# Item-level Usage Statistics - a review of current practices and recommendations for normalization and exchange

Christine Merk
Christine.Merk@uni-konstanz.de
University of Konstanz

Frank Scholze
Frank.Scholze@mwk.bwl.de
Ministry of Science, Research and the Arts Baden-Württemberg

Nils Windisch
windisch@sub.uni-goettingen.de
Göttingen University and State Library

**Abstract**

Purpose

The JISC Usage Statistics Review Project aims to formulate a fundamental scheme for recording usage data and to propose a standard for its aggregation to provide meaningful and comparable item-level usage statistics for electronic documents like e.g. research papers and scientific resources.

Approach

A core element of the project has been a stakeholder workshop. This workshop was held in Berlin on July 7th and 8th 2008. Representatives of key stakeholder groups (repositories, libraries, COUNTER, IRStats, JISC, LogEc, MESUR, OA-Statistics and other Open Access projects) were invited. During the workshop a fundamental scheme for the recording and the exchange of log files was discussed as well as the normalization of data collected.

Findings

The following mandatory elements describing usage events were agreed upon during the stakeholder workshop: Who: Identification of user/session, What: Item identification and type of request performed (e.g. full-text, front-page, including failed/partially fulfilled requests), When: Date and time, Usage event ID. The following elements were regarded as optional: From where: Referrer/the referring entity and Identity of the service.

Usage events should be exchanged in the form of OpenURL Context Objects using OAI. Automated access (e.g. robots) should be tagged. The definition of automated access has to be straightforward with an option of gradual refinement. Users have to be identified unambiguously, but without recording personal data to avoid conflicts with privacy laws.

Policies on statistics should be formulated for the repository community as well as the publishing community. Information about statistics policies should be available on services like OpenDOAR and RoMEO.

Originality

This paper is based on the detailed project report to the JISC, available at http://ie-repository.jisc.ac.uk/250/

## 1 Introduction

The promise of usage data for scholarly evaluation lies in its timeliness. Unlike citation-based statistics, counts and rankings can be updated soon after the actual usage event. Usage data can be the basis for calculating the visibility and hence the impact of electronic publications. Metrics based on usage data can provide transparent information for authors about how visible or how popular their work is. This can support authors' decisions about where to publish. Readers can evaluate a publication or an author on this basis. However usage data is not used widely in the daily routine of scholarly evaluation. Bollen (Bollen et. al., 2007) lists four main issues as reason for this, the most basic being the lack of standards for recording and aggregating usage data. To address this, the JISC Usage Statistics Review Project aims to formulate a fundamental scheme for recording usage data and to propose a standard for its aggregation to provide meaningful and comparable item-level usage statistics for electronic documents like research papers and journal articles.

The project was funded by JISC and conducted by a research consortium in which the Humboldt University Berlin (Computer and Media Service), the Göttingen University and State Library, the Library of the University of Konstanz, the Saarland State and University Library and Stuttgart University Library worked together.

A core element of this review has been a stakeholder workshop. It was held in Berlin on July 7th and 8th 2008. Representatives of key stakeholder groups (repositories, libraries, COUNTER, IRStats, JISC, LogEc, MESUR, OA-Statistics and other Open Access projects) were invited.

This paper gives a short overview of current projects, services and initiatives in the field of collecting usage data for online scientific publications. This is followed by a presentation of the workshop results dealing with the technical aspects of normalizing, exchanging and mapping usage data. Then legal and policy issues about the recording of usage data are described, and finally the future development of services in this field is envisioned.

## 2 Review of current practices

The DRIVER inventory study (Eijndhoven and van der Graaf, 2007) of 114 repositories in 17 European countries showed that about 70 percent of those repositories logged download and access data, but only 30 percent offered item-level usage statistics for their end-users in 2006. About half the repositories which had such a service were based in the UK and a quarter of all repositories in the survey were planning their introduction.

In the context of this review, we selected a number of existing approaches and projects on measuring electronic usage and had a look at their practices.

**IRStats** (Interoperable Repository Statistics) is a project at the University of Southampton aimed at the design of a usage statistics module for repositories using eprints or Dspace. The result of the project is a pilot version of the statistics tool at the eprints-repository at the School of Electronics and Computer Science at the University of Southampton. The tool excludes multiple clicks within 24 hours and uses the AWStats-robots list to delete non-human access. So far, the package cannot be implemented on a broad level and can only be operated as an add-on to AWStats

**LogEc** is a free online service which complements the metadata aggregator RepEc, which specializes in economic literature. LogEc provides the statistics for items available from the participating services of RepEc: IDEAS, EconPapers, Socionet Personal Zone and Inomics. It also creates rankings of these items by the number of abstract views and the download frequency. The statistics for every available service i.e. abstract views or downloads for each item can be accessed via its front-page. This free online service is hosted by the Swedish Business School at Örebro University. Multiple clicks are excluded using IP addresses for re-visits within one month. The log analyzer is run locally by the participating services, and processed data is then uploaded to LogEc. Non-human access is excluded by the use of robots.txt combined with the algorithm that users who access more than 10% of all the items on RepEc within a month are not counted.

**DINI e.V.** (German Initiative for Networked Information) supports co-operation among, and the standardization of, information and communication services in Germany. For repositories it sets standards with its DINI-Certificate. So far, its criteria for this certification do not include a comprehensive checklist for repository statistics; but with the projects such as OA-Statistics and the OA-Network, it is currently working on the implementation of a pilot version for a network of Open Access repositories.

**Australian Benchmark Statistics Service (BEST)** is a project which was conducted by the Australian Partnership for Sustainable Repositories (APSR) between September and December 2007. It makes recommendations for the introduction of item-level statistics in the Australian context (Benchmark Statistics Project, 2007). For the exchange of usage data they incorporated OAI-PMH. Its approach to identify non-human access is similar to LogEc but adds the AWStats-list.

**MESUR** (Metrics from Scholarly Usage of Resources) is the most advanced project in the field of usage statistics for scholarly work. It conducts a survey of more than 60 different metrics of scholarly impact, formulating guidelines and recommendations. It uses a database with about 1 billion usage events to test different usage metrics and to generate a network of journals and a network of items. The tracking of users' click streams from document to document creates relational data. Based on this data, relationship matrices were for example used to design the prototype of a recommender system at the article-level (Bollen and Van de Sompel, 2006).

The data from different sources – i.e. publishers, aggregators, and institutions – was collected between 2002 and 2007. Necessary pieces of information for subsequent processing are: unique identification of the event, identification of the user or the session, a persistent identifier for the item and the time of the event (Bollen et al., 2007). In cases where a unique identifier is not available, MESUR uses a 'bag of identifiers' approach to de-duplicate. It starts out with the ISSN; then the record is compared to the bag of identifiers, which contains versions and abbreviations of journal titles. The items below the journal level are matched using the year of the publication and the first 25 characters of the title. Non-human access is identified by its typical behavioural characteristics. Part of the data – the linking server log files from the California State University system – was packed into OpenURL Context Objects and harvested via OAI-PMH.

The Open Source analytical software **AWStats** is widely used in the repository context. Multiple clicks within one hour are excluded via the IP address. Non-human access is excluded using AWStats' own robots list. This list sets some kind of a de facto standard for robot exclusion, even though different versions of the list are used.

The **COUNTER** (Counting Online Usage of NeTworked Electronic Resources) code of practice for journals and databases is the well-established standard for journal usage statistics. In August 2008 its third release was published. The major changes in this release are the mandatory use of SUSHI for harvesting, the requirement to provide the reports in XML-format and the exclusion of non-human access. For the latter a reference list containing 36 robots and crawlers is given (COUNTER, 2008: Appendix K), which is open for updates. Multiple clicks to an HTML document by a single user are not counted if they happen within 10 seconds. For PDF documents this time span is 30 seconds. Users are either identified by their IP address or session cookies. Only successful full-text-requests, i.e. with the http return codes 200 and 304 (COUNTER, 2008: Appendix D), are counted. COUNTER uses journals as the lowest level of granularity for its statistics. Project PIRUS (Publisher and Institutional Repository Usage Statistics) unites the publishing and the repository communities in an effort to formulate common standards for the recording of item-level usage statistics which are applicable within both contexts. First results can be expected in the beginning of 2009 (PIRUS, 2008).

The **United Kingdom Serials Group (UKSG),** as an interest group for publishers and librarians, is also involved in PIRUS. Besides that, COUNTER Executive Peter Shepherd (2007) undertook a stakeholder survey on the introduction of a Usage Factor (UF) for journals in 2007, commissioned by the UKSG. The UF puts the journal usage from the COUNTER reports in relation to the number of articles published online; both variables are in reference to a specified period (Shepherd, 2007). In the near future, the UKSG plans to issue a request for proposals for a framework for the UF.

The **International Federation of Audit Bureaux of Circulations (IFABC)** sets standards for the measurement of the usage of online content from non research-related commercial providers. Though its standards do not deal with scholarly content, it is included in this overview because its guidelines represent one prominent way of measuring internet traffic. Page impressions and visits per domain are counted (IFABC WWW Standards, 2001). The IFABC sets minimum standards; it is in fact a framework for different approaches. For example, users can be identified either through their IP address and the user agent, a cookie or a registration ID. The standards of the German member organization of the IFABC – the IVW – are more specific. They prescribe the use of tracking bugs or pixels. Multiple clicks are excluded within a time span of 30 minutes (IVW, 2008). The British ABCe implements the IFABC rules and uses a list of robots for the exclusion of non-human access. The documentation for the exclusion of robots and the harvesting of usage data is not publicly available because the IFABC and its members are commercial services which provide documentation only for their members.

## 3 Fundamental scheme for recording item-level usage data

### *3.1 Data sources*

The request for a document in itself says nothing about its actual use i.e. reading or citing it; more accurately it should be called access data. Nevertheless, the term 'usage data' is predominantly used. At the workshop, the definition of an access was agreed as the successful request for an item or its front-page. Inclusion in the result list of a search does not constitute a usage event because it does not indicate clearly enough the user's interest in the item.

Log files come in different formats depending on the source. There are differences between web servers in general, and also between licence servers, link resolvers, and repository software packages. In order to make them comparable, they have to be parsed and converted into a normalized format. The access to web pages is primarily recorded by web server logs, but can also be derived from link resolver logs, licence servers and repository software packages.

Linking servers can also be used to record usage data. In order to do this, it is essential to create a digital library infrastructure that enables linking servers to record as much usage data as possible. This is not the case in many libraries or research institutions in Europe. Whereas licensed publications can be tracked elegantly, content from repositories is not always covered by link resolver systems. Furthermore, documents cannot be accessed only via link resolvers; they are also available directly via a persistent identifier on a publisher site, for example.

Web server logs have a higher, i.e. full, coverage because they are always record the event when an item is requested. On the other hand, they are not always available to the institution which provides the licence for access, but does not deliver the content. These usage events can be logged by the link resolver. The inherent problems in

these two options make a combined approach the most promising. This means that usage data from web server logs and from link resolvers needs to be merged.

### *3.2 The fundamental scheme*

Before records can be aggregated, they have to be made comparable between different web servers and between different repositories. Therefore, the recording function has to be harmonized; this is called the 'fundamental scheme' throughout this report. During the workshop the participants agreed on the basic elements for recording usage. These are the identification of the user, the session, the item requested, the type of request, and the date and time of the request. These elements should be an integral part of the log files. They provide information about the basic questions 'Who?' 'What?' and 'When?' But beyond these entries, the schema should be open to extension – the identity of the service, the referrer, and the referring entity would especially lend themselves to extended the format.

#### *3.2.1 User / Session / Usage event ID*

The user who requests a resource has to be identified to make the exclusion of multiple clicks viable and to facilitate the tracking of the click stream through the online content. The first option for this is the IP address; but as it can be tracked back to the user, privacy issues arise. It is questionable whether the IP address should be recorded at all. Another problem is the issue of proxy servers or a network of computers which share the same IP address.

The first alternative is session IDs for every visit, i.e. a UUID (Universally Unique Identifier) for a continuous click stream. This allows to identify multiple requests of a document by a user during one session. It also shows the user's path from one document to another when the requests for one session ID are ordered according to their time stamp.

The second option is the use of session cookies or cookies. A session cookie is deleted when the browser is closed. A cookie, which expires after (for example) one month, makes return visits within this time span identifiable and it extends the length of the recorded click stream. A disadvantage is the non-selection of users, who disable cookies. Various studies report that about 5 to 10 percent of internet users turn off cookies. Another downside of cookies is the high cost of implementation because they have to be configured for the various software programmes in use.

A unique identifier for every event or session is also a necessary element. The identifier facilitates de-duplication and prevents the double counting of events, which were triggered by a link resolving event and therefore recorded twice. So far, the technical implementation of session IDs for linking servers is difficult and rarely available. The MESUR project used an extension for SFX. Whether other link resolver services are technically able to implement session IDs is unclear.

#### *3.2.2 Item*

The item has to be identifiable for the aggregation and the de-duplication of items and -- if event IDs are not available -- for the de-duplication of events. Multiple hits for HTML documents which contain images, and are therefore recorded multiple times in the log file, have to be excluded, too. The items have to be identified on the basis of their metadata.

Ideally, every document is assigned a unique identifier like a DOI (Digital Object Identifier), a URN (Universal Resource Name) or some other similar persistent identification solution. Unfortunately, different repositories adhere to different standards for unique item identification; a unique identifier for all items has to be made a pre-requisite for the exchange of usage data. In cases where this is not possible, fuzzy matching based on the metadata is an option in order to aggregate the events for a single item (Bollen and Van de Sompel 2005). Such an approach has also been adopted by MESUR and the OA-Network. This solution is also viable for the identification of documents which are part of more than one repository and have been assigned different unique identifiers. The de-duplication of items and events has to be addressed by the aggregator in order to be coherent.

### *3.3 Non-human access*

Log files are not only a record of human accesses, but also of the activity of spiders or web crawlers. Log file entries for spiders can be excluded using different strategies (like e.g. robots.txt, robots lists, atypical usage patterns) which have already been mentioned in section 2. The click stream for a human user is supposed to be shorter and to follow a meaningful pattern, while automated accesses randomly retrieve all the documents. This avoids the dependence on potentially incomplete or outdated lists of robots. Several participants argued against the deletion of the records of non-human accesses. The log file entries from non-human accesses could either be tagged at the local level or by the aggregator. Tagging at the local level would mean the additional effort of standardising repositories' practices on an international level. The deletion or tagging of non-human access is therefore most effectively done on a centralized level. This would allow for the consistent exclusion of robots.

The general consensus is that there will always be non-human accesses and that not all of these can be identified. But it is essential to have common guidelines for any such exclusions in order to make the data comparable. The identification should be pragmatic and easy to implement on a wide variety of systems (like the example of RepEc). It should also be possible to refine and adapt the identification gradually.

## 4 Exchange of usage data

### *4.1 A common protocol for harvesting usage data*

The two competing approaches for the harvesting of usage data are OAI-PMH and SUSHI. The harvesting is not just undertaken in one direction from content provider

to aggregator, but also vice versa. The content provider (e.g. a repository) can harvest aggregated statistics or even metrics from the aggregator in order to supply its users with the information needed.

The SUSHI-protocol is very well defined, but the experiences with its implementation are ambiguous and it is not as well embedded in the repository community as OAI-PMH. SUSHI is the standard for exchanging statistics in the publishing community. The workshop's participants were in favour of OAI-PMH. There will have to be crosswalks between the two options if OAI-PMH instead of SUSHI is implemented. This is particularly important on the semantic level. Therefore, it is very important that the results of this workshop are fed into the development of COUNTER regarding the extension of its standards to the item-level.

The implementation of FTP as a transfer protocol is technically less complicated, but harvesting is less accurate than with OAI-PMH. Files which have already been harvested cannot be tagged and the deletion of files is not uniformly recorded, as FTP does not provide a naming standard for files.

Workshop participants agreed upon the usage of OpenURL Context Objects as containers for the XML payload. The advantages of OpenURL Context Objects are that they can be easily extended and that the data is highly compressible.

### *4.2 Mapping of the usage data*

Expression of usage data as OpenURL Context Objects has already been proposed by Bollen and Van de Sompel (Bollen and Van de Sompel, 2006). We have adapted their proposal and adjusted it slightly to the needs of the fundamental scheme, ensuring that we have taken the OpenURL Standard (ANSI/NISO Z39.88-2004) into account. Starting out with the entries in a log file in the left-hand column, the respective elements within a Context Object are identified in Table 1 in the right-hand column.

**Table 1: Mapping of usage data to the data structure of OpenURL Context Objects**

| Log file entry | OpenURL Context Object |
|---|---|
| document identifier | *Referent* |
| time of access | *Timestamp* |
| event identifier | *Context Object Identifier* |
| IP address | *Requester* |
| session identifier | *By-Value metadata of the requester* |
| user agent | *Referrer* |
| http status code | *ServiceType* |

The *Referent* is the requested item, which is to be identified by some kind of persistent identifier, which could be the URN for example. *By-Value Metadata* can be added to the *Referent* in case the items have to be de-duplicated. By using the *By-Reference Metadata* instead, the event data could be merged back with the metadata if necessary. The time of the access and the event identifier are part of the header of the XML Context Object.

The user's IP address or an encoded version of it can be mapped onto the *Requester* field. A separate entry for a session ID does not exist, but it can be included into the *Requester*'s metadata. The format for this entry is yet to be defined (Bollen and Van de Sompel, 2006: 301). The user agent is mapped onto the *Referrer* field.

The mapping of the HTTP status code onto the *ServiceType* is not yet possible. Possible values for the *ServiceType* entity are 'fulltext', 'abstract', 'citation', 'holdings', 'ill', and 'any' according to the OpenURL registry. To include the information from the HTTP response code, new values would have to be defined. The information as to whether the *Referent* is a full-text or an abstract should be included in the Context Object too, although it might not be available from the log files. The content provider has to provide this information and either add it to the log files or write it into the By-Value Metadata of the ServiceType field. In order to have both kinds of *ServiceTypes* the Context Object can be filled with two *ServiceType* entries.

Context Objects are harvested using OAI-PMH but as they are in the XML-format they should also be harvestable with SUSHI. The OAI-header has to be replaced by the SUSHI header. SUSHI itself does not require the XML payload to have a specified format (ANSI/NISO Z39.93-2007). Technical interoperability cannot be assessed further at this point, but the effort should be made in the future to stay in line with developments from the publishing community.

## 5 Comparison of the fundamental scheme with other practices

A comparison between the workshop proposal and COUNTER measures according to the third release of their Code of Practice is not viable. The granularities are too different at this point. As COUNTER is working on the extension of its code to the article-level, the results from the PIRUS project should be taken into account as soon as they are available.

The LogEc scheme is just slightly different as two aspects of the fundamental scheme are not yet defined in more detail: the normalization of the data and the time criterion for the exclusion of multiple clicks. LogEc does not harvest with a pull-mechanism but the locally-analyzed usage events are uploaded using FTP.

Making the statistics comparable to the IFABC standards is a little more difficult as the robots list is not freely available. The definition of what is counted as access is tailored to websites.

**Table 2: Comparison of the JISC usage statistics review proposal with relevant parts of other schemes**

| Criteria | Fundamental scheme | COUNTER | LogEc | IFABC |
|---|---|---|---|---|
| **Granularity** | Item-level | Journal-level | Item-level | Page impression, visit |
| **Definition of usage event** | Successful abstract views and downloads | Successful full-text requests | Successful abstract views and downloads | Number of page impressions per domain, number of visits |
| **De-duplication of multiple clicks** | No specification | HTML 10 sec., PDF 30 sec. | One month | 1 hour |
| **Identification of non-human access** | No specification (usage patterns) | Robots list | Robots.txt, dynamic criteria | Robots list |
| **Harvesting** | OAI-PMH | SUSHI | Locally-analyzed usage data uploaded via | No specification |

## 6 Legal constraints for recording and aggregating log files

Privacy laws can be infringed by the recording and processing of IP addresses or the use of cookies, but regulations vary strongly between countries. The EU legislation is stricter than the US regulations but it is less strict than German legislation. The legal situation in the UK is mainly dominated by EU regulations.

In the US, the Federal Privacy Protection Act (1974) applies only to public authorities. The Telecommunications Act (1996) was mainly designed for the context of access provision; it does also not apply to content providers. Users' privacy protection relies heavily on self-regulation; privacy protection laws for the private electronic communication sector do not exist in the US. The transfer of personal data from the EU to the US is not permitted under European law. The only exception is companies which belong to the safe harbour system. The participating US companies have to comply with EU standards.

Germany has very comprehensive regulations concerning privacy. The constitution grants the right of informational self-regulation. The German Telemedia Act (TMG) defines the boundaries: personal data must not be collected or processed unless it is for the purpose of providing a service or billing. In the German legal context the recording of IP addresses is strongly restricted and the current interpretation of the Act is ambiguous. To avoid legal problems, it would be best to pseudonymize IP addresses shortly after the usage event or to implement some sort of session identification, which does not record IP addresses.

Copyright for statistics is a debatable issue: the basic question is whether statistics have enough inherent originality to have them protected. If they lack originality, they would be categorized as facts and would as such not be protected by copyright laws. Under the EU Database Directive or the German Copyright Act the storing and arranging of data implies intellectual input which can however be protected.

## 7 Usage statistics policies

It was discussed during the workshop whether usage statistics should be open access and, if yes, to what extent. There was a common understanding that the raw data should not be publicly available as privacy might easily be breached. Access to raw data should therefore be strictly regulated and limited (e.g. for metrics research).

There was less unanimity about the status of processed data, i.e. usage statistics. Many repositories are, on the one hand, part of the Open Access movement, and therefore do not want to contradict its ideals. On the other hand, the infrastructure for services has to be financed. Usage statistics would be a valuable service. They can be used for research evaluation and they are the pre-condition for the introduction of recommender systems. A third option besides a freely-available or a fee-based service is a partially publicly-available service. Basic measures can be made

available in Open Access while the access to more sophisticated measures and recommender systems can be restricted. In this layered approach the content providers can decide what level of statistics they offer.

A repository policy should contain guidelines for the collection and processing of the data. The Berlin Principles on the Ranking of Higher Education Institutions can be taken as an orientation for the creation of a policy on statistics (Berlin Principles on Ranking, 2006).

The aggregation and processing of usage data, as well as quality assurance, has to be undertaken by trusted and neutral third parties. Publicly-available usage statistics might interfere with the economic interests of publishers. Centralization minimizes potential noise due to divergent local practices. The central service should also install an auditing process to ensure the quality of data recording (as is already practiced by COUNTER or IFABC). Existing institutions which lend themselves to the role of aggregator are, for example, DRIVER, DINI or JISC Collections. The choice will in the end depend on the overall scope of the initiative: it can either result in national or international usage statistics.

## 8 Future development of usage statistics services

In the UK, the next step towards comparable item-level usage statistics will be taken by the recently-initiated PIRUS project. Its aim is to formulate a COUNTER-compliant standard for publishers and repositories for the measurement of the usage of journal articles. The standard will also be designed to be applicable in a repository context. PIRUS has been funded by JISC from September to December 2008; it marks a joint effort between publishers and repository representatives.

In the German context, the results of the current report will primarily be taken further by the DFG-funded project OA-Statistics. It will develop a pilot version for a statistics service for the project partners' repositories; it focuses on issues about infrastructure and not on metrics. Important tasks, like the de-duplication of articles, will be done centrally by the OA-Network. Those two projects co-operate strongly and plan to develop a common user interface for searching repository content and for delivering usage statistics

Within the Knowledge Exchange (Knowledge Exchange, 2007), the national organizations DEFF, DFG, JISC, and SURF are already sharing their views on usage statistics; this should be continued. Even more important is the uptake of the national projects' results in the DRIVER context. OA-Network is the German contribution to the DRIVER project. Its modules – and therefore also OA-Statistics – are designed to be compatible with the DRIVER platform. The adaptation of the OA-Statistics guidelines within the DRIVER guidelines might be possible; respective contacts have already been established. Co-operation between the national projects would also guarantee broad support within the DRIVER community.

Beyond technical and organizational implementation, it is important to provide repositories and authors with information and support in dealing with usage statistics; this will increase the acceptance of usage statistics as well as compliance with the standards. So far, the OpenDOAR policy tool has helped repositories to formulate policies on metadata, data, content, submission, and preservation. This tool could be extended to provide possible policies for usage statistics. OpenDOAR is an authoritative directory of academic Open Access repositories. It could also in the future give an overview of the different usage statistics policies in repositories. The users could easily find out whether the repository publishes usage statistics and what their granularity or format is. In addition, the SHERPA/RoMEO service listing the Open Access Policies of publishers could be extended to display information about which statistical data is available from publishers under which conditions.

## 9 References


ANSI/NISO Z39.93-2007 (2007), *The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol,* www.niso.org/workrooms/sushi (accessed 2 October 2008)

ANSI/NISO Z39.88-2004 (2004), *The OpenURL framework for context-sensitive services,* www.niso.org/standards/resources/Z39_88_2004.pdf (accessed 2 October 2008)

Benchmark Statistics Project (BEST) (2007), *Guiding Principles: Version 4.* Australian Partnership for Sustainable Repositories (APSR), http://www.apsr.edu.au/best/best_gp_4.pdf (accessed 2 October 2008)

Berlin Principles on Ranking (2006), *Berlin Principles on Ranking of Higher Education Institutions*, http://www.che.de/downloads/Berlin_Principles_IREG_534.pdf (accessed 2 October 2008)

Bollen, J., Rodriguez, M.A. and Van de Sompel, H. (2007), „MESUR: usage-based metrics of scholarly impact", *JCDL '07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, p. 474, http://www.mesur.org/Documentation_files/JCDL07_bollen.pdf (accessed 2 October 2008)

Bollen, J. and Van de Sompel, H. (2006), „An architecture for the aggregation and analysis of scholarly usage data", *JCDL '06: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries,* Chapel Hill, NC, USA, pp. 298-307.

COUNTER – Counting Online Usage of NeTworked Electronic Resources (2008), *The COUNTER code of practice: Journals and Databases – Release 3,* http://www.projectcounter.org/r3/Release3D9.pdf (accessed 2 October 2008)

Eijndhoven, K. and van der Graaf, M. (2007), *Inventory study into the present type and level of OAI compliant Digital Repository activities in the EU,* http://www.driver-support.eu/documents/DRIVER%20Inventory%20study%202007.pdf (accessed 2 October 2008)

IFABC WWW Standards (2001), http://www.ifabc.org/standards.htm (accessed 2 October 2008)

IVW (2008), *Anlage 1 zu den IVW-Richtlinien für Online-Angebote: Definitionen und technische Erläuterungen, Version 2.1,* http://daten.ivw.eu/download/pdf/Online_RichtlinienV2_1_Anlage1.pdf (accessed 2 October 2008)

Knowledge Exchange (2007), *Institutional repositories workshop strand report, strand title: Usage statistics,* http://www.knowledge-exchange.info/Default.aspx?ID=164 (accessed 2 October 2008)

PIRUS (2008), *Publisher and Institutional Repository Usage Statistics,* http://www.projectcounter.org/PIRUS_Project_plan.doc (accessed 2 October 2008)

Shepherd, P.T. (2007), *Final report on the investigation into the feasibility of developing and implementing journal usage factors,* http://www.uksg.org/usagefactors/final (accessed 2 October 2008)