

## Wie konfiguriere und "bearbeite" ich meine Logfiles, um bei OA-Statistik mitzumachen?

### Workshop OA-Statistik:

„Was zählt? - Nutzungsstatistiken  
als alternative Impact-Messung“  
Göttingen, 21. Januar 2010

### Sabine Henneberger

Humboldt-Universität zu  
Berlin

[shenneberger@cms.hu-berlin.de](mailto:shenneberger@cms.hu-berlin.de)

Initiiert durch:

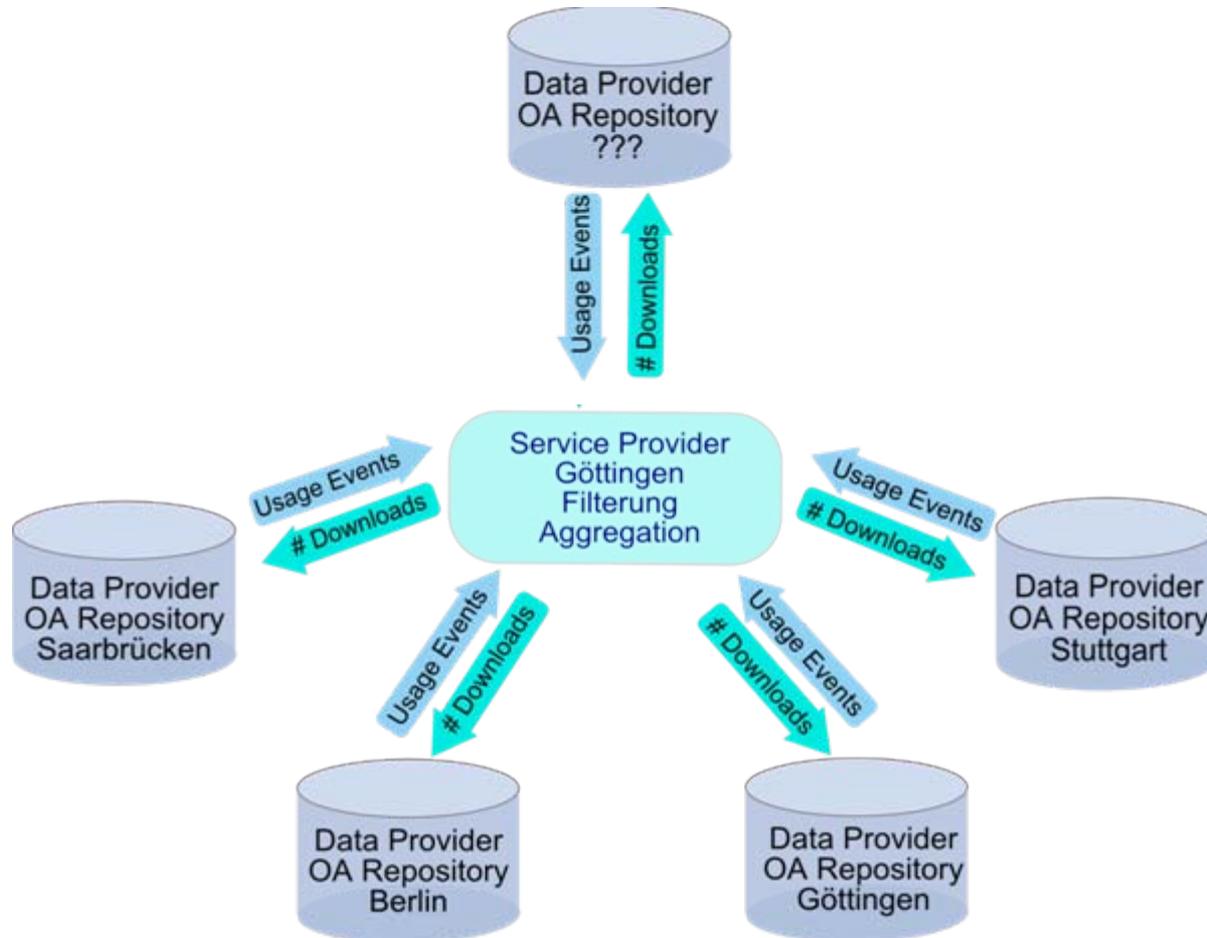


Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**

# Was heißt „Bei OA-Statistik mitmachen“?



# Wie wird ein Repository zum Data Provider für OA-Statistik?

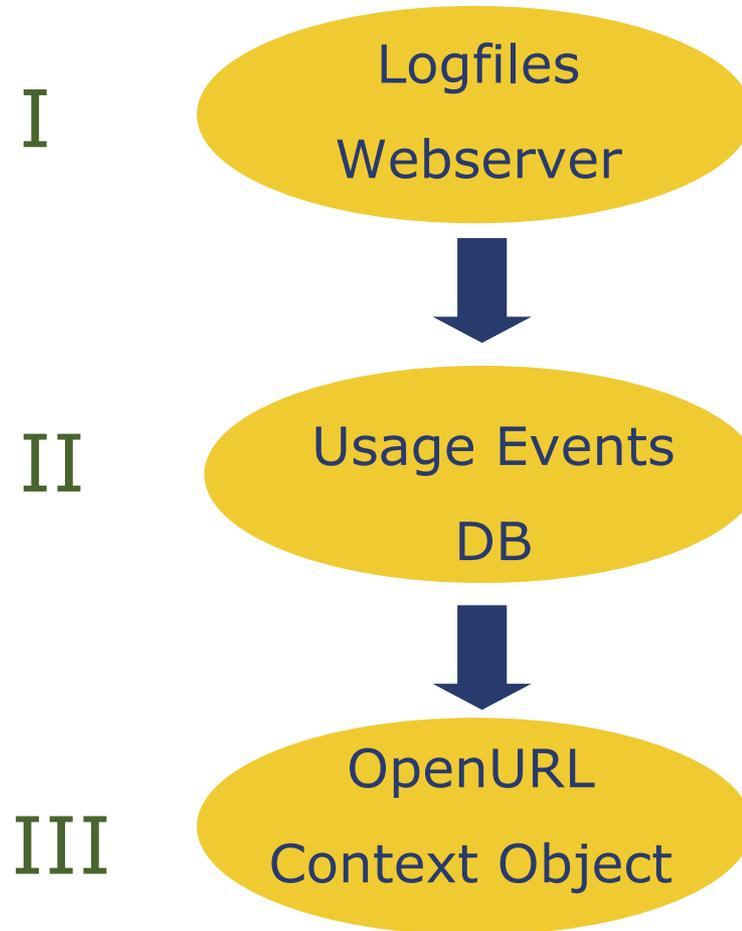
---

Bereitstellung von Usage Events, die der Service Provider verarbeiten kann:

- I. Erzeugung von Logfiles
- II. Aufbereitung der Logfiles, Speicherung in einer Datenbank
- III. Anbieten von OpenURL Context Objekten über OAI-PMH-Web-Schnittstelle

# Wie wird ein Repository zum Data Provider für OA-Statistik?

---



# I Die Logfiles des Webservers

---

## □ Konfiguration in httpd.conf:

```
# The following directives define some format nicknames for use with
# a CustomLog directive (see below).
#
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\"" combined
LogFormat "%h %l %u %t \"%r\" %>s %b" common
```

# I Die Logfiles des Webservers

## □ Ergebnis:

```
141.20.193.10 - - [30/Jun/2006:08:07:55 +0200] "GET  
/dissertationen/seifarth-joerg-2004-07-07/PDF/seifarth.pdf  
HTTP/1.1" 200 1200
```

```
"http://www.google.de/search?q=wann+wurde+die+erste+schul  
e+im+yukon+territory+errichtet&hl=de&lr=&start=10&sa=N"
```

```
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; Crazy  
Browser 1.0.5; .NET CLR 1.1.4322),"
```

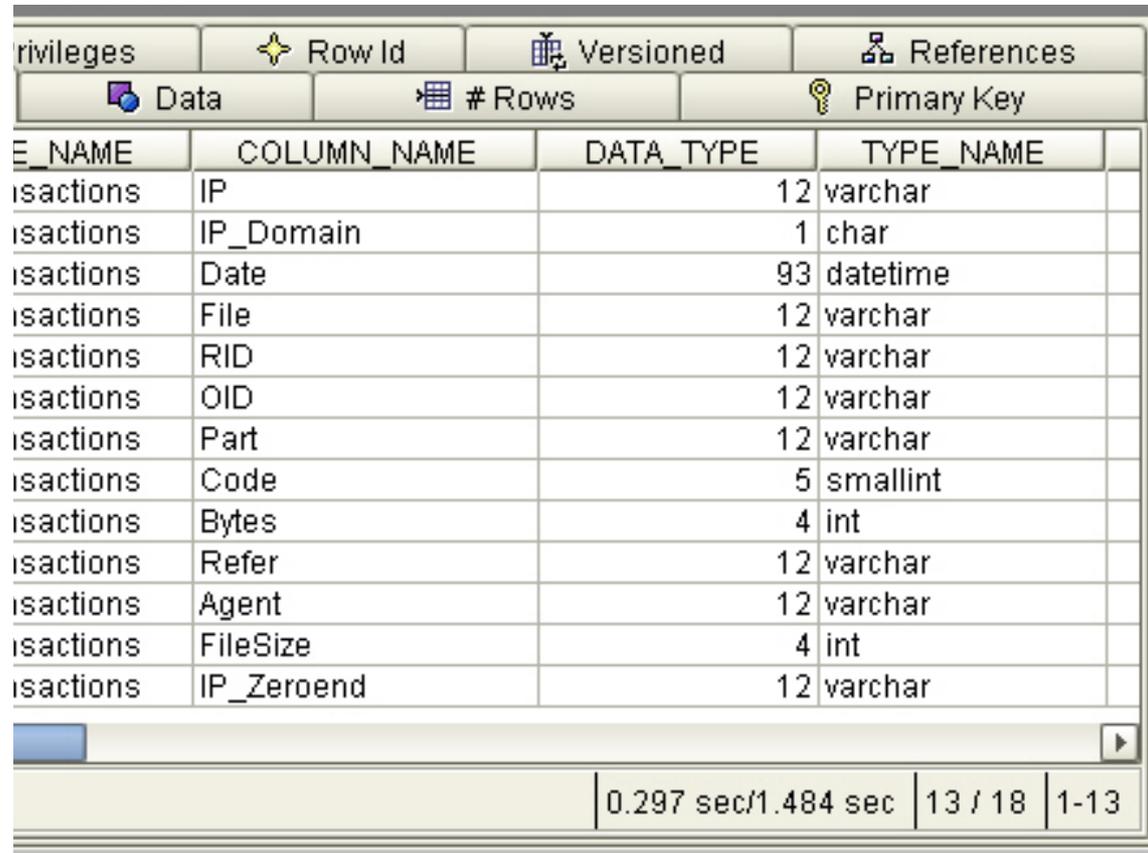
```
%h      %t   %or      %s      %b      %{Referer}      %{User-Agent}  
Host - - Date Request Status Byte      Referrer      Agent
```

## II Aufbereitung der Logfiles, Speicherung in einer Datenbank

---

- Pseudonymisierung Host
- Bestimmung der ID
- Bestimmung Filesize
- Speicherung in geeigneter Form

## II Aufbereitung der Logfiles, Speicherung in einer Datenbank



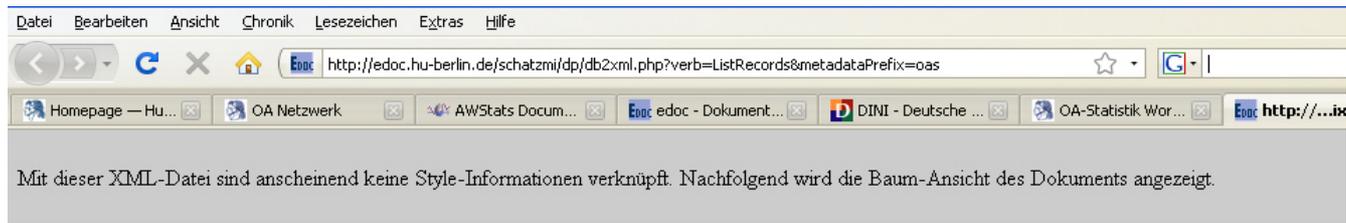
The screenshot shows a database management interface with a table structure for 'Logfiles'. The table has four columns: E\_NAME, COLUMN\_NAME, DATA\_TYPE, and TYPE\_NAME. The table contains 13 rows of data. The interface also shows a toolbar with icons for 'Privileges', 'Row Id', 'Versioned', 'References', 'Data', '# Rows', and 'Primary Key'. At the bottom, there is a status bar showing '0.297 sec/1.484 sec | 13 / 18 | 1-13'.

E_NAME	COLUMN_NAME	DATA_TYPE	TYPE_NAME
Logfiles	IP	12	varchar
Logfiles	IP_Domain	1	char
Logfiles	Date	93	datetime
Logfiles	File	12	varchar
Logfiles	RID	12	varchar
Logfiles	OID	12	varchar
Logfiles	Part	12	varchar
Logfiles	Code	5	smallint
Logfiles	Bytes	4	int
Logfiles	Refer	12	varchar
Logfiles	Agent	12	varchar
Logfiles	FileSize	4	int
Logfiles	IP_Zeroend	12	varchar

# II Aufbereitung der Logfiles, Speicherung in einer Datenbank

Spalte	Inhalt	Beispiel
IP	IP oder DNS-Name pseudonymisiert	54b82b80f8ddb44fabbcbbadc953e4e92e75410c835b05e032268f99179dc416
IP_Domain	Domain Name	hu-berlin.de
Date		2006-06-30 08:07:55
File		/dissertationen/seifarth-joerg-2004-07-07/PDF/seifarth.pdf
RID	ID Repository	26482
OID	ID OA-Netzwerk	2378
Part	Mime Type	application/pdf
Code	Status Code	200
Bytes	Download Size	1200
Refer	Referrer	http://www.google.de/search?q=wann+wurde+die+erste+schule+im+yukon+territory+errichtet&hl=de&lr=&start=10&sa=N
Agent	User Agent	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; Crazy Browser 1.0.5; .NET CLR 1.1.4322
FileSize	Filesize Filesystem	4084975
IP_Zeroend	C-class pseudonymisiert	b547e3ccd53d1f1e88d0f31317d74ee9ebb86e2cf7167df3129fa07300dd7152

# III Anbieten von OpenURL Context Objekten über OAI-PMH-Web-Schnittstelle



```
- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2010-01-18T15:14:27Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oas">http://edoc.hu-berlin.de/</request>
- <ListRecords>
  - <record>
    - <header>
      <identifier>oai:hu-berlin.de</identifier>
      <datestamp>2010-01-18</datestamp>
    </header>
    - <metadata>
      - <context-objects xsi:schemaLocation="info:ofi/fmt:xml:xd:ctx http://www.openurl.info/registry/docs/xsd/info:ofi/fmt:xml:xd:ctx">
        - <context-object timestamp="2008-06-05T03:03:00Z">
          - <administration>
            - <oa-statistics>
              <status_code>200</status_code>
              <size>30625</size>
              <document_size>30625</document_size>
              <format>application/octet-stream</format>
              <service>http://edoc.hu-berlin.de/</service>
            </oa-statistics>
          </administration>
        - <referent>
          <identifier>http://edoc.hu-berlin.de/ebind/dtd/ebind.dtd</identifier>
          <identifier>oai:HUBerlin.de:0</identifier>
        </referent>
      - <requester>
```

# III Anbieten von OpenURL Context Objekten über OAI-PMH-Web-Schnittstelle

```
<context-object timestamp="2006-06-30T08:07:00Z">
  <administration>
    <oa-statistics xmlns="http://dini.de/namespace/oas-info">
      <status_code>200</status_code>
      <size>1200</size>
      <document_size>4084975</document_size>
      <format>application/pdf</format>
      <service>http://edoc.hu-berlin.de/</service>
    </oa-statistics>
  </administration>
  <referent>
    <identifier>http://edoc.hu-berlin.de/dissertationen/seifarth-joerg-2004-07-07/PDF/seifarth.pdf</identifier>
    <identifier>oai:HUBerlin.de:26482</identifier>
  </referent>
  <referring-entity>
    <identifier>
      http://edoc.hu-berlin.de/http://www.google.de/search?q=wann+wurde+die+erste+schule+im+yukon+territory+errichtet&hl=de&lr=&start=10&sa=N</identifier>
    </referring-entity>
  <requester>
    <metadata-by-val>
      <format>http://dini.de/namespace/oas-requesterinfo</format>
      <metadata>
        <requesterinfo xmlns="http://dini.de/namespace/oas-requesterinfo">
          <hashed-ip>54b82b80f8ddb44fabbcbbadc953e4e92e75410c835b05e032268f99179dc416</hashed-ip>
          <hashed-c>b547e3ccd53d1f1e88d0f31317d74ee9ebb86e2cf7167df3129fa07300dd7152</hashed-c>
          <hostname>hu-berlin.de</hostname>
          <user-agent>Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; Crazy Browser 1.0.5; .NET CLR 1.1.4322</user-agent>
        </requesterinfo>
      </metadata>
    </metadata-by-val>
  </requester>
</context-object>
```

**Vielen Dank für Ihre Aufmerksamkeit!**

## **Workshop OA-Statistik:**

„Was zählt? - Nutzungsstatistiken  
als alternative Impact-Messung“  
Göttingen, 21. Januar 2010

## **Sabine Henneberger**

Humboldt-Universität zu  
Berlin

[shenneberger@cms.hu-berlin.de](mailto:shenneberger@cms.hu-berlin.de)

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**