



DigiZeitschriften
DAS DEUTSCHE DIGITALE ZEITSCHRIFTENARCHIV

Open Access Statistik Workshop

DigiZeitschriften e.V.

12. September 2012

Was ist DigiZeitschriften?

DigiZeitschriften ist ein gemeinnütziger Verein von derzeit 14 Bibliotheken, die wissenschaftliche Zeitschriftenliteratur verschiedenster Fachgebiete zur Verfügung stellen

Angebot von retrospektiv digitalisierten Zeitschriftenbeständen überregionaler Sammelschwerpunkte

Zeitschriften mit langer Forschungstradition und stetem Erscheinungsverlauf

Infrastrukturmaßnahme zum Nutzen der deutschen Wissenschaft und des deutschen wissenschaftlichen Publikationswesens

Vereinsmitglieder

Folgende 14 Bibliotheken sind Mitglieder des Vereins:

- Staatsbibliothek zu Berlin
- ULB Bonn
- SLUB Dresden
- UB Erlangen-Nürnberg
- UB Frankfurt/Main
- DIPF Frankfurt/Main
- UB Freiberg
- SUB Göttingen
- TIB Hannover
- UB Heidelberg
- ZBW Kiel
- USB Köln
- BSB München
- UB Tübingen

Zeitschriftenumfang

Bereitstellung aktuell von

- 200 Zeitschriftentiteln aus
- 19 Fachgebieten mit
- ca. 600.000 Beiträgen auf
- insgesamt rund 6 Mio. Seiten
- davon ca. 50 % im Open Access Bereich

Zielgruppen

- 2 Verwertungsgesellschaften (VG)
Wort und Bild-Kunst
- 37 Verlage
- 220 Abonnenten (In- und Ausland)
- Contentlieferanten Open Access Zeitschriften
(DFG-Projekte von SSG-Bibliotheken)

Auswertungen pro Jahr sind Grundlage der
Lizenzgebührenabrechnung an die VGs.

Anforderungen der Zielgruppen

Verwertungsgesellschaften

- Gesamtzahl der jährlichen Zugriffe pro Zeitschrift und Verlag
- Zugriffe nach Erscheinungsjahr: vor 1925; 1926-1995; ab 1996
- Seitenimporte pro Jahr jeder Zeitschrift

Verlage

- Zugriffszahlen auf die eigenen Zeitschriften
- Vergleich mit anderen Titeln innerhalb eines Fachgebietes

Anforderungen der Zielgruppen

Abonnenten

- Zugriffe der eigenen Nutzer auf Gesamtangebot
- Autoren für eigene Publikationen

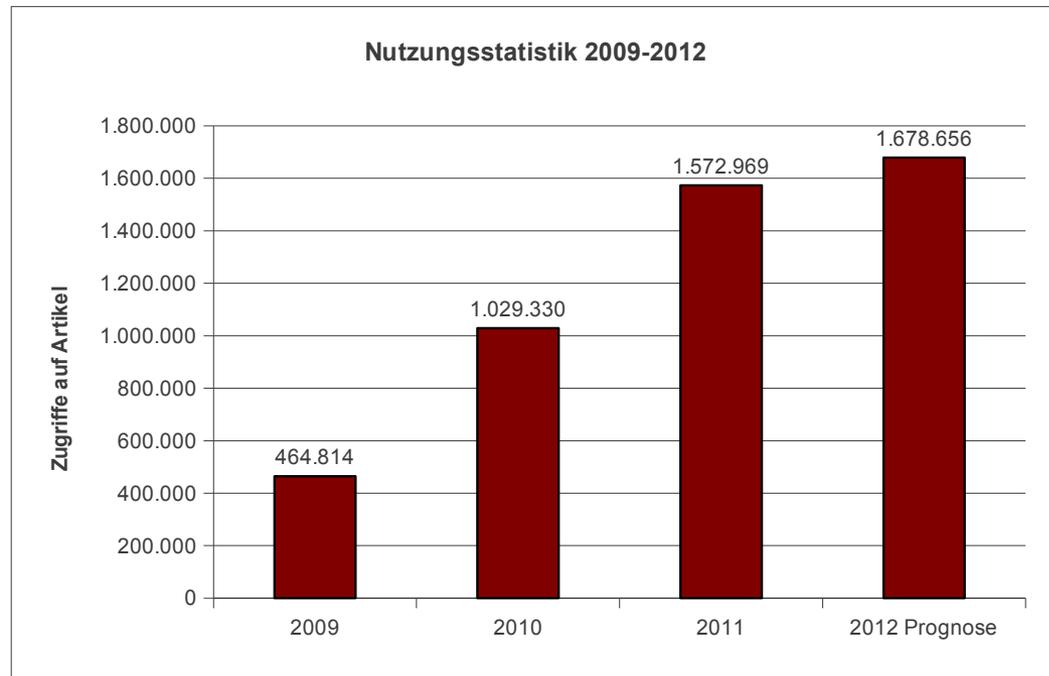
Contentlieferanten (Bibliotheken)

- Zugriffe auf die gelieferten Zeitschriften (Open Access)
- Vergleich mit anderen Titeln innerhalb eines Fachgebietes

Nutzungsstatistik 2009-2012

Nutzungsstatistik

- für jeden Verlag und Abonnent mit Passwort individuell abrufbar



Ziele

- Veraltete proprietäre Software ersetzen
- Gewinnung vergleichbarer Nutzungsstatistiken
- Dauerhafte Infrastruktur
- Datenschutzkonforme Speicherung und Auswertung
- Entlastung der eigenen Infrastruktur

Zahlen

- 2,5 Mio Zeilen / Monat
- 700.000 Zeilen Content Zugriffe
- 300.000 Zeilen bereinigt
- 150.000 „echte“ Zugriffe
- Für diese Zugriffen werden monatlich ca. 280 unterschiedliche Auswertungen produziert

Entwicklungsstand

- Ausbau der OAI-Schnittstelle, es werden Datensätze (Identifier) für alle Beiträge geliefert (Zeitschrift, Band, Heft, Artikel,...)
- Aufbau einer Identifier Datenbank (harvest_identifiers.php) mit täglicher Aktualisierung (aktuell ca 800.000 Datensätze)
- Entwicklung eines OAS Parsers auf Basis der bestehenden Parser
- Einrichtung eines OAS Data Providers
- Testlauf mit bestehender Logdatei

Zwischenergebnis Context Objects

Identifizier für:

- Zeitschrift
- Band
- Artikel
- Kollektion
- ...

Classification für:

- institutionellen
oder
 - Open Access
- Zugriff

```
<context-object timestamp="2012-04-01T13:38:04Z" >
  <administration>
    <oa-statistics>
      <status_code>200</status_code>
      <size>0</size>
      <format>text/html</format>
      <service>http://www.digizeitschriften.de/</service>
    </oa-statistics>
  </administration>
  <referent>
    <identifier>http://localhost:8086/gcs/gcs?action=metImage&format=jpg&metFile=PPN509862098_0001&divID=phys358&width=800&rotate=0</identifier>
    <identifier>digizeitschriften.de:PERIODICAL:PPN509862098</identifier>
    <identifier>digizeitschriften.de:VOLUME:PPN509862098_0001</identifier>
    <identifier>digizeitschriften.de:YEAR:1866</identifier>
    <identifier>digizeitschriften.de:ACL:Steiner</identifier>
    <identifier>digizeitschriften.de:ACL:free</identifier>
    <identifier>digizeitschriften.de:ACL:Philology</identifier>
    <identifier>digizeitschriften.de:DC:400.Philology</identifier>
    <identifier>oai:www.digizeitschriften.de:PPN509862098_0001/dmdlog46</identifier>
  </referent>
  <referring-entity>
    <identifier>http://www.google.com/search</identifier>
  </referring-entity>
  <service-type>
    <metadata-by-val>
      <format>info:ofi/fmt:xml:xsd:sch_svc</format>
      <metadata>
        <any>yes</any>
      </metadata>
    </metadata-by-val>
  </service-type>
  <requester>
    <metadata-by-val>
      <format>http://dini.de/namespace/oas-requesterinfo/</format>
      <metadata>
        <requesterinfo>
          <hashed-ip>3c078440c3337f94165bf6470b370ae1f885c025e77202d8144b26c7392dec82</hashed-ip>
          <hashed-c>c2990696b8e1c6542c7307c3aa5990629f5c702ab5e4dcff8a2977d0a3b3f7ba</hashed-c>
          <hostname>18.160</hostname>
          <classification>extern</classification>
          <user-agent>Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/534.51 (KHTML, like Gecko; Google Web Preview) Chrome/12.0.742 Safari/534.51</user-agent>
        </requesterinfo>
      </metadata>
    </metadata-by-val>
  </requester>
</context-object>
```

Zwischenergebnis JSON

```
{
  "from": "2012-04-01",
  "to": "2012-04-07",
  "entrydef": ["identifizier", "date", "counter", "LogEc", "IFABC", "RobotsCounter", "RobotsLogEc", "RobotsIFABC"],
  "entries": [

    ...
    {"identifizier": "digizeitschriften.de:ACL:free", "date": "2012-04-02", "counter": 932, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0,
    "RobotsLogEc": 0, "RobotsIFABC": 0},
    ...
    {"identifizier": "digizeitschriften.de:DC:340.Law", "date": "2012-04-02", "counter": 1157, "LogEc": 0, "IFABC": 0, "RobotsCounter": 4,
    "RobotsLogEc": 4, "RobotsIFABC": 4},
    ...
    {"identifizier": "digizeitschriften.de:PERIODICAL:PPN345574974", "date": "2012-04-02", "counter": 1134, "LogEc": 0, "IFABC": 0,
    "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    ...
    {"identifizier": "digizeitschriften.de:VOLUME:PPN345574974_0031", "date": "2012-04-02", "counter": 143, "LogEc": 0, "IFABC": 0,
    "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    ...
    {"identifizier": "digizeitschriften.de:YEAR:1976", "date": "2012-04-02", "counter": 152, "LogEc": 0, "IFABC": 0, "RobotsCounter": 1,
    "RobotsLogEc": 1, "RobotsIFABC": 1},
    ...
  ]
}
```

ToDo

- Erstellung der verschiedenen Statistiken für Kollektionen, Zeitschriften, Verlage, Open Access,... aus den JSON Objekten in Austauschformaten (Counter.org, XML, Excel)
- Inwieweit lassen sich die Zahlen zu den Identifiern zu neuen Ergebnissen verknüpfen
- Evaluation der Ergebnisse, Vergleich mit Zahlen aus dem alten System
- Ggf. Anpassung der Context Objects und der Identifier Datenbank
- Weitere Testläufe
- Regelbetrieb mit weitestgehender Automation

Wunschzettel

- COUNTER 3
- Web 2.0
- XML
- Excel
- Diagramme
- ...





DigiZeitschriften
DAS DEUTSCHE DIGITALE ZEITSCHRIFTENARCHIV

Danke

<http://www.digizeitschriften.de>

Caren Schweder <schweder@sub.uni-goettingen.de>

Jochen Kothe <kothe@sub.uni-goettingen.de>