Open-Access-Statistik

Open Access Repositories & Interoperable Usage Statistics: Current Developments in Germany and Europe

International Seminar on Standardization of IR Usage Statistics: How we count the access to institutional repositories

National Institute of Informatics, Tokyo
January 11, 2011

Initiated by:

DINI
DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

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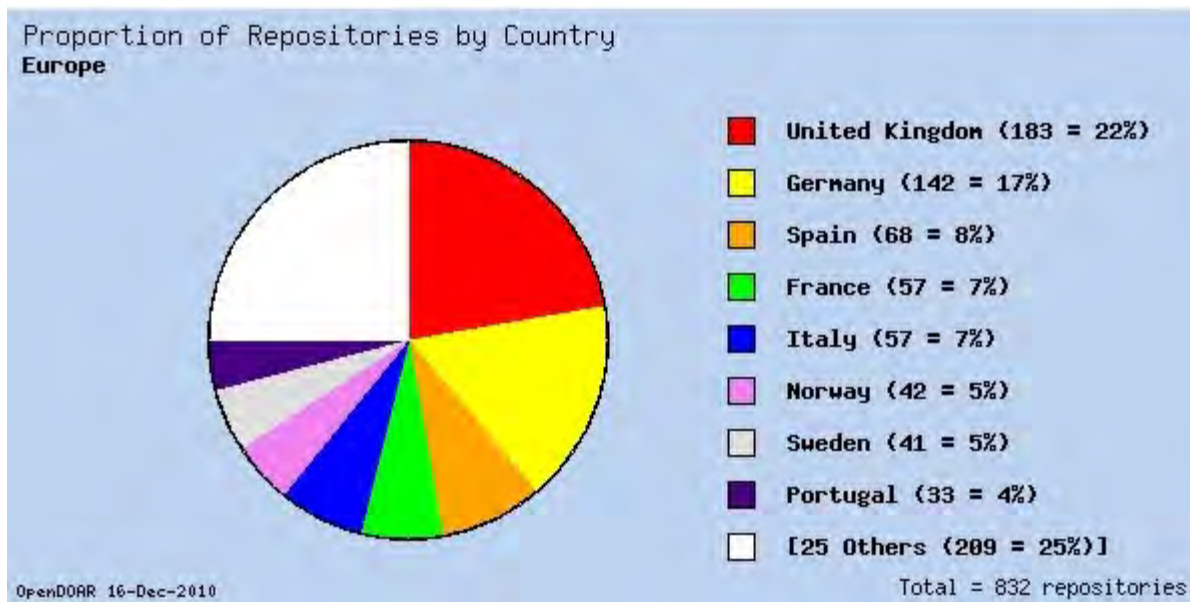u.herb@sulb.uni-Saarland.de

Funded by:

Deutsche
Forschungsgemeinschaft

DFG

- ❏ IR development in Europe and Germany

- ❏ Impact measures
  - Citation vs. Usage

- ❏ Usage Metrics: Standards?

- ❏ Open Access Statistics (OAS)
  - Aims
  - Technical infrastructure
  - Results & outlook
  - Repository usage statistics: The European perspective

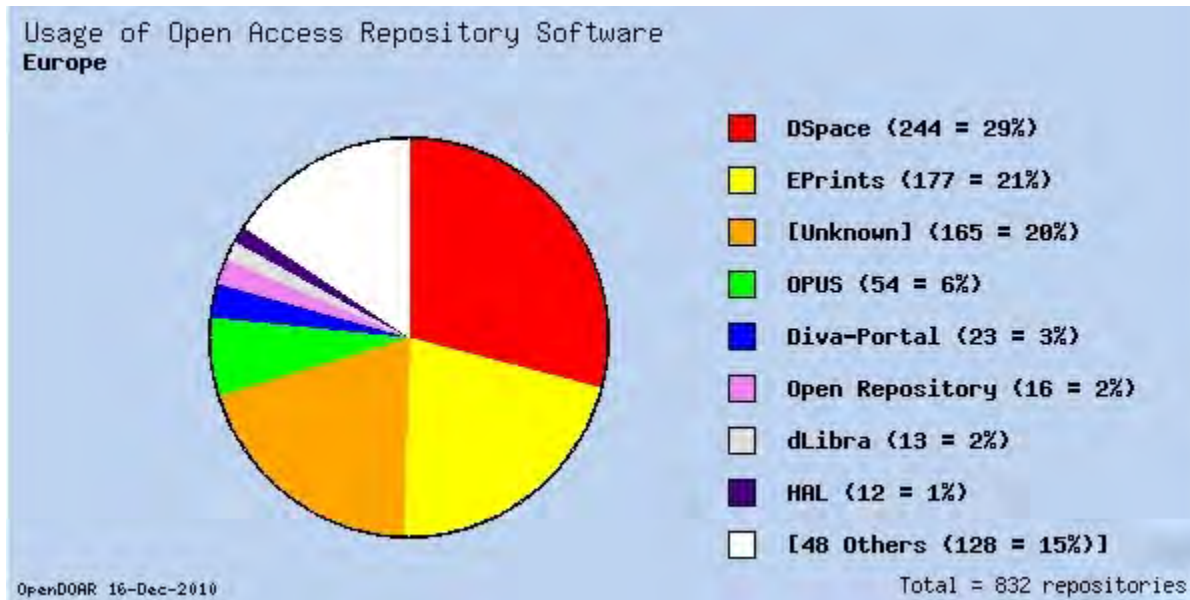**International Seminar on Standardization of IR Usage Statistics:**
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

IR development in Europe and Germany

Open-Access-Statistik

IR development in Europe and Germany

Proportion of Repositories by Country
Europe

- United Kingdom (183 = 22%)
- Germany (142 = 17%)
- Spain (68 = 8%)
- France (57 = 7%)
- Italy (57 = 7%)
- Norway (42 = 5%)
- Sweden (41 = 5%)
- Portugal (33 = 4%)
- [25 Others (209 = 25%)]

OpenDOAR 16-Dec-2010

Total = 832 repositories

taken from www.opendoar.org

**International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011**

oas
Open-Access-Statistik

IR development in Europe and Germany

Usage of Open Access Repository Software
Europe

DSpace (244 = 29%)
EPrints (177 = 21%)
[Unknown] (165 = 20%)
OPUS (54 = 6%)
Diva-Portal (23 = 3%)
Open Repository (16 = 2%)
dLibra (13 = 2%)
HAL (12 = 1%)
[48 Others (128 = 15%)]

OpenDOAR 16-Dec-2010

Total = 832 repositories

taken from www.opendoar.org

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

IR development in Europe and Germany

UK

EPprints	**44%**
DSpace	19%
proprietary tools	16%
Open Repository	5%

France

proprietary tools	**47%**
HAL	19%
EPrints	18%
DSpace	11%

Germany

OPUS	**38%**
proprietary tools	25%
EPrints	11%
DSpace	5%

Italy

EPrints	**51%**
DSpace	32%
proprietary tools	12%

Spain

DSpace	**59%**
proprietary tools	25%
DigiBib	10%
EPrints	6%

Netherlands

proprietary tools	**39%**
DSpace	22%
EPrints	9%

taken from www.opendoar.org

IR development in Europe and Germany

- Heterogeneous software landscape with some "local heroes"

- Creating incentives (metrics, scientific capital)

- European repository community takes strong efforts in interlinking and integration of repositories

  … both on the technical and the service layer (DRIVER, COAR) and at the level of funding agencies (Knowledge Exchange)

- Integration into

  … academic workflows (SWORD, SONEX)

  … academic and administrative information systems (current research information systems, project databases)

  … Social Networks (ResearchGate, Mendeley, …)

  … do repositories really need an user interface?

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

IR development in Germany

- 200+ institutional and disciplinary repositories
  - Various repository platforms operational
  - Large differences in design, size of collections, and coverage
  - Heterogenous types of content

- Enhancement of content visibility on a national and international level by various means
  - Widespread implementation of OAI-PMH, but still deficits in standardization and data harmonization
  - Prominent repository registries, repository collaborations, search engines

Registry of Open Access Repositories (ROAR)

OpenDOAR

C AR
Confederation of Open Access Repositories

Google scholar beta

BASE
Bielefeld Academic Search Engine

Medienzentren Bibliotheken **DINI** Rechenzentren Wissenschaft

driver
Digital Repository Infrastructure Vision for European Research

oan
Netzwerk von
Open-Access-Repositorien

OAIster
...find the pearls

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

IR development in Germany

- **Open-Access-Network**
  German research institutions interlink their Open Access repositories and create an overarching collection of publications through the information infrastructure of OA Network
  http://www.dini.de/projekte/oa-netzwerk/

- **Standardization and stimulation of IR development**
  DINI Certificate for document and publication services
  DINI = German Initiative for networked information
  http://www.dini.de/english/dini-certificate/

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Impact Measures

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Open-Access-Statistik

Impact measures: relevance

Individual level: *publish or perish*

> If you do not publish you do not have any scientific capital, reputation or impact

> Without any impact, you won't make your 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

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Open-Access-Statistik

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 Edu*

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
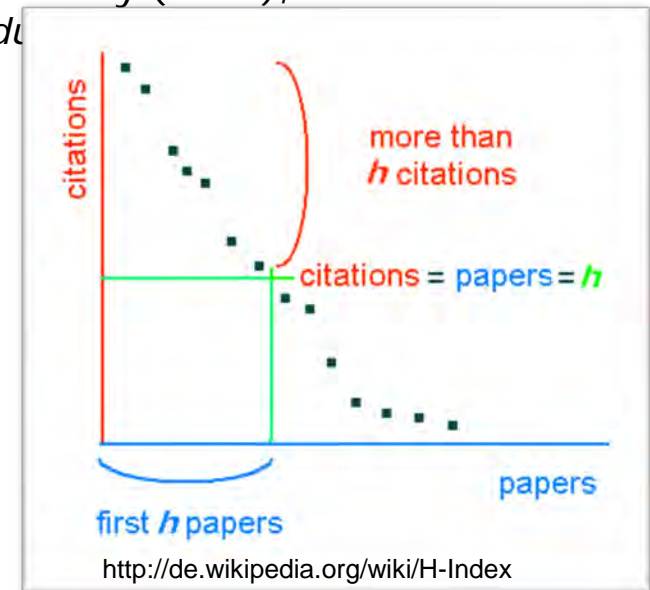
Citation impact: some critique

- ☐ Restricted scope, exclusion of many publication types
- ☐ Based exclusively on journal citation report / web of science
- ☐ 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

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Open-Access-Statistik

Impact measures: a categorization

☐ 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

Open-Access-Statistik

Impact measures: a categorisation, pt. II
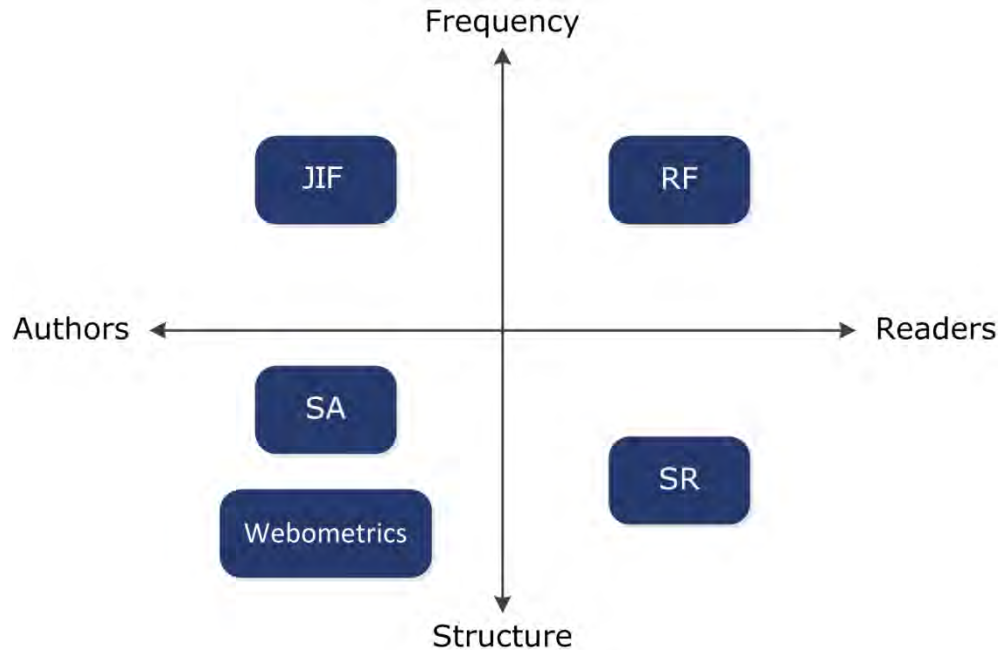
JIF = Journal Impact Factor

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

Usage Metrics: Standards?

Open-Access-Statistik

International Seminar on Standardization of IR Usage Statistics: How we count the access to institutional repositories. National Institute of Informatics, Tokyo. January 11, 2011

Usage based impact: standardisation?

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

http://www.projectcounter.org

□ *LogEc*

http://logec.repec.org/

□ IFABC

http://www.ifabc.org/

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Open-Access-Statistik

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?"

oas
Open-Access-Statistik

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

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

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Project:
Open Access Statistics

oas
Open-Access-Statistik

Open Access Statistics (OAS)

☐ 07/2008 – 02/2010
☐ Project partners
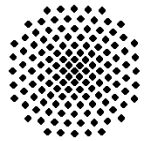
Universität Stuttgart

HUMBOLDT-UNIVERSITÄT ZU BERLIN

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

SAARLÄNDISCHE
UNIVERSITÄTS-UND
LANDESBIBLIOTHEK

Initiated by:

Funded by:

DINI
DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

DFG

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

**International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011**

oas
Open-Access-Statistik

OAS: Aims

- ☐ A common standard to exchange usage date 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

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

OAS: Associated projects

□ Open Access Statistics

□ DOARC
(Distributed Open Access Reference and Citation Services)

□ Open Access Network

**International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011**

Technical Infrastructure

Open-Access-Statistik

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

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

OAS: Data provider
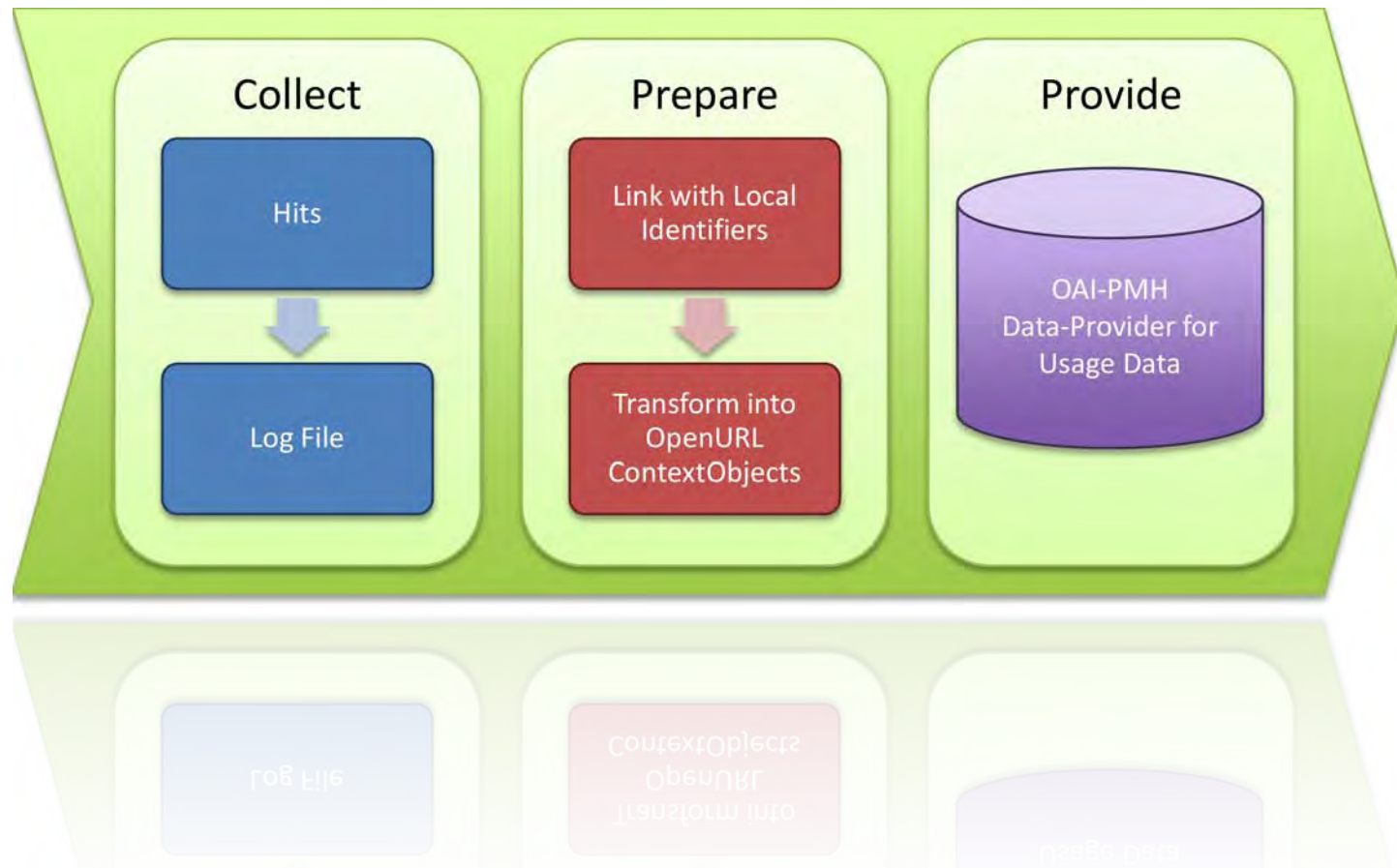
OAS: Service provider
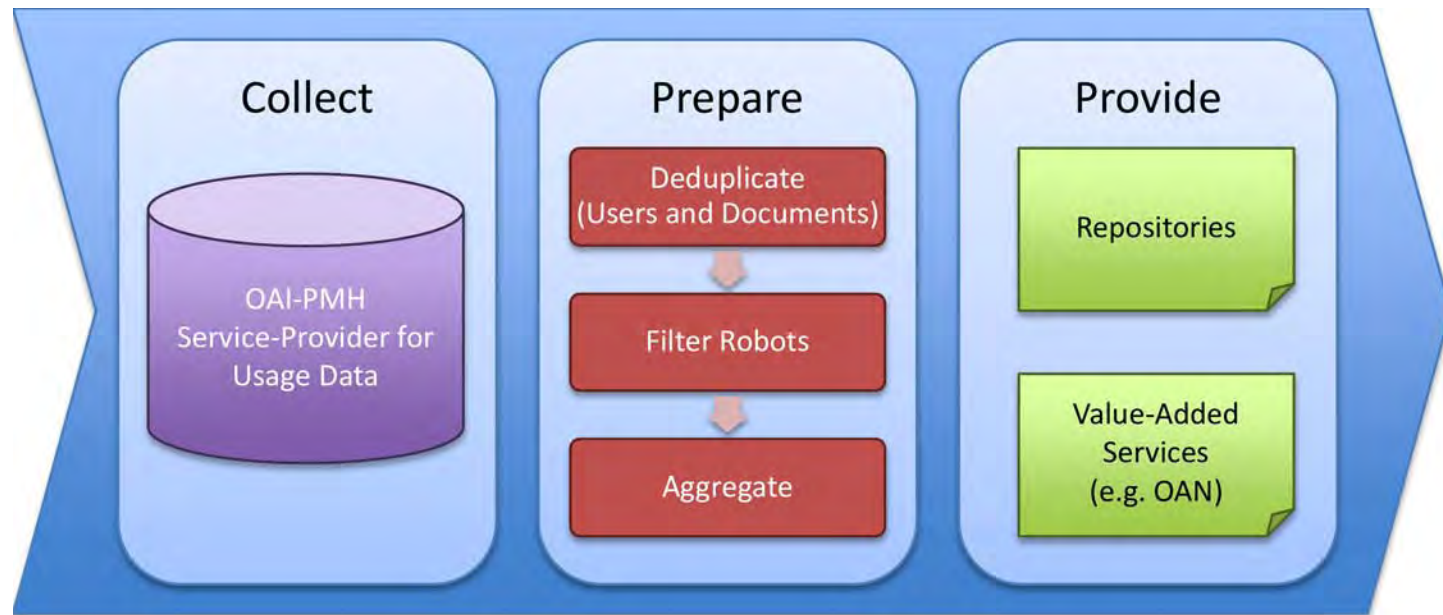
OAS: Repository integration

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

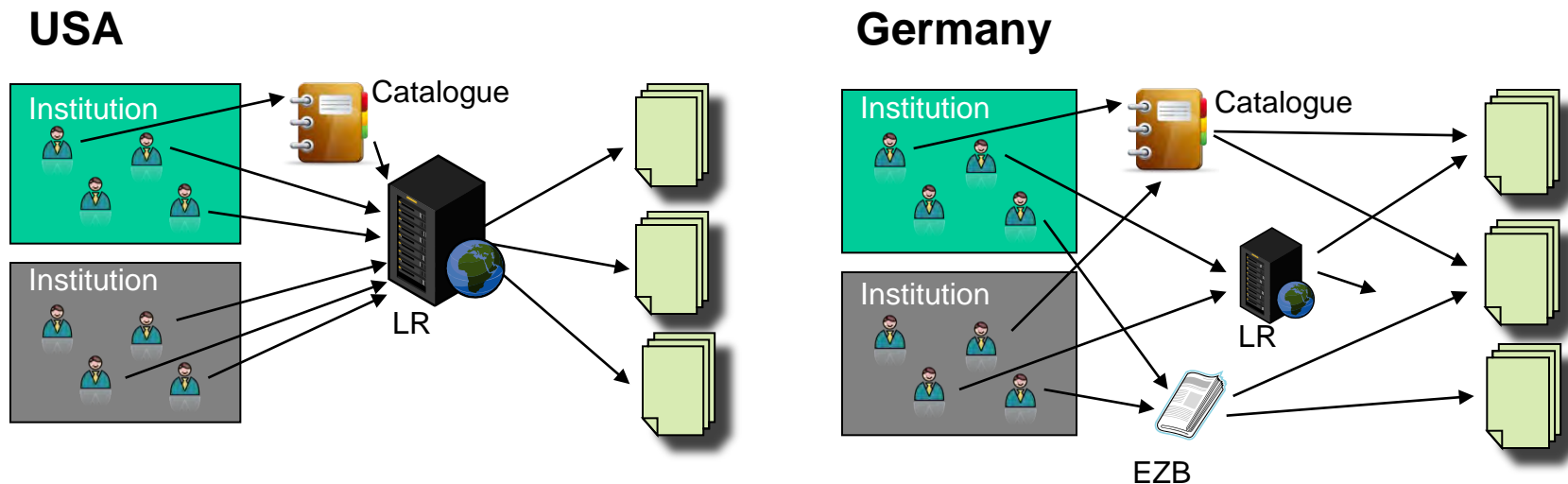# Results and Outlook

Open-Access-Statistik

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 data

USA

Institution

Institution

Catalogue

LR

Germany

Institution

Institution

Catalogue

LR

EZB

oas
Open-Access-Statistik

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
(http://www.dini.de/fileadmin/oa-statistik/projektergebnisse/Specification_V5.pdf)

Online demo
(http://oa-statistik.sub.uni-goettingen.de/statsdemo)

Website with further information
(http://www.dini.de/projekte/oa-statistik/english/)

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

OAS: Further plans → OAS 2

Aims for a possible second funding:

☐ Clarification of privacy issues

☐ Opening the OAS infrastructure to offer standardized usage statistics

☐ Evaluation of metrics
  a) based on the pure frequency of usage
  b) more sophisticated approaches

☐ Cooperation for international comparable usage statistics

☐ Offer a suitable service infrastructure

oas
Open-Access-Statistik

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

OAS: International cooperations

- ☐ SURFSure Statistics on Usage of Repositories, NL

- ☐ PIRUS Publisher and Institutional Repository Statistics, UK

- ☐ Knowledge Exchange Usage Statistics Group

  Denmark's Electronic Research Library (DEFF)
  German Research Foundation (DFG)
  Joint Information Systems Committee (JISC) UK
  SURFfoundation, Netherlands

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

- ☐ COUNTER, NEEO, PEER, OAPEN …

International Seminar on Standardization of IR Usage Statistics:
How we count the access to institutional repositories.
National Institute of Informatics, Tokyo. January 11, 2011

Thanks for your attention!

... and thanks to my colleagues Maxi Kindling, Daniel Beucke & Stefan Buddenbohm for their support!

Contact:

u.herb@sulb.uni-saarland.de

+49 681 302 2798

International Seminar on Standardization of IR Usage Statistics: How we count the access to institutional repositories. National Institute of Informatics, Tokyo. January 11, 2011