

# Was machen mit den Zahlen?

---

## OAS-Workshop

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

## Tobias Schäfer

Niedersächsische Staats- und  
Universitätsbibliothek Göttingen

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**

# Agenda

---

1. Standards
2. Verarbeitung
3. Darstellung
4. Auslieferung und Einbindung
5. Live-Demonstration
6. Ausblick

# Standards: Übersicht

---

- ❑ COUNTER
- ❑ IFABC
- ❑ LogEc
- ❑ Andere: AWStats, IRStats

# Standards: COUNTER

---

- ❑ Counting **O**nline **U**sage of **N**etworked **E**lectronic **R**esources
- ❑ Messung der Zugriffszahlen von Onlinepublikationen
- ❑ Standard mit der größten Relevanz für OA-Statistik



# Standards: Weitere

---

## □ IFABC

- International Federation of Audit Bureaux of Circulations
- In Deutschland vertreten durch IVW
- Messung der Verbreitung von Werbeträgern



## □ LogEc

- Zählstandard des Projektes RePEc
- Messung der Zugriffe auf gehostete Publikationen

*LogEc*

# Verarbeitung: Übersicht

---

## **Nutzerdeduplizierung**

Ausfilterung von Mehrfachzugriffen durch denselben Nutzer innerhalb eines bestimmten Zeitfensters

## **Robotfilterung**

Ausfilterung von automatisierten Zugriffen durch Robots, Spider und Crawler mittels Blacklist und anderer Verfahren

## **Speicherung**

Tageweise aggregierte Nutzungszahlen werden zentral gespeichert und sind jederzeit abrufbar

# Verarbeitung: Deduplizierung

Standard	COUNTER	LogEc	IFABC
<b>Zählkriterium</b>	HTTP- Statuscode 200 oder 304	HTTP- Statuscode 200, 206, 301, 302 oder 304	Tracking Pixel, mehr als 95 % der Datei über- tragen
<b>Multiklick- zeitspanne</b>	10 s für HTML, 30 s für PDF	ein kalenda- rischer Monat	30 min
<b>Nutzer- identifikation</b>	IP, Session	IP	IP und Useragent, Session

# Verarbeitung: Robotfilterung

---

Standard	COUNTER	LogEc	IFABC
<b>Definition</b>	Robots, Prefetches, Caching	Robots, automatisier- te Downloads	Robots, automatisier- te Downloads
<b>Identifikation</b>	Useragent, HTTP-Header	Zugriff auf robots.txt, Zugriffs- verhalten	Useragent



# Verarbeitung: Datengranularität

---

- Rohdaten liegen für alle erfassten Publikationen sekundengenau vor
- Aggregation derzeit tageweise
- Auslieferung ursprünglich monatsweise geplant (für LogEc notwendig), aber auch tageweise möglich

# Darstellung: Übersicht

---

- Nutzerbedürfnisse
- Verlaufsdiagramme
- Relationen

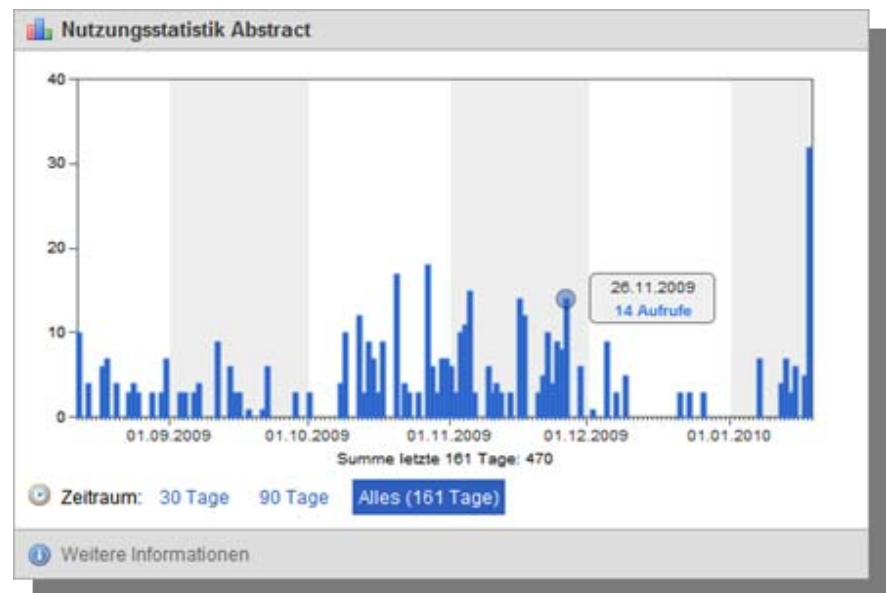
# Darstellung: Nutzerbedürfnisse

---

- ❑ Welche Darstellung bietet dem Nutzer echte Vorteile?
- ❑ Nutzer sind unterschiedlichste Besucher der Repositorien
- ❑ Präsentation erfolgt gemäß COUNTER, Zahlen lassen sich aber auch nach anderen Standards darstellen

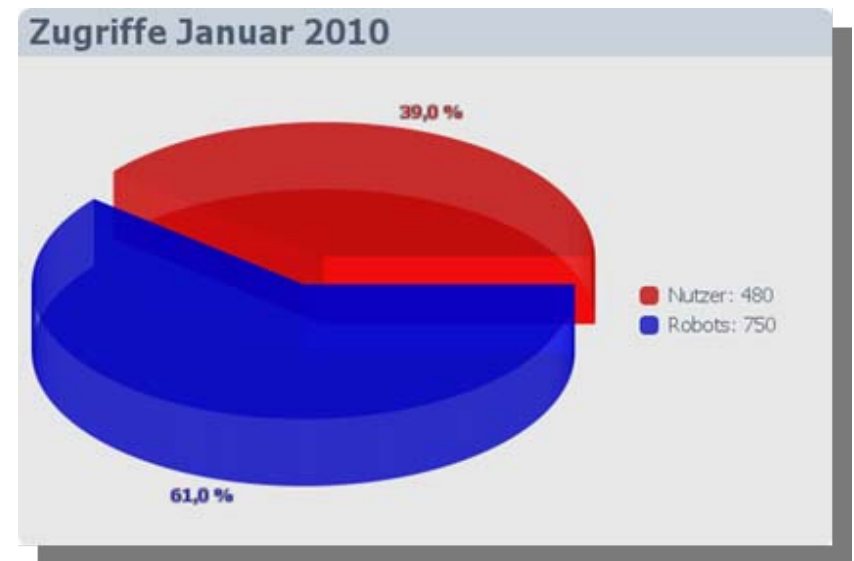
# Darstellung: Verlaufsdiagramme

- Säulen-/Balkendiagramm, Liniendiagramm, evtl. Histogramm
- Beschränkung auf für den Nutzer nützliche Daten und Darstellungen



# Darstellung: Relationen verschiedener Zugriffstypen

- Campuszugriffe vs. Gesamtzugriffe
- Robots vs. Nutzer
- Abstract vs. Volltext
- Auch als Verlauf darstellbar



# Auslieferung und Einbindung: Übersicht

---

- Performance
- Auslieferung
  - Grafikdatei
  - Inlineframe
  - Rohdaten

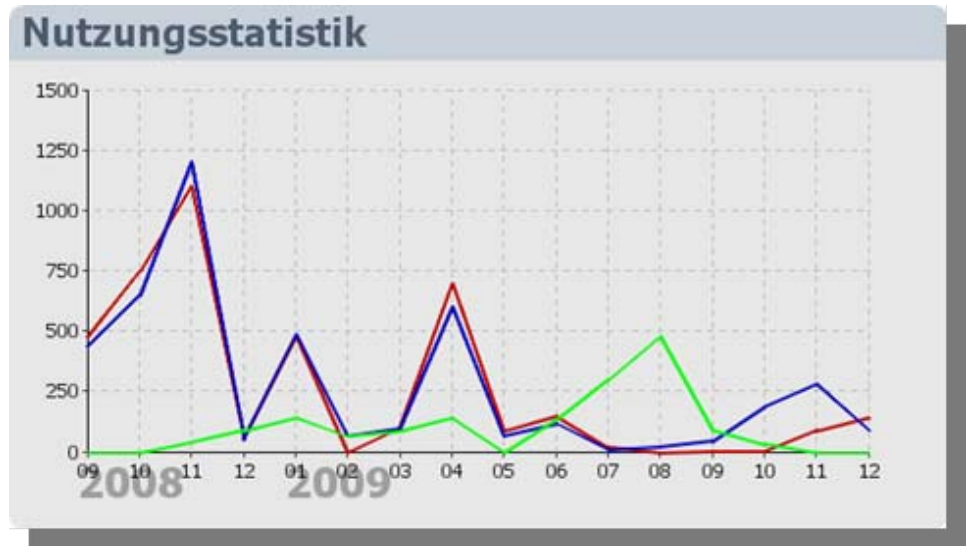
# Auslieferung und Einbindung: Performance

---

- Datenaggregation erfolgt zentral für jedes Dokument
- Erste Generierung on-request, aber auch tageweise Gesamtauswertung möglich
- Aggregierte Daten und generierte Grafiken werden gecachet, dadurch sehr gute Performance auch bei hoher Nutzerlast

# Auslieferung und Einbindung: Grafikdatei

- Format:  
PNG oder JPEG
- Sehr einfache  
Implementation
- Maximale  
Kompatibilität
- Keine interaktiven Elemente
- Einbindung mit nur einer Zeile Code:  
``





# Auslieferung und Einbindung: Inlineframe

---

- ❑ Vollständige Webseite wird in `<iframe>` geladen
- ❑ Einfache Implementation
- ❑ Einsatz von Scriptsprachen erfordert entsprechende Kompatibilität des Webbrowsers
- ❑ Keine Einschränkungen hinsichtlich interaktiver Elemente
- ❑ Einbindung als kurzer Codeblock

# Auslieferung und Einbindung: Rohdaten

---

- ❑ Webservice stellt aggregierte Daten geordnet nach Tagen und Zählstandards bereit
- ❑ Formate: CSV, XML oder direkt als HTML-Tabelle
- ❑ Repositorienbetreiber können grafische Aufbereitung übernehmen

# Live-Demonstration: GOEDOC

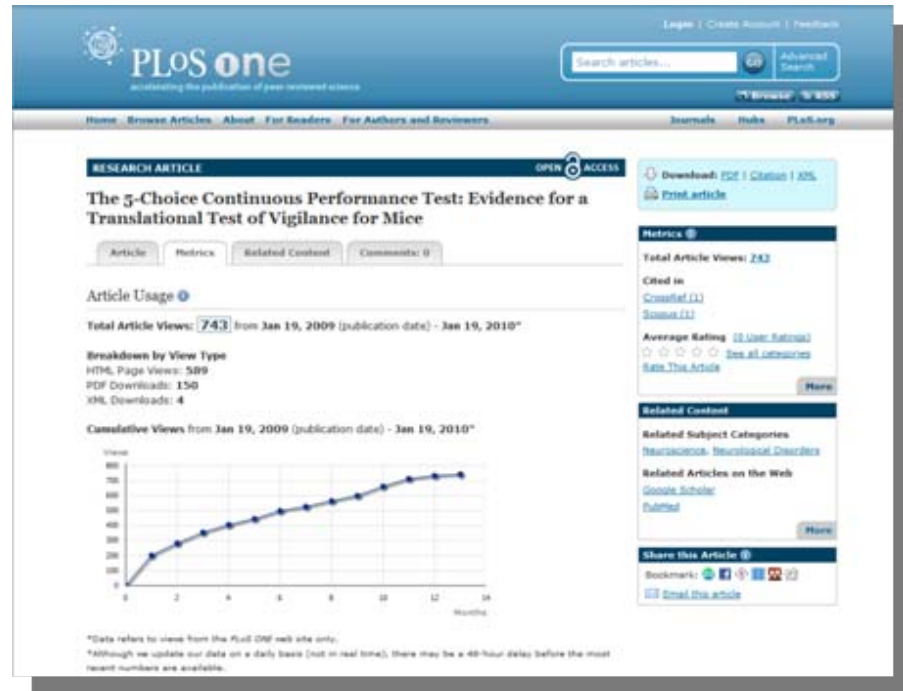
---

- Dokumentenserver der Georg-August-Universität Göttingen
- Betrieben von der SUB Göttingen
- Bietet allen Angehörigen der Uni Göttingen die organisatorischen und technischen Rahmenbedingungen zur elektronischen Publikation wissenschaftlicher Dokumente, vorrangig Dissertationen
- Empfehlung an die Autoren: Open Access



# Ausblick: Was machen die anderen?

- PLoS ONE bietet Article Level Metrics



Quelle: <http://www.plosone.org>

- BibTip: Recommendersystem als kostenpflichtige Dienstleistung, entwickelt am Karlsruhe Institute of Technology

# Ausblick: Neue Funktionalitäten

---

- Geolocation:  
Aufschlüsselung  
nach Herkunft der  
Nutzer



- Recommendersystem mit frei verfügbaren  
Rohdaten
- Auf Repositorien zugeschnittener  
Zählstandard

# Quellen

---

- ❑ COUNTER - Online Usage of Electronic Resources  
<http://www.projectcounter.org/>
- ❑ LogEc: Download Statistics  
<http://logec.repec.org/>
- ❑ International Federation of Audit Bureaux of Circulations  
<http://www.ifabc.org/>
- ❑ The Web Robot Pages (Robot Exclusion Standard)  
<http://www.robotstxt.org/orig.html>
- ❑ GOEDOC - Dokumentenserver der Georg-August-Universität Göttingen (SUB Göttingen)  
<http://webdoc.sub.gwdg.de/>
- ❑ PLoS ONE: accelerating the publication of peer-reviewed science  
<http://www.plosone.org/>
- ❑ BibTip  
<http://www.bibtip.org/>

# Vielen Dank für Ihre Aufmerksamkeit!

**Tobias Schäfer**

Niedersächsische Staats- und  
Universitätsbibliothek Göttingen

Initiiert durch:



Gefördert durch:

Deutsche  
Forschungsgemeinschaft

**DFG**