



Leitfaden für die Teilnahme an OA-Statistik

"Vom Projekt zum Dienst"

Marco Recke
recke@gbv.de

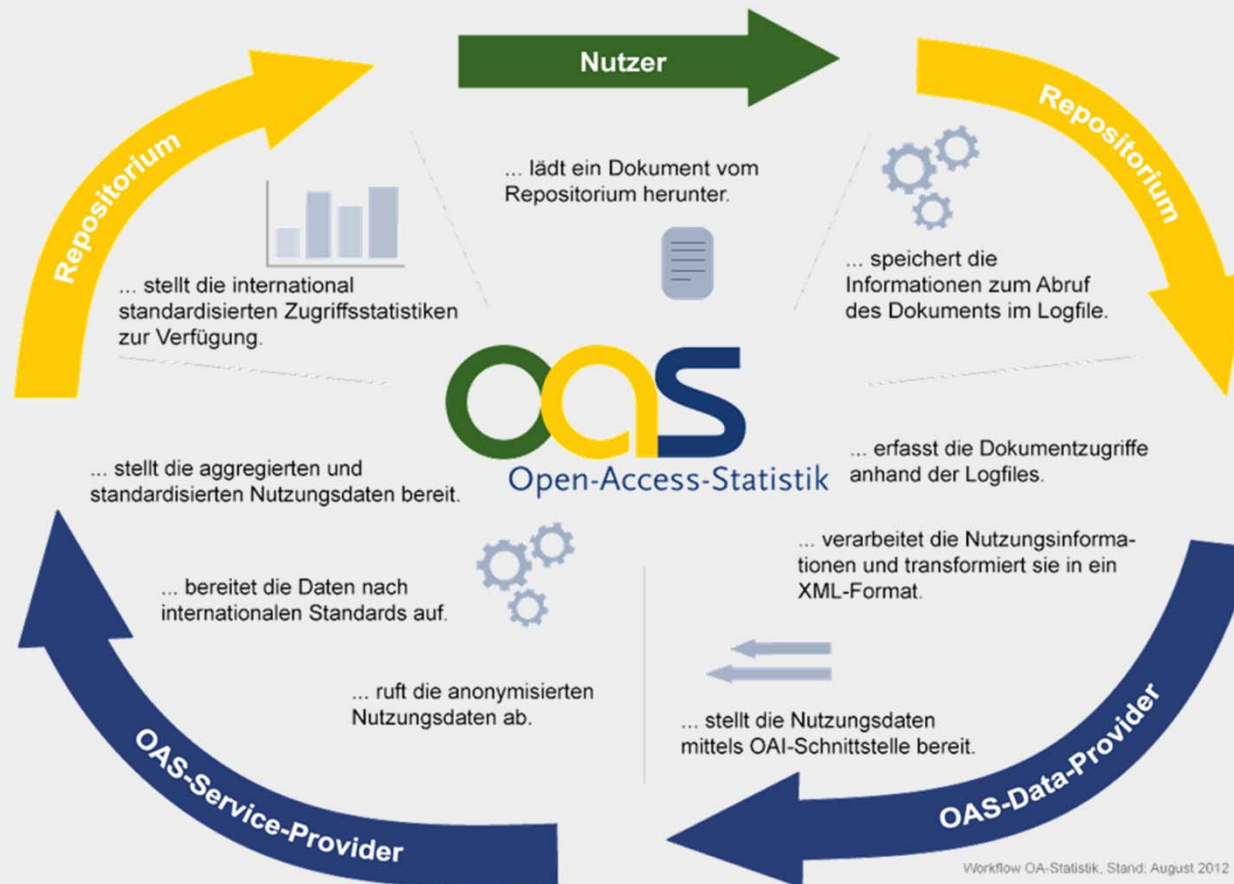
Verbundzentrale des GBV (VZG)



- Bibliotheksverbund
- Ziel: Nachhaltigkeit
mit Ende der Förderlaufzeit 2013
OAS-Infrastruktur in einen stabilen
Dienst überführen
- Hardware / Kosten
- Support



Workflow von OA-Statistik



Projektstand

Bisher wurden folgende Dataprovider (DP) integriert:

- Econstor / ZBW Kiel (Datenkontrolle)
- Dokumenten- und Publikationsservice Qucosa, Dresden
- PUB Publikationsdatenservice der Universität Bielefeld
- GoeScholar SUB Göttingen

Testphase:

- Edoc-Server HU Berlin
- GBV Göttingen
- SULB Saarbrücken



In 8 Schritten zur Teilnahme an OA - Statistik

1. Softwarepakete von der OA-Statistik-Projektwebsite herunterladen
2. Anpassung des OAS-Data-Providers an das eigene Repository
IDENTIFIER / Apache Logformat
3. Aktualisierung des SALTs
4. Nutzung der Datei Robots.txt
5. Geschützter Ort für den OAS-Data-Provider mit .htac
6. Implementation eines Opt-Outs (Empfehlung)
7. Registrierung Ihres OAS-Data-Providers
8. Rückführung der aggregierten Nutzungszahlen



1. Softwarepakete auswählen und herunterladen

- LINUX, APACHE, MySQL, PHP
- Paket für DSpace / Webdoc:
Anpassbar für viele Repositorien
- Paket OPUS



DINI OA-Statistik Website:
<http://www.dini.de/projekte/oa-statistik/die-projektergebnisse/software-download/>



2. Anpassung an das eigene Repository



- Der Logfile-Parser wandelt Server-Logfiles in OpenURL Context Objects um
- LOG-Format **combined** einstellen

```
/etc/apache2/apache2.conf:  
LogFormat  
"%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\"" combined
```

```
und unter  
/etc/apache2/sites-available  
CustomLog /var/log/apache2/access.log combined
```

2. Anpassung an das eigene Repository



- Identifier müssen je nach Repository angepasst werden.

OAI Record Header

OAI Identifier oai:www.digizeitschriften.de/:digizeit_201204.99 [oas](#) [formats](#)

Datestamp 2012-07-23T13:42:58Z

OAI Record Header

OAI Identifier oai:www.digizeitschriften.de/:digizeit_201204.171 [oas](#) [formats](#)

Datestamp 2012-07-23T13:43:12Z

2. Anpassung an das eigene Repository



- Der Data-Provider stellt die OpenURL Context Objects über OAI zur Verfügung.

```
context-objects "info:ofi/fmt:xml:xsd:ctx"
├─ context-object "2010-12-26T23:03:35Z"
│   └─ administration
│       └─ referent http://dms.gbv.de/dms/goettingen/236118692.pdf
│           └─ identifier http://dms.gbv.de/dms/goettingen/236118692.pdf
│               └─ referring-entity http://www.gbv.de/dms/goettingen/
│                   └─ identifier http://www.gbv.de/dms/goettingen/
│                       └─ requester
│                           └─ metadata-by-val http://dini.de/namespace/oas-requesterinfo
│                               └─ format http://dini.de/namespace/oas-requesterinfo
│                                   └─ metadata
│                                       └─ requesterinfo "http://dini.de/namespace/oas-requesterinfo" 001a3056feccf4ca3ea76b183e8794268c14248cd7fe1c8fdea67bb8d48c447c
│                                           └─ hashed-ip 001a3056feccf4ca3ea76b183e8794268c14248cd7fe1c8fdea67bb8d48c447c
│                                               └─ hashed-c e25af4fb3a646f9482b7c95b466e868ae8f31911ddb47df6011987b80796b5a9
│                                                   └─ hostname net.ma
│                                                       └─ user-agent Mozilla/4.0 (compatible; MSIE 8.0;
```

3. Salted Hash Verschlüsselungsverfahren



- Datenschutz / Verschlüsselung
- IP-Adresse = Personenbezogene Daten
- Empfehlung der ZENDAS



- Rechtliche Prüfung des Datenschutzes beim Sammeln der Nutzungsdaten.
- Rechtliche Prüfung der Weitergabemöglichkeiten der ermittelten Informationen.

3. Salted Hash Verschlüsselungsverfahren



- Hash (IP + Salt)
- IP Adressen sind auf dem OAS-Data-Provider pseudonymisiert und auf dem OAS-Service-Provider (der den Salt nicht kennen darf) anonymisiert

```
<metadata-by-val>  
  <format>http://dini.de/namespace/oas-requesterinfo</format>  
  <metadata>  
    <requesterinfo>  
      <hashed-ip>0a102e45f7d661138b9b480207ba89b48bfd56f0f711d9b6f85dcadfa18b36a7</hashed-ip>  
      <hashed-c>ad9f3de2d7a82be0617622c3864ce575432dda63b42fc8f6e4dd4d7d2068008a</hashed-c>  
      <hostname>uni-muenchen.de</hostname>  
      <classification>digizeitschriften.de:institution:112</classification>  
      <user-agent>Mozilla/5.0 (Windows NT 5.1; rv:10.0) Gecko/20100101 Firefox/10.0</user-agent>  
    </requesterinfo>
```

3. SALT regelmäßig aktualisieren



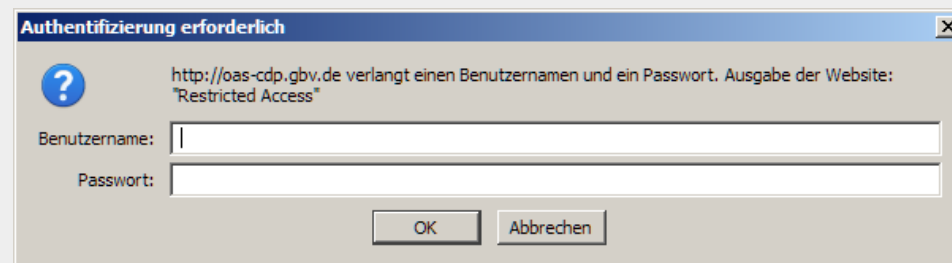
- Der Salt wird wechselt monatlich und wird von der Saarländischen Universitäts- und Landesbibliothek (*SULB*) bereitgestellt.
- Kontakt: oas@sulb.uni-saarland.de

4. Nutzung der Datei Robots.txt

- Bitte legen Sie ein Datei robots.txt an
- Zur Erkennung von automatisierten Zugriffen
- („Gute“ Roboter greifen darauf zu)

5. Geschützter Ort für den OAS-Data-Provider mit .htaccess

- Aus Datenschutzgründen muss die Schnittstelle Ihres OAS-Data-Providers mit .htaccess geschützt werden.
- Bitte teilen Sie uns die Zugangsdaten telefonisch mit



6. Implementation eines Opt-Outs

- Berücksichtigung der Vorgaben aus dem Datenschutz (Zendas Gutachten)
- Bei Nutzungserfassungen muss Nutzern ein „Opt-Out“ angeboten werden.
- Z.B. ein Button, über den ein Cookie abgelegt wird. Dieser wird von Ihrem Webserver ausgelesen und verhindert, dass die IP-Adressen Ihrer Nutzer aufgezeichnet werden.
- Empfehlung von OA-Statistik

7. Registrierung Ihres OAS-Data-Providers

- Registrierung über die OA-Statistik Website:

<http://www.dini.de/projekte/oa-statistik/die-technik/registrierung/>

Registrierung

Allgemein

Data Provider URL:

Name des Repositories:

DINI-Zertifikat: DINI 2010
 DINI 2007
 DINI 2004
 keines

Institution

Name:

Anschrift:

Web-Adresse:

Ansprechpartner

Anrede: Herr Frau

Titel:

Name:

Telefon:

Email:

Senden

15

Berechnung nach COUNTER

- Gezählter Zugriff nach COUNTER:
 - "http Status Code" von 200 oder 304
 - Feld Servicetype den Wert "fulltext"
- Doppelklickspanne: 10 sec.
- 30 sec. für Dokumente wenn <format> = application/pdf
- Export: wenn ein Feld ungleich 0
- Roboterfilterung nach unserer Liste
- Tägl. Auswertung nach Identifiern für die letzten 8 Tage

8. Rückführung der aggregierten Nutzungszahlen

Rohdaten bereitstellen:

- JSON (JavaScript Object Notation)
- XML

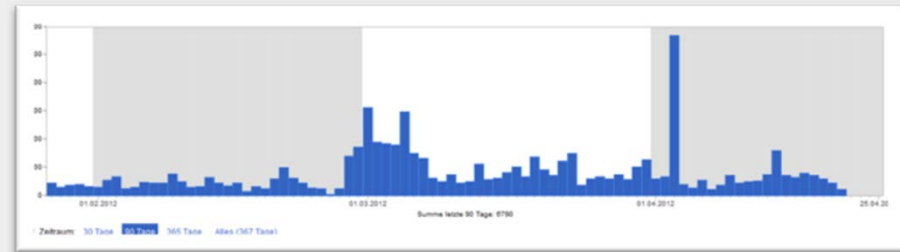
SUSHI geplant

Visualisierung:

- Graphik mit iFrame einbinden

```

    "from": "2012-03-01",
    "to": "2012-03-07",
    "entrydef": ["identifier", "date", "counter", "LogEc", "IFABC", "RobotsCounter", "RobotsLogEc", "RobotsIFABC"],
    "entries": [
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-01", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-02", "counter": 1, "LogEc": 0, "IFABC": 1, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-03", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-04", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-05", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-06", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-07", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0},
      {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-07", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0}
    ]
  
```





- Vielen Dank für Ihre Aufmerksamkeit und einen guten Start!
- Thank you for your attention and have a good start!

Fragen / Kommentare ?



Ansprechpartner / contact person

Marco Recke



Verbundzentrale des GBV
Abteilung Digitale Bibliothek
Platz der Göttinger Sieben 1
37073 Göttingen

Tel.: 0551/3912323

E-Mail: recke@gbv.de