IRUS-UK: Improving Understanding of The Value And Impact of Institutional Repositories

Ross MacIntyre

Presenter

Abstract

Institutional repositories (IRs) are important research management tools that can give increased visibility to the institution's scholarly outputs. Although statistics were previously available through the various repository interfaces, without an agreed standard it was not possible to measure usage across a range of IRs accurately. IRUS-UK is a national aggregation service, containing details of all content downloaded from participating IRs in the UK. Through collecting raw usage data and processing them into item-level usage statistics IRUS-UK provides comparable and authoritative standards-based data and also acts as an intermediary between UK repositories and other agencies.

KEYWORDS *usage statistics, COUNTER, institutional repositories*

Introduction

Many educational institutions have repositories for research outputs. The number of items available through institutional repositories is growing, and is expected to continue to do so due to requirements for outputs from public-funded research to be Open Access. But

how much usage are institutional repositories and their individual items getting?

IRUS-UK Service

The Jisc-funded[1] service 'IRUS-UK'[2] is designed to help institutions understand more about the usage of their institutional repositories. It follows on from the successful PIRUS2 project[3], which demonstrated how COUNTER-compliant article-level usage statistics[4] could be collected and consolidated from publishers and institutional repositories. (Some of the underlying technical principles were taken forward from the substantial work done in LANL's Mesur project[5] and in the European Knowledge Exchange Working Group on Usage Statistics[6].)

IRUS-UK collects raw usage data from participating repositories and processes these into COUNTER-compliant statistics. This provides repositories with comparable, authoritative, standards-based data and opportunities for profiling and benchmarking. It enables institutions to run reports at both repository level (e.g. total download figures) and at item level. IRUS-UK has developed a taxonomy of 25 'Item Types'[7] which all items are mapped to so that usage across specific item types can be consistently compared. Item Types are currently: *Art/Design Item; Article; Audio; Book; Book Section; Conference Papers /Posters; Conference Proceedings; Conference or Workshop Item – Other; Dataset; Exam Paper; Image; Learning Object; Moving Image; Music/Musical Composition; Other; Patent; Performance; Preprint; Report; Show/Exhibition; Text; Thesis or Dissertation; Unknown; Website; Working Paper*. However, the original metadata is retained, so the data can be remapped if required.

IRUS-UK utilises a robust, multistage ingest process, validating data, stripping out robot and unusual accesses[8], and filtering out double clicks, to transform raw usage data

into COUNTER-compliant statistics. However, COUNTER has recognized, in an increasingly Open Access environment, that further work is required to filter out robots and 'rogue usage'. IRUS-UK is involved in the recently formed COUNTER Working Group on Robots, devising more sophisticated - but practical - algorithms to improve filtering. Outcomes will eventually become incorporated into the COUNTER standard and adopted by IRUS-UK.

IRUS works by adding a small piece of code to repository software which employs the 'Tracker Protocol'[9]. Tracker code is currently available for DSpace, EPrints, and implementation guidelines are available for Fedora. Whilst not in IRUS-UK scope, it is worth noting that the Tracker Protocol has also been successfully deployed to create similar statistics services for: OAPEN Library[10] (using ARNO software), which offers freely accessible academic books and CORE[11], which aggregates millions of scholarly articles from many Open Access repositories.

The tracker gathers basic raw data for each download and sends it to the IRUS-UK server. On a daily basis, the data goes through a number of filters and checks before being added to the portal, and is then reviewed again at the end of each month. These processes help ensure robot activity or other unusual activity is spotted and removed to improve the accuracy of the data.

Once the data has been checked, it is accessible via the web user interface in the IRUS-UK portal, which is accessed via UK Access Management Federation[12] authentication. Within the portal, there are a wide range of views, showing the data in various different forms of reports. All the reports are available for viewing in HTML format, or for download as CSV or TSV files for further analysis (e.g. to use in Microsoft Excel).

Reporting

The data from IRUS-UK can be used to provide information for management reporting, for usage monitoring, and for external reporting (such as annual statistics for UK academic libraries).

Summary reports, which give an overview of downloads from our participating repositories, include:

1. An overall summary of downloads for all participating repositories.
2. Total number of downloads for each individual participating repository.
3. A breakdown of repository participation and number of downloads by selected countries in the UK (England, Scotland and Wales).
4. A breakdown of repository participation and number of downloads by platform used (DSpace, EPrints or Fedora).
5. Numbers of each type of item downloaded and number of downloads of each type of item for all participating repositories.
6. Numbers of each type of item downloaded and the number and percentage for each item type which have DOIs available in the metadata that we harvest.
7. An analysis of the data ingest process for each repository showing raw data, exclusions for robots and double clicks, and the resulting number of downloads showing in IRUS-UK.

The Item-level reports include:

1. ‘Item Report 1’ provides the number of successful item download requests by month and repository identifier for a selected repository.
2. ‘Item Report 2’ provides the number of successful item download requests by month and item type for a selected repository.
3. ‘Article Report 4’ provides the number of successful article downloads by month for participating repositories. The report can be filtered to limit the results to a selected journal or repository.
4. ‘Book Report 1’ provides the number of successful book downloads by month for a selected repository.
5. ‘Book Report 2’ provides the number of successful book section downloads by month for a selected repository.
6. ‘Electronic Thesis or Dissertation Report 1’ provides the number of successful thesis or dissertation download requests by month and repository identifier for a selected repository. For each thesis or dissertation, it shows the item URL, EThOS ID (British Library’s Electronic Theses Online Service)[13] if available, title, author and total downloads by month and in total for the period selected.
7. ‘Journal Report 1’ provides the number of successful Full-Text Article Requests by Month and Journal for participating repositories. The report can be filtered to limit the results to a selected journal or repository.
8. ‘Repository Report 1’ enables you to view the number of successful item downloads by month for all participating repositories. The report can be filtered to limit the results to a selected item type, Jisc Band and/or Country.

IRUS-UK is not just limited to producing statistical reports. The service can also offer metadata quality related reports, e.g. a report identifying duplicate or incorrectly catalogued DOIs within repository records. There are possibilities for metadata enrichment, by interacting with APIs offered by other services/agencies such as CrossRef, RCUK, OpenAIRE[14], etc.

IRUS data can also be accessed via a SUSHI service and a SUSHI Lite API, which is currently being developed in collaboration with the NISO SUSHI Lite Technical Report Working Group. The API allows retrieval of statistics to embed in other places (e.g. in the repository).

There are currently 87 institutional repositories participating in IRUS-UK, which has currently recorded downloads for over 400,000 different items across the participating repositories, with the total number of downloads currently at just over 40,000,000 (data taken in August 2015).

Future Development

IRUS-UK recognises that institutional repository usage statistics may not represent total usage, and is investigating ways to combine usage to support institutions in their overall usage statistics. One development is usage figures for electronic theses and dissertations; IRUS-UK has demonstrated the possibility of consolidating usage statistics from institutional repositories from the British Library's Electronic Theses Online Service (EThOS) to show total usage.

IRUS-UK is also working with and contributing to other groups and initiatives involved in a range of activities relating to usage statistics. These include: the Distributed

Usage Logging/CrossRef DOI Event Tracker Working Group, the COAR Interest Group 'Usage Data and Beyond'[15], and OpenAIRE2020.

IRUS-UK is a community-driven development, responding to user needs. User feedback has always been core to the service, and continues to develop with the establishment of a Community Advisory Group. Excellent feedback has been received from users on the ease and speed of set-up, and the value of having consistent, standards-based usage statistics and benchmarking opportunities.

Conclusion

IRUS-UK provides a usage statistics service for UK repositories, based on the COUNTER standard, which enables them to expose credible, authoritative and trustworthy usage figures for item downloads, on the same basis as – and therefore comparable with – the majority of publishers, in an extremely cost-effective manner.

By providing a nationwide view of UK repository usage, it also benefits national organizations (such as Jisc and SCONUL) and offers opportunities for benchmarking as well as the ability to act as an intermediary between UK repositories and other agencies. We hope that IRUS-UK will act as a model which can be adopted in other countries and regions around the world.

Finally, it may help to inform the current debate, taking place in the absence of reliable or comprehensive usage data, about the value of repositories and their place and significance in the dissemination of OA research literature.

NOTES

CONTRIBUTOR NOTES

Ross MacIntyre is Jisc Head of Library Analytics Services, UK.

Citations

1. Jisc, accessed May 18, 2015, http://www.jisc.ac.uk/.

2. IRUS-UK, accessed May 18, 2015, http://irus.mimas.ac.uk/.

3. “PIRUS2 Final Report”, COUNTER, accessed May 18, 2015, http://www.projectcounter.org/News/Pirus2_oct2011.pdf.

4. COUNTER Code of Practice, accessed May 18, 2015, http://www.projectcounter.org/code_practice.html.

5. Johan Bollen and Herbert Van de Sompel, “An Architecture for the Aggregation and Analysis of Scholarly Usage Data”, (arXiv:cs/0605113, Cornell University Library), accessed May 18, 2015, http://arxiv.org/abs/cs/0605113.

6. Knowledge Exchange, accessed May 18, 2015, http://www.knowledge-exchange.info/.

7. “IRUS-UK Item Type Mappings Report”, accessed May 18, 2015, http://www.irus.mimas.ac.uk/help/toolbox/IRUS_item_type_report_v3.3.pdf.

8. “IRUS-UK position statement on the treatment of robots and unusual usage”, accessed May 18, 2015, http://www.irus.mimas.ac.uk/news/IRUS-UK_position_statement_robots_and_unusual_usage_v1_0_Nov_2013.pdf.

9. “IRUS Tracker protocol”, accessed May 18, 2015, http://www.irus.mimas.ac.uk/help/toolbox/TrackerProtocol-V3-2014-04-22.pdf.

10. OAPEN-UK, accessed May 18, 2015, http://oapen-uk.jiscebooks.org/.

11. CORE, accessed May 18, 2015, http://core.ac.uk/.

12. UK Access Management Federation, accessed May 18, 2015, http://www.ukfederation.org.uk/.

13. British Library eTheses Online Service, accessed May 18, 2015, http://ethos.bl.uk/.

14. OpenAIRE, accessed May 18, 2015, http://www.openaire.eu/.

15. Confederation of Open Access Repositories Interest Group: Usage Data and Beyond, accessed May 18, 2015, https://www.coar-repositories.org/activities/repository-interoperability/usage-data-and-beyond/.