Usage data: Workshop Objectives from the perspective of the DINI - DFG - JISC projects

Frank Scholze
Nils Windisch

JISC Usage Statistics Workshop
Humboldt University Berlin,
Erwin Schrödinger Center, 7.-8. July 2008
Cluster of projects

- open-access.net
- OA Repository Network
- OA Statistics
- OA Citations
- DRIVER
- JISC usage statistics statistics review!
DINI certificate on usage statistics

• Minimum standards
  – Every document repository must keep consistent access statistics (web server log files)
  – Web server log files must be anonymized for long-term archiving purposes
  – Criteria used to collect or filter the data must be documented

• Recommendations
  – Access to documents by automated agents, robots or similar is filtered out (and documented)
  – Web server log files are edited according to the Counter Code of Practice
  – Access statistic is attached to the document as dynamic metadata and visible to the end-user
What do we count?

- Practical definition of meaningful items
  - Files?
  - Publications (journal articles etc.)?

- Identification of meaningful items
  - Checksums
  - Persistent identifiers
  - Distributed heterogeneous publication network
How do we count?

- Practical and pragmatic definition of usage
  - Access
  - Click spans
  - Definition of non-human access
  - Pseudonymization
  - Deleting or tagging
  - Sessions
How do we aggregate?

- Technically
  - processing on which level
  - OpenUrl ContextObjects
  - SUSHI

- Organisationally
  - DRIVER
  - OA Repository Network
  - ...

- Co-operation
What do we report?

- Access over time
- Sources of aggregation
- Standards for processing

➔ Transparency on the what and the how of counting

Johan will tell us more
Infrastructure for aggregating usage data

- Log Repository
- OpenURL ContextObjects
- Log DB
- Data Mining
- Filtering
- Metrics
- Services
- Webserver - Log
- Rewrite module
- Normalise (optional) -> Robots, pseudonymization

Based on: Bollen and Van de Sompel, OAI4, Geneva
Open Access Statistics

realize what others had in mind…
Project

• Funded by: DFG (German Research Foundation)

• 18 months 2008-07-01 – 2009-12-31

• Partner: Berlin (CMS) Göttingen (SUB), Saarbrücken (SUUB), Stuttgart (UB)
Context

• One of three projects:
  • Open Access Network of Repositories (OA-N)
  • Open Access Citation (DOARC, Distributed Open-Access Reference Citation services)
  • Open Access Statistics (OA-S)
Objectives

• Aggregate and normalize usage data locally
• Act as data provider
• Collect data at service provider level
• Process data to provide added values services
Protein Kinase D regulates several aspects
Maier, Dieter; Nagel, Anja C.; Gloc, Helene

Originalveröffentlichung: (2007) BMC Developmental Biology
http://www.biomedcentral.com/1471-213X/7/57

Statistiken:
Gedruckte Ausgabe:
Dokument empfehlen:
Dokument in Google Scholar suchen:
SWD-Schlagwörter:
Institut:
DDC-Sachgruppe:

URN: urn:nbn:de:bsz:93-opus-32590
URL: http://elib.uni-stuttgart.de/opus/volltexte/2007/3259
Deal with usage date

• Different sources
  • License server (HAN-Server)
  • Link resolver (SFX)
  • Repository software (DSpace, OPUS, e-doc, etc.)
Repository software (DSpace)

JISC Usage Statistics Workshop, Berlin 2008-07-07/08, Nils K. Windisch (windisch@sub.uni-goettingen.de)
Link resolver (SFX)
Map usage data

• Use established formats and rule sets
  • COUNTER
  • IFABC
  • LogEc
from A to B

• SUSHI vs. OAI-PMH + OpenURL CO
  • *project objectives include evaluation of each approach*
• SUSHI/OAI-OMH just a transport container/vehicle
• *(Usage) data as XML payload*
Context

• What about the JISC Usage Statistics Project and Workshop?
• Provide inside information
• Build on experience and expert opinions
• Re-use existing technologies