PIRUS Publisher and Institutional Repository Usage Statistics

PIRUS2: The Demonstrator

Paul Needham, Cranfield University  
PIRUS2 Project Manager & WP4 leader

PIRUS2 End of Project Seminar - 23 February 2011

Funded by: JISC
The technical challenge:

- Is it possible to capture usage events from various sources and consolidate them to show the overall usage of an article???

- i.e. can we:
  - Gather... usage data and statistics
  - Consolidate...
  - Re-expose... to authorized third parties

- Yes ... mostly!
PIRUS2: The Demonstrator

What the demonstrator isn’t:

- It isn’t intended to go into production
- It isn’t particularly pretty
- It isn’t a finished product
PIRUS2: The Demonstrator

What the demonstrator is:

- A MySQL database
- A set of perl scripts for processing & getting data in
- A web user interface written in php
- A proof of concept
- ... and a playground!
PIRUS2: The Demonstrator

PIRUS2 database v0.2
PIRUS2: The Demonstrator

Perl scripts to process and load data into db

- Publisher statistics
  - All loaded from AR1 (or similar) reports
  - Article and Journal authority, Statistics
  - Reasonably straightforward

- Repository usage data
  - Some loaded
  - Not so straightforward
PIRUS2: The Demonstrator

Perl scripts to process and load data into db

Repository usage data


Perl scripts to process and load data into db

- **Repository usage data**
  - Parsed and sorted into Local ID & datetime
  - robots and double clicks eliminated
  - Grouped into monthly accesses per article
  - Equivalent to AR1 from publishers
  - Entries either have
    - DOI
    - Article title/Author surname
PIRUS2: The Demonstrator

Perl scripts to process and load data into db

- Repository records with DOI
  - Known to the system - loaded
  - Unknown to the system – queued for further processing
PIRUS2: The Demonstrator

Perl scripts to process and load data into db

- **Repository records without DOI**
  - Article title/Author surname
    - Lookup DOI in the local system
    - Lookup DOI in CrossRef
  - Repository records - DOI found
    - Known to the system - loaded
    - Unknown to the system – queued for further processing
  - Repository records - no DOI found
    - Queued for further processing
The php web interface: Demonstration

- http://cclibweb-4.dmz.cranfield.ac.uk/pirus2/
Lessons learnt:

- Where there’s a DOI to be found, consolidation is easily achievable
- Where there isn’t, need some other reliable identifier(s)
- Publishers
  - If you don’t allocate DOIs – please do!
- Repositories
  - Quality of metadata matters!
  - If a DOI exists for an article – catalogue it!
  - Typos, errors, missing metadata cause headaches!
- Publishers and Repositories
  - Sort out versioning and peer-review status metadata!
Next steps

- Update the db with queued repository records where we’ve found a DOI and that DOI is new to the system
- Give further consideration to and play with a records where there isn’t a DOI
  - ‘Bucket’ of identifiers?
  - CCH to mint its own?
  - Come up with some recommendations
PIRUS2: The Demonstrator

Thank you for listening!

For more information:
http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/