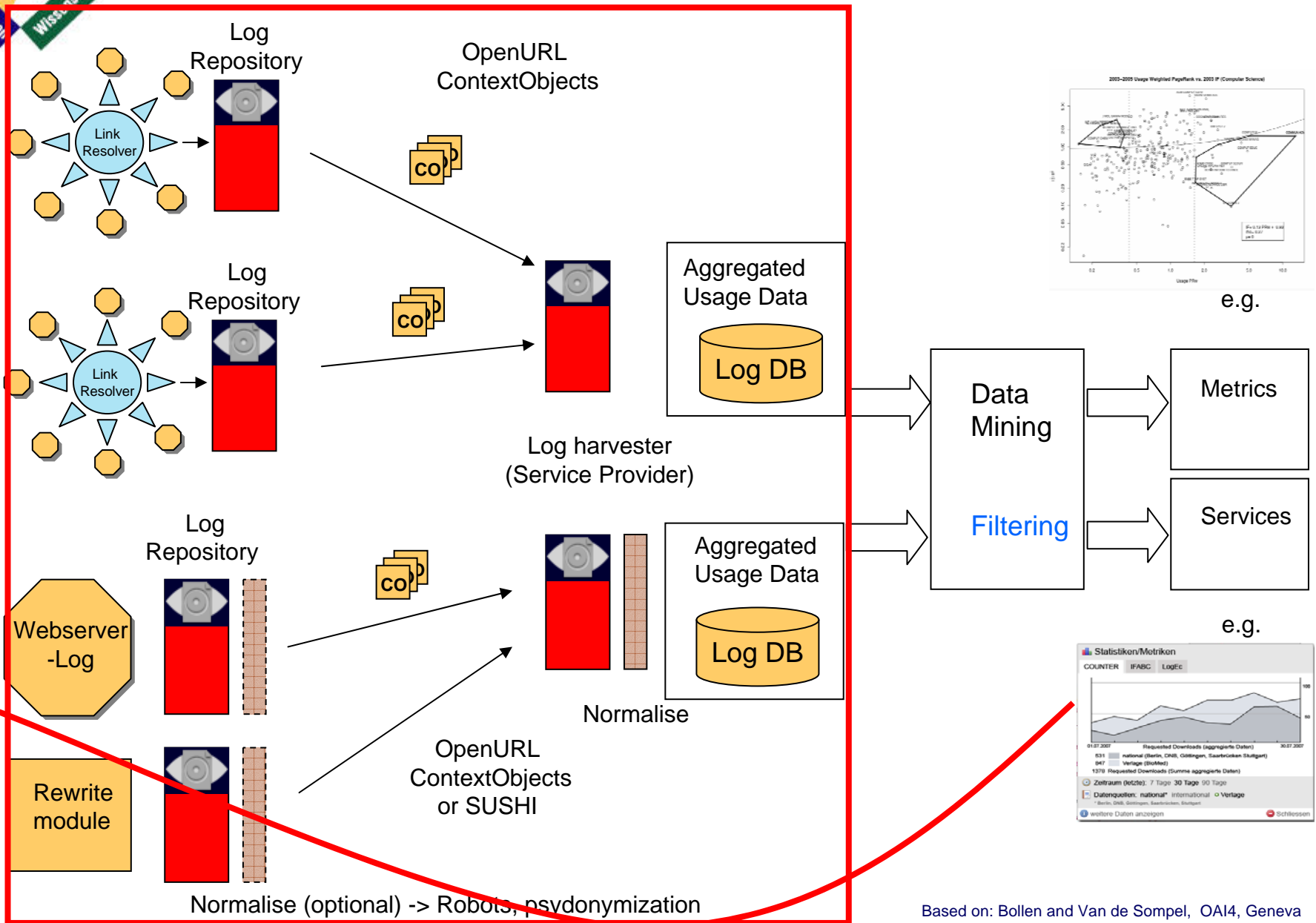
Johan will tell us more



➔ Transparency on the what and the how of counting



# Infrastructure for aggregating usage data





# Open Access Statistics

realize what others had in mind...

# Project

- **Funded by: DFG (German Research Foundation)**
- **18 months 2008-07-01 – 2009-12-31**
- **Partner: Berlin (CMS) Göttingen (SUB), Saarbrücken (SUUB), Stuttgart (UB)**

# Context

- One of three projects:
  - Open Access Network of Repositories (OA-N)
  - Open Access Citation (DOARC, Distributed Open-Access Reference Citation services)
  - Open Access Statistics (OA-S)

# Objectives

- Aggregate and normalize usage data locally
- Act as data provider
- Collect data at service provider level
- Process data to provide added values services

Aufsatz zugänglich unter  
URN: [urn:nbn:de:bsz:93-opus-32590](http://nbn-resolving.org/urn:nbn:de:bsz:93-opus-32590)  
URL: <http://elib.uni-stuttgart.de/opus/volltext.php?id=11111>


**Protein Kinase D regulates several aspects of cell growth and survival**

Maier, Dieter ; Nagel, Anja C. ; Gloc, Helmut


Originalveröffentlichung: (2007) BMC Developmental Biology  
<http://www.biomedcentral.com/10.1186/1471-2219-8-111>

pdf-Format: Dokument 1.pdf

**Statistiken:** 

**Gedruckte Ausgabe:**  Print-out

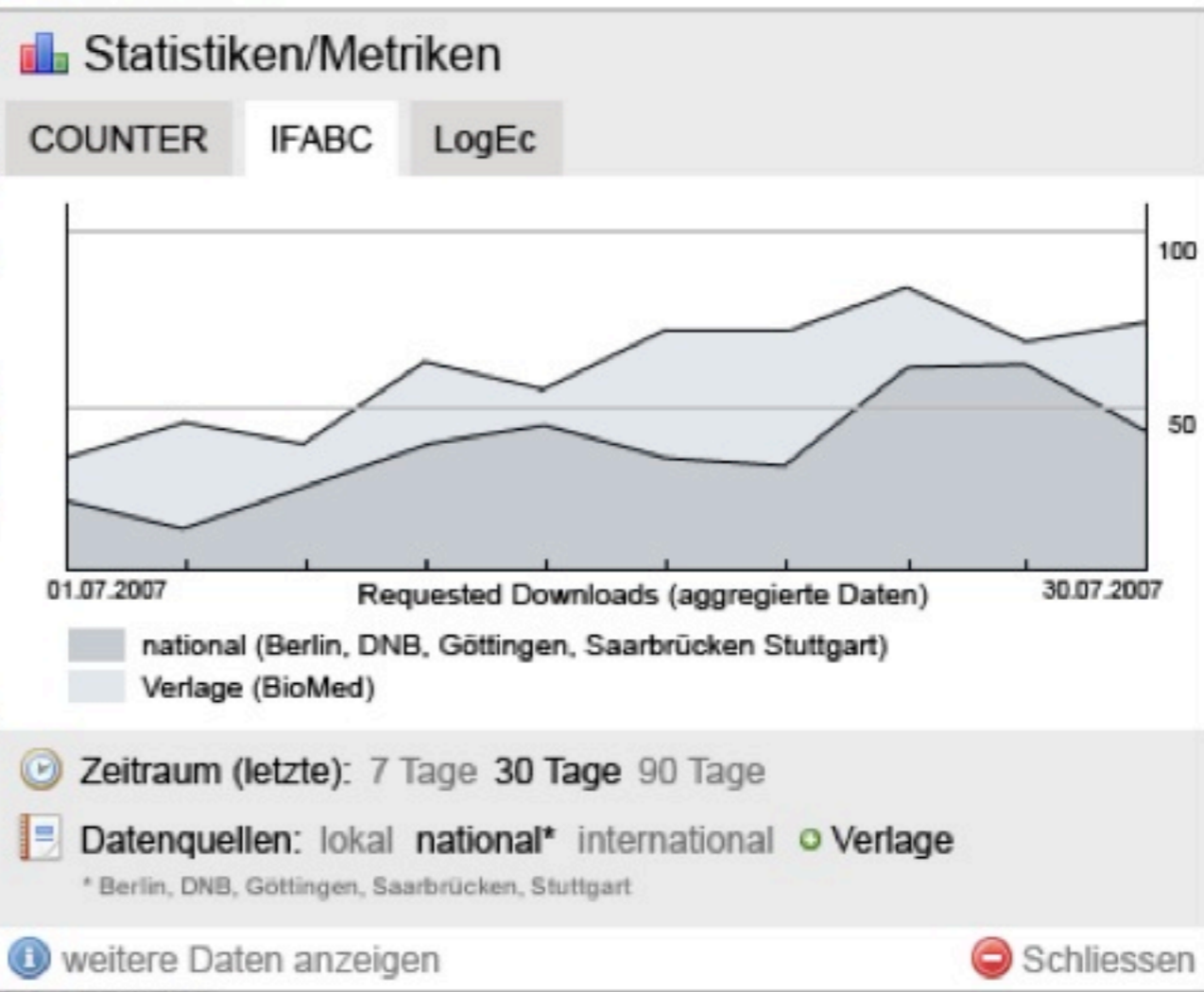
**Dokument empfehlen:** 

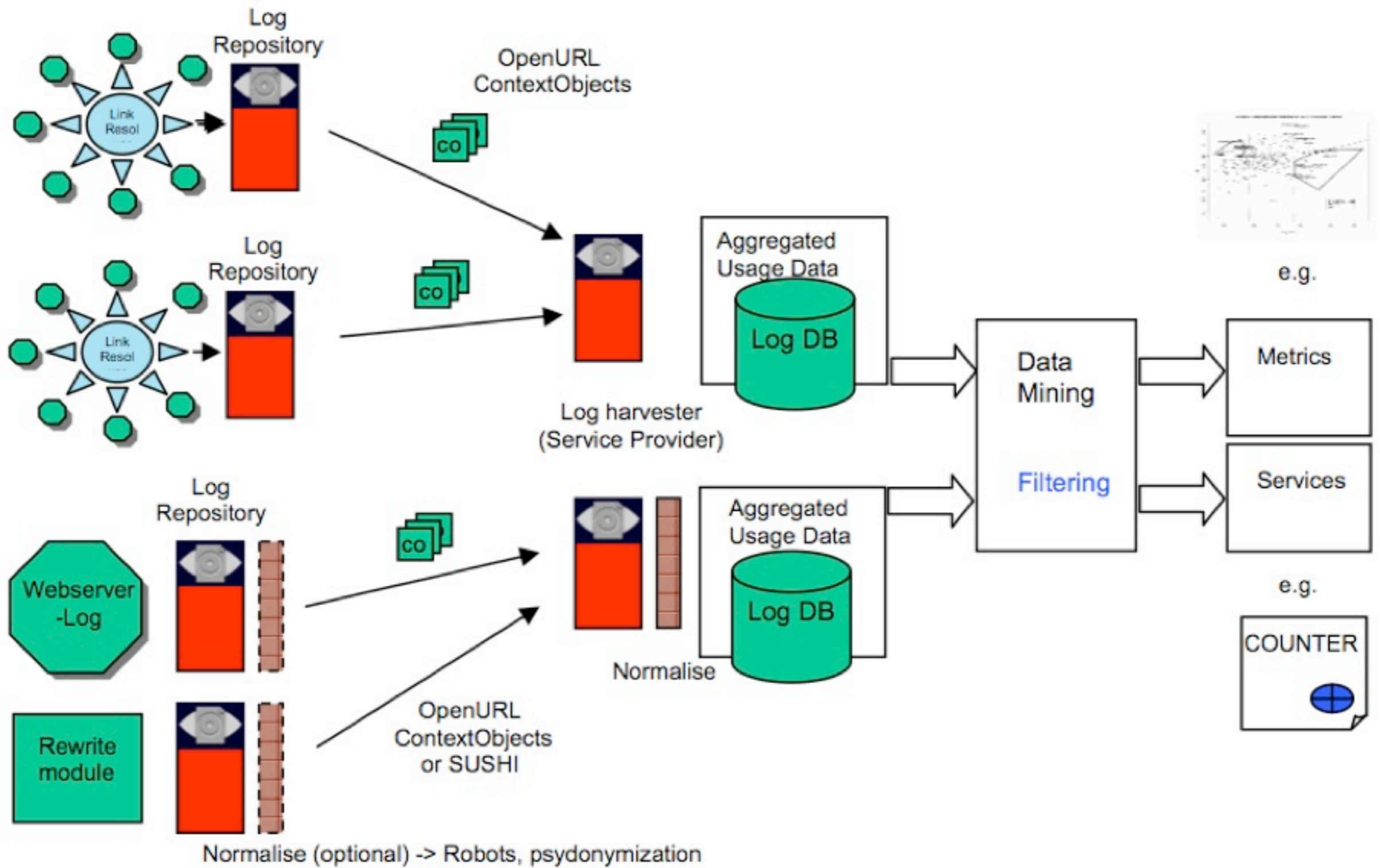
**Dokument in Google Scholar suchen:** 

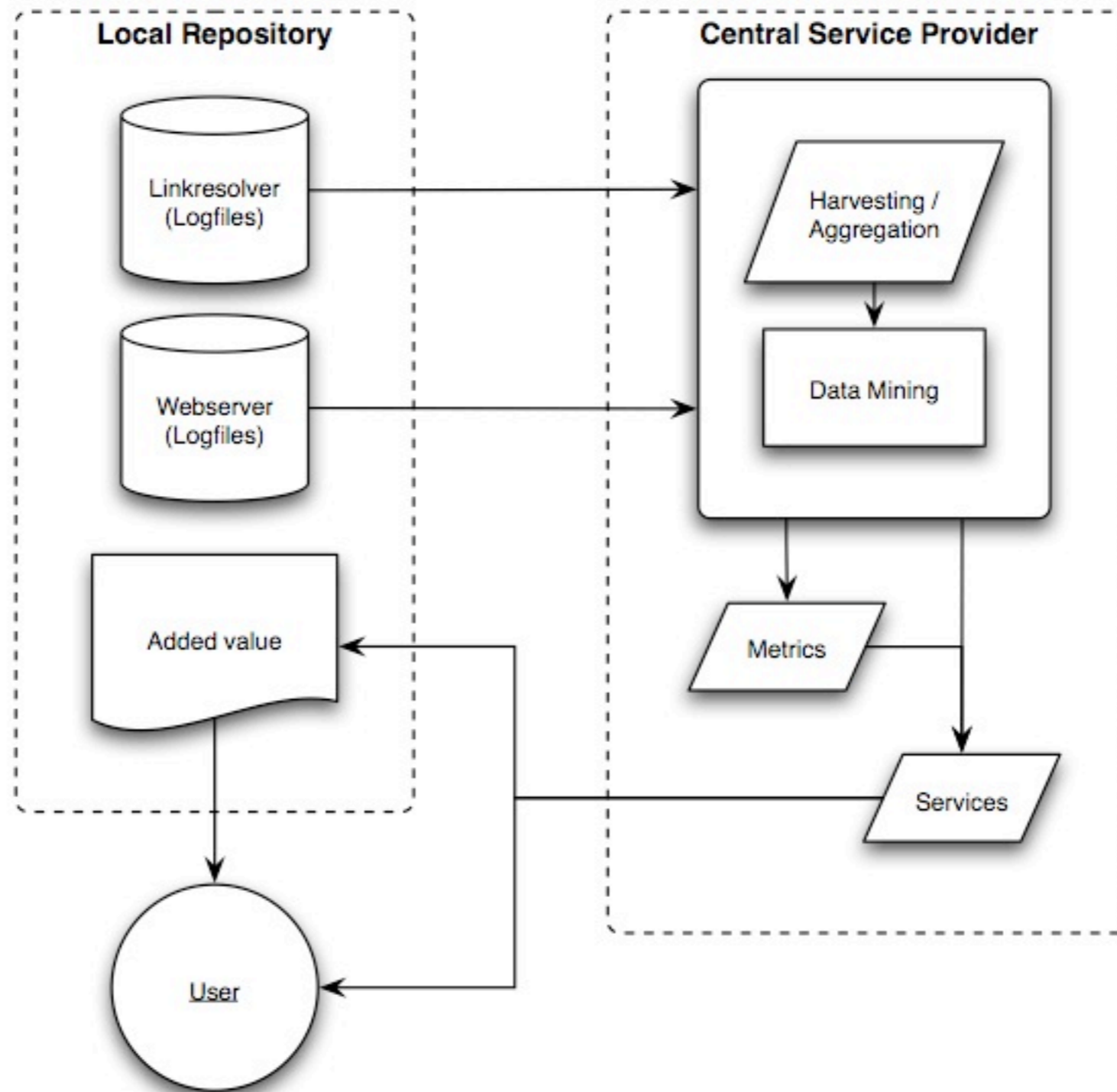
**SWD-Schlagwörter:** Proteinkinase D

**Institut:** Institut für Zellbiologie und Immunologie

**DDC-Sachgruppe:** Biowissenschaften, Biologie



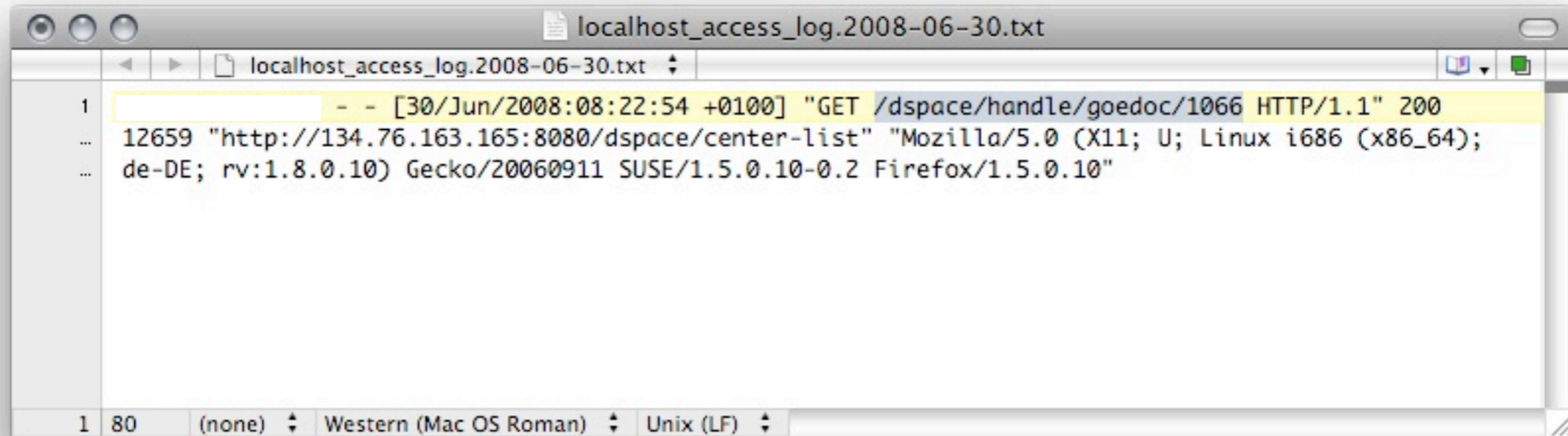




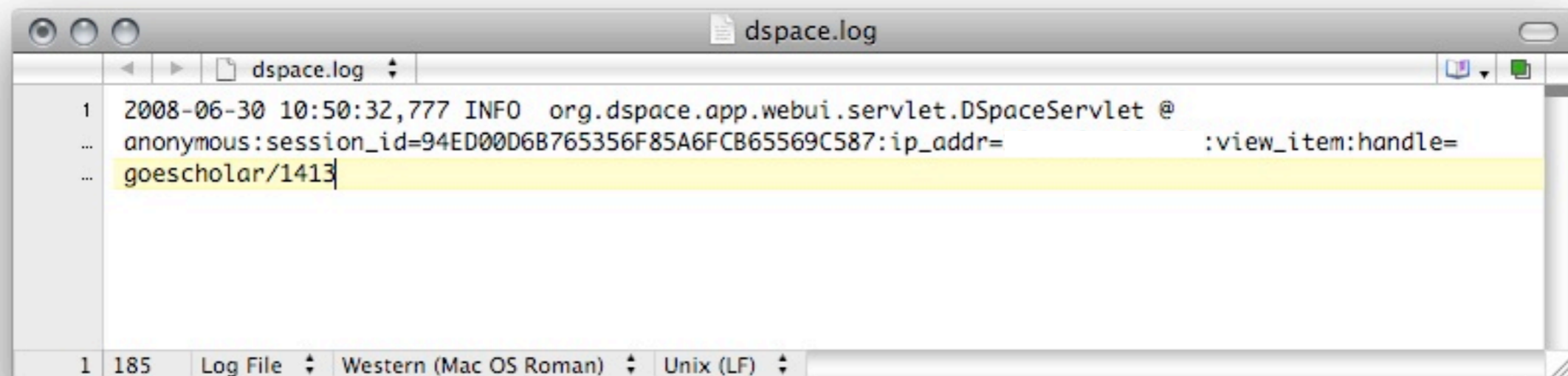
# Deal with usage data

- Different sources
  - License server (HAN-Server)
  - Link resolver (SFX)
  - Repository software (DSpace, OPUS, e-doc, etc.)





A screenshot of a text editor window titled "localhost\_access\_log.2008-06-30.txt". The window shows a single log entry on line 1, which is highlighted in yellow. The entry is: `- - [30/Jun/2008:08:22:54 +0100] "GET /dspace/handle/goedoc/1066 HTTP/1.1" 200 12659 "http://134.76.163.165:8080/dspace/center-list" "Mozilla/5.0 (X11; U; Linux i686 (x86_64); de-DE; rv:1.8.0.10) Gecko/20060911 SUSE/1.5.0.10-0.2 Firefox/1.5.0.10"`. The status bar at the bottom shows "1 80 (none) Western (Mac OS Roman) Unix (LF)".

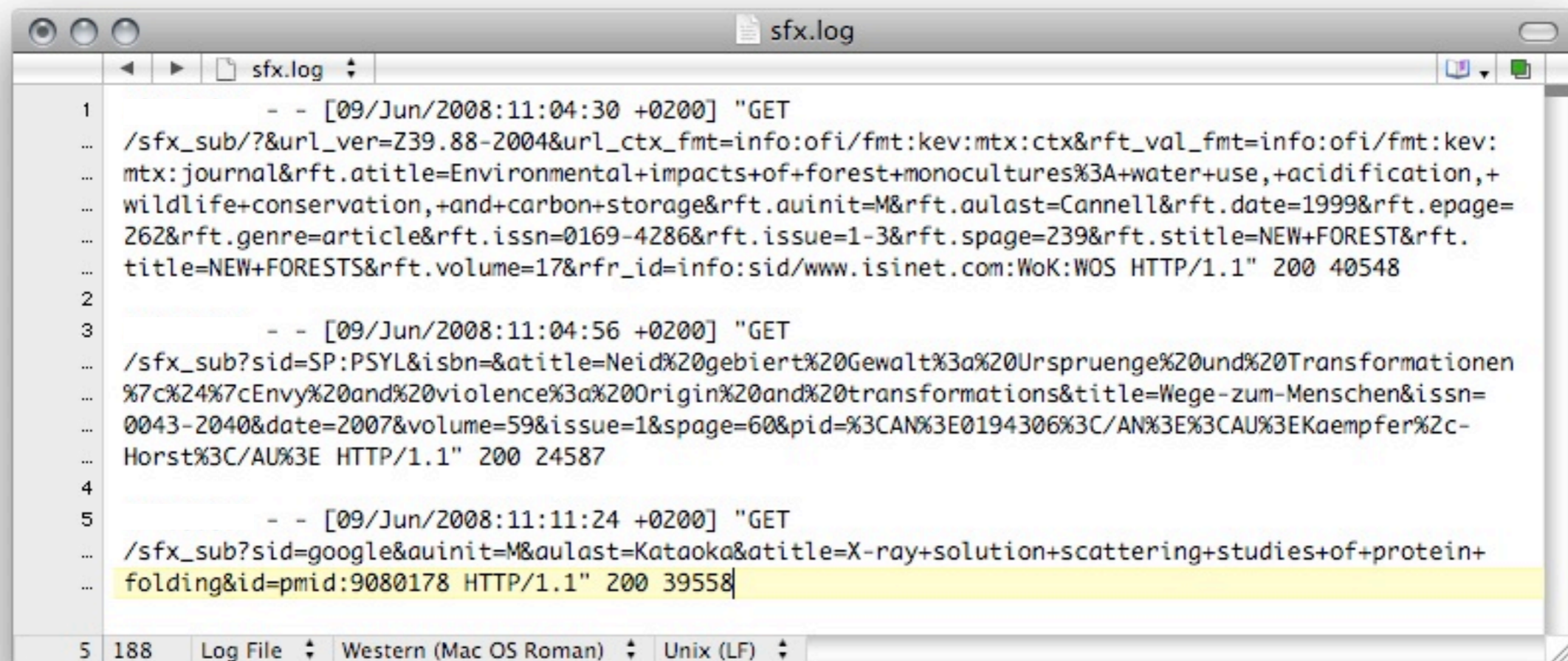


A screenshot of a text editor window titled "dspace.log". The window shows a log entry on line 1, which is highlighted in yellow. The entry is: `2008-06-30 10:50:32,777 INFO org.dspace.app.webui.servlet.DSpaceServlet @ anonymous:session_id=94ED00D6B765356F85A6FCB65569C587:ip_addr= :view_item:handle= goescholar/1413`. The status bar at the bottom shows "1 185 Log File Western (Mac OS Roman) Unix (LF)".

# Repository software (DSpace)

```
han.log
1 ;4548;00071715088;09/Jun/2008:11:20:39;GET;/han/ISIWebofScience/apps.isiknowledge.com/gratuitous.jpg?URL=http%3A//han.sub.uni-goettingen.de/han/ISIWebofScience/apps.isiknowledge.com/summary.do&PageStart=1213003230890&LoadTime=688
2
3 ;4268;-;09/Jun/2008:11:20:40;GET;/ezb/start?ezbid=794&title=Archives+of+Microbiology+%28Formerly%3A+Archiv+f%FCr+Mikrobiologie%29&publisher=Springer&zdbid=1458451&eissn=1432-072X&subject=W&l=%3B%3B%3B%3B%3B%3B%3B%3Bnds_springer&url=http%3A%2F%2Fwww.springerlink.com%2Flink.asp%3Fid%3D100459&hash=58d5f94ae0be12ace541cde17b3e86e1
4
5 ;4268;UNIGOE;09/Jun/2008:11:20:40;GET;/han/794
6 ;4268;UNIGOE;09/Jun/2008:11:20:40;GET;/han/794/www.springerlink.com/link.asp?id=100459
7 ;6564;UNIGOE - SERVER JURISTISCHE FAKULTAET;09/Jun/2008:11:20:40;GET;/han/PsycInfo/web.ebscohost.com/ScriptResource.axd?d=wANNmv2XogJUceTRBZqLh0S0RFuIUiMfvNnZ1S221Mk595AkCW-eHlpag0Ju78AzwMHEWVg7k1ZnX8W-7qhE1CxjZ4MBk4Y-X-3a0xWHm5w1&t=633456741902995000
```

# License server (HAN)



```
sfx.log
1      - - [09/Jun/2008:11:04:30 +0200] "GET
...  /sfx_sub/?url_ver=Z39.88-2004&url_ctx_fmt=info:ofi/fmt:kev:mtx:ctx&rft_val_fmt=info:ofi/fmt:kev:
...  mtx:journal&rft.atitle=Environmental+impacts+of+forest+monocultures%3A+water+use,+acidification,+
...  wildlife+conservation,+and+carbon+storage&rft.auinit=M&rft.aulast=Cannell&rft.date=1999&rft.epage=
...  262&rft.genre=article&rft.issn=0169-4286&rft.issue=1-3&rft.spage=239&rft.stitle=NEW+FOREST&rft.
...  title=NEW+FORESTS&rft.volume=17&rfr_id=info:sid/www.isinet.com:WoK:WOS HTTP/1.1" 200 40548
2
3      - - [09/Jun/2008:11:04:56 +0200] "GET
...  /sfx_sub?sid=SP:PSYL&isbn=&atitle=Neid%20gebiert%20Gewalt%3a%20Urspruenge%20und%20Transformationen
...  %7c%24%7cEnvy%20and%20violence%3a%20Origin%20and%20transformations&title=Wege-zum-Menschen&issn=
...  0043-2040&date=2007&volume=59&issue=1&spage=60&pid=%3CAN%3E0194306%3C/AN%3E%3CAU%3EKaempfer%2c-
...  Horst%3C/AU%3E HTTP/1.1" 200 24587
4
5      - - [09/Jun/2008:11:11:24 +0200] "GET
...  /sfx_sub?sid=google&auinit=M&aulast=Kataoka&atitle=X-ray+solution+scattering+studies+of+protein+
...  folding&id=pmid:9080178 HTTP/1.1" 200 39558
5 188  Log File  Western (Mac OS Roman)  Unix (LF)
```

# Link resolver (SFX)

# Map usage data

- Use established formats and rule sets
  - COUNTER
  - IFABC
  - LogEc

# from A to B

- SUSHI vs. OAI-PMH + OpenURL CO
  - *project objectives include evaluation of each approach*
- SUSHI/OAI-OMH just a transport container/vehicle
- (Usage) data as XML payload

# Context

- What about the JISC Usage Statistics Project and Workshop?
- Provide inside information
- Build on experience and expert opinions
- Re-use existing technologies