

# Welcher Aufwand ist beim Repository nötig?

**Workshop OA-Statistik:**  
„Was zählt? -  
Nutzungsstatistiken als  
alternative Impact-Messung“  
Göttingen, 21. Januar 2010

**Björn Mittelsdorf**

Saarländische Universitäts- und  
Landesbibliothek

[b.mittelsdorf@sulb.uni-saarland.de](mailto:b.mittelsdorf@sulb.uni-saarland.de)

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**

# Ablauf

---

1. Ressourcen und Infrastruktur
2. Webserverkonfiguration
3. Parsing der Logdatei
4. Erstellung des Transferformats
5. Rückmeldung der Nutzungskennzahlen

# Konfiguration – Webserver

---

- Apache Webserver konfigurieren
  - mod\_log\_config:
    - Mindestens: Extended (Combined) Logformat
  
  - Optional: /etc/sysconfig/apache2
    - Mod\_usertrack
    - Mod\_headers
  
  - Optional: HTTP\_X\_Forwarded\_For, etc.

# Konfiguration – Webserver - Module

---

<VirtualHost 1.1.1.1>.conf (bzw. httpd.conf)

- Mod\_log\_config
  - CustomLog /var/log/apache2/scidok\_access\_log combined\_proxy
  
- Mod\_usertrack
  - CookieTracking on
  - CookieExpires "2 months"
  
- Mod\_headers
  - Header set Expires 0
  - Header set Cache-Control no-cache
  - Header add Cache-Control no-store

# Einrichten der lokalen Skripte

---

- Skript Interpreter einrichten
  - PHP oder PERL
    - Hash-Algorithmus SHA-256 bereitstellen
    - PHP: Memory Limit erhöhen auf mindestens 16MB
    - Eventuell: DNS-Lookup einrichten
    - Optional: PHP Session ID

# Anpassung der lokalen Skripte

---

- Skripte installieren (PHP und MySQL)
  - Skripte konfigurieren
    - Reguläre Ausdrücke anpassen
      - [http://psydok....../frontdoor.php?source\\_opus=12](http://psydok....../frontdoor.php?source_opus=12)
      - oai:psydok.uni-saarland.de:12
    - Zeitplan aufstellen (crontab)
  - Geschützter Ordner für Parser-Skripte
    - Logdatei für Skriptmeldungen anlegen
  - Htdocs-Ordner für die OAI-PMH-Schnittstelle

# Datenfluss sicherstellen

---

1. Webserver
2. Lokaler OA-S Datenspeicher
  1. Ordner für Logdateien
  2. Ordner/Datenbank für Context-Objects
3. OA-S Data Provider Schnittstelle
  1. Stabile URL
  2. Htaccess einrichten (Zugriffskontrolle)
    - Optional: Kompression des ausgehenden Datenstroms zum SP (ob\_gzhandler)

# Anpassung des Logformats (Minimum)

---

LogFormat (combined):

%h

%l %u (HTTP-Authentifizierung: unwichtig)

%t

"%r"

%>s

%b

"%{Referer}i"

"%{User-Agent}i"

# Anpassung des Logformats (Erweiterung)

---

LogFormat (combined):

- %h
- ...
- "%{User-Agent}i"

LogFormat (combined):

- %h
- ...
- "%{User-Agent}i"
  
- Wahre Dateigröße
- Teilstückinformation
- Wahre IP-Nummer
- Session Tracking

# Anpassung des Logformats (Praktische Umsetzung)

---

## LogFormat

- ❑ %h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\"
- ❑ \"%{Content-Type}o\"
- ❑ \"%{Content-Length}o\"
- ❑ \"%{Content-Range}o\"
- ❑ \"%{X\_FORWARDED\_FOR}i|{%VIA}i|{%X\_NOKIA\_IPADDRESS}i\"
- ❑ \"%{cookie}n\"

# Bearbeitung der Logdaten

---

- Pseudonymisierung
  - IP
  - C-Klasse-Netz
- Isolation der Dokumentidentifikation
  - OPUS Ident-Nummer
  - URL
  - URN
- Optional:
  - Geolocation
    - GeoIP Extension (für PHP)
    - GeoIP Database (z.B. Maxmind GeoIP Database)
  - User IP (Firewalls, Thin Clients etc.)

# Datentransfer zum Service Provider

---

- OAI-PMH-Container
  - OpenURL Context Objects-Container
    - OpenURL Context Object
    - OpenURL Context Object
    - ...

# OpenURL Context Object

---

Ein Context Object enthält im Fall von OpenAccess-Statistik folgende Informationen in den Feldern:

- Referent: genutztes Dokument (aktueller Fokus)
- ReferringEntity: Herkunftsdocument
- Requester: eindeutige und anonyme Besucheridentifikation
- ServiceType: Art der Nutzung (Volltext, Abstractseite)
- Resolver: benutzter Linkresolverserver (falls zutreffend)
- Referrer: Erzeuger des Context Object

# OpenURL Context Object (Beispiel) I

---

```
<administration>  
  <oa-statisticsxmlns="...">  
    <status_code>200</status_code>  
    <size>234500</size>  
    <document_size>234567</document_size>  
    <format>application/pdf</format>  
    <service>http://webdoc.sub.gwdg.de/</service>  
  </oa-statistics>  
</administration>
```

# OpenURL Context Object (Beispiel) II

---

<referent>

<identifizier>.../diss/2000/kirchner/ii-2.Pdf</identifizier>

<identifizier>urn:nbn:de:7-webdoc-12345</identifizier>

<identifizier>oai:elib.uni-stuttgart.de-opus:100</identifizier>

</referent>

# OpenURL Context Object (Beispiel) III

---

```
<referring-entity>  
  <identifier>  
    http://.../goescholar/handle/goescholar/2088  
  </identifier>  
  <identifier>urn:isbn:3-7643-6612-5</identifier>  
</referring-entity>
```

# OpenURL Context Object (Beispiel) IV

---

<requester>

...

<hashed-ip>

cfbf2a774df596123dd001cee172dae5519660b6014056f53a

</hashed-ip>

<hashed-c>

519660b6014056f5346d05e629c508bdcfbf2a774df596123d

</hashed-c>

<hostname>uni-saarland.de</hostname>

<user-agent>

mozilla/5.0 (...) gecko/20061204

</user-agent>

...

</requester>

# OpenURL Context Object (Beispiel) V

---

```
<service-type>
  <metadata-by-val>
    <format>info:ofi/fmt:xml:xsd:sch_svc</format>
    <metadata>
      <abstract xmlns="info:ofi/fmt:xml:xsd:sch_svc">
        yes
      </abstract>
    </metadata>
  </metadata-by-val>
</service-type>
```

# OpenURL Context Object (Beispiel) VI

---

<resolver>

    <identifier>http://sfx.gbv.de:9004/sfx\_sub/</identifier>

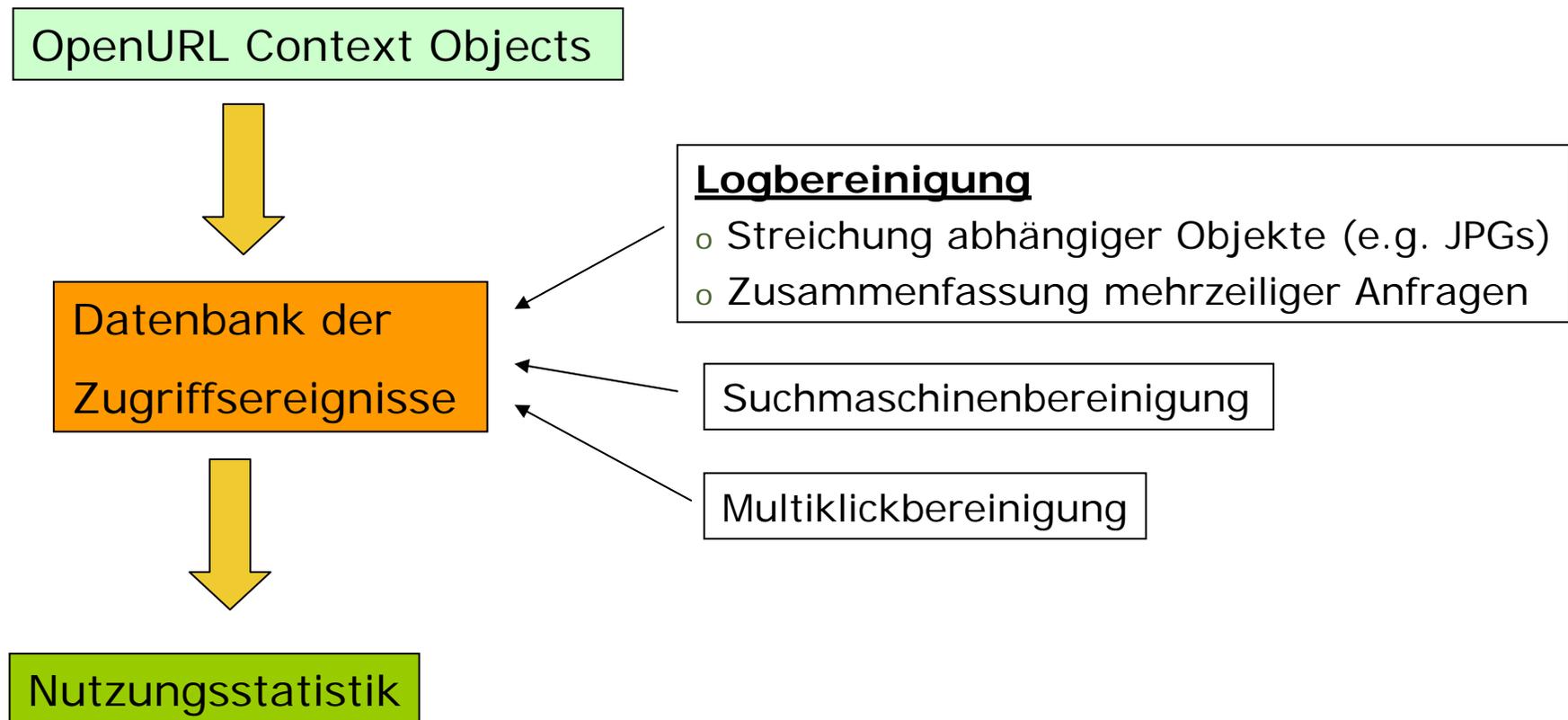
</resolver>

<referrer>

    <identifier>info:sid/dlib.org:dlib</identifier>

</referrer>

# Datenverarbeitung im Service Provider



# Rücktransfer zum Repository

---

- Transportprotokolle
  - OAI-PMH
  - RSS
  
- Syntax
  - XML
  
- Zeitliche Auflösungsebene:
  - Monate
  
- Objektebene
  - Dokumente (Volltexte)

# Datentransfer zum Repository (Beispiel)

---

```
<frequencies>
```

```
  <frequency for="urn:nbn:de:123" year="2009" month="May" standard="COUNTER">
```

```
    42
```

```
  </frequency >
```

```
  < frequency for="urn:nbn:de:..." year="..." month="..." standard="...">
```

```
    n
```

```
  </frequency >
```

```
  ...
```

```
</frequencies >
```

# Vielen Dank für Ihre Aufmerksamkeit!

**Workshop OA-Statistik:**  
„Was zählt? -  
Nutzungsstatistiken als  
alternative Impact-Messung“  
Göttingen, 21. Januar 2010

**Björn Mittelsdorf**

Saarländische Universitäts- und  
Landesbibliothek

[b.mittelsdorf@sulb.uni-saarland.de](mailto:b.mittelsdorf@sulb.uni-saarland.de)

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**

# Literatur

---

## □ Apache Dokumentation

- (<http://httpd.apache.org/docs/>)

## □ OpenURL Context Object

- (<http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=ListSets>)

## □ PHP Dokumentation

- (<http://www.php.net/docs.php>)

## □ RFC 2616: Hypertext Transfer Protocol -- HTTP/1.1

- (<http://www.w3.org/Protocols/rfc2616/rfc2616.html>)