

Alternative impact measures for open access documents? An examination how to generate interoperable usage information from distributed open access services

IFLA 2010

Session 72 - Statistics and Evaluation

Göteborg, 11.08.2010

Initiated by:



Ulrich Herb
Saarland University and State Library, Germany
u.herb@sulb.uni-saarland.de

Funded by:



Overview

- Impact measures:
 - relevance
 - a categorisation

- Usage-based impact measures: standardisation?

- Project: Open Access Statistics
 - Aims
 - Technical infrastructure
 - Results
 - Outlook

Impact Measures

„The ‚impact factor‘ is the most commonly used assessment aid for deciding which journals should receive a scholarly submission or attention from research readership. It is also an often misunderstood tool.“
Dong et al. 2005

Impact measures: relevance

□ Individual level: *publish or perish*

- If a scientist does not publish she/he does not have any scientific capital, reputation or impact
- Without any impact, she/he won't make her/his career

□ Organisational level: evaluation

- Evaluation results determine prospective resources of institutes *and* the future main research
- Criteria: number of doctoral candidates, amount of third party funds, publications

From publications to impact

- ❑ Scientific reputation (or scientific capital) is derived from publication impact
- ❑ Impact is calculated mostly by citation measures
 - Journal impact factor (JIF)
 - Hirsch-index (h-index)

Especially within the STM domain

Citation impact: calculation

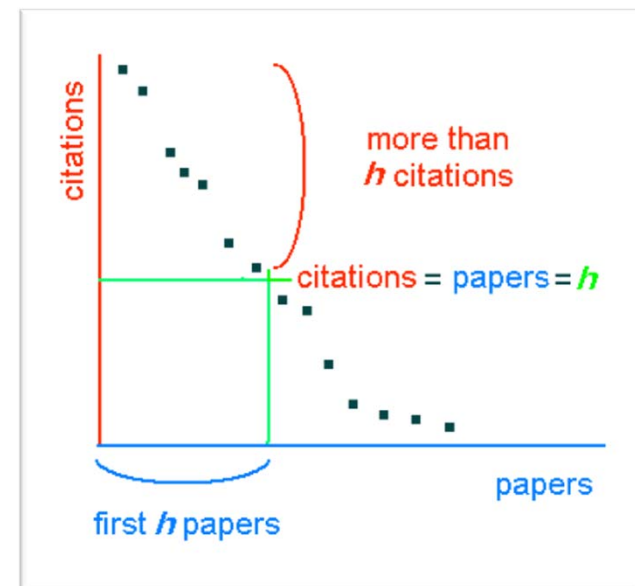
JIF

In year X , the impact factor of a journal Y is the average number of citations to articles that were published in Y during the two years preceding X

Garfield: „We never predicted that people would turn this into an evaluation tool for giving out grants and funding.“ From: Richard Monastersky (2005), *The Number That's Devouring Science* *The Chronicle of Higher Education*

H-index

A scientist has index h if h of N papers have at least h citations each, and the other $(N - h)$ papers have less than h citations each



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

Citation impact: critical points

- ❑ Restricted scope, exclusion of many publication types
 - ❑ Based exclusively on journal citation reports / web of science (JIF) or other databases
 - ❑ Language bias: items in English language are overrepresented within the database, so they reach higher citation scores
 - ❑ JIF focuses on journals: few articles evoke most citations
 - ❑ JIF discriminates disciplines with lifecycles of scientific information > 2 years
- Mixture of quality and popularity

Impact measures: a categorisation

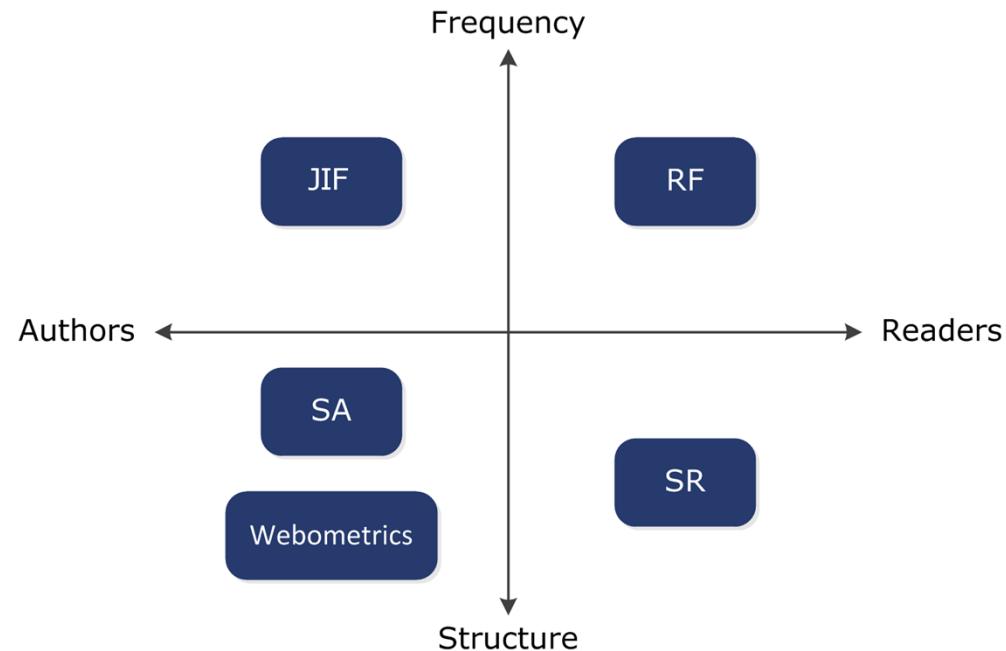
□ Citation based measures

- Author-centred
- Delayed measurement: at first in the following generation of publications
- Impact of a separate object is mostly not described

□ Usage based measures

- Reader-centred
- Measuring: on-the-fly and consecutive
- Impact of a separate object can be described
- Automated measurement is possible

Impact measures: a categorisation, pt. II



JIF = Journal Impact Factor

RF = Reading Factor

SA = Structure Author

- based on networks built by authors and their activities, e.g. Google PageRank, citation graphs, webometrics

SR = Structure Reader

- based on document usage and its contextual information, e.g. recommenders, download graphs

Bollen, J. 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>



Standards

„An important issue, however, was the lack of standards on how to produce and report the usage data in a way that could be compared“

Baker et al. 2008

Usage based impact: standardisation?

-  **COUNTER** Counting Online Usage of NeTworked Electronic Resources

<http://www.projectcounter.org>

- *LogEc*

<http://logec.repec.org/>

- 

<http://www.ifabc.org/>

Usage based impact: standardisation?

- The models mentioned differ in many aspects
 - Detection and elimination of non-human access (robots, automatic harvesting)
 - Definition of double click intervals
 - ...

- General problems
 - Ignorance of context information
 - Detection of duplicate users
 - Detection of duplicate information items
 - Ignorance of philosophical questions like: “What degree of similarity makes two files the same document?”

Alternative impact measures: conclusion

- ❑ Alternative impact measures are possible
- ❑ But: very little standardisation
- ❑ Promising, but complex examples/models like MESUR
<http://www.mesur.org>
- ❑ Requirement: sophisticated infrastructure to generate and exchange interoperable usage information within a network of several different servers



Project: Open Access Statistics

Open Access Statistics (OAS)

- 07/2008 – 02/2010
- Project partners:



Universität Stuttgart

HUMBOLDT-UNIVERSITÄT ZU BERLIN



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



SAARLÄNDISCHE
UNIVERSITÄTS-UND
LANDESBIBLIOTHEK



Initiated by:



DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

Funded by:

DFG

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

Open Access Statistics: motivation

- ❑ open access publications are often excluded from citation based impact measures
 - repository documents by definition
 - articles in open access journals due to their short citation history and often also due to their language
- ❑ citation based impact measures are revealing several deficiencies
- ❑ citation based impact measures should be complemented by usage based impact measures
 - because a multi-faceted approach could remedy some of their deficiencies
 - because the latter ones could create an incentive to use open access services

OAS: aims

- ❑ A common standard to exchange usage data between different services
- ❑ An infrastructure to collect, process and exchange usage information between different services
- ❑ Usage information should be processed according to the standards of COUNTER, LogEc and IFABC
- ❑ Additional service for repositories
- ❑ Implementation guidelines

OAS: associated projects

- Open Access Statistics



- DOARC

(Distributed Open Access Reference and Citation Services)



- Open Access Network



OAS: associated Projects

- ❑ Open Access Statistics addresses usage description
- ❑ DOARC address the issue of tracking citations between electronic publications
- ❑ Open Access Network
 - intends to build a network of repositories
 - will bundle the results of DOARC and Open Access Statistics in one user interface
 - offers services for DOARC and Open Access Statistics, e.g. deduplication of documents (based on a asymmetric similarity of fulltext documents)



Technical Infrastructure

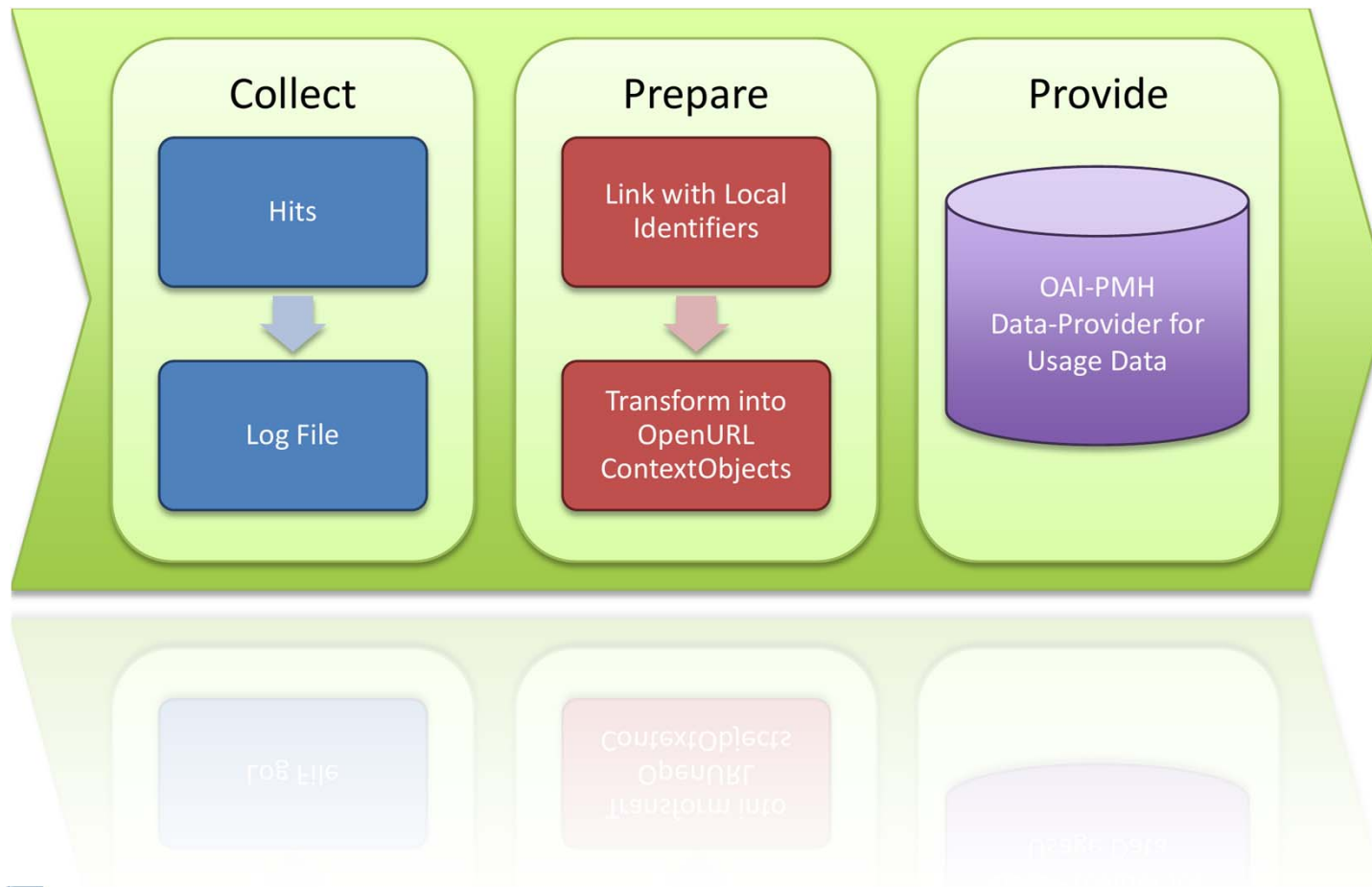
„Collecting, processing, and interpreting usage data is a challenge for libraries, big and small“

Manoff et al. 2006

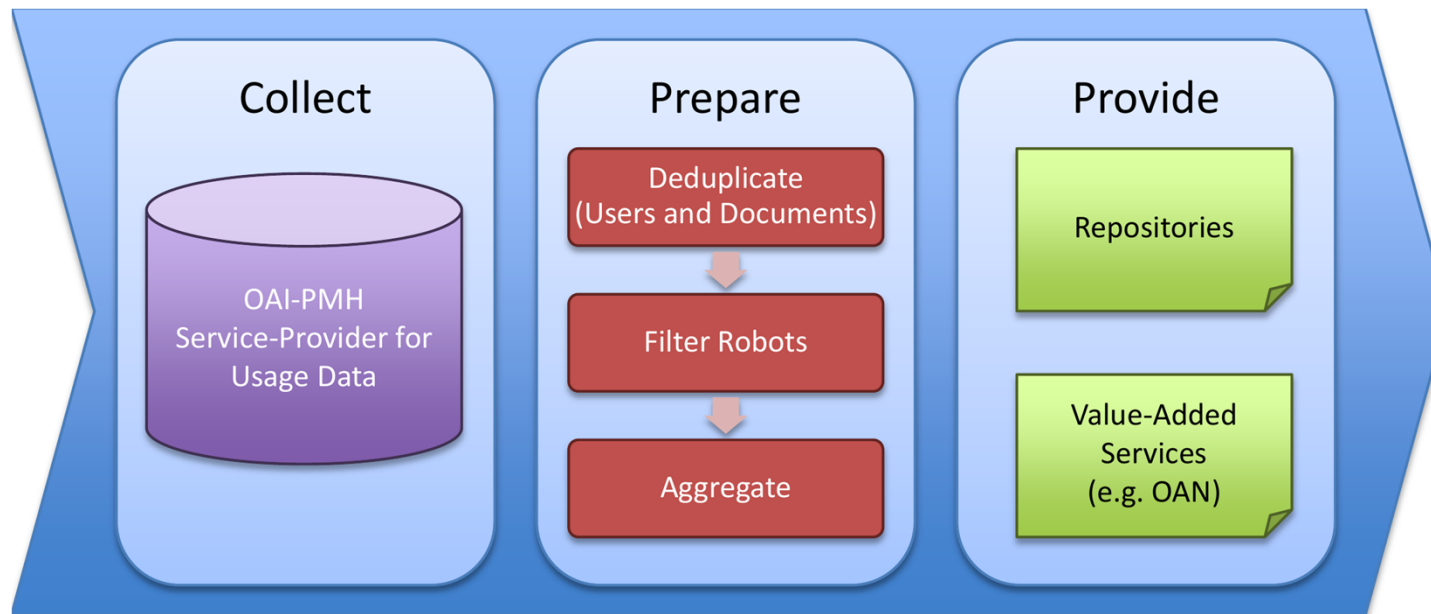
OAS: background

- ❑ Data pools at partner institutions
- ❑ Aggregation of usage events in a central service provider
- ❑ Services provided by the central service provider
- ❑ Usage data will be retransferred to the local data pools and to the Open Access Network Service

OAS: data provider



OAS: service provider



OAS: repository integration

Georg-August-Universität Göttingen

Georg-August-Universität Göttingen

Impressum | Kontakt | Feedback

GoeScholar > Naturwissenschaften, Mathematik und Informatik > Fakultät für Chemie > Zeitschriftenartikel - Fakultät Chemie >

Suche
 Erweiterte Suche

Browsen
» Fakultäten
» Zentren
» Titel
» Autor
» Jahr

Mein GoeScholar
» Dokument(e) anmelden
» Checkliste
» Meine Publikationsliste
» Benutzerprofil
» DEMO

Infos & Hilfe
» Hilfe & FAQ
» Rechtseinräumung

Login:
GWDG-User:
Passwort:

Verlinken Sie auf bzw. zitieren Sie dieses Dokument mit der folgenden permanenten URL:
<http://resolver.sub.uni-goettingen.de/purl?goescholar/3377>

Nutzungsstatistik

COUNTER IFABC LogEc

Datum	Aufrufe
28.09.2009	20
30.09.2009	13
01.10.2009	10
04.10.2009	8

Summe letzte 7 Tage: 89

Zeitraum: 7 Tage 30 Tage 90 Tage

Serie/Report Nr.: Organometallics, Vol. 12, 4, 1193-1200

Zusammenfassung: The set of alkali metal solid-state structures of $\text{Ph}_3\text{CM-nL}$ [$\text{M} = \text{K}, \text{Rb}, \text{Cs}$; $\text{L} = (\text{ligands}) \text{PMDTA}$ ($\text{N,ZV,}^{\text{JV}}\text{JV}^{\text{JV}}$ -pentamethyldiethylenetriamine), THF (tetrahydrofuran)] provides instructive comparisons. $\text{Ph}_3\text{CK-THF-PMDTA}$ (1) crystallizes as a monomeric contact ion pair: the K^+ cation is symmetrically σ -coordinated to one of the phenyl rings, but not to the deprotonated central carbon. Both [$\text{Ph}\&\text{Rb-PMDTA}$] (2) and $\text{P}^{\sim\sim}\text{CCWPMDTA}$ (3) form one-dimensional polymers and eschew THF. The Rb cations in 2 bridge the triphenylmethyl moieties by $+x^{\text{m}}\text{xm}$ tion to separate phenyl rings. This gives rise to a zigzag chain. In 3, each Cs cation also bridges two carbanions, but in a somewhat different fashion. While Cs^+ is located rather symmetrically (119 above the phenyl ring of one trityl moiety a "propeller"-like coordination to a second trityl anion

OAS: usage scenarios

data may be used

- ❑ from an user perspective as a criterion to estimate the relevance of a document (e.g. rankings)
- ❑ from an author perspective as an indicator for the dissemination of a concept
- ❑ from a service provider perspective:
 - as additional metadata for search engines, databases ...
 - as a recommender service
- ❑ from a repository perspective:
 - as a recommender service
 - as additional metadata for users



Results and Outlook

OAS: lessons learned

linkresolvers are rarely offering suitable information

- external services (ovid) don't offer usage information
- SFX-logs are very heterogenous
 - target may be a splash page or a fulltext
- hardly any information about open access documents

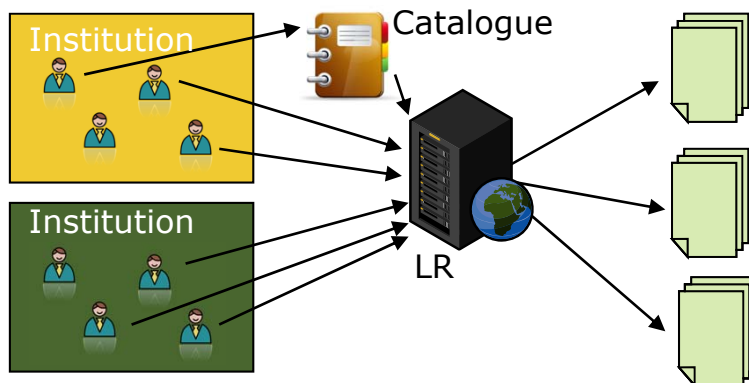
document deduplication seems difficult

- a given document may have more than one IDs
cause: multiple fulltext deposit on several repositories
- a given document may have several splash pages on different servers pointing at one fulltext on one single server
cause: metadata harvesting
- ...

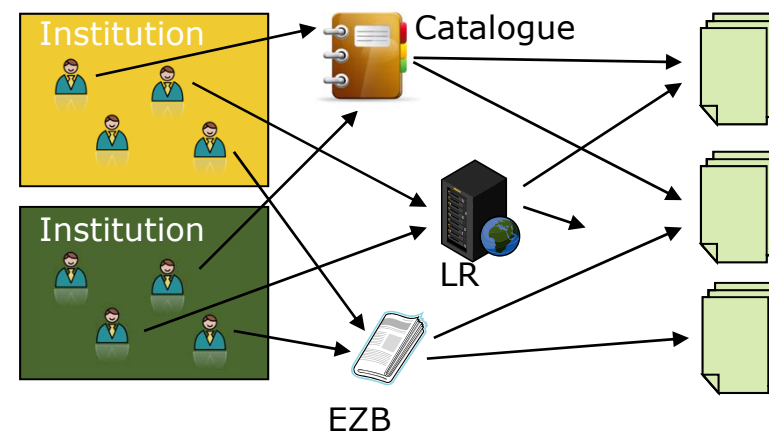
OAS: lessons learned

- ❑ The requirement for a central clearing house
- ❑ A lot of unnecessary data (OpenURL CO)
→ increase of the data size by factor ~ 10
- ❑ Different situation with Linkresolver

USA



Germany



OAS: results

- ❑ Infrastructure for exchange usage statistics
- ❑ Modules for OPUS- and DSpace-based repositories, other products can be configured easily (<http://www.dini.de/projekte/oa-statistik/english/software/>)
- ❑ Specification of the data format and exchange
- ❑ Online demo (<http://oa-statistik.sub.uni-goettingen.de/statsdemo>)
- ❑ Website with further information (<http://www.dini.de/projekte/oa-statistik/english/>)

OAS: further plans → OAS 2

Aims for a possible second funding:

- ❑ Opening the OAS infrastructure to offer standardised usage statistics
- ❑ Evaluation of metrics more sophisticated than the calculation of pure usage frequencies
- ❑ Cooperation for international comparable usage statistics
- ❑ Offer a suitable service infrastructure

OAS: international cooperation

- SURFSure
- COUNTER
- PIRUS
- Knowledge Exchange – Usage Statistics Group
- NEEQ
- PEER
- OAPEN

Thanks for your attention!

IFLA 2010

Session 72 - Statistics and Evaluation

Göteborg, 11.08.2010

Initiated by:



Ulrich Herb
Saarland University and State Library, Germany
u.herb@sulb.uni-saarland.de

Funded by:

