New COUNTER-based metrics?
- Journal Usage Factor and PIRUS

Peter Shepherd

COUNTER

Journal Usage Factor: The challenge.......

ISI's Impact Factor compensates for the fact that larger journals will tend to be cited more than smaller ones

- Can we do something similar for usage?

- In other words, should we seek to develop a "Usage Factor" as an additional measure of journal quality/value?

Journal Usage Factor

Journal Usage Factor (JUF)=

Total usage over period 'x' of articles published during period 'y'

Total articles published during period 'y'

Usage Factor Phase 2
Modelling and Analysis

○ Real journal usage
data analysed by
John Cox Associates
and Frontline GMS

○ Participating
publishers:-
- American Chemical
Society
- Emerald
- IOP
- Nature Publishing
- OUP
- Sage
- Springer

The data

- 326 journals
  - 38 Engineering
  - 32 Physical Sciences
  - 119 Social Sciences
    - *29 Business and Management*
  - 35 Humanities
  - 102 Medicine and Life Sciences
    - *57 Clinical Medicine*

- c.150,000 articles

- 3350 spreadsheets

- 1GB of data

Results

- Content Type

  - In social sciences JUFs were higher for **<u>non-article content</u>**

  - In medicine and life sciences JUFs were higher for **<u>article content</u>**

  - In humanities, physical sciences, and business & management, JUF differences between article and non-article content were not significant

Results

- Article Version

  - In physical sciences the JUF was significantly (sometimes dramatically) **lower** when calculations were confined to the Version of Record

  - In all other subjects the JUF was significantly **higher** when calculations were confined to the Version of Record
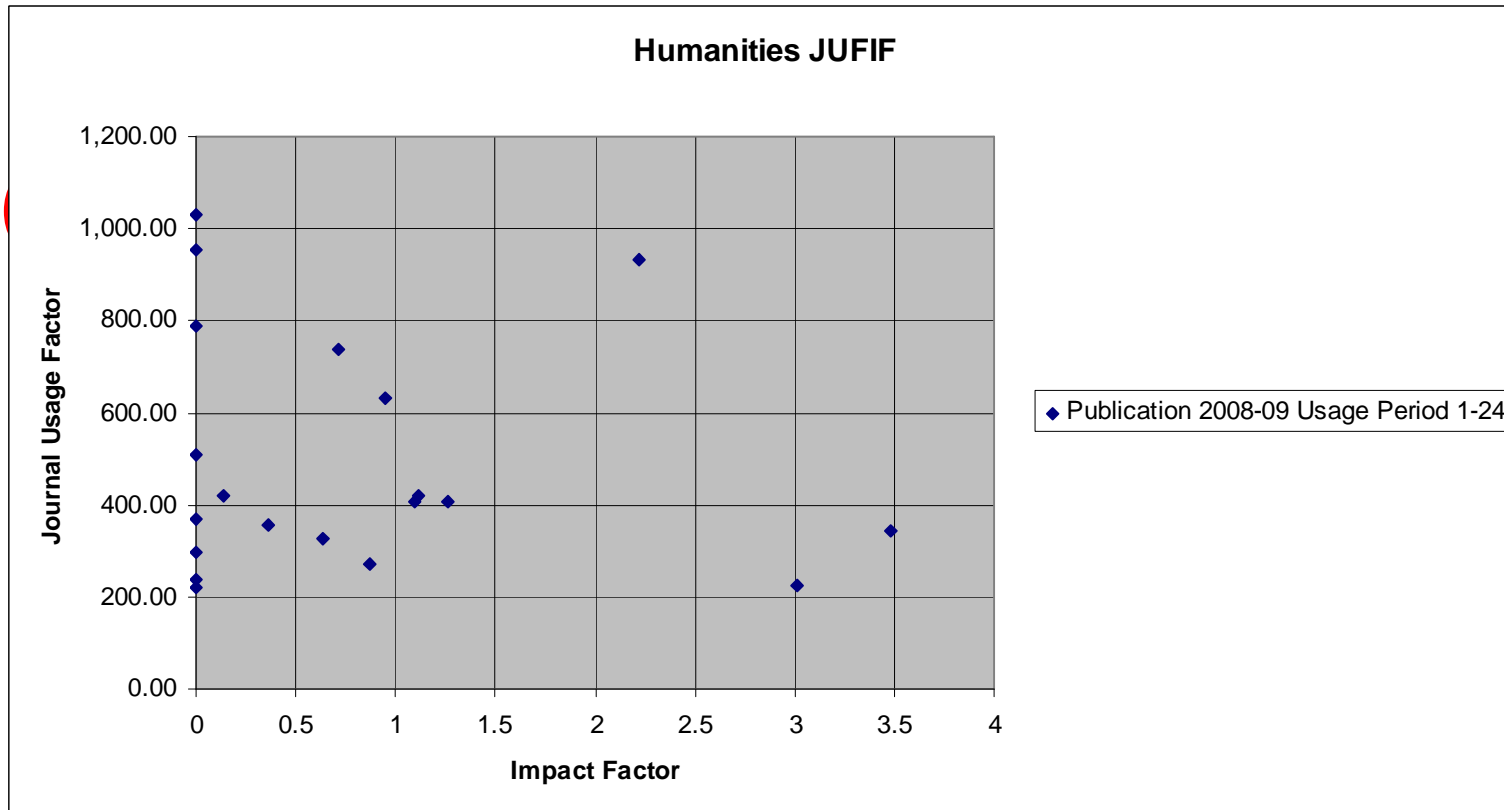
Results

- JUF and Impact Factor

  - Little correlation apart from the Nature branded titles

  - Some titles with no or very low impact factors have very high JUFs

Journal Usage Factor Project – some initial results

Humanities JUFIF

Journal Usage Factor (y-axis): 0.00 to 1,200.00

Impact Factor (x-axis): 0 to 4

Legend: ◆ Publication 2008-09 Usage Period 1-24

Recommendations – the metric

- The most promising JUF metric for further testing will be based on:-
  - All content types except standing matter
    - Non-article matter is published for a purpose and its usage forms part of the usage of the journal as a whole
    - Item type control is difficult to manage
  - All versions published
    - For simplicity and completeness
  - Publication period: 2 years
    - For a greater "smoothing" effect on occasional unexplained peaks and troughs in usage
    - To reduce the effect of pre-"Version of Record" publication
  - Usage period: 2 years contemporaneous with publication period
    - To capture peak post-publication usage
    - To keep the metric as current as possible

Journal Usage Factor

Journal Usage Factor (JUF)=

Total usage over period 'x' of articles published during period 'y'

Total articles published during period 'y'

X = 24 months

Y = maximum of 24 months

Recommendations - infrastructure

- Development of systems to automate the extraction and collation of data needed for JUF calculation is essential if calculation of this metric is to become routine

- Development of an agreed standard for content item types, to which journal specific item types would be mapped, is desirable as it would allow for greater sophistication in JUF calculation

- Development or adoption of a simple subject taxonomy to which journal titles would be assigned by their publishers

Recommendations - infrastructure

- Publishers should adopt standard "article version" definitions based on NISO recommendations

- But no specific recommendations for the labelling or making available of these versions

Next steps

- Article published in March 2011 issue of *Serials*
- CIBER are undertaking more detailed analysis of the large volume of usage data collected
  - To validate the results of the original study
  - To extend the analysis to cover al the journals in the project for each subject field
  - To asses whether the 24-month usage period proposed in the Cox report could be shortened without compromising the reliability of the metric
  - To investigate the impact of different gaming/fraud scenarios
  - To suggest other usage-based metrics that could provide insights into the status/value/prestige of individual journals
- Issues relating to subject taxonomies, metadata and automation of publisher processes need to be resolved prior to further development of JUF
- Organizational models for implementation of JUF to be investigated

Journal Usage Factor Project
- lead by UKSG

- Sponsors of this latest phase:-

- GOLD

- SILVER
  - ALPSP
  - American Chemical Society
  - STM
  - Nature Publishing Group
  - Springer

Journal Usage Factor

For further information:

http://www.uksg.org/usagefactors/

PIRUS: why now?

Increasing interest in article-level usage

- Authors and funding agencies are increasingly interested in a reliable, global overview of usage of individual articles
- Online usage becoming an alternative, accepted measure of article and journal value
  - Knowledge Exchange report recommends developing standards for usage reporting at the individual article level
  - Usage-based metrics already provided by PLoS and other publishers
- More journal articles hosted by aggregators, institutional and subject repositories

PIRUS: why now?

Article-level usage metrics now more practical

- Implementation by COUNTER of XML-based usage reports makes more granular reporting of usage a practical proposition
- Implementation by COUNTER of the SUSHI protocol facilitates the automated consolidation of usage data from different sources.

PIRUS: the challenge

○ An article may be available from:-

- The main journal web site
- Ovid
- ProQuest
- PubMed Central
- Authors' local Institutional Repositories

○ If we want to assess article impact by counting usage, how can we maximise the actual usage that we capture?

PIRUS: mission and project aims

Mission

> To develop a global standard to enable the recording, reporting and consolidation of online usage statistics for individual journal articles hosted by Publishers, Institutional Repositories, and other entities

Project aims

- Develop COUNTER-compliant usage reports at the individual article level

- Create guidelines which, if implemented, would enable any entity that hosts online journal articles to produce these reports

- Propose ways in which these reports might be consolidated at a global level in a standard way.

PIRUS1(Sept 2008-Jan 2009): -outcome

The PIRUS1 project demonstrated that it is *technically* feasible to create, record and consolidate usage statistics for individual articles using data from repositories and publishers, despite the diversity of organizational and technical environments in which they operate. If this is to be translated into a new, implementable COUNTER standard and protocol, further research and development will be required, especially into the technical, organizational and economic issues.

PIRUS2: issues addressed

- **Technical:** further tests, with a wider range of repositories and a larger volume of data, will be required to ensure that the proposed protocols and tracker codes are scalable/extensible and work in the major repository environments.

- **Organizational**: the nature and mission of the central clearing house/houses proposed by PIRUS1 has to be developed, and candidate organizations identified and tested

- **Economic:** assess the costs for repositories and publishers of generating the required usage reports, as well as the costs of any central clearing house/houses; investigate how these costs could be allocated between stakeholders

- **Political:** the broad support of all the major stakeholder groups (repositories, publishers, authors, etc) will be required.

PIRUS2 (Oct 2009-Feb 2011): - aims and objectives

The aim of PIRUS2 is to address these issues and by doing so specify standards, protocols, an infrastructure and an economic model for the recording, reporting and consolidation of online usage of individual articles hosted by repositories, publishers and other entities.

Specific objectives:

○ Develop a suite of free, open access programmes to support the generation and sharing of COUNTER-compliant usage data and statistics that can be extended to cover any and all individual items in repositories

○ Develop a prototype article-level publisher/repository usage statistics service the Central Clearing House (CCH)

○ Define a core set of standard useful statistical reports that repositories should produce for internal and external consumption
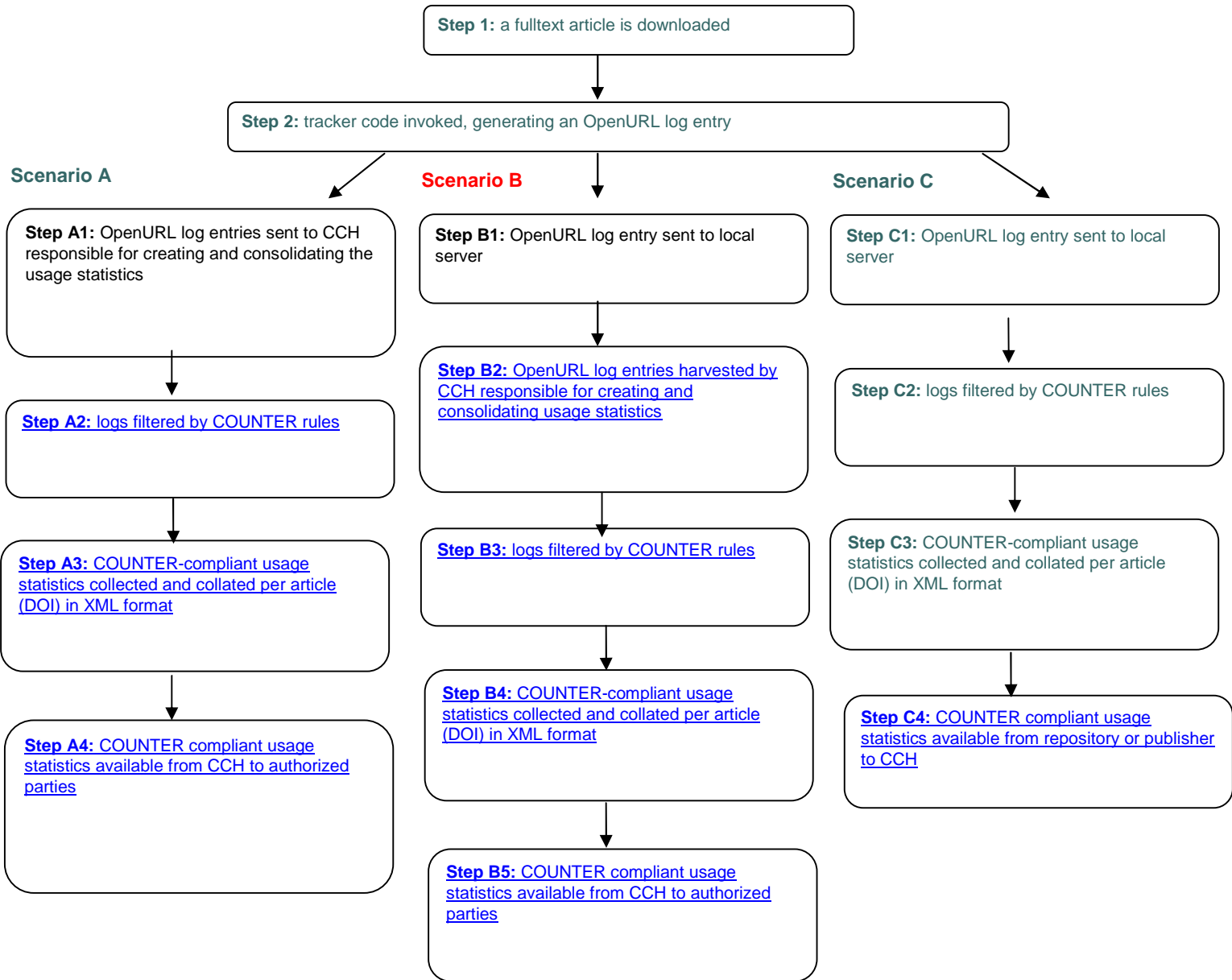
PIRUS2: organizational issues

- Specifications for the Governance of PIRUS, going forward
- define the nature and mission of the Central Clearing House(s) (CCH) in more detail, in discussion with publishers and repositories
- Develop a specification for the technical, organizational and business models for the CCH
- identify candidate organizations for involvement in the CCH

PIRUS2: nature and mission of the Central Clearing House

- **One global CCH**
  - Cost-effective
  - Industry is global, with global standards
  - Easier to set and modify standards
  - Simpler interface with publishers and repositories
  - Can be outsourced
  - Existing organizations exist with the required capabilities
- **Scenarios to be supported**
  - See next slide……..

Step 1: a fulltext article is downloaded

Step 2: tracker code invoked, generating an OpenURL log entry

Scenario A

Step A1: OpenURL log entries sent to CCH responsible for creating and consolidating the usage statistics

Step A2: logs filtered by COUNTER rules

Step A3: COUNTER-compliant usage statistics collected and collated per article (DOI) in XML format

Step A4: COUNTER compliant usage statistics available from CCH to authorized parties

Scenario B

Step B1: OpenURL log entry sent to local server

Step B2: OpenURL log entries harvested by CCH responsible for creating and consolidating usage statistics

Step B3: logs filtered by COUNTER rules

Step B4: COUNTER-compliant usage statistics collected and collated per article (DOI) in XML format

Step B5: COUNTER compliant usage statistics available from CCH to authorized parties

Scenario C

Step C1: OpenURL log entry sent to local server

Step C2: logs filtered by COUNTER rules

Step C3: COUNTER-compliant usage statistics collected and collated per article (DOI) in XML format

Step C4: COUNTER compliant usage statistics available from repository or publisher to CCH

PIRUS2: outputs from the CCH

- Usage reports for publishers
- Usage reports for repositories
- Usage reports for research institutions

Key requirements:

- Set of core reports
- Flexibility in outputs

PIRUS2: Implementation of the CCH

In view of the technical challenges that the CCH faces, its strong dependency on other initiatives, such as ORCID and institutional identifier and the requirements for publishers to re-engineer some of their processes, it may be prudent to implement the CCH in two Stages:

○ Stage 1: gather and consolidate usage data only from repositories and provide the usage statistics generated by the CCH to publishers and other authorised bodies

○ Stage 2: and collect usage data from publishers that wish to use the CCH service for this purpose

PIRUS2: organizations that could play a role in the CCH

- Setting the standards for usage reporting and specifications for the usage reports – COUNTER
- data gathering – existing vendor (eg ScholarlyiQ, MPS Technologies, etc
- DOI information- CrossRef
- counting and reporting- existing vendor (ScholarlyiQ, MPS Technologies, etc)
- final report compilation – existing vendor (ScholarlyiQ, MPS Technologies, etc)
- auditing of the CCH – ABCe
- management of the CCH – existing vendor (ScholarlyiQ, MPS Technologies, etc.,) supervised by a PIRUS management board
- customer service/account management – existing vendor (ScholarlyiQ. MPS Technologies, etc)

PIRUS2: economic issues

- estimate the costs of running statistical aggregator services

- assess the costs to repositories and publishers for generating the required usage reports;

- propose a model for the allocation of costs to stakeholders

PIRUS2: model for recovering costs

Possible sources of revenues to support the CCH:

○ membership fees that give members the right to use the services of the CCH

○ transaction-based fees:

- From repositories, who provide raw data to the CCH and obtain usage statistics from the CCH
- From repositories, who provide raw data to the CCH and obtain usage statistics from the CCH
- from publishers, who obtain usage statistics from the CCH for consolidation into their own usage reports
- from organizations, such as Thomson ISI or Elsevier (SciVal), who could use the data from the CCH to enhance the citation and usage based performance reports that they provide to institutions.
- from research institutions, who want global usage reports for content produced by their researchers and departments

PIRUS2: political issues

- Support of stakeholder groups
  - Authors
  - Publishers
  - Repositories
  - Research institutions
  - Research funding agencies
- Principle of reporting article-level usage
- Organizational model
- Economic model
- Reports output by the CCH
- Intellectual property and privacy issues

PIRUS2: Further work on organizational/political issues

More feedback from stakeholder groups on:

- acceptability of the proposed organizational model
  - Governance
  - Structure
  - Participating organizations
- acceptability of the proposed economic model
  - Level of costs
  - Cost allocation model
- the proposed usage reports to be generated by the CCH
  - Different reports for different stakeholders

PIRUS
- sponsored by JISC

Further information:

PIRUS2 website

http://www.cranfieldlibrary.cranfield.ac.uk/pirus2