



Open Access
Infrastructure
for Research in Europe

OpenAIRE

Usage Statistics - nationale Projekte und OpenAIRE

Jochen Schirrwagen, Universitätsbibliothek Bielefeld

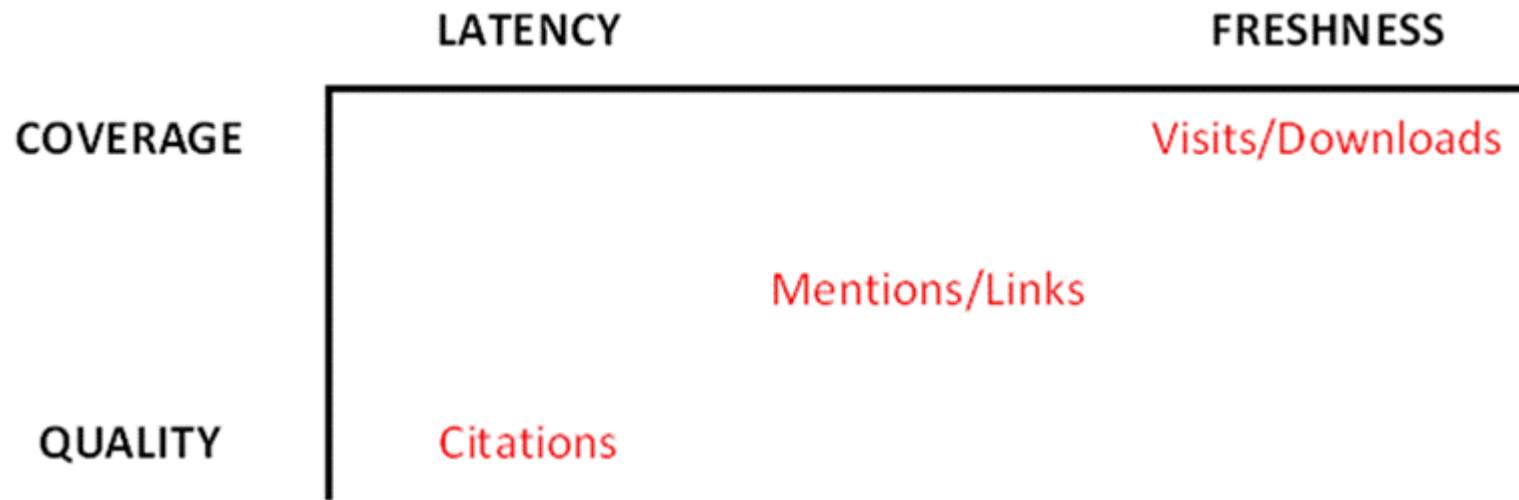
06. Oktober 2011, OA-Statistik-Workshop, Regensburg

Nutzungsstatistiken

- Liefern Indikatoren über den Impact wissenschaftlicher Publikationen
- Beziehen sich auf: wissenschaftliche Publikationen (**Artikel-Ebene**) und jegl. digitale Inhalte in Repositories
- Adressieren: Wissenschaftler (Rezipienten wie Autoren), Hoster (Bibliotheken, Verlage), Förderorg.
- Benötigen Interoperabilität: Standards für Übertragungsprotokolle, Austauschformat, Filter- und Normalisierungsregeln
 - *für Vergleichbarkeit und Konsolidierung*
- Status: Forschung bzgl. Indikatoren, Betriebsmodelle, Mehrwertdienste, (nationale) Pilotprojekte bzgl. Aggregatoren & Statistik-Services

Metriken - ein Vergleich

- Nutzungsstatistiken als alternative (komplementäre) Metrik zu Bibliometrie, Webometrie



feature of metrics

KE: Usage Statistics Arbeitsgruppe

MEASURING OPEN ACCESS USAGE AN OVERVIEW OF COOPERATING EUROPEAN PROJECTS

Tobias Schäfer, Hans-Werner Hilke, Daniel Beucke
Göttingen State and University Library, Germany

**SURE
NETHERLANDS**

The project SURE (Statistics on the Usage of Repositories) is led by a group of three Dutch universities, and is funded by SURFfoundation. The main objective of the project is to implement a national infrastructure for the accumulation and the exchange of usage data. The results of statistical analyses of these data will become available via web services, and will also be exposed via the national Dutch research portal NARCIS.

**PIRUS
UNITED KINGDOM**

PIRUS builds on the standards already established by COUNTER, and on the results of the original PIRUS project (Publisher and Institutional Repository Usage Statistics), which demonstrated that it is technically feasible to create, amend and consolidate usage statistics for individual articles using data from repositories and publishers, despite the diversity of organisational and technical environments in which they operate.

**OAS
GERMANY**

OAS (Open Access Statistics) has created a sustainable infrastructure to collect and exchange usage data between different agents, focusing on open access repositories. It established a reliable standard in order to gather statistics, based on rule sets such as COUNTER. By now OAS strives for a second funding phase which will focus on international cooperation and the extension of the OAS infrastructure by integrating further repositories.

**KNOWLEDGE EXCHANGE
DENMARK, GERMANY, NETHERLANDS, UNITED KINGDOM**

Knowledge Exchange (KE) is a cooperative of Denmark's Electronic Research Library (DEFF), the Deutsche Forschungsgemeinschaft (DFG), German Research Society), SURFfoundation in the Netherlands, and the Joint Information Systems Committee (JISC) in the UK. These partners share the vision of bringing European research together by making a layer of scholarly and scientific content openly available and creating the building blocks for an outstanding European information environment.

The members of the KE Usage Statistics Workgroup, which also include the projects RePEC and NEEQ, have met in several workshops where they addressed the usage and interoperability of statistics regarding the retrieval of scholarly information, articulated studies and cooperative initiatives, and agreed to a common set of guidelines. An important element of these guidelines is a consolidated robot filtering list to be used by all related projects, which is already available online.

DEff
Denmarks Elektroniske
Teg og Forskningsbibliotek

Deutsche
Forschungsgemeinschaft
DFG

KE
Knowledge Exchange

JISC

SURF
FOUNDATION

September 2011

KE: Aktivitäten

- Wissens- und Erfahrungsaustausch über
 - *Mehrwert(-dienste) von Nutzungsstatistiken*
 - *Guidelines für Vergleichbarkeit und Austausch*
 - *Rohdaten vs. Statistiken vs. Interpretation der Statistiken*
- Expertentreffen (Genf (2009), Göttingen (2010), Berlin (2010), Leiden (2011))
- Unterstützung der Vernetzung von Nutzungs-Statistik Projekten: OA-Statistik, PIRUS-2, SURFsure
- Ergänzt durch Praxiserfahrungen in LogEc, Mesur, NEEQ, PEER, PLoS, OAPEN

KE: Usage Statistics Guidelines

- gemeinsame Guidelines über Protokolle (OAI-PMH, SUSHI) und Austauschformat (OpenURL Context Objects) <http://purl.org/REP/standards/KE>
 - *COUNTER Standard als gemeinsamer Nenner*
 - *verbessern die Vergleichbarkeit von Nutzungsstatistiken über Ländergrenzen*
 - *offenbaren aber auch unterschiedliche Datenschutzrichtlinien in der EU*
 - *verpflichtende und optionale Elemente; erlauben (projektspezifische) Erweiterungen*

Aufstellung gemeinsamer Regeln

■ Robot-Detection und Filterung

- basierend auf gemeinsamer Liste: <http://purl.org/robotslist/current/robotlist.xml>

■ Erkennung von Double-Clicks, z.B. je nach Standard verschieden:

COUNTER	10 sec (HTML), 30 sec (PDF)
LogEc	1 month
AWStats	1 hour
IFABC	30 minutes

PIRUS: Phase I

(**Publisher** and **Institutional Repository** Usage Statistics)

- Laufzeit ~ 2008 - 2009
- geleitet von COUNTER
- Ergebnisse:
 - *Prototyp XML basierter Usage Reports auf Artikelebene*
 - *Tracker-Code zur Integration in Repositories*
 - *Kriterien für eine zentrale Clearing-Stelle für das Sammeln und Konsolidieren von Nutzungsdaten*

PIRUS: Phase 2

- ~ 2009 - 2010
- Usage Reports kompatibel zu COUNTER auf Artikel-Ebene
- Guidelines für Verlage & Repositories um Usage Reports zu erzeugen
- Möglichkeiten zur Konsolidierung der einzelnen Reports auf globaler Ebene (z.B. Aggregator)
- Entwicklung von OpenSource Software für Erzeugung/Austausch von Nutzungsdaten und Statistiken gemäß COUNTER Standard
 - *die alle digitalen Objekte in Repositories erfassen kann*
- Entwicklung eines prototypischen Usage-Statistics-Service auf Artikel-Ebene
- Entwicklung einer Grundmenge von Statistik-Reports, erzeugt durch Repositories für internen/externen Gebrauch
- Szenarien für Protokoll-Unterstützung: Tracker-Code (ähnl. Google Analytics, OAI-PMH {Repositories}, SUSHI {Publisher})
- Plugins für : DSpace, EPrints, Fedora

(Statistics on the Usage of Repositories)

■ Zielgruppe: Repositories in den Niederlanden

■ Phase 1:

- *Implementierung von Standards und Tools zum Aufzeichnen, Aggregieren und Visualisieren von Nutzungsstatistiken*
- *u.a. Erstellung von Service Katalog und Betriebsmodell <http://wiki.surffoundation.nl/display/statistics/Project+results>*

■ Phase 2:

- *Entwicklung eines Statistik Dashboard*
- *Gewinnung weiterer Repositories als Log-Provider*
- *Integration von Statistiken im nationalen Portal NARCIS*
- *Open Source Software zur Erzeugung von OpenURL ContextObjects und Aggregator*
- *Ergebnisse zur Open Access Woche 2011 erwartet*

OA-Statistik

■ Phase I:

- *Guidelines (Protokoll, Format)*
- *Software-Plugins (OPUS, DSpace, WebDoc)*
- *Log-Aggregator & Statistik-Service*

■ Als Besonderheiten seien hervorgehoben:

- *Kompatibilität mit Standards: implementiert: **COUNTER, IFABC**; geplant: **LogEc***
- *Anonymisierung aller Aspekte, die Rückschlüsse auf konkrete Nutzer zulassen könnten (IP-Adresse, ...)*

■ DINI-OAS-Erweiterungen in den Guidelines

- **Requester** *Klassifizierung (intern, administrativ, institutionell)*
- **oas-requesterinfo** *Identifier für die komplette Session eines gegebenen Nutzers*
- **dini:user-agent** *vollständiger User-Agent String*

OpenAIRE - Impact Metriken

- Umsetzung OA-Pilot der EC
 - *d.h. Publikationsnachweise zu FP7-Projekten*
- Teilaspekt: Impact-Messung
- Komplementärer Ansatz - Bibliometrie, Webometrie, Usage Statistics
- Randbedingungen:
 - *enges Zeitfenster*
 - *begrenzt auf SC39 Publikationen*
- Nutzungsstatistiken: Anwendungsfall eines Statistik-Aggregators auf globaler Ebene
- Adaptiert die KE Usage Statistics Guidelines, berücksichtigt "lokale" Besonderheiten
- Aufbauend auf den Projekterfahrungen im KE-USG-Kontext

OpenAIRE - Usage Statistics Service

Statistics Service

Home page **Statistics**

Metrics & Patterns

- Publication Statistics & Metrics
- Repository Statistics & Metrics
- Repository Session Statistics
- Top Publications per Repository
- Cumulative Daily Repository Traffic
- Cumulative Weekly Repository Traffic
- FPT Projects output
- FPT Programmes output

Terms explanation

View Count (VC):
Number of times that a publication's metadata has been browsed.

Download Counts (DC):
Number of times that a publication has been downloaded.

Hit Index (HI):
Sum of VC and DC.

Popularity Index (PI):
Weighted combination of VC and DC, putting more emphasis on DC.

Publication Statistics & Metrics

Set parameters

Custom parameters
 Publication Id:
 Time Unit:

Examples
 Time Unit: week, Publication Id: oai:uva.pub:209057

Visualisation

motionchart
 table

Metric	Value (approx.)
Hit Index	95
Download Counts	80
Popularity Index	65
View-metadata Counts	18

Week 47, 2010

Aggregator Institution	Bielefeld University Library
Repository Name	Publications of the Universiteit van Amsterdam
Repository Link	http://dare.uva.nl
Authors	Koning, H. de Does, R.J.M.M. Mast, J. de
Title	Lean Six Sigma
Identifier	http://dare.uva.nl/record/209057

© National and Kapodistrian University of Athens (NKUA)

Ausblick und ToDos

- verlässliche und aussagekräftige Statistiken, wenn
 - *hoher Abdeckungsgrad an Repositories + Verlage*
- Korrelation Nutzungsstatistiken, Zitationen, Verlinkungen, Suchmaschinentreffer, Ranking
- De-Duplication von Publikationen und Logdaten
- Nutzungsstatistiken über komplexe Objekte (z.B. nach Publikationen und Zusatzmaterialien)
- Lösung hinsichtlich unterschiedlicher Datenschutzrichtlinien
 - *Anwendung verschiedener Anonymisierungs-/Pseudonymisierungsverfahren*