

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**

PIRUS2: The Demonstrator

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

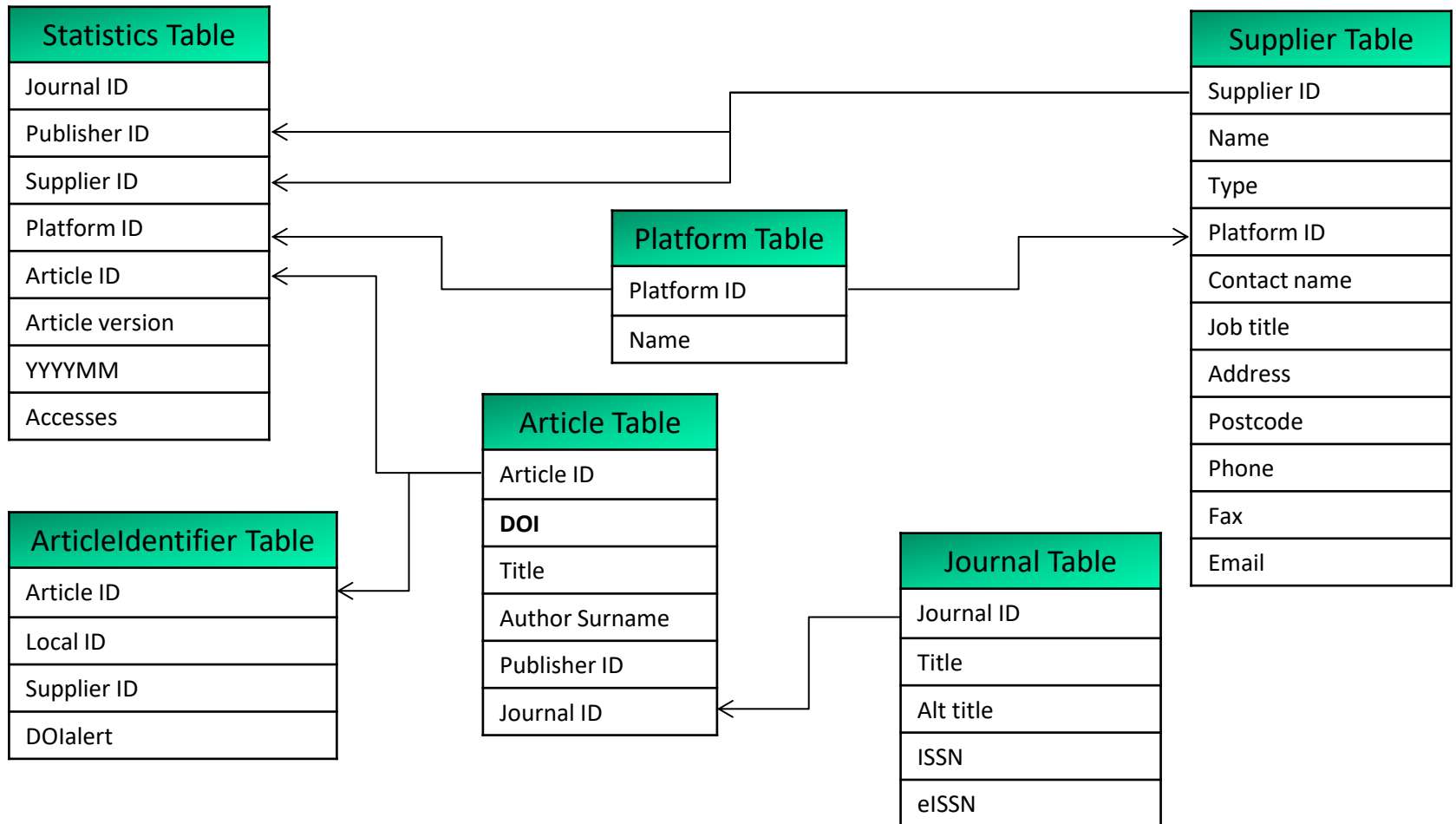
PIRUS2: The Demonstrator

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

```
138.250.13.22 - - [17/Oct/2010:04:04:44 +0100] "GET /tracker/?url_ver=Z39.88-2004&req_id=e02db545fbefd7d19bf24302a57f93ac&req_dat=Mozilla%2F5.0+%28compatible%3B+Googlebot%2F2.1%3B+%2Bhttp%3A%2F%2Fwww.google.com%2Fbot.html%29&rft.artnum=http%3A%2F%2Fdspace.lib.cranfield.ac.uk%2Fhandle%2F1826%2F3228&svc_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Adc&svc.format=application%2Fpdf&svc_dat=Unknown&rft_id=dspace.lib.cranfield.ac.uk&url_tim=2010-10-17T03%3A04%3A42Z&rft_id=info%3Adoi%3Ahttp%3A%2F%2Fdx.doi.org%2F10.1016%2Fj.istr.2008.10.006 HTTP/1.1" 200 635 "-" "Java/1.6.0_21"
```

```
152.78.189.11 - - [20/Feb/2011:04:16:47 +0000] "HEAD /tracker/?url_ver=Z39.88-2004&url_tim=2011-02-20T04%3A16%3A45Z&req_id=urn%3Aip%3A219.219.127.3&req_dat=Mozilla%2F4.0+(compatible%3B+MSIE+6.0%3B+Windows+NT+5.1%3B+)&rft.artnum=http%3A%2F%2Feprints.ecs.soton.ac.uk%2Fid%2Feprint%2F8671&svc.format=application%2Fpdf&rft_id=eprints.ecs.soton.ac.uk&rft.date=2003-12&rft.aulast=Schwanecke&rft.volume=91&rft.atitle=Broken+Time+Reversal+of+Light+Interaction+with+Planar+Chiral+Nanostructures HTTP/1.1" 200 - "-" "EPrints 3.2.5 (Stollen) [Born on 2011-01-17]"
```

PIRUS2: The Demonstrator

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

PIRUS2: The Demonstrator

The php web interface: Demonstration

- <http://cclibweb-4.dmz.cranfield.ac.uk/pirus2/>

PIRUS2: The Demonstrator

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!

PIRUS2: The Demonstrator

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/>