Institutional Repository Usage Statistics

RSP Webinar
20 March 2013
Paul Needham, Cranfield University
Ross Macintyre, Mimas
Angela Conyers, Evidence Base, BCU

IRUS-UK

- Funded by Jisc as part of UK RepositoryNet+

- Led by Mimas

- Primary Project Team Members:
  - Mimas
  - Cranfield University
  - EvidenceBase, BCU

- IRUS-UK: Institutional Repository Usage Statistics - UK

IRUS-UK

IRUS-UK: history

- Emerged as an outcome of PIRUS2
    - Publisher and Institution Repository Usage Statistics project
    - http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/
    - Aimed 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 others
    - Proved it was technically feasible, but thwarted by organisational and political issues
    - However, crisis/opportunity… we now knew it was possible to…

IRUS-UK

IRUS-UK: aim

- Enable UK IRs to share/expose usage statistics based on a global standard – COUNTER
  - Produced on the same basis as publishers
  - Filtered to remove robots and double clicks
  - Comparable
  - Reliable
  - Trustworthy
  - Authoritative

IRUS-UK

irus.mimas.ac.uk

IRUS-UK: objectives

- Collect raw usage data from UK IRs for all item types within repositories
  - Downloads not record views

- Process those raw data into COUNTER-compliant statistics

- Return those statistics back to the originating repositories for their own use

- Give Jisc (and others) a nation-wide picture of the overall use of UK repositories
  - demonstrate their value and place in the dissemination of scholarly outputs

- Offer opportunities for benchmarking

- Act as an intermediary between UK repositories and other agencies
  - e.g. global central clearinghouse, national shared services, Research Councils, SCONUL, OpenAIRE

IRUS-UK

irus.mimas.ac.uk

IRUS-UK: how to gather data?

- We considered 2 primary mechanisms for gathering data from IRs
  - 'Tracker' code
    - Think of it as a server-side 'Google Analytics' for item downloads
    - Pushes metadata to a third-party server as OpenURL Key/Value strings whenever a download occurs
  - OAI-PMH harvesting
    - Used to by third parties to 'pull' metadata from repositories
    - Repurposed to expose usage events as OpenURL Context Objects

- We opted to employ the Tracker
  - We receive notification of download events as and when they occur
  - It's just easier than employing OAI-PMH, which becomes difficult to manage as the number of participating IRs grows

IRUS-UK

IRUS-UK: Tracker OpenURL strings

- What does the tracker actually send to IRUS-UK when a download happens?

  - http://www.jusp.mimas.ac.uk/counter/?url_ver=Z39.88-2004&url_tim=2012-07-05T22%3A59%3A59Z&req_id=urn%3Aip%3A86.15.47.114&req_dat=Mozilla%2F5.0+(iPhone%3B+U%3B+CPU+iPhone+OS+5_1_1+like+Mac+OS+X%3B+en-us)+AppleWebKit%2F534.46.0+(KHTML%2C+like+Gecko)+CriOS%2F19.0.1084.60+Mobile%2F9B208+Safari%2F7534.48.3&rft.artnum=oai%3Aeprints.hud.ac.uk%3A8795&svc_format=application%2Fpdf&rfr_id=eprints.hud.ac.uk

IRUS-UK: Tracker OpenURL strings

- The OpenURL key/value pairs
  - url_ver=Z39.88-2004
  - url_tim=2012-07-05T22%3A59%3A59Z
  - req_id=urn%3Aip%3A86.15.47.114
  - req_dat=Mozilla%2F5.0+(iPhone%3B+U%3B+CPU+iPhone+OS+5_1_1+like+Mac+OS+X%3B+en-us)+AppleWebKit%2F534.46.0+(KHTML%2C+like+Gecko)+CriOS%2F19.0.1084.60+Mobile%2F9B208+Safari%2F7534.48.3
  - rft.artnum=oai%3Aeprints.hud.ac.uk%3A8795
  - svc_format=application%2Fpdf
  - rfr_id=eprints.hud.ac.uk

IRUS-UK

IRUS-UK: deploying tracker code

- There are a (growing) number of software platforms offering Institutional Repository functionality, these include:
  - DSpace, Eprints, Fedora, intraLibrary (Intrallect), Digital Commons (Bepress), PURE Portal…
  - …And the list goes on

- Each of these platforms works differently under the hood meaning a single solution for deploying tracker functionality is not possible

- So, we've focused, initially, on DSpace and Eprints, which account for about two-thirds of UK repositories
  - Plugins/patches available for DSpace (1.8.x and 3.x) and Eprints (3.3.x)
  - However, we are interested in working with other platforms and would welcome dialogue with interested vendors & developers

IRUS-UK

IRUS-UK: processing data (1)

- Data gets stored in daily log files

- The Log for any given day is usually processed the following day

- We have a three step ingest process

- Step 1: Perl script parses the logs
  - Processes entries from recognised IRs
  - Sorts and filters entries following COUNTER rules
    - Removing robot entries and double-clicks
  - Filters entries using additional IRUS-UK filters
    - There are a lot of robots out there!
  - Consolidates raw usage data for each item into daily statistics
  - Outputs to intermediate file

IRUS-UK

IRUS-UK: processing data (2)

- Step 2: Perl script processes the intermediate file
  - Intermediate file entries
    - 5   2   oai:usir.salford.ac.uk:9967   2013-03-18   1
    - 5   2   oai:usir.salford.ac.uk:9970   2013-03-18   6
    - 5   2   oai:usir.salford.ac.uk:9972   2013-03-18   2
  - Using the OAI identifier, it looks up each item against the Item Authority table in the IRUS DB to see if we already know about it
    - If we do, all well and good – there's an existing IRUS Item Identifier
    - If not, it adds a stub-entry to the Item Authority table – minting a new IRUS Item Identifier and adding the repository identifier, platform and OAI identifier
    - The rest of the metadata is set to 'unknown' at this stage
  - Finally, the script adds the download statistics associated with each IRUS Item Identifier to the Daily Statistics table

IRUS-UK

IRUS-UK: processing data (3)

- Step 3: Perl script obtains the "unknown" metadata
  - Queries the DB to find the 'known unknowns'
  - Using the OAI identifiers, issues OAI-PMH GetRecord calls to retrieve OAI_DC metadata
  - Parses the OAI records
  - Updates the metadata – Title, Author, Item Type, etc., in the Item Authority Table in the DB
  - Additionally maps the Item Type, as given by the source repository, to a smaller (more manageable list) of IRUS Item Types

- Step 4: A monthly Perl script
  - Consolidates the Daily Statistics into a Monthly Statistics table
    - The 'traditional' COUNTER granularity
  - So, as a service, we can easily work with statistics at either granularity

IRUS-UK

IRUS-UK: community engagement

- Growing number of repositories sending data to IRUS-UK

- Currently 16 participants:

  - Bournemouth, City, UEA. Greenwich, Huddersfield, LSE, NERC, Northumbria, Open, Reading, Salford, Sussex, Warwick (Eprints)

  - Aberystwyth, Cranfield, RGU (DSpace)

  - Others in the pipeline

IRUS-UK

IRUS-UK: COUNTER-compliant

- Reliable, Trustworthy, Authoritative

"It's great that it's COUNTER"

"Authoritative download stats, using a standard we already work with."

"One of values of IRUS will be verifying data, doing data cleaning. How many downloads are genuine and how many are robots?"

"Reassurance that the data we have is as close as possible to a true representation of the use of the system as possible."

IRUS-UK

IRUS-UK: OPEN data

- Comparable data
  - For benchmarking:

"Being able to see some usage data from other institutions within IRUS, and knowing that the same methodology has been used to arrive at these numbers, will be really valuable to us."

  - For tracking authors or subjects in different IRs:

"The 'Search' is interesting in that shows downloads of the same author or subjects across repositories."

IRUS-UK

IRUS-UK: overall use of UK IRs

- Demonstrates scale of downloads across UK IRs:

"The summary data which demonstrates the scale of repository access across the registered repositories is very informative."

"The way in which it demonstrates the value of IRs in providing Green OA to high demands."

- As at 15 March 2013:

  16 repositories – 2 million+ items downloaded since IRUS-UK started

- For the future:

  150+ repositories – how many items downloaded?

IRUS-UK

IRUS-UK: helping increase content?

- Reliable data, benchmarking opportunities as means of encouraging more deposit?

  "The quality data …allows me to confidently sell our services to our researchers to gather further content for the repository."

  "We've found statistics are helpful in encouraging our staff and postgraduate research students to deposit their work, so if it's easier for us to gather this data (or direct our users to it), the IRUS-Stats will be really useful to us."

IRUS-UK

IRUS-UK: how to join

- If you are a UK repository:
  - Contact us at irus.mimas.ac.uk to register your interest
  - Answer a few questions on the type of repository you have and the version you are running
  - Get advice from us on what work will be involved depending on your repository type and version
  - Implement any changes advised and then see your usage data instantly in IRUS-UK with no more work from you

"The set up was quick and painless, which is always a delight!"

"Consistent collection of statistics without me having to do it!"

IRUS-UK

IRUS-UK: Exposing statistics

- Through the User Interface - The Portal

- Various reports are available to Institutions
  - For humans, downloads as CSV/Excel spreadsheet files
    - Monthly and daily granularity
  - For machines, harvesting via SUSHI
    - Monthly granularity

- Usage statistics for incorporation into Repositories
  - Via an API/Web Service
    - Initial version in place
    - Further requirements to be determined

IRUS-UK

irus.mimas.ac.uk

IRUS-UK: next steps

- More of what we're already doing…

- Community Engagement
    - Getting more IRs on-board
    - Defining and evaluating user requirements
        - including surveys, interviews, monitoring feedback, etc.

- Processes to support
    - institutional liaison, data collection, monitoring and responding to future requirements, ensuring interoperability
    - development and production environments including monitoring, back-up, authentication and archiving polices and processes

- Formal application for COUNTER-compliance

IRUS-UK

irus.mimas.ac.uk

IRUS-UK: future webinars

- IRUS-UK is planning to run more webinars:

    - 17 April 14.00 – aimed primarily at repository managers

    - 22 May 14.00 – aimed primarily at repository technical managers

    - 17 July 14.00 – end of project webinar

- More details to follow

IRUS-UK

Contacts & Information

- For general enquiries, please contact
    - support@repositorynet.ac.uk

- If you are a UK repository wishing to participate in IRUS-UK, please contact
    - irus@mimas.ac.uk

- Project web site:
    - http://www.irus.mimas.ac.uk/

- Thank you!

IRUS-UK