Open Access Statistics:
Interoperable Usage Statistics for
Open Access Documents

Public Knowledge Project / PKP
Scholarly Publishing Conference 2011
Berlin, 28.09.2011

Initiated by:
Ulrich Herb
Saarland University and State Library, Germany
beucke@sub.uni-goettingen.de

Funded by:
Overview

- Impact measures:
  - relevance
  - a categorisation

- Usage-based impact measures: standardisation?

- Project: Open Access Statistics
  - Aims
  - Technical infrastructure
  - Results
  - Outlook
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
Usage based impact: standardisation?

- [COUNTER](http://www.projectcounter.org)
- [LogEc](http://logec.repec.org/)
- [IFABC](http://www.ifabc.org/)

[Open-Access-Statistik](http://www.ifabc.org/)
[CC BY-SA](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
- ...

[Open Access Statistics]
Alternative impact measures: conclusion

- Alternative impact measures can be designed

- 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
Alternative impact measures: conclusion

„Our results indicate that the notion of scientific impact is a multi-dimensional construct that can not be adequately measured by any single indicator, although some measures are more suitable than others. The commonly used citation Impact Factor is not positioned at the core of this construct, but at its periphery, and should thus be used with caution.“

„Usage-based measures such as Usage Closeness centrality may in fact be better ,consensus\measures.“

Open Access Statistics (OAS)

Initiated by:

Funded by:
OAS – Fact Sheet

- First term 2008 - 2010

- Project partners:
  - Göttingen State- and University Library
  - Computer and Media Service, Humboldt-Universität zu Berlin
  - Saarland University and State Library
  - University Library of Stuttgart

- Funded by German Research Foundation (DFG)

- [http://www.dini.de/projekte/oa-statistik/english/](http://www.dini.de/projekte/oa-statistik/english/)
Challenges

- Log files as a result of the use of repositories
- Gathering and aggregation of usage events through a central service provider
- Value-added services provided by the central service provider
- Usage data then to be retransferred to the repositories
Aims

- A common standard to enable the exchange of usage data between different services (e.g. repositories)

- An infrastructure to collect, process and exchange this usage data

- Usage data has to be provided by the repositories according to standards (COUNTER, LogEc and IFABC)

- Value-added services for repositories

- Implementation guidelines
Technical Infrastructure
Data provider
Service provider

**Collect**
- OAI-PMH Service-Provider for Usage Data

**Prepare**
- Deduplicate (Users and Documents)
- Filter Robots
- Aggregate

**Provide**
- Repositories
- Value-Added Services (e.g. OAN)
Results and Outlook
Lessons Learned

- Need for a central clearing house

- An amount of unnecessary data (OpenURL CO) → increase of the data size by factor \( \sim 10 \)

- Potential legal problems with German laws on privacy issues
Results

- Infrastructure for exchange of usage statistics
- Modules for OPUS- and DSpace-based repositories, other products can be configured easily, [http://www.dini.de/projekte/oa-statistik/english/software/](http://www.dini.de/projekte/oa-statistik/english/software/)
- Online demo [http://oa-statistik.sub.uni-goettingen.de/statsdemo](http://oa-statistik.sub.uni-goettingen.de/statsdemo)
- Website with further information [http://www.dini.de/projekte/oa-statistik/english/](http://www.dini.de/projekte/oa-statistik/english/)
Open Access Statistics 2 (OAS 2)

Initiated by: GBV VZG

Funded by: DFG
OAS 2 – Aims

Start in April 2011 – for 2 years:

- Clarification of legal questions (laws on privacy protection)
- Opening the OAS infrastructure to offer standardised usage statistics
- Evaluation of metrics
  a) based on the pure frequency of usage
  b) more sophisticated approaches
- Cooperation to facilitate international comparable usage statistics
- Offer a functional service infrastructure
  - Sustainability report after the first year
International cooperation

- PIRUS Publisher and Institutional Repository Statistics, UK
- SURFSure Statistics on Usage of Repositories, NL
- Knowledge Exchange Usage Statistics Group
  - Denmark’s Electronic Research Library (DEFF)
  - German Research Foundation (DFG)
  - Joint Information Systems Committee (JISC)
  - SURFfoundation, Netherlands

- Common sense
  - Exchange format: OpenUrl ContextObjects
  - Transfer via OAI-PMH
  - Infrastructure based on a data provider – service provider system
  - Normalisation: Robots-Detection

- COUNTER, NEEO, PEER, OAPEN ...
Thanks for your attention!

Public Knowledge Project / PKP
Scholarly Publishing Conference 2011
Berlin, 28.09.2011

Initiated by:

Ulrich Herb
Saarland University and State Library, Germany
beucke@sub.uni-goettingen.de

Funded by: