

Open Access Statistics

From a Project to a Service

Workshop “Usage Statistics and Beyond”

22-23 April 2013, Berlin

Daniel Beucke | beucke@sub.uni-goettingen.de



Open Access Statistics

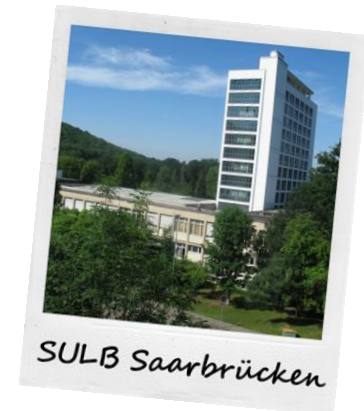


OAS 1

- May 2008 – December 2010

OAS 2

- April 2011 – July 2013



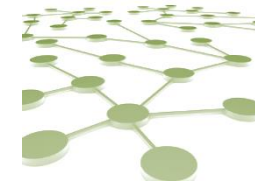
DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

Objectives

Internationally **standardised usage statistics**



Sustainable **infrastructure** for collecting and evaluating usage statistics



Increase of the **acceptance** of Open Access among authors and users of scholarly publications

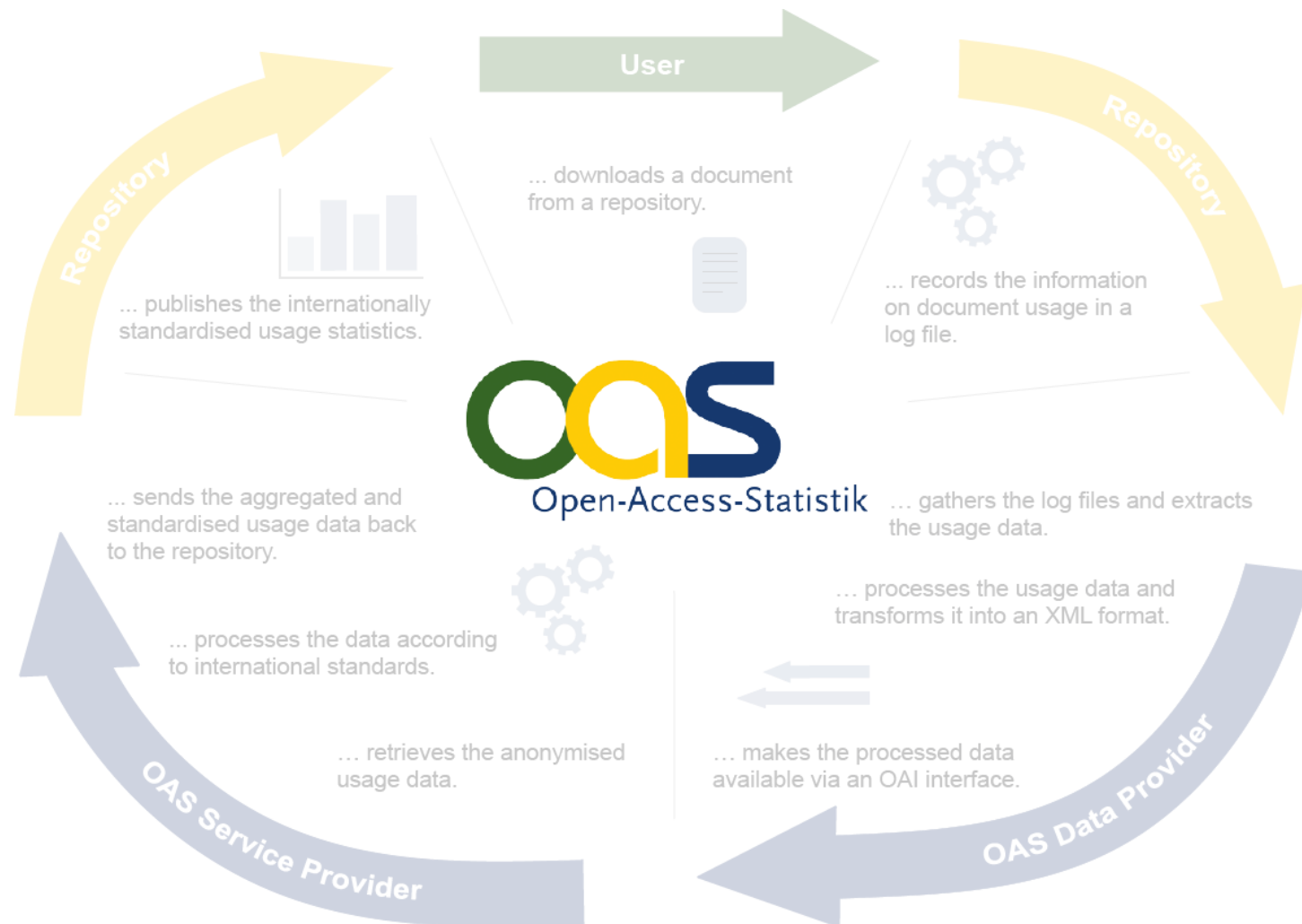


Standardised Usage Statistics

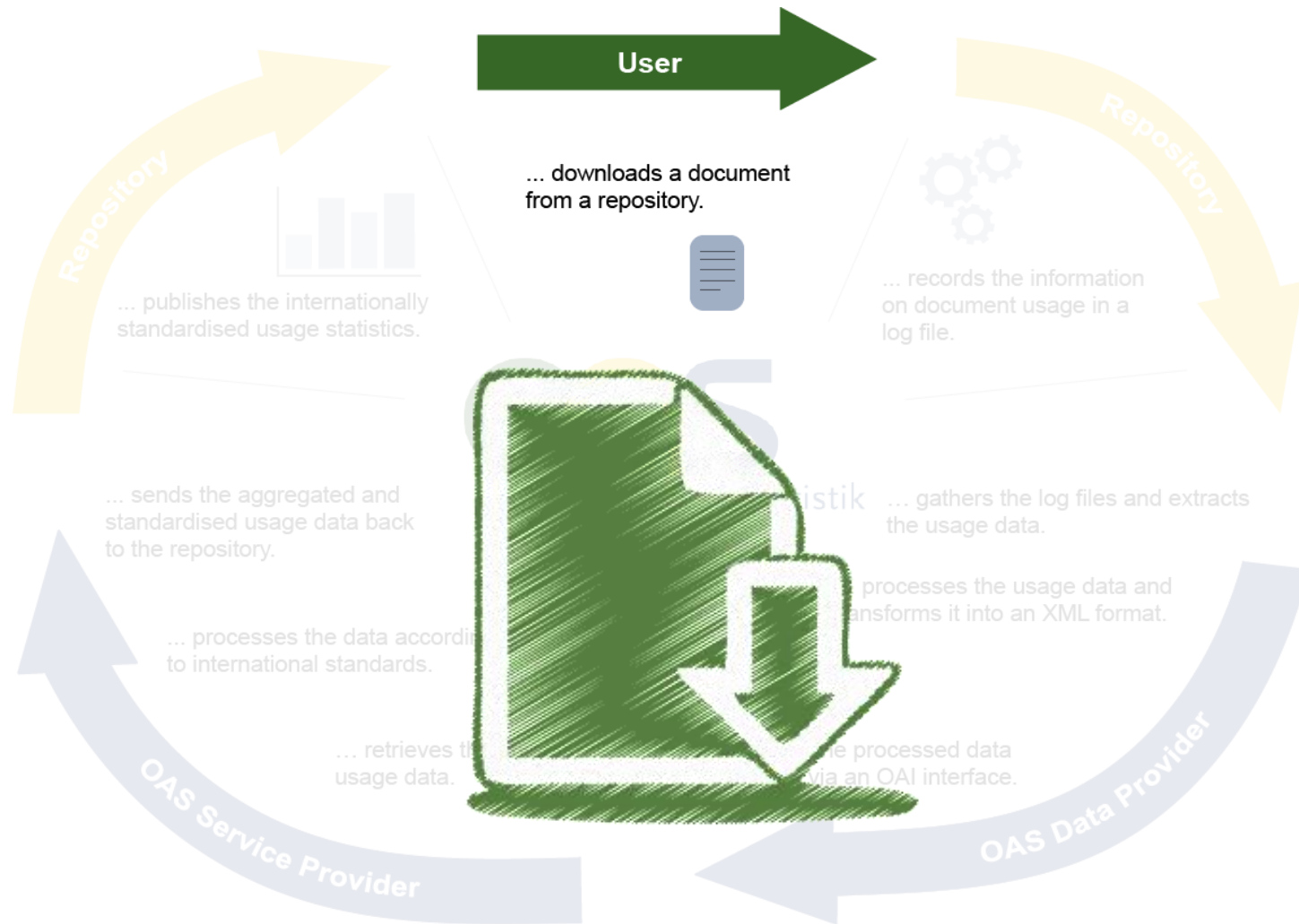
- Uniform **standards** for the gathering, exchange, and analysis of usage data
 - OpenURL ContextObjects
 - OAI-PMH
 - COUNTER

- **Tools**
 - Guidelines (OAS, KE)
 - Robot list

Infrastructure



Infrastructure



Infrastructure

Combined Log Format

LogFormat "%h %l %u %t \"%r\" %>s %b
 \"%{{Referrer}}i\" \"%{{User-agent}}i\""

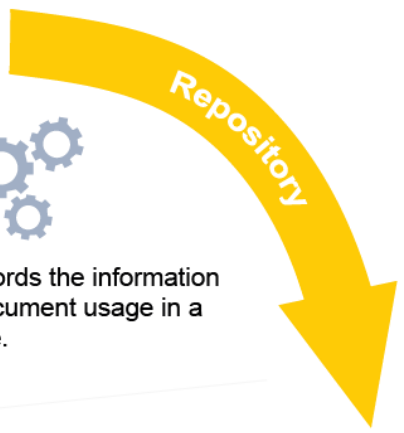
publishes the internationally
 combined standardised usage statistics.



... downloads a document
 from a repository.



... records the information
 on document usage in a
 log file.



... sends the aggregated and

... gathers the log files and extracts

```
176.9.38.xxx - [15/May/2012:00:00:38 +0100] "GET /goescholar/themes/goescholar/lib/style.css HTTP/1.1" 200 26069 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:38 +0100] "GET /goescholar/themes/goescholar/images/ginkgo.gif HTTP/1.1" 200 18746 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:38 +0100] "GET /goescholar/themes/goescholar/lib/print.css HTTP/1.1" 200 8755 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:38 +0100] "GET /goescholar/themes/goescholar/lib/checkbox.css HTTP/1.1" 200 4012 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:38 +0100] "GET /goescholar/themes/goescholar/lib/jquery.autocomplete.css HTTP/1.1" 200 797 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:38 +0100] "GET /goescholar/themes/goescholar/js/jquery.js HTTP/1.1" 200 63848 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:42 +0100] "GET /goescholar/static/images/openaire-compliance.png HTTP/1.1" 200 6393 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:42 +0100] "GET /goescholar/static/images/draussen.gif HTTP/1.1" 200 79 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:42 +0100] "GET /goescholar/themes/goescholar/images/rss2.gif HTTP/1.1" 200 187 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:42 +0100] "GET /goescholar/themes/goescholar/images/rss2.gif HTTP/1.1" 200 202 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
127.0.0.xxx - [15/May/2012:00:00:44 +0100] "POST /solr/statistics/update?wt=json&version=2.2 HTTP/1.1" 200 40 "-" solr[org.apache.solr.client.solrj.impl.CommonsHttpSolrServer] 1.0"
127.0.0.xxx - [15/May/2012:00:00:44 +0100] "POST /solr/statistics/update HTTP/1.1" 200 40 "-" solr[org.apache.solr.client.solrj.impl.CommonsHttpSolrServer] 1.0"
176.9.38.xxx - [15/May/2012:00:00:44 +0100] "GET /goescholar/themes/goescholar/images/uni_header_transp.gif HTTP/1.1" 200 7281 "http://goedoc.uni-goettingen.de/goescholar/themes/goescholar/lib/style.css" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
176.9.38.xxx - [15/May/2012:00:00:44 +0100] "GET /goescholar/static/images/dm1-zertifikat-klein.jpg HTTP/1.1" 200 25186 "http://goedoc.uni-goettingen.de/goescholar;jsessionid=1d2f80370e829929f117535b274040f0" Mozilla/5.0 (X11; Linux i686; rv:7.0) Gecko/20100101 Firefox/7.0"
95.108.240.xxx - [15/May/2012:00:00:44 +0100] "GET /goescholar/handle/goescholar/4145 HTTP/1.1" 200 21054 "-" Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)"
66.249.72.xxx - [15/May/2012:00:00:44 +0100] "GET /goescholar/handle/goescholar/37/browse?type=author&value=Luther&C=stefan HTTP/1.1" 200 21842 "-" Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
134.76.163.xxx - [15/May/2012:00:00:46 +0100] "GET /jsapi/resolvexml/webdoc-2245 HTTP/1.1" 200 377 "-" Jakarta Commons-HttpClient/2.0.2"
134.76.10.xxx - [15/May/2012:00:00:53 +0100] "GET /goescholar HTTP/1.0" 200 15423 "-" "Whatsup Professional/1.0"
66.249.72.xxx - [15/May/2012:00:00:57 +0100] "GET /goescholar/handle/goescholar/1785/browse?rpp=20&offset=0&sort_by=2&etal=1&type=dateissued&order=DESC HTTP/1.1" 200 89104 "-" Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
66.249.72.xxx - [15/May/2012:00:00:57 +0100] "GET /goescholar/handle/goescholar/2922/submit HTTP/1.1" 302 "-" Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
123.126.50.xxx - [15/May/2012:00:01:06 +0100] "GET /goescholar/handle/1/6045/browse?value=rosch&C=k&Z=80&N=90&type=author HTTP/1.1" 200 23994 "-" "Sogou web spider 4.0(+http://www.sogou.com/docs/help/webmasters.htm#07)"
134.76.21.xxx - [15/May/2012:00:01:06 +0100] "GET /jsapi/resolvexml/gr_e11b HTTP/1.1" 200 373 "-" Jakarta Commons-HttpClient/2.0.2"
66.249.72.xxx - [15/May/2012:00:01:10 +0100] "GET /goescholar/handle/goescholar/37/browse?type=author&value=Reilerts&C=Armin HTTP/1.1" 200 24422 "-" Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
180.76.5.xxx - [15/May/2012:00:01:23 +0100] "GET /goescholar/browse?type=author&value=woiters&C=DA HTTP/1.1" 200 18955 "-" Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)"
66.249.72.xxx - [15/May/2012:00:01:23 +0100] "GET /goescholar/handle/goescholar/39/browse?type=author&value=Sarker&C=MD HTTP/1.1" 200 21904 "-" Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"
```

... removes the anonymised
 usage data.

... makes the processed data
 available via an OAI interface.



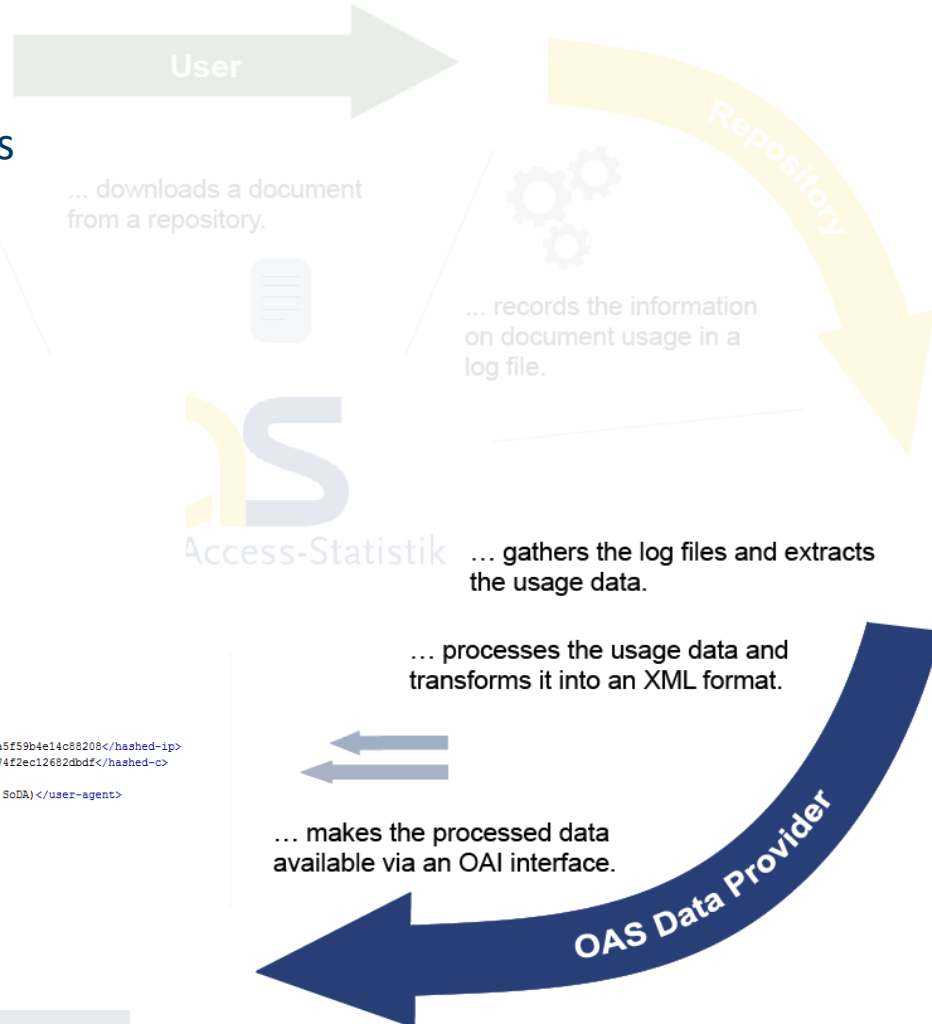
Infrastructure

- Filter pdf and html views
- Link with local identifier
- Salt and hash IP

... publishes the internationally standardised usage statistics.

```

1 <context-object timestamp="2013-03-22T08:46:46Z" >
2   <administration>
3     <oa-statistics>
4       <status_code>200</status_code>
5       <size>35214</size>
6       <document_size>0</document_size>
7       <format>application/pdf</format>
8       <service>http://www.gbv.de</service>
9     </oa-statistics>
10  </administration>
11  <referent>
12    <identifier>http://www.gbv.de/dms/zbw/588981982.pdf</identifier>
13    <identifier>oa1:gbv.de:588981982</identifier>
14  </referent>
15  <referring-entity>
16    <identifier>http://www.gbv.de/dms/zbw</identifier>
17  </referring-entity>
18  <requester>
19    <metadata-by-val>
20      <format>http://dini.de/namespace/oas-requesterinfo</format>
21    </metadata-by-val>
22    <metadata>
23      <requesterinfo>
24        <hashed-ip>b16a0fdb2e6b67d28e6bc35ed717603a5865f16157996abda5f59b4e14c88208</hashed-ip>
25        <hashed-c>bd152ff33cc3fa7f9db802ae98f99c1e5aeba23ed2c0d066a74f2ec12682dbdf</hashed-c>
26        <hostname>com.cn</hostname>
27        <user-agent>Mozilla/4.0 (compatible: MSIE 5.00; Windows 98; SoDA)</user-agent>
28      </requesterinfo>
29    </metadata>
30  </requester>
31  <service-type>
32    <metadata-by-val>
33      <format>info:ofi/fmt:xml:xsd:sch_svc</format>
34    </metadata-by-val>
35    <metadata>
36      <fulltext>yes</fulltext>
37    </metadata>
38  </service-type>
39 </context-object>
  
```



Infrastructure

Data formats of aggregated usage data

- JSON
- XML
- CSV
- iFrame

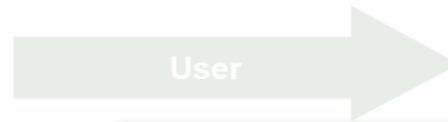
... publishes the internationally standardised usage statistics.

... sends the aggregated and standardised usage data back to the repository.

... processes the data according to international standards.

... retrieves the anonymised usage data.

OAS Service Provider

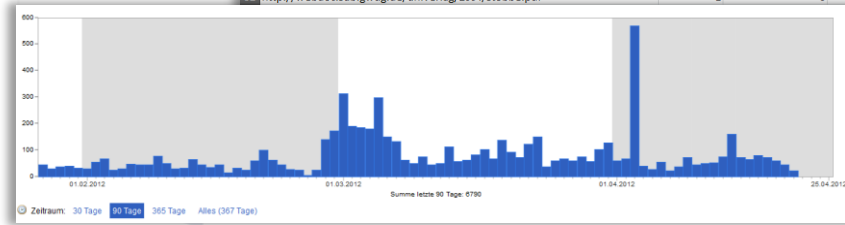


```

{
  "from": "2012-03-01",
  "to": "2012-03-07",
  "entityDef": ["identifier", "date", "counter", "LogEc", "IFABC", "RobotsCounter", "RobotsLogEc", "RobotsIFABC"],
  "entities": [
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-01", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-02", "counter": 1, "LogEc": 0, "IFABC": 1, "RobotsCounter": 1, "RobotsLogEc": 1, "RobotsIFABC": 1},
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-03", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-04", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-05", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-06", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0},
    {"identifier": "oai:econstor.eu:10419/20681", "date": "2012-03-07", "counter": 0, "LogEc": 0, "IFABC": 0, "RobotsCounter": 0, "RobotsLogEc": 0, "RobotsIFABC": 0}
  ]
}
    
```



	B	C	D	E
1	identifier	counter	counter_abstract	robots
2	http://webdoc.sub.gwdg.de/univerlag/2003/garmann/pages_1-144.pdf	2	0	0
3	http://webdoc.sub.gwdg.de/univerlag/2003/garmann/pages_1-144.pdf	1	0	0
4	http://webdoc.sub.gwdg.de/univerlag/2003/garmann/pages_1-144.pdf	1	0	0
5	http://webdoc.sub.gwdg.de/univerlag/2003/garmann/pages_144-288.pdf	2	0	0
6	http://webdoc.sub.gwdg.de/univerlag/2003/garmann/pages_144-288.pdf	1	0	1
7	http://webdoc.sub.gwdg.de/univerlag/2003/garmann/pages_144-288.pdf	1	0	0
8	http://webdoc.sub.gwdg.de/univerlag/2004/GeoBio.pdf	3	0	0
9	http://webdoc.sub.gwdg.de/univerlag/2004/GeoBio.pdf	5	0	1
10	http://webdoc.sub.gwdg.de/univerlag/2004/GeoBio.pdf	5	0	0
11	http://webdoc.sub.gwdg.de/univerlag/2004/GeoBio.pdf	3	0	0
12	http://webdoc.sub.gwdg.de/univerlag/2004/GeoBio.pdf	1	0	0
13	http://webdoc.sub.gwdg.de/univerlag/2004/GeoBio.pdf	1	0	0
14	http://webdoc.sub.gwdg.de/univerlag/2004/kersten_MathBioGeo3.pdf	3	0	0
15	http://webdoc.sub.gwdg.de/univerlag/2004/kersten_MathBioGeo3.pdf	5	0	0
16	http://webdoc.sub.gwdg.de/univerlag/2004/kersten_MathBioGeo3.pdf	1	0	0
17	http://webdoc.sub.gwdg.de/univerlag/2004/kersten_MathBioGeo3.pdf	1	0	0
18	http://webdoc.sub.gwdg.de/univerlag/2004/adnan.pdf	1	0	0
19	http://webdoc.sub.gwdg.de/univerlag/2004/dbd.pdf	1	0	0
20	http://webdoc.sub.gwdg.de/univerlag/2004/geisler.pdf	1	0	0
21	http://webdoc.sub.gwdg.de/univerlag/2004/jacobs.pdf	2	0	0
22	http://webdoc.sub.gwdg.de/univerlag/2004/jacobs.pdf	3	0	0
23	http://webdoc.sub.gwdg.de/univerlag/2004/jacobs.pdf	1	0	0
24	http://webdoc.sub.gwdg.de/univerlag/2004/schaaf.pdf	1	0	0
25	http://webdoc.sub.gwdg.de/univerlag/2004/schaaf.pdf	1	0	0
26	http://webdoc.sub.gwdg.de/univerlag/2004/schaaf.pdf	2	0	0
27	http://webdoc.sub.gwdg.de/univerlag/2004/schaaf.pdf	1	0	0
28	http://webdoc.sub.gwdg.de/univerlag/2004/stachelhaeuter.pdf	1	0	0
29	http://webdoc.sub.gwdg.de/univerlag/2004/stachelhaeuter.pdf	2	0	1
30	http://webdoc.sub.gwdg.de/univerlag/2004/stachelhaeuter.pdf	1	0	1
31	http://webdoc.sub.gwdg.de/univerlag/2004/stachelhaeuter.pdf	1	0	0
32	http://webdoc.sub.gwdg.de/univerlag/2004/stobbe.pdf	1	0	0



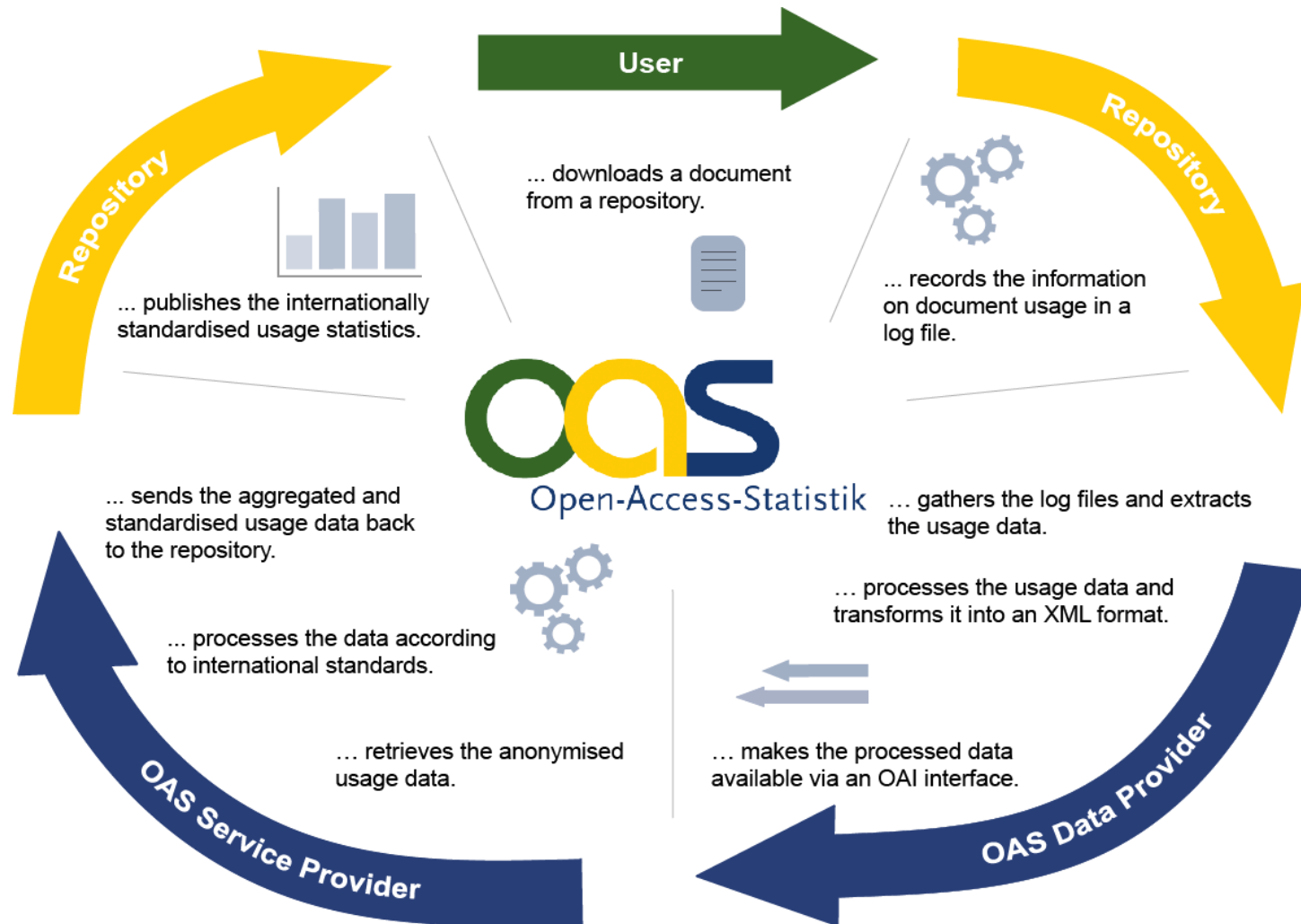
Infrastructure



The image shows three overlapping screenshots of web portals:

- ECONSTOR:** A search results page for a document titled "Deskriptive Statistik: Formen, Aufgaben, Klausurübung" by Peter von der Lippe. It shows metadata like authors, issue date (2006), and ISBN (978-3-486-57963-8). It also displays usage statistics, including "No. of Downloads: 802" and "last Month: 3166".
- FACHPORTALpaedagogik.OE:** A page titled "Download-Statistik für Dokument 7678" showing a table of download statistics. The table has columns for "Datei", "Downloads", "im Monat", and "im Jahr". The data shows 7.3 downloads in April 2013 and 52.4 in the year 2013.
- Georg-August-Universität Göttingen:** A page showing a bar chart titled "Alle Besuche pro Monat" for the document "The Nature and Persistence of Publications in Human Musical Rhythms". The chart shows monthly visit counts from October 2012 to April 2013. Below the chart is a table of "Top Lander" with countries like Russia, United States, Germany, and others.

Infrastructure



Acceptance of Open Access

Usage data as a value-added service

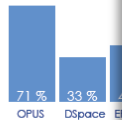
24%

Usage Statistics⁶

The bigger a repository, the more likely it offers usage statistics.

56%

Bibliographic Exports



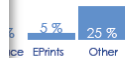
The smaller a repository, the more likely it offers bibliographic exports.

Paul Vierker, <http://nbn-resolving.org/urn:nbn:de:hbz:5:1-30000-p0011-8>

- 5) Bibliographic Exports
- 6) Usage Statistics
- 7) Checksums
- 8) A functional RSS Feed
- 9) Social Bookmarking
- 10) Social Network Functions

24%

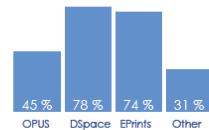
Checksum⁷



The smaller a repository, the more likely it offers a checksum.

48%

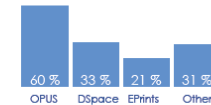
RSS Feed⁸



There is no relation between the size of a repository and its support of RSS.

45%

Social Bookmarking⁹



The bigger a repository, the more likely it does not support social bookmarking.

11%

Social Networks¹⁰



The bigger a repository, the more likely it has integrated social network functions.

David, Mathias Lösch: *2012 Census of Open Access Repositories in Germany*, 11

- 1) is available on item or collection level.
- 2) is available for unregistered users on item level.
- 3) is available on item level.
- 4) is available on item level or browsing page.
- 5) is available on item level.
- 6) is available on item level.
- 7) is available on item level.
- 8) is available on item level.
- 9) is available on item level.
- 10) is available on item level.

Lessons Learned

Lessons Learned I

Implementation of infrastructure is expensive

- amount of data
- heavy traffic
- permanent maintenance



Data privacy act

- our assumption: IP address is personal data
- strict data privacy
- salt and hash



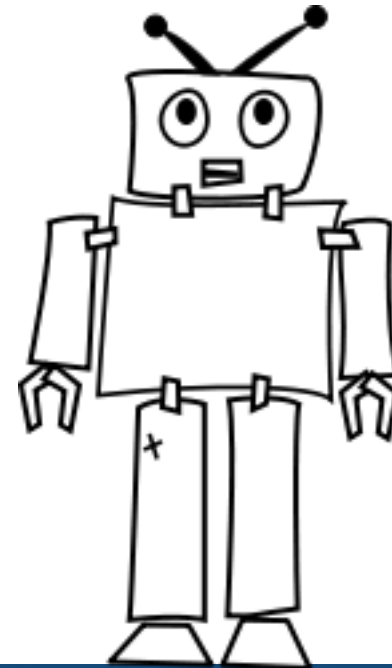
Lessons Learned II

Standards for analysis

- COUNTER \neq COUNTER – depending on the robot list for example

Robots

- 30% - 50% of all users are robots



Outcomes

Infrastructure

OAS Data Provider

- [Demo OAS Data Provider](#)
- [Software](#)



OAS Service Provider

Involved partners

- Digizeitschriften
- Econstor
- EDOC HU Berlin
- OJS - CeDis Berlin
- Pedocs
- Psydoc SULB Saarbrücken
- PUB Bielefeld
- Qucosa
- GoeScholar
- Universitätsverlag Göttingen
- VZG



Guidelines

Guideline for participation in OA Statistics

Specification: Data Format and Exchange for OA Statistics

Cooperation with KE Usage Statistics Guidelines *Guidelines for the aggregation and exchange of Usage Data*

Recommendations in *DINI Certificate – Document and Publication Services 2010*



Sustainable Infrastructure

Feasibility study



Data privacy report



OAS Validator

Validator
zur Überprüfung der XML-Struktur des OA-Statistik-Data-Provider (OAS-DP) nach den Schemata
OALPMH ContextObjects oa-statistics requestInfo

Geben Sie die URL des OAS-DP und optional Benutzernamen und Passwort an:

Dateiname: *

Benutzername:

Passwort:

Die angegebene URL verweist auf eine existierende Datei.
Die geladene Datei ist nicht leer.
Die geladene Datei hat die richtige Dateieindung.

Validierung
Die Validierung kann einige Minuten dauern. Die Prüfung erfolgt maximal bis zum resumptionToken mit pos=100.
Klicken Sie im Fehlerfall auf den Button!

Identify
Die Ausgabe entspricht dem OAS-DP-Schema.

ListIdentifiers
Das resumptionToken ist gültig.
resumptionToken von 0-10: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 10-20: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 20-30: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 30-40: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 40-50: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 50-60: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 60-70: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 70-80: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 80-90: Die Ausgabe entspricht dem OAS-DP-Schema.
resumptionToken von 90-100: Die Ausgabe entspricht dem OAS-DP-Schema.

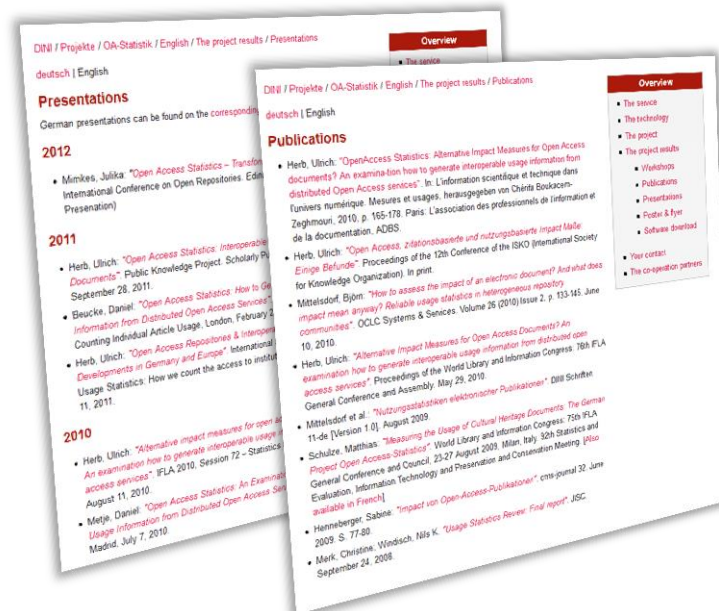
Project Publications

Project publications on project website

- [Publications](#)
- [Presentations](#)
- [Poster](#)
- [Workshops](#)

In summer 2013

- *Usage Statistics and Beyond*
Version 2.0
also available in English



From a Project to a Service



Repository manager indicates interest, service will be offered by the VZG from summer 2013



... Marco Recke

Thank you for your attention. Questions?

E-Mail: beucke@sub.uni-goettingen.de

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

License:

