

# OA-Statistik: Standardisierte Nutzungsanalysen als alternative Impact-Messungen wissenschaftlicher Publikationen

Ulrich Herb, Saarländische Universitäts- und Landesbibliothek (SULB)

Workshop zu Standardisierung und Vernetzung von  
Open-Access-Repositoryn für die Wissenschaft  
Göttingen, 19.02.2009



## Übersicht

- Impact: Relevanz
- Impact Maße: Die Referenzen
- Zitationsbasierte Impact Maße und Open Access
- Zitationsbasierte und nutzungsbasierte Impact Maße: Kategorisierungsversuch
- Nutzungsbasierte Impact Maße: Standardisierung und Fazit
- DFG-Projekt: Open Access Statistik



## Impact: Individuelle Relevanz

Karrierechancen der Wissenschaftler hängen von ihrer Reputation ab.

Reputation wird in aller Regel über *Impact* zu bestimmen versucht.

Impact wird ermittelt über Zitationen

- a) bezogen auf Zeitschriften, in denen Wissenschaftler publizieren  
z.B. Journal Impact Factor
- b) bezogen auf Artikel eines Wissenschaftlers  
z.B. Hirsch Index



## Impact: Organisationale Relevanz

### **Hochschulevaluation** und **Leistungsorientierte Mittelvergabe (LOM)**

entscheiden über zukünftige materielle und personelle Ausstattung und durch Entzug oder Erhöhung von Mitteln auch über die Wahl von Forschungsschwerpunkten.

### **Kriterien**

1. Anzahl der Promotionen
2. Umfang der eingeworbenen Drittmittel
3. **Publikationsverhalten**  
-> bestimmt über Zitationsmaße, v.a. JIF, seltener h-Index



# Impact-Maße: Referenz I – Der Journal Impact Factor JIF

## Berechnung

Zahl der Zitate im laufenden Jahr auf Artikel (eines Journals) der  
vergangenen zwei Jahre

-----

Zahl der Artikel des Journals der vergangenen zwei Jahre

We never predicted that people would turn this into an evaluation tool for giving out grants and funding.  
Eugene Garfield

Aus:  
Richard Monastersky (2005) The Number That's Devouring Science *The Chronicle of Higher Education*



# Journal Impact Factor

## Kritikpunkte Teil 1:

- Begrenzter Scope/ Ausschluss kompletter Dokumentarten: graue Literatur, Bücher, Großteil der Web-Publikationen.
- Berücksichtigt werden nur im Journal Citation Report JCR indizierte Journals.
- Sprachbias zugunsten englischsprachiger Journals: Zeitschriften in anderen Sprachen haben einen niedrigeren IF.
- JIF bezieht sich auf Journale, nicht Artikel: I.d.R. führt eine geringe Anzahl sehr häufig zitierter Artikel zu einem hohen Wert für das Journal



# Journal Impact Factor

## Kritikpunkte Teil 2:

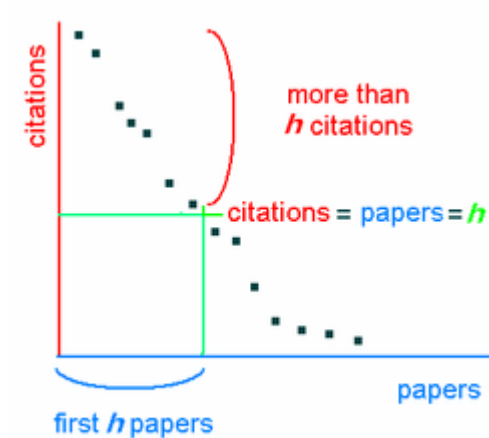
- Ignoranz der Verwertungszyklen in unterschiedlichen Disziplinen:  
Benachteiligung der Journale aus Disziplinen mit Verwertungszyklen  $> 2$  Jahren  
(z.B. Mathematik, Geisteswissenschaften)
- Ignoranz des Mehrautorenproblems
- Vernachlässigung kontextueller Aspekte (Gewichtung)
- Konfundierung von Popularität und Qualität



## Impact-Maße: Referenz II – Der Hirsch-Index (h-index)

Hirsch (h-) Index bezieht sich auf Autoren und nicht auf Journals

Berechnung: Ein Autor hat einen Index  $h$ , wenn  $h$  von seinen insgesamt  $N$  Veröffentlichungen mindestens jeweils  $h$  Zitierungen haben und die anderen  $(N-h)$  Publikationen weniger als  $h$  Zitierungen.



<http://de.wikipedia.org/wiki/H-Index>

Ein Autor hat einen h-Index von 8, wenn er 8 Schriften veröffentlicht hat, die jeweils mindestens 8 Mal zitiert worden sind. Ein h-Index von 12 setzt die Publikation von 12 Schriften voraus, die mindestens 12 Mal zitiert wurden.





# h-Index

## Vergleich h-Index/JIF

Vorteil: Zitationen einer einzigen, vielzitierten Veröffentlichung schlagen sich nicht nieder.

Nachteil: innovative Ansätze werden nicht berücksichtigt.

## Kritik des h-Index:

- vgl. JIF-Kritik: Vernachlässigung von Dokumentengattungen und nicht-englischer Publikationen, Mehrautorenproblematik, Messung von Popularität oder Qualität? ...
- Trennscharfe Autorenidentifikation in der Datenbasis (Web of Science) ist nicht sichergestellt
- h-Index hängt von Alter des Autors und Disziplin ab



## Zitationsbasierte Impact Maße und Open Access

Traditionelle, zitationsbasierte Impact Maße liefern Argumente pro Open Access:

Open-Access-Dokumente werden im Vergleich zu lizenzpflichtigen signifikant häufiger heruntergeladen, genutzt und zitiert (Lawrence 2001, Brody & Harnad 2004, Sietmann 2006).

Downloadhäufigkeit scheint mit Zitationshäufigkeit zu korrelieren und diese vorherzusagen (Brody, Harnad & Carr 2005).

Impact Messung des Open-Access-Angebots erfolgt meist mittelbar über Zitationsdatenbanken (Journal Citation Report, Web of Science) und mit Bezug zum Journal.



**Sind alternative Impact-Maße modellierbar?**



# Eine Kategorisierung

## ***Zitationsbasierte Maße***

- autorenzentriert
- Messung erst in den nachfolgenden Publikationsgenerationen möglich
- Messung erfolgt i.d.R. auf Journal- oder Autorenebene, Impact eines Textes/Objekts wird nicht abgebildet

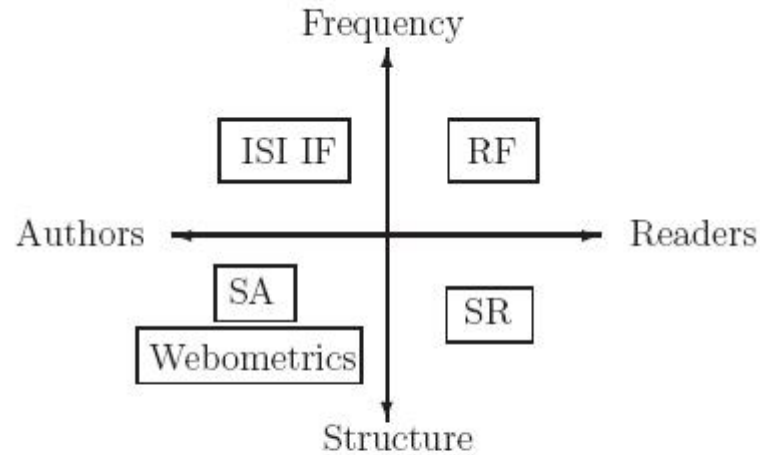
## ***Nutzungsbasierte Maße***

- leserzentriert
- Messung live und fortlaufend möglich
- Impact eines Textes/Objekts wird abgebildet
- Stark automatisierte Messung möglich



# Eine Kategorisierung

- Zitationsbasierte und nutzungsbasierte Maße**



ISI IF = Journal Impact Factor

RF = Reading Factor

SA = Structure Author

- basiert auf Netzwerk, das durch Autorenhandlung gebildet wird

- Bsp: Google PageRank, Zitationsgraphen, Webometrics

SR = Structure Reader

- basiert auf Kontextinformationen der Dokumentnutzung

- Recommendersysteme

- Bsp: Downloadgraphen

Aus: Bollen et al. 2005



## Nutzungsbasierte Verfahren: Standardisierung?

**COUNTER**, <http://www.projectcounter.org/>

**LogEc**, <http://logec.repec.org/>

**International Federation of Audit Bureaux of Circulations (IFABC)**,  
<http://www.ifabc.org/>

**AWStats**, <http://awstats.sourceforge.net/>

**MESUR: METrics from Scholarly Usage of Resources**,  
<http://www.mesur.org/MESUR.html>



## Nutzungsbasierte Verfahren: Fazit

- Alternative Impact-Maße (abseits der Zitationsmaße) sind modellierbar
- Aber: bislang wenig Standardisierung
- Vielversprechende, teils komplexe Modelle, v.a. im MESUR-Vorgänger
- Aufwändige Infrastruktur zur Herstellung und zum Austausch interoperabler Daten zwischen Servern erforderlich



# Projektskizze „Open-Access-Statistik“





## DFG-Projekt Open-Access-Statistik (OAS): Partner

Drittmittelgeber: Deutsche Forschungsgemeinschaft DFG

<http://www.dfg.de>

Projektpartner:

- Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Universitätsbibliothek Stuttgart
- Humboldt Universität zu Berlin (Computer- und Medienservice)
- Saarländische Universitäts- und Landesbibliothek

Laufzeit:

07/2008 bis 12/2009

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



## OAS: Motivation

### Ziel:

Infrastruktur für standardisierte, interoperable Nutzungsstatistiken unter besonderer Berücksichtigung von Open-Access-Angeboten

### Motivation:

- Open-Access-Angebote werden von etablierten Impact-Modellen nicht/selten erfasst -> Impact = Anreiz, der OA attraktiv machen kann
- etablierte, zitationsbasierte Impact Maße weisen Mängel auf
- elektronische Dokumente ermöglichen es nutzungsbasierte Maße darzustellen
- vielversprechende Evaluierungen nutzungsbasierter Impact-Modelle (Bollen et al. 2005)



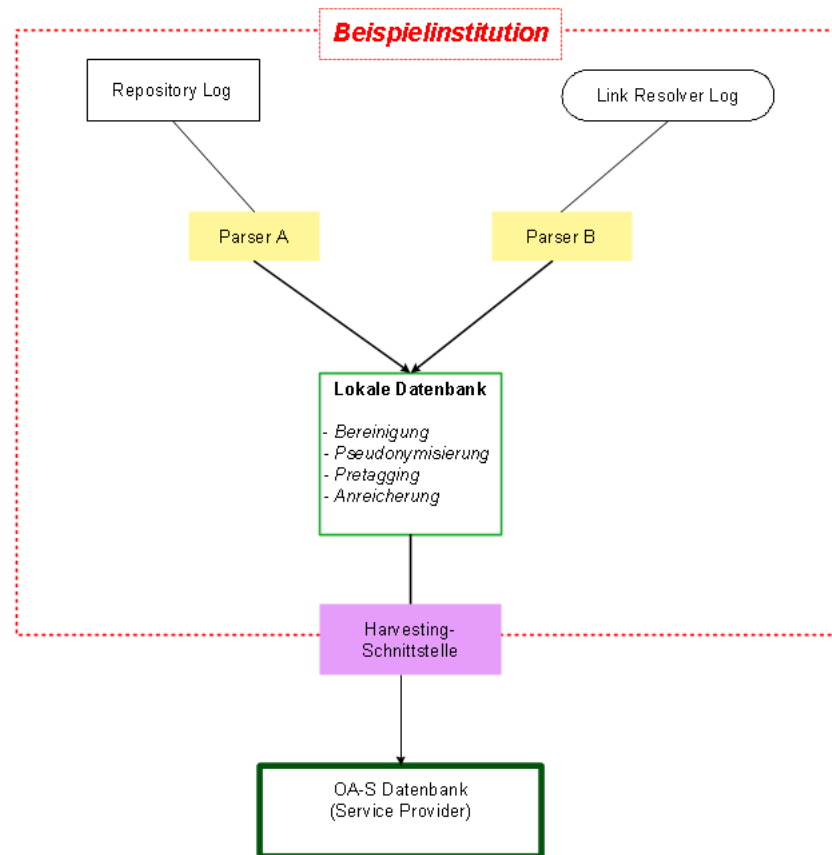
## OAS: Skizze

### Skizze:

- Aggregation von Nutzungsdaten verschiedener Open-Access-Angebote/ Server in einem Serviceprovider (v.a. Webserverlogs und Linkresolverlogs)
- Bereinigung von Verzerrungen (maschinelle Zugriffe durch Spider, Dublettenerkennung etc.)
- Aufbereitung der Nutzungsdaten anhand verschiedener Standards (COUNTER, LogEc, IFABC)
- Rückspielen der ermittelten Kennwerte in die verteilten Angebote und Verwendung als Metadatum (ggf. Ranking)



# OAS: Infrastruktur zur Aggregation von Nutzungsdaten



## OAS: Szenarien

### Welche Verwendung können die Daten finden?

- Entwicklung aufbauender Services, z.B. Recommender
- Zusatzservice für wissenschaftliche Suchmaschinen und Zusatzmetadatum für Datenbankanbieter
- Verwendung als szientometrische Information in Szenarien unterschiedlicher Komplexität (reine Häufigkeiten als Vorhersage von Zitationswerten bis hin zum MESUR-Ansatz)
- Kennzahlen basierend auf der Häufigkeit (Usage Factor) oder auf strukturellen Zusammenhängen (Usage Page Rank als Quantifizierung kontextuellen Nutzungsverhaltens) von Nutzungsevents
- Entwicklung und Evaluierung unterschiedlicher szientometrischer Verfahren





DEUTSCHE INITIATIVE  
FÜR NETZWERKINFORMATION E.V.

# Konzeptentwurf OA-Netzwerk Suche mit integrierten Mehrwertdiensten

Startseite

Über DINI

Mitgliedschaft

Wiss. Publizieren

Arbeitsgruppen

Dokumente

Veranstaltungen

**Service**

Kalender

Intern

DINI / Service / OA-Netzwerk Suche

## OA-Netzwerk Suche

Suche:  Alle Quellen  nach Fächern

### Treffer

Intrakranielle Volumenänderungen im Magnetresonanzt  
neuropsychologische Veränderungen bei Patienten mit  
Dörnte, Jan (Göttingen, 2006) - [ 87 mal zitiert ] [ viel gel

Expression von pro- und antiinflammatorischen Zytokin  
unter Normal- und Entzündungsbedingungen  
Wirth, Annika (Stuttgart, 2005) - [ 78 mal zitiert ] [ viel gel

Der Hedgehog Rezeptor Patched bei der Tumorentstehung  
Uhmann, Anja (Berlin, 2005) - [ 56 mal zitiert ] [ viel gel

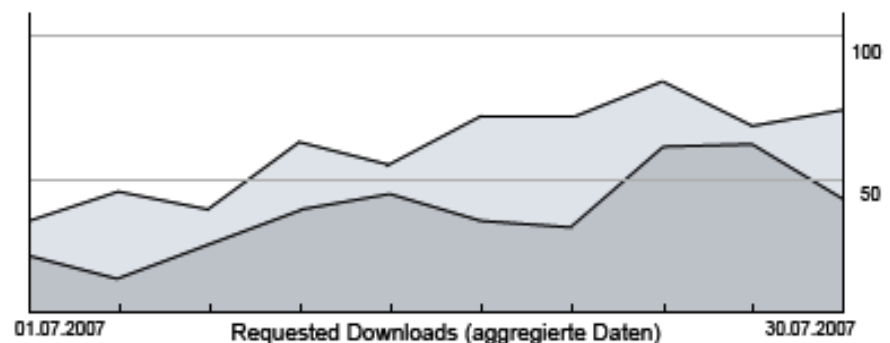
Die Rolle von Interleukin-6 beim Myelinabbau durch Ma  
Hilbert, Sören-Wibo (Berlin, 2007) - [ 42 mal zitiert ] [ re

Sortierung nach:

- Zitationshäufigkeit
- Usage factor

### Statistiken/Metriken

COUNTER IFABC LogEc



531 national (Berlin, DNB, Göttingen, Saarbrücken, Stuttgart)  
847 Verlage (BioMed)  
1378 Requested Downloads (Summe aggregierte Daten)

Zeitraum (letzte): 7 Tage 30 Tage 90 Tage

Datenquellen: national\* international Verlage

\* Berlin, DNB, Göttingen, Saarbrücken, Stuttgart

weitere Daten anzeigen

Schliessen



Your search:

All BASE content sources  Free content sources  Search

# Konzeptentwurf Integration in BASE

Current Search: **future (119776)**

119776 documents found in 0.8726 seconds

## 1. Future Fiscal and Budgetary Shocks

»Open in new browser window

**Title:** Future Fiscal and Budgetary Shocks

**Author:** Hian Teck Hoon ; Edmund S Phelps

**Description:** We study here the effects of future tax and budgetary shocks on present levels of economic activity and real interest rates in a nonmonetary and possibly non-Ricardian economy. The paper first takes up an (unanticipated) temporary tax cut to be effective on a given future date-a delayed "debt bomb." The sudden prospect of this future-dated shock causes at once a drop in the (unit) value placed on the firms' business asset, the customer, and accordingly on the price of ...

[https://mercury.smu.edu.sg/rsrchpubupload/4644/Future\\_budgetary.pdf](https://mercury.smu.edu.sg/rsrchpubupload/4644/Future_budgetary.pdf) (2.5k) [HTML]

Content provided by RePEc: Research Papers in Economics

» Check this title in Google Scholar

## 2. Symposium on the Future of the Universe and the Future of our Civilization

»Open in new browser window

**Title:** Symposium on the Future of the Universe and the Future of our Civilization

**Author:** Burdyuzha, V ; Kohzin, G

**Keywords:** Astrophysics and Astronomy

**Description:** Burdyuzha, V ; Kohzin, G ; Symposium on the **Future** of the Universe and the **Future** of our Civilization ; Astrophysics and Astronomy

**Year of Publication:** 1999-06-06T22:00:00Z

**Language:** en

<http://www.uniciv-future.ru> (0.7k) [HTML]

Content provided by CERN Document Server (CDS)

» Check this title in Google Scholar

### Sort your Results ?

Sort by:

relevance (standard)  
usage factor\* (COUNTER)

### Refine your Results ?

usage factor\* (IFABC)  
usage factor\* (LogEc)

by source:

author, Z-A

by subject:

document size, largest first

by document:

title, Z-A

by document:

date of publication, ascending

by year of publication:

by language:

by file type:

by author:

### Search History ?

- future (119776)

### \* Please note ?

Usage factor calculation data is provided by the central

Diskussion

Fragen?

Anregungen?



Kommentare?





# Literatur

Bollen, Johan et al. (2005): Toward alternative metrics of journal impact: A comparison of download and citation data. In: Information Processing and Management 41(6): S. 1419-1440. Preprint Online: <http://arxiv.org/abs/cs.DL/0503007>

Brody, Tim and Harnad, Stevan (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. Dlib Magazine, 10, Nr. 6, <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

Brody, Tim, Harnad, Stevan and Carr, Les (2005). Earlier Web Usage Statistics as Predictors of Later Citation Impact. Journal of the American Association for Information Science and Technology (JASIST). <http://eprints.ecs.soton.ac.uk/10713/>

Lawrence, Steve (2001). Free online availability substantially increases a paper's impact. Nature, 411, S. 521-522 <http://www.nature.com/nature/journal/v411/n6837/full/411521a0.html>

Sietmann, Richard (2006): Über die Ketten der Wissensgesellschaft. In: c't Magazin für Computer und Technik(12): S. 190-199.

Suber, Peter (2007): Why we need OA to citation data. In: Open Access News. News from the open access movement. Online: <http://www.earlham.edu/~peters/fos/2007/12/why-we-need-oa-to-citation-data.html>

