COUNTER

Counting Online Usage of NeTworked Electronic Resources

PIRUS 2

Creating a common standard for measuring online usage of individual articles

Paul Needham, Cranfield University
Peter Shepherd, **COUNTER**

Madrid

7 July 2010

PIRUS Publisher and Institutional Repository Usage Statistics

- **Sponsored by JISC**
  - UK Joint Information Systems Committee
- **PIRUS 1** completed in January 2009
  - Lead by COUNTER
  - Report available at: http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus_finalreport.pdf
- **PIRUS 2**, October 2009-December 2010
  - Lead by Mimas and Cranfield University
  - Primary project team members: Mimas, Cranfield, COUNTER, CrossRef, Oxford University Press

Usage statistics and journal metrics

- **COUNTER**
  - Sets the standard for vendor-generated online usage statistics
  - Covers over 15,000 full-text online journals

  http://www.projectCounter.org
- **MESUR**
  - Enriches the toolkit used for the assessment of the impact of scholarly communication items with usage data
  - Has created a map of science based on usage data

  http://www.mesur.org/
- **Journal Usage Factor**
  - Assess the feasibility of Journal Usage Factor as an alternative metric to Journal Impact Factor

  http://www.uksg.org/usagefactors
- **PIRUS**
  - Aims to provide, publishers, repositories and other organizations with a common standard for measuring usage at the individual article (item) level

PIRUS: why now?

Increasing interest in article-level usage

- More journal articles hosted by Institutional and other Repositories
- 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 being considered as a tool for use in the UK Research Excellence Framework and elsewhere.

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.

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 Project Mission

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

PIRUS 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.

PIRUS: benefits

- Reliable usage data will be available for journal articles, wherever they are held
- Repositories will have access to new functionality from open source software that will allow them to produce standardised usage reports from their data
- Digital repository systems will be more integral to research and closely aligned to research workflows and environments
- The authoritative status of PIRUS2 usage statistics will enhance the status of repository data and content
- The standard can be extended to cover other categories of content stored by repositories

PIRUS1: publisher response

- Majority enthusiasm for concept
- All surveyed publishers use DOIs to identify all versions of a single published work
- Minority concern that article level reporting to institutional customers is our goal
  - It isn't
- Concern about size of any reports providing usage data at article level.
  - Not the intention of the project to recommend publishers produce reports relating to more than one article at a time

PIRUS1: repository response

GOOD NEWS

- The overwhelming majority of respondents add DOIs to their records - where they are available.

BUT…….

- No standard process for allocating DOIs in IRs
- Great variation in the metadata element used to store them:-
  - dc.description
  - dc.identifier
  - dc.identifier type DOI
  - dc.identifier.citation
  - dc.relation.isreferencedby
  - dc.rights
  - DOI
  - relation

PIRUS1: outputs

1. A proof-of-concept COUNTER-compliant XML prototype for an individual article usage report
2. A tracker code, to be implemented by repositories, that sends usage data as OpenURL Context Objects to either:
   - An external party
   - The local repository server
3. A set of scenarios for collecting usage data in different repository environments
4. A set of criteria for a central Clearing House that will create (where required), or collect and consolidate the usage statistics

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

PIRUS2: progress so far:-
WP 4: software, standards and protocols

- Technical aspects of project
- Gathering … usage data and statistics
  - For full-text article downloads (not record/abstract views)
  - From repositories and publishers
- Consolidating …
  - In an article-level usage statistics demonstrator portal
  - Experiment and illustrate possibilities
- Re-exposing …
  - To authorized third parties

PIRUS2: progress so far:-
WP 4: software, standards and protocols

- Three scenarios for gathering …
  - (A) 'tracker' code – a server-side 'Google Analytics' for full-text article downloads
  - (B) OAI-PMH harvesting – protocol familiar to repositories
  - (C) SUSHI - Standardized Usage Statistics Harvesting Initiative Protocol – familiar to publishers

PIRUS2: progress so far:-
WP 4: software, standards and protocols

- Usage data from Repositories
- Scenarios (A) Tracker & (B) OAI-PMH
  - Usage data are exposed as:
    - (A) OpenURL Key-Value Pair Strings
    - (B) OpenURL Context Objects.
  - OpenURL approach first suggested by MESUR. Taken forward in Europe under 'Knowledge Exchange' – an initiative involving DEFF, DFG, JISC and SURFfoundation, see: http://wiki.surffoundation.nl/display/standards/OpenURL+Context+Objects
- Usage data must be:
  - filtered according to COUNTER rules to eliminate Robots and Double clicks
  - Processed into monthly statistics

PIRUS2: progress so far:-
WP 4: software, standards and protocols

- Usage statistics from Publishers
- Scenario (C) SUSHI
  - SUSHI - a SOAP-based web service used by publishers to expose COUNTER Release 3 compliant usage statistics to institutions and consortia
  - Currently operates at journal level, e.g. JR1 report: Number of Successful Full-Text Article Requests by Month and Journal
- PIRUS2 has devised a proposed COUNTER Article Report 1 (AR1) Report: Number of Successful Full-Text Article Requests by Month and DOI
- Usage statistics are pre-filtered according to COUNTER rules

PIRUS2: progress so far:-
WP 4: software, standards and protocols

- PIRUS2 Repository software plug-ins/extensions
  - Dspace – developed by @mire
  - Eprints – developed by Tim Brody, Southampton University
  - Fedora – developed by Ben O'Steen, Oxford University
  - Links and downloads on PIRUS2 project web site
- PIRUS2 AR1 Report
  - SUSHI ultimately
  - Currently working with AR1 reports in MS Excel/CSV format from participating publishers
  - Draft AR1 report in MS-Excel and XML available on PIRUS2 project web site

PIRUS2: progress so far:-
WP 4: software, standards and protocols

- Current situation
  - Loaded data from 4 publishers
    - Over 450,000 articles
    - From 5,500 journals
  - Gathering data via tracker from 3 repositories
    - Working on scripts to process and load data
  - Creating user interface to demonstrate possibilities
- Next
  - Load data from another 4 publishers
  - Extend participation by repositories
  - Ongoing development and testing of user interface
  - Develop SUSHI server to re-expose statistics

PIRUS2: progress so far
WP5: prototype service

- Tests of publisher usage data
  - Usage data from 8 publishers flowing in
- Define functions to be fulfilled by a Central Clearing House
  - Collect, collate and store usage data
- Define capabilities required of a Central Clearing House
  - Conversion of logfiles, storage, access control, etc
- Define organizational options for a Central Clearing House
  - Global vs. local; identify candidate organizations

PIRUS 2: primary project team

- Ross MacIntyre (Mimas, Manchester University)
- Paul Needham (Cranfield University)
- Richard Gedye (Oxford University Press)
- Ed Pentz (CrossRef)
- Peter Shepherd (COUNTER)

For more information……

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

Thank you!