

IRUS-UK: making scholarly statistics count in UK repositories

PAUL NEEDHAM

Research & Innovation Manager

Kings Norton Library, Cranfield University

GRAHAM STONE

Information Resources Manager

University of Huddersfield

Abstract

IRUS-UK is a new national standards-based statistics aggregation service for institutional repositories in the UK. The service processes raw usage data from repositories, consolidating those data into COUNTER-compliant statistics by following the rules of the COUNTER Code of Practice - the same code adhered to by the majority of scholarly publishers. This will, for the first time, enable UK repositories to provide consistent, comparable and trustworthy usage data, as well as supporting opportunities for benchmarking at a national level. This article provides some context to development, benefits and opportunities offered by the service, an institutional repository perspective and future plans.

Introduction to IRUS-UK

Institutional repositories (IRs) have attracted much attention over the last decade and there has been considerable interest in the growing number of repositories and their contents. However, until now there has been a lack of comprehensive information about actual use of these items.

Although most IRs provide statistics that purport to show usage, you can't count on them – not entirely. Different types of software – out-of-the-box, add-ons, Google Analytics and other third-party solutions – process raw usage data in different ways, making it impossible to compare like for like across repositories. There is currently no agreed standard to measure usage across repositories.

IRUS-UK is a new national aggregation service that responds to this problem by providing standards-based statistics for all content downloaded from participating UK IRs. The service will collect usage data from participating repositories, process the data into COUNTER-compliant statistics and then present statistics back to originating repositories to be used in a variety of ways. It will provide opportunities for benchmarking at a national level by enabling UK IRs to access and share comprehensive and comparable usage data.

Eventually, IRUS-UK will provide a nationwide view of UK institutional repository use to help demonstrate the importance and value of IRs. There is also potential for the service to act as an intermediary between UK repositories and other agencies.

IRUS-UK is a wave 1 component of UK RepositoryNet+, the JISC-funded repository and infrastructure service which aims to increase the cost effectiveness of repositories of open access (OA) literature¹. The service is being developed by a Consortium involving Mimas, Cranfield University and Evidence Base at Birmingham City University. The team is also responsible for development of the Journal Usage Statistics Portal (JUSP), which provides a 'one-stop shop' for libraries to view, download and analyse their journal usage reports from multiple publishers². Consequently, the team members have significant skills and expertise in managing and developing usage statistics products and services.

Background to development of the service and PIRUS2

IRUS-UK builds on the work of the successful PIRUS2 project³, which demonstrated how COUNTER-compliant⁴ article-level usage statistics could be collected and consolidated from publishers and institutional repositories. The primary aims and objectives of PIRUS2 were to assess the feasibility of and develop the technical, organizational and economic models for the recording, reporting and consolidation of usage of journal articles hosted by publishers, institutional repositories and subject repositories⁵.

PIRUS2 achieved its aims by delivering a prototype statistics aggregation service, comprising:

- usage data and statistics from publishers and institutional repositories
- a practical organizational model based on co-operation between data processing suppliers
- data management and auditing services that meet the requirement for an independent, trusted and reliable service
- an economic model that provides a cost-effective service and a logical, transparent basis for allocating costs among the different users of the service.

PIRUS proposed the establishment of a global central clearing house (CCH) to deliver such a service. Unfortunately, it became clear from a survey conducted at the end of the project that the majority of publishers were not, largely for economic reasons, yet ready to implement or participate in such a service.

Nevertheless, this work has been used to inform the development of a COUNTER PIRUS Code of Practice, which will provide specifications for the recording and reporting of usage at the individual article level that are based on and are consistent with the main COUNTER Code of Practice for e-Resources. The first draft of the new PIRUS Code of Practice will be made available on the COUNTER website for public consultation towards the end of 2012.

Furthermore, the project found that usage of articles hosted by institutional repositories is substantial. Over the 7-10 month period of the project during which usage data was collected for articles hosted by the six participating repositories, there were over half a million downloads of 6,000+ articles; an average of 86 downloads per article.

As a result of this, a second set of aims and objectives emerged: to develop the technical, organizational and economic models for the standardized recording and reporting of usage at the individual item level – regardless of content type – for items hosted by institutional repositories and subject repositories (IRUS).

To support these extra objectives, a secondary demonstrator service was developed, which focused solely on repositories. It revealed that significant numbers of other item types (theses, conference papers, reports, etc.) were also being regularly downloaded.

This additional work ultimately led to the establishment of IRUS-UK, which will adhere to both the COUNTER and PIRUS Codes of Practice.

Currently, IRUS-UK is working with a small group of participating repositories⁶ (our 'pioneers'), but we are receiving considerable interest from other institutions and expect the number involved to significantly increase by the end of 2012.

Benefits of a shared service and community-driven developments

It is anticipated that IRUS-UK will include repositories from the majority of UK higher education (HE) institutions. By eliminating duplication of effort, IRUS-UK will offer cost and time efficiencies for participating institutions together with the benefits of a shared national service.

In theory, every institution could produce its own COUNTER-compliant statistics for its repository. The rules for eliminating robot accesses and double-clicks and for counting downloads are not that difficult to understand or implement. However, there is more to COUNTER compliance than simply following the COUNTER Code of Practice. In order to become truly COUNTER compliant, it is necessary to go through a regular auditing process. By the time registration, annual membership and report auditing fees are taken into account, this can potentially cost several thousands of pounds per year per IR.

By collecting and processing download data into COUNTER statistics on behalf of IRs, IRUS-UK can substantially reduce these costs; in this scenario, only IRUS-UK itself needs to be audited, the individual IRs do not!

IRUS-UK is about more than just collecting and processing usage data for items hosted by IRs. In order to provide a coherent and comprehensive service, we also harvest the metadata associated with those items. This creates a number of opportunities. Looking across the mass of data, we are able to identify inconsistencies and errors in cataloguing, and provide feedback about issues to participating IRs.

We are also in a position to act as an intermediary between UK IRs and other agencies, such as OpenAIRE⁷, which has an interest in obtaining usage statistics for research outputs funded under the European Seventh Framework Programme (FP7)⁸. Having a single point of access to FP7 article statistics for the UK will be a lot easier to manage than collecting those statistics from all the relevant individual repositories, so there is a clear benefit in terms of time and effort to both OpenAIRE and the UK-IRs.

IRUS-UK is not being developed by a team living in an ivory tower. Its design and development are shared endeavours, driven by the collective requirements of all its stakeholders: participating institutions, JISC, UK RepositoryNet+, SCONUL, research funders and other national and international agencies.

All this will result in a user-centric, shared service that will allow institutions to save time and money and make evidence-based decisions on management of their repositories.

IRUS from the institutional repository perspective

The University of Huddersfield is one of the IRUS-UK pioneer repositories and Graham Stone offers his perspective on the developing service below.

The University of Huddersfield has been running an EPrints-hosted institutional repository since 2006⁹. For a number of years, we were without any reliable usage statistics for items held in the repository and had to rely on Google Analytics. In recent years, we enabled the IRStats¹⁰ add-on from EPrints and have taken part in the PIRUS2 project, which successfully demonstrated that reliable usage data could be taken from EPrints repositories according to COUNTER rules.

The repository was also integral to a JISC-funded project to establish HOAP (Huddersfield Open Access Publishing)^{11, 12}, a low cost journals platform for the publication of open access (OA), peer-reviewed University Press journals. In its final report, the HOAP project recommended the development of IRUS to support both repositories and University OA publishing.

Since the launch of IRStats in early 2010, the repository has been able to provide detailed reports both to Schools and researchers on the usage of their research and has been a major factor in attracting full-text deposits to the repository. PIRUS2 allowed us to add an extra dimension to this. However, it was limited to journal articles only, which only represents around 40-50% of repository content¹³. Due to the different way IRStats and PIRUS2 statistics were compiled, we were not able to compare the two sets of data.

Huddersfield is very excited to be one of the five repositories testing IRUS-UK. The demonstrator is already giving us an 'Item Report 1' (IR1) report. In addition, we are testing the new IRStats2 module from EPrints, which is now using the PIRUS COUNTER specification and is therefore giving comparable statistics, meaning that reporting is potentially very powerful.

As an EPrints user, we are in a fortunate position in having both IRStats and IRUS up and running. However, when talking to repository managers running other software, it is apparent that they often lack reliable statistics for full-text downloads. IRUS-UK is being tested with EPrints and DSpace repositories and will therefore allow all UK repositories to collate total usage and really show the true impact of OA at a national level. Logically, there seems to be a role for the JISC JUSP in future, as they are perfectly placed to run custom reports via SUSHI reports that are already available from IRUS-UK. Commercial packages such as 360 COUNTER or UStat could also successfully import this data, meaning the repository downloads could be analysed alongside other e-resource usage data.

It will be interesting to see how IRUS-UK will be viewed by gold OA publishers, such as Biomed Central (BMC), who allow the published PDF to be deposited. In the BMC member pages¹⁴, Huddersfield has five articles listed as 'highly accessed' – but *are* they, as BMC only lists usage on its site? IRUS could potentially allow us to see the real picture by giving us total usage at article level.

There are other implications for projects such as OAPEN-UK¹⁵, who need to measure repository downloads for chapters that are available on OA. Unlike PIRUS, which looked at article-level metrics, IRUS-UK counts items and therefore includes content such as book chapters, enabling comparisons to be made with the published content¹⁶.

Finally, from an OA publishing perspective, a proper understanding of usage using IRUS will assist in the identification of 'hot topics' published in the HOAP journals, which will help to map out future

directions for those titles. Analysis of usage will also show potential return on investment for the journals, e.g. a cost per download figure could be established by measuring usage against the ongoing production costs. This type of key performance indicator (KPI) could also be applied to the content of the whole repository.

Future development plans and next steps

IRUS-UK is currently funded by JISC until March 2013. During this period, our activities will focus on transitioning from the original demonstrator, developed as part of PIRUS2, to a fully-fledged UK RepositoryNet+ service.

This will result in significant expansion throughout 2012 and into 2013, steadily increasing the number of UK repositories participating in the service. At the same time, we will continue to work on refining ingest procedures; developing the user interface in collaboration with all our stakeholders through consultation and evaluation activities; expanding reports created by the service; and developing a new application programming interface (API) to support the machine-to-machine retrieval of statistics which can then be embedded directly in repository and other user interfaces, or used by other services and agencies as they see fit.

With the long-term future in mind, development of a sustainability plan to ensure continuation of the service post-project funding will also be a key activity.

Conclusion

IRUS-UK will provide a usage statistics service for UK repositories, based on the COUNTER standard, which will enable 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 will also benefit organizations such as JISC and SCOUNL, and will offer opportunities for benchmarking as well as the ability to act as an intermediary between UK repositories and other agencies.

We hope, also, 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.

References

1. UKRepositoryNet+:
<http://www.repositorynet.ac.uk/project-summary/> (accessed 28 September 2012).
2. JUSP:
<https://www.jusp.mimas.ac.uk/> (accessed 28 September 2012).
3. PIRUS2:
<http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/> (accessed 28 September 2012).
4. COUNTER Code of Practice:
http://www.projectcounter.org/code_practice.html (accessed 28 September 2012).

5. PIRUS2 Final Report:
http://www.projectcounter.org/News/Pirus2_oct2011.pdf (accessed 28 September 2012).
6. IRUS Participants:
<http://www.irus.mimas.ac.uk/participants/> (accessed 28 September 2012).
7. OpenAIRE:
<http://www.openaire.eu/> (accessed 28 September 2012).
8. Seventh Framework Programme (FP7):
http://cordis.europa.eu/fp7/home_en.html (accessed 28 September 2012).
9. University of Huddersfield Repository:
<http://eprints.hud.ac.uk/> (accessed 28 September 2012).
10. IRStats:
<http://wiki.eprints.org/w/IRStats> (accessed 28 September 2012).
11. Stone, G, White, S, Robinson, D, Pitchford, I and Edmunds, C, *Huddersfield Open Access Publishing final report*, University of Huddersfield, Project Report, 2012, Huddersfield.
<http://eprints.hud.ac.uk/13278/>
(accessed 28 September 2012).
12. Stone, G, Huddersfield Open Access Publishing. In: *Social Shaping of Digital Publishing: Exploring the Interplay Between Culture and Technology - Proceedings of the 16th International Conference on Electronic Publishing*, 2012, Amsterdam, IOS Press, p.118.
<http://eprints.hud.ac.uk/13909/>
(accessed 28 September 2012).
13. Content Types in OpenDOAR Repositories – Worldwide:
<http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&IID=&potID=&rSoftWareName=&search=&groupby=ct.ctDefinition&orderby=Tally%20DESC&charttype=bar&width=600&caption=Content Types in OpenDOAR Repositories - Worldwide>
(accessed 28 September 2012).
14. Biomed Central member: University of Huddersfield:
<http://www.biomedcentral.com/inst/4651>
(accessed 28 September 2012).
15. OAPEN-UK:
<http://oapen-uk.jiscebooks.org/> (accessed 28 September 2012).
16. Milloy, C, Stone, G and Collins, E, OAPEN-UK: an Open Access Business Model for Scholarly Monographs in the Humanities and Social Sciences. In: *Social Shaping of Digital Publishing: Exploring the Interplay Between Culture and Technology - Proceedings of the 16th International Conference on Electronic Publishing*. 2012, Amsterdam, IOS Press, p. 122.
<http://eprints.hud.ac.uk/13912/>
(accessed 28 September 2012).

Article © Paul Needham and Graham Stone

Corresponding author: Paul Needham, Research & Innovation Manager, Kings Norton Library

Cranfield University, 19 Albion Street, Swindon, Wiltshire, SN1 5LJ, UK

Tel: +44 (0)1793 511358 | E-mail: p.a.needham@cranfield.ac.uk