Metrics in an Open Access environment: an infrastructure for collecting and aggregating usage data

Frank Scholze
Stuttgart University Library

3rd London Conference on
Opening Access to Research Publications, 11.6.2007

Background

- Metrics as a research policy issue
- Assessment and evaluation of research
  - Appointment decisions
  - Funding decisions
  - Monitoring trends
  - Prioritize activities and attention
- Qualitative vs quantitative methods

Rankings are here to stay, and it is therefore worth the time and effort to get them right.
Alan Gilbert (President University of Manchester)

From: D Butler (2007) Academics strike back at spurious rankings, *Nature* 447, pp 514-515

Open Access and metrics

- Possibility to collect and process quantitative data on electronic publications
  - Usage
  - Citations

  } become open access as well !

- Possibility to construct new indicators to measure different aspects of research impact

- Possibility to enhance and complement existing metrics

Citations and the journal impact factor

Journal Impact Factor: mean 2-year citation rate

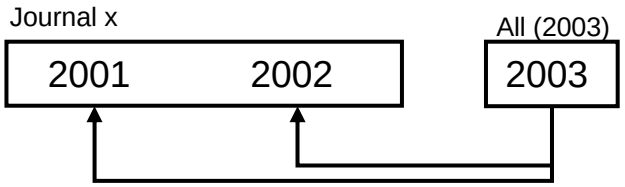
2003 citations to 2001 and 2002 articles in X

Divided by

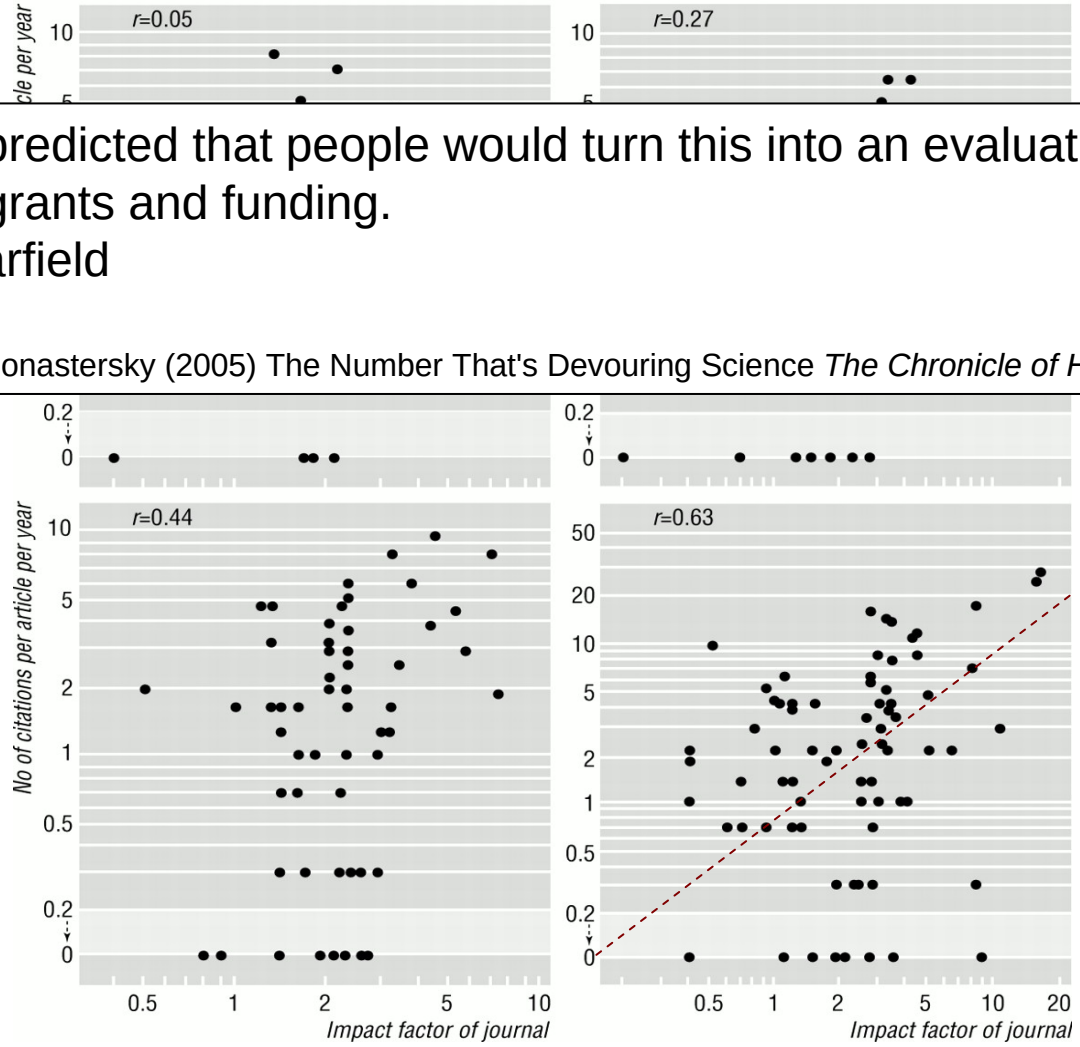
number of articles published in X in 2001 and 2002

Widely applied in research evaluation

- Fair approximation of journal “status”,…but
- Used to rank authors, departments, institutions, regions, nations, etc.
- Now common in tenure, promotion and other evaluation procedures!



Correlation between article citation rate and journal impact factor



We never predicted that people would turn this into an evaluation tool for giving out grants and funding.
Eugene Garfield

From: Richard Monastersky (2005) The Number That's Devouring Science *The Chronicle of Higher Education*

From: Seglen, P.O. (1997) Why the impact factor of journals should not be used for evaluating research *BMJ* 314

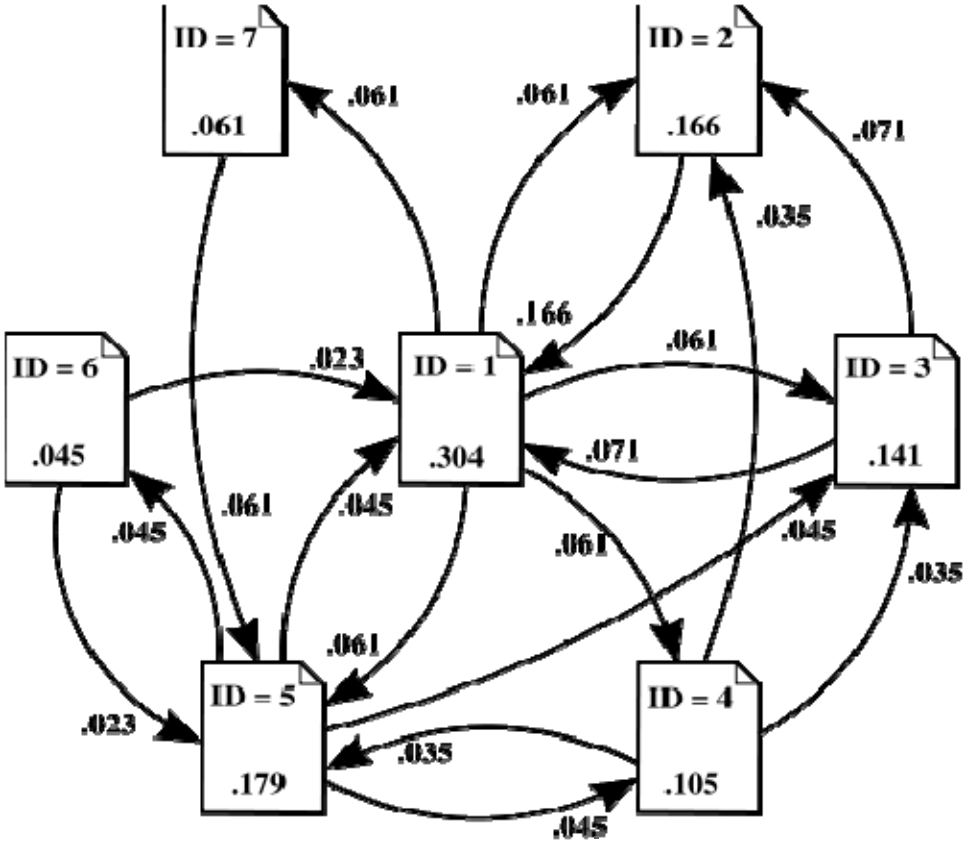
Page rank

Reciprocal voting of nodes in a network

Can be applied to any collection of entities with reciprocal references

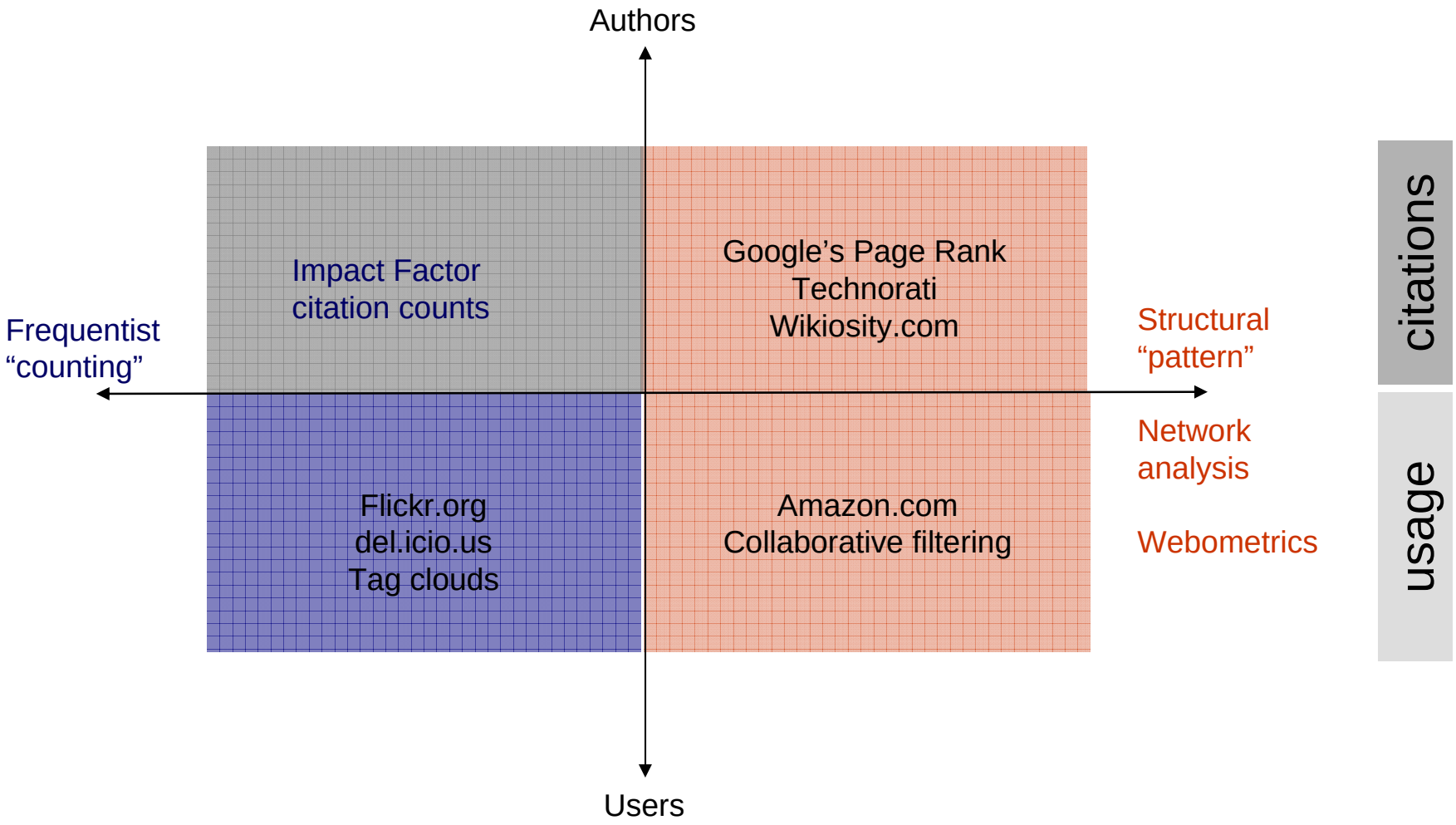
Related approaches in network analysis include authority and hub values

(HITS algorithm)



$$PR_i = \frac{1-d}{N} + d \sum_{\forall j \in \{(j,i)\}} \frac{PR_j}{C_j}$$

A taxonomy of metrics



Based on: Bollen, Johan and Van de Sompel, Herbert (2005) A framework for assessing the impact of units of scholarly communication based on OAI-PMH harvesting of usage information. *OAI4*, Geneva

Multivariate Metrics

- Citations: Impact Factor, Co-Citation, Immediacy Factor, h-index, Citation PageRank, Weighted In-degree, Weighted Out-degree, In-Degree entropy, Out-Degree entropy …

- Usage: Usage Factor, Usage Impact Factor, Usage PageRank, Weighted In-degree, Weighted Out-degree, In-Degree entropy, Out-Degree entropy …

See also: Harnad, Stevan (2007) Open Access Scientometrics and the UK Research Assessment Exercise. *11th Annual Meeting of the International Society for Scientometrics and Informetrics*, Madrid

Measuring publication impact: the elements (schematic)

Services (ranking, evaluation, recommender, collection building and management...)

analyse

Metrics (structural, frequentist)

Data mining

collect

Compare, aggregate

Link-Resolvers

Collect
Aggregate

Logfile analysis

Collect
Aggregate

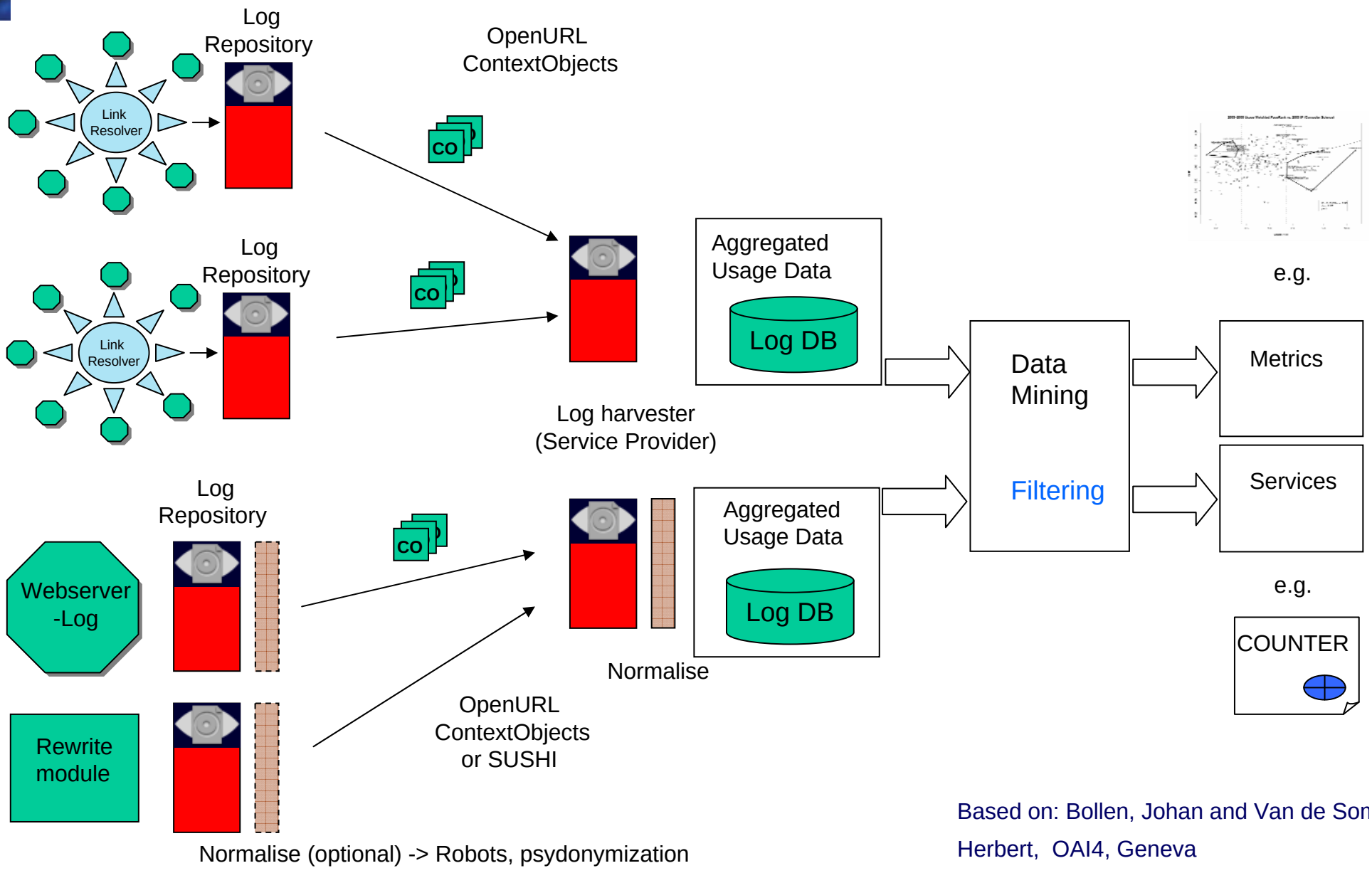
Citation analysis

Parse
Extract
Aggregate

choose

Basic set of documents (Journals, repositories, primary data etc.)

Infrastructure for collecting usage data



Based on: Bollen, Johan and Van de Son
Herbert, OAI4, Geneva

Ongoing work

- LANL
  - bX (with CalState, ExLibris)
  - MESUR
  - ...
- UK
  - University of Southampton (Citebase, IRS, EPStats …)
  - University College London
  - ...
- Germany (DINI / DFG)
  - Göttingen State and University Library
  - Stuttgart University Library
  - Computer and Media Service Humboldt University Berlin
  - Saarbrücken State and University Library

Examples: MESUR

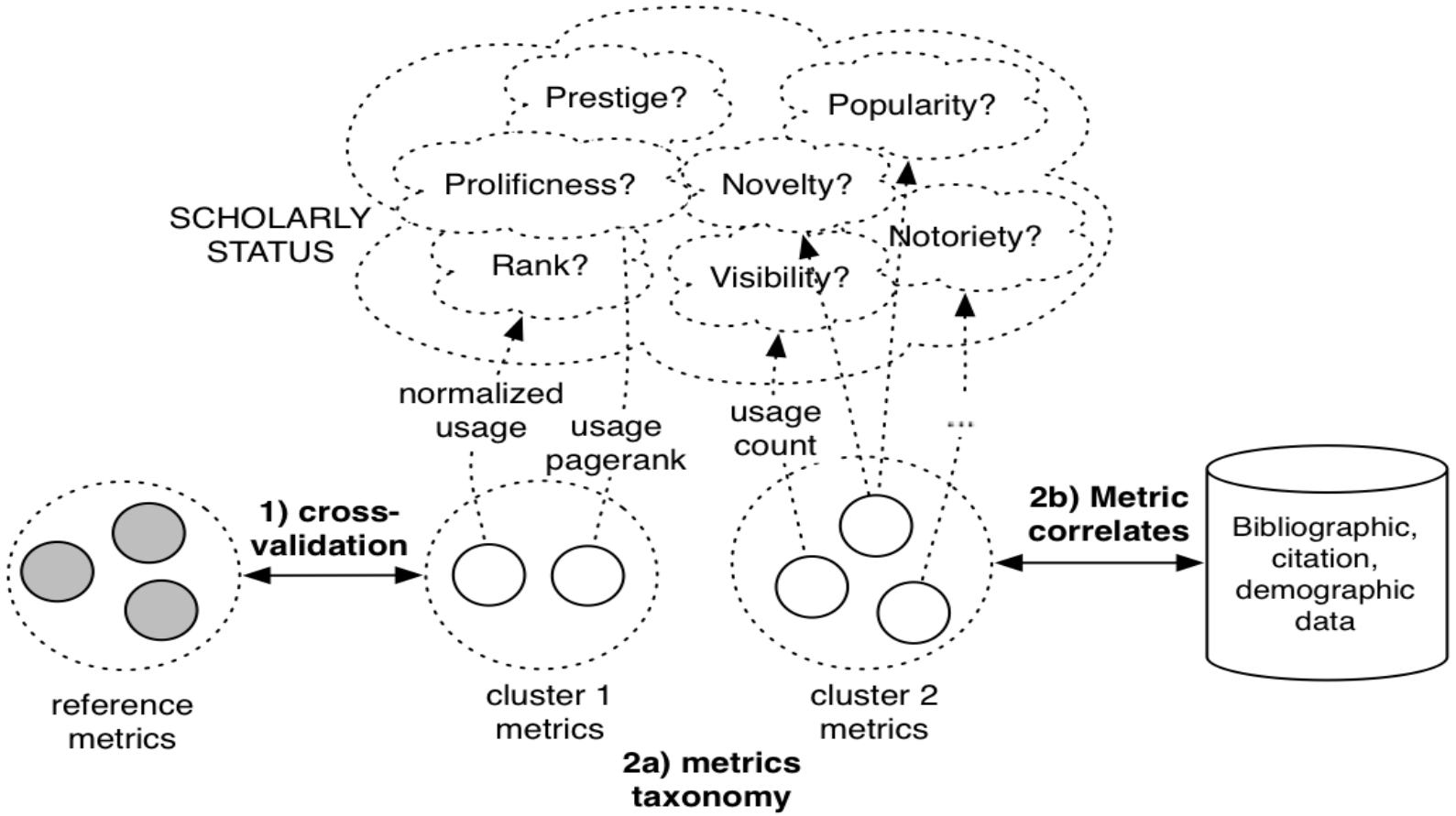
	Usage PR	IF (2003)	Title (abbv.)
1	60.196	7.035	PHYS REV LETT
2	37.568	2.950	J CHEM PHYS
3	34.618	1.179	J NUCL MATER
4	31.132	2.202	PHYS REV E
5	30.441	2.171	J APPL PHYS

LANL Los Alamos

	Usage PR	IF (2003)	Title (abbv.)
1	78.565	21.455	JAMA-J AM MED ASSOC
2	71.414	29.781	SCIENCE
3	60.373	30.979	NATURE
4	40.828	3.779	J AM ACAD CHILD PSY
5	39.708	7.157	AM J PSYCHIAT

Cal. State U.

Examples: MESUR II



From: Bollen, Johan (2007) MESUR: metrics from scholarly usage of resources *OAI5*, Geneva

Examples: DINI

- Cluster of proposals to the DFG
  - Network of certified open access repositories 2y ✓
  - Usage statistics demonstrator ✓
  - Distributed open access reference citation service demonstrator ✓
- Co-operation with German collecting society for copyright charges (VG Wort)
  - Statistics based payout to authors (METIS)

Conclusion

- Infrastructure for collecting and aggregating usage data conceptually available, has to be deployed and implemented in practice on a large scale (DINI/DFG)
- Investigating metrics for different needs and purposes (MESUR)