

# Technische Umsetzung der Datenerfassung und -übermittlung am Beispiel von OPUS und DSpace

Open-Access-Tage 2009  
in Konstanz

**Björn Mittelsdorf**

SULB, Saarbrücken

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**

# Ablauf

---

1. Webserverkonfiguration
2. Parsing der Logdatei
3. Erstellung des Transferformats
4. Verarbeitung im Service Provider
5. Rückmeldung der Nutzungskennzahlen

# Anpassung der lokalen Infrastruktur

---

## Apache Webserver konfigurieren

- mod\_log\_config: Extended (Combined) Logformat
- Optional: Apache mod\_usertrack
- Optional: PHP Session ID
- Optional: HTTP\_X\_Forwarded\_For, etc.
- Skript Interpreter einrichten
  - PHP oder PERL
    - PHP: Memory Limit erhöhen auf mindestens 16MB
    - Eventuell: DNS-Lookup einrichten
- Skripte installieren (PHP und MySQL)
- Datenfluss sicherstellen
  1. Webserver
  2. Lokaler OA-S Datenspeicher
  3. OA-S Data Provider Schnittstelle (beschränkter Zugriff)
    - Optional: Kompression des ausgehenden Datenstroms zum SP

# Anpassung der lokalen Infrastruktur (Beispiel)

---

LogFormat:

%h

%l %u (HTTP-Authentifizierung: unwichtig)

%t

"%r"

%>s

%b

"%{Referer}i"

"%{User-Agent}i"

"%{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
  - 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>

<session-id>gvi8276</session-id>

<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>

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

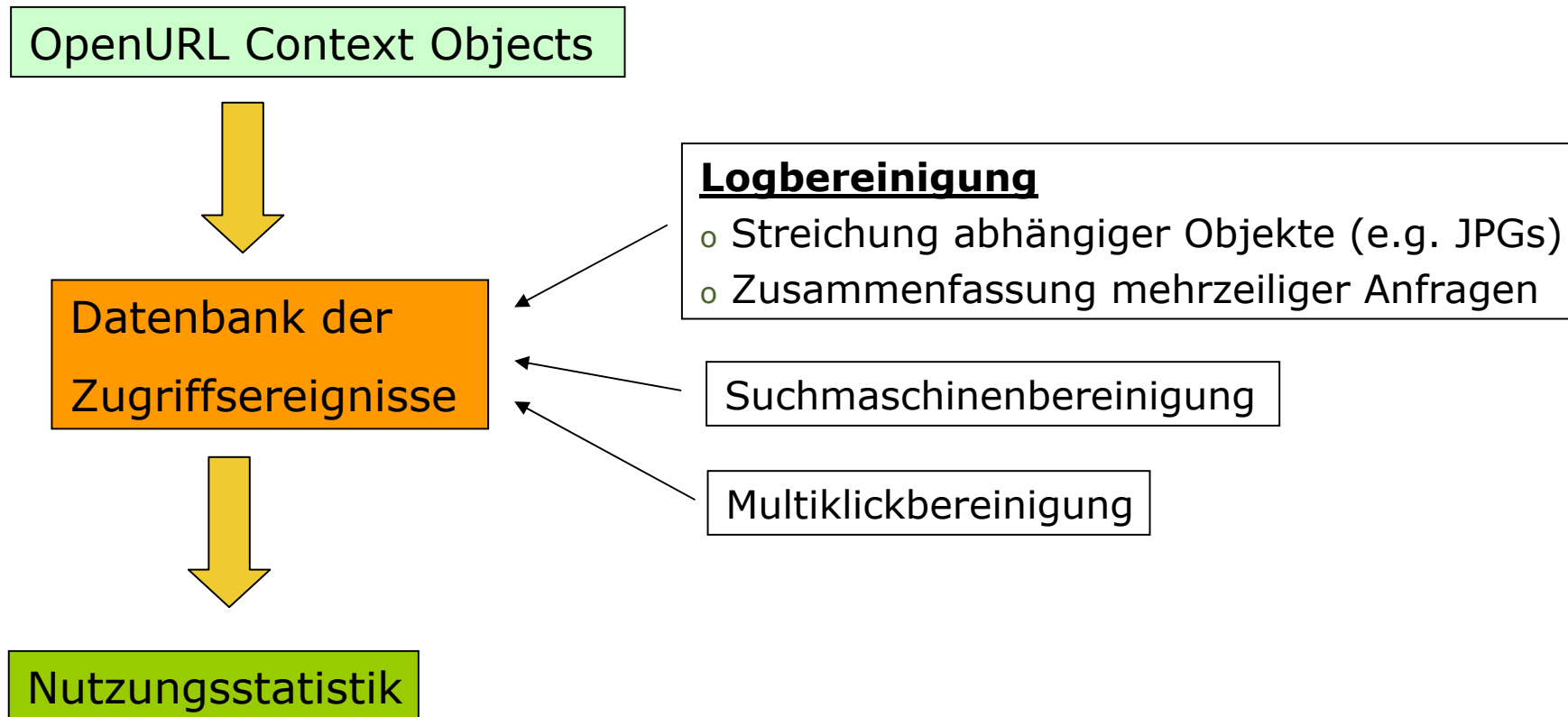
</resolver>

<referrer>

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

</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="oai:psydok...:100" year="2009" month="May" standard="COUNTER">
```

```
    42
```

```
  </frequency >
```

```
  < frequency for="oai-id" year="..." month="..." standard="...">
```

```
    n
```

```
  </frequency >
```

```
  ...
```

```
</frequencies >
```



# Vielen Dank für Ihre Aufmerksamkeit!

Open-Access-Tage 2009  
in Konstanz

**Björn Mittelsdorf**

SULB, Saarbrücken

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

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**