

Learned Publishing, 24:109–114
doi:10.1087/20110205

CASE STUDY

JUSP in time: a partnership approach to developing a journal usage statistics portal



Chris Cradock



Paul Meehan



Paul Needham

Introduction

At a time of economic constraint, the use of COUNTER¹ (Counting On-line Usage of Networked Electronic Resources)-compliant data is essential in making a compelling case about the value of journal subscriptions. The Journal Usage Statistics Portal (JUSP)² is a new service, built in response to demand from UK higher-education libraries. It enables librarians to access their usage data quickly and efficiently, and supports publishers in providing access to statistics for their customers. Essentially JUSP provides a central point where higher-education libraries can access usage statistics reports for the NESLi2³ publisher deals to which they subscribe.

The portal is being developed by a consortium involving JISC Collections,⁴ Mimas⁵ at the University of Manchester, Evidence Base⁶ at Birmingham City University and Cranfield University.⁷

Background

In 2008, a funding call from JISC proposed the technical design and prototyping of a usage statistics portal. Over the next year, we worked with a sample group comprising five libraries, three publishers, and an intermediary; this prototype demonstrated that such a portal would be feasible, and led to subsequent funding from the JISC Information Environment Programme to develop the prototype into a fully functional service for the UK higher education community. Service development commenced in April 2010, and is due to be completed in December 2011, at which point we hope to include all NESLi2 publishers and approximately 150 UK higher-education institutions.

Benefits for publishers and libraries

JUSP provides a number of benefits for

Chris CRADOCK^a, Paul MEEHAN^b and Paul NEEDHAM^c

^aOxford University Press

^bMimas

^cCranfield University

ABSTRACT. *The Journal Usage Statistics Portal (JUSP) service is expanding rapidly, and aims to encompass all NESLi2-participating publishers by the end of 2011, together with all UK higher-education institutions subscribing to deals from those publishers. A key part of the success of JUSP will be the implementation of the Standardized Usage Statistics Harvesting Initiative (SUSHI) protocol for automatic harvesting of usage statistics directly from publishers. This article outlines development of the service and describes how JUSP is working with Oxford University Press to deliver usage statistics to libraries. We also describe future developments at JUSP, including the creation of our own SUSHI server, and explain how new initiatives and enhancements are developed and rolled out in conjunction with our participating institutions.*

© Chris Cradock, Paul Meehan and Paul Needham 2011

the ability to view aggregated usage statistics from publishers and intermediaries provides a more complete picture of journal usage

libraries and publishers. For libraries, obtaining and analysing COUNTER-compliant data can be extremely labour intensive. JUSP saves time in that it avoids the need to visit multiple publisher websites to view usage statistics, and allows libraries quickly and easily to compare usage across various publishers and years via the Web portal. JUSP also brings a range of benefits to publishers. The portal makes the delivery and analysis of usage statistics more efficient for your customers, whilst the ability to view aggregated usage statistics from publishers and intermediaries provides a more complete picture of journal usage. JUSP also removes the need for libraries to maintain separate passwords to visit each publisher website to download their data, and thus provides greater efficiency.

Publishers and SUSHI

By the end of January 2010, eight publishers and intermediaries were working with JUSP; these included Elsevier, Oxford University Press (OUP), Springer, The John Hopkins University Press for Project Muse, the American Institute of Physics (AIP), Nature Publishing Group, Publishing Technology (ingentaconnect), and Swets. These publishers have recognized the benefits that the portal can offer their customers, particularly with regard to an automated, streamlined, and effective gathering of monthly statistics into one central place.

The implementation of SUSHI (Standardized Usage Statistics Harvesting Initiative)⁸ has the potential to provide great benefits to institutions, and is a significant part of JUSP. The SUSHI protocol is a machine-to-machine way of gathering COUNTER-compliant usage statistics; this removes the need for institutions to login to publisher websites to download reports, and also provides for the automated monthly gathering of data. Participating sites thus have no further need to gather statistics, as this will be done directly by JUSP. COUNTER compliance and the automated monthly harvesting of publisher statistics are fundamental to JUSP, and we are hopeful that all publishers will adopt this in the near future.

JUSP has developed SUSHI clients for a

number of publishers; much of this work is pioneering, as some publishers did not have SUSHI repositories, and others had little experience of working with a third party acting on behalf of a range of institutions. We have implemented SUSHI clients for OUP, Springer, Publishing Technology (ingentaconnect), AIP and Project Muse and are actively working with others. The work carried out with OUP on setting up a SUSHI client and setting JUSP up to act as a third party 'agent' forms the basis of this case study.

JUSP and SUSHI

Developing a SUSHI client should be a breeze – right? After all, SUSHI is a specified standard, and the NISO SUSHI website offers pointers to documentation, example client software, software development toolkits, FAQs, and a checklist of things to consider when developing a client.⁹ However, in reality, we found life a little more difficult than we had first hoped.

The example clients and toolkits were either written under the Microsoft .NET Framework and designed to work in the Windows environment, or used Java (a cross-platform but complex programming language), whereas what we were really hoping for was a simple client using the Perl – the programming language most familiar to us – that would work on the JUSP Unix server and which we could adapt to our needs with relative ease.

Furthermore, our early research and investigations soon revealed that, although SUSHI is a specified standard, there are in practice variations in its implementation among publishers, both in the use of the required report request parameters and in authentication methods.

We therefore came to the conclusion that we would have to develop our own Perl client from scratch, and that a pragmatic approach would be to begin by developing a client that would work with a single publisher for a single institution.

That publisher was Oxford Journals and the institution was Cranfield University – where the JUSP test server is located. Oxford Journals requires only a couple of

mandatory fields in the SUSHI report request: the Customer Reference ID, which is the same as the username used to login to the Oxford Journals usage statistics system, and the Report type being requested – JR1, JR1a, etc. The contents of all other fields are optional. Furthermore, their authentication requirements are straightforward: all that is necessary is to register the SUSHI client's public IP address using the 'SUSHI Admin' tab, available on their usage statistics login page.¹⁰

Armed with the Cranfield username, and with IP authentication in place, we began to put a Perl client together using a combination of the NISO SUSHI documentation and code snippets gleaned from the Web, until we had a first, best-guess client. With hope, rather than expectation, we ran it. Almost inevitably, we received an error message back. The next couple of weeks involved a seemingly endless round of coding tweaks and changes in response to each new error message as we re-ran the client to see what happened, until . . . success! The JUSP client was talking to the Oxford Journals server and returning our first SUSHI JR1 report.

After that, the next steps were relatively straightforward. Oxford Journals added JUSP as an available third-party agent to its SUSHI Admin page, institutions participating in JUSP supplied us with their Customer Reference IDs, and we were finally in a position where we could harvest SUSHI usage statistics on their behalf.

Of course, that was just one publisher. Over the ensuing months, we have continued to develop new clients, one per publisher, and currently have more than half a dozen working SUSHI clients. Not surprisingly, as we become more familiar with working with the SUSHI protocol, with one or two exceptions, the development process has tended to become easier and easier. Moreover, as our SUSHI expertise has grown, it has become apparent to us that we will be able to reassemble our various clients into one generic client able to handle the majority of publishers, driven by the various report request parameters and authentication requirements stored in a database. It is our intention to develop this generic client

over the coming months and to make it available as free, open-source software by the end of the JUSP project.

Oxford Journals and SUSHI

Background

As part of OUP, Oxford Journals has provided Web usage statistics for around ten years, the last six following COUNTER specification. As the volume of data has increased we undertook an exercise to rework the infrastructure that converts the Web logs into usage statistics. This new infrastructure came online early in 2010, and offered statistics from January 2009 onwards.

In order to be fully COUNTER compliant, Oxford Journals offers an automated interface to collect statistics using a SOAP (originally defined as Simple Object Access Protocol) server following SUSHI specification ('the SUSHI service'). This interface was built as an integral part of the same Web statistics infrastructure. Indeed, the queries used against the database to extract the data on each report in the SUSHI service are the same as the client-driven interface, each of which produce the COUNTER-compliant reports.

A key aspect of statistics gathering that was not included in the SUSHI specification, and which was essential for Oxford Journals to put in place, was control over which parties could access our server and retrieve usage statistics.

Authentication and access regulation

SUSHI specification includes a field to identify the data requester. However this is an open field, with no mandatory supporting authentication, thus opening up the possibility of someone acquiring a valid requester ID and connecting to our server without any controlled authentication. This may be acceptable on an internal network, but is not acceptable to us when considering offering the service globally. For a small publisher one could contemplate limiting access to the SUSHI server on a case-by-case basis, but Oxford Journals deals with a large number of customers, and regulating them cost-effec-

the development process has tended to become easier and easier

tively therefore means asking the customers to authenticate themselves.

As SUSHI is intended to function in an automated fashion, we are not in a position to change the SUSHI standard, or to introduce any additional connection mechanics required before the SUSHI request is transmitted. One possible solution would be to place the SOAP server behind an HTTPS Secure Sockets Layer/Transport Layer Security (SSL/TLS) secure service, and have this service demand an X509 digital certificate from the customer. However this would add an extra complication for customers who would need to arrange client X509 certificates with each publisher. We therefore decided not to impose this on our customers.

Having discounted this possibility, we are left with a need to verify access permission transparently from the information gleaned from the requesting server. The way we achieve this is via IP address authentication. We require the customer to register the IP address of the machine that will send the SUSHI requests to our server. In most cases, the IP address will not be that of the individual customers themselves, as we generally expect customers to delegate statistics collection to computers that were not necessarily their own; for example a subject librarian may be a registered user, but the requests will be sent from a computer centre rather than that user's own computer.

Similarly, we anticipated receiving requests from third-party organisations such as JUSP. To aid our customers we allow third parties to tell us they are acting in this manner. A customer is subsequently provided with a list of any available third-party agents, and can select one or more directly through their own administration interface; thus each user can choose who will collect the usage stats on their behalf. The agent or third party maintains the appropriate IP address(es) of the requesting machines independently.

Oxford Journals Server and Software

At Oxford Journals, Web statistics are built, managed, and queried using Marketwave's Pilot Hitlist software.¹¹ This is a Windows-based service which stores and queries statis-

tical data using a SQL server. All SUSHI requests are directed initially at our Apache webserver: this server has been programmed to collect the SUSHI request, and inspect it to determine the customer ID. Using this customer ID, we can then determine if the IP address from which the request originates has been listed by the customer or belongs to an agent it has specified. If the address does not match customer records, the request is terminated with an authentication failure known as a SOAP exception. If the authentication succeeds, the Apache server replays the SOAP request to our Hitlist service, and its response is queued and replayed to the calling customer's program.

One advantage to the replay arrangement is that our Hitlist server is not subject to delays in waiting for the response to be transmitted back to the customer across the Internet, and hence reduces the period that the Hitlist server is tied up per request. As we host of the order of 250 journals, and some of our consortia represent thousands of individual organisations, some reports extractable from SUSHI can be extremely large (hundreds of megabytes), potentially taking hours to send across the Internet. Our dual-server arrangement allows requests to be processed much more quickly as the single server is not tied up. This process is transparent to the requesting server.

Oxford Journals and JUSP

The excellent working relationship and practices established with Oxford Journals will provide a basis for future work between JUSP and other publishers. Oxford Journals has also found considerable benefits from the relationship.

We are excited by the aggregated content concept that the initiative aims to deliver. Usage stats are becoming an increasingly important tool for librarians in judging the value of subscribed content and OUP is committed to making such statistics easily accessible in order to benefit both library and publisher understandings of journals usage. (Colin Meddings, Senior Library Marketing Manager at Oxford University Press)

Integrating customer feedback into the development cycle

The JUSP portal is under continual development, and is version controlled, with fixed releases at regular intervals. The team use 'agile' software development processes, built on the foundation of iterative and incremental development. We aim to deliver customer satisfaction by rapid, continuous delivery of useful software.

A crucial part of the development cycle is gathering user feedback and implementing suggestions. In November and early December 2010 we conducted a user survey of participating institutions, asking them to provide feedback on various portal aspects, such as analysing individual reports, pointing out particular strengths and perceived weaknesses, and suggesting further improvements, additional reports and other enhancements.

Survey feedback was almost unanimously positive: 16 of the 22 institutions participating at the end of November 2010 responded, with almost everyone praising the portal for its clear interface and ease of use, and SUSHI automation was highlighted by several as one of its most successful aspects. Indeed, the main limitation perceived by the respondents was that JUSP is currently restricted to NESLi2 publishers and hence not all institutional usage can be incorporated – this may change at a later date. Following the user survey, a number of 'quick wins' were implemented; these included an ability to search by journal title keyword, and improved sort and export facilities.

JUSP developments rely on a continued dialogue with publishers and intermediaries to ensure development of this valuable resource. We collect and respond to ongoing feedback via a dedicated mailing list and a publicly visible helpdesk address: jusp@mimas.ac.uk. We liaise closely with our participating sites and a number of portal developments are being investigated in conjunction with individuals at these sites.

Future development plans

In the short term we envisage considerable

expansion of JUSP. Institutions will be added to the portal at the rate of ten per month throughout 2011, with all higher-education libraries taking NESLi2 deals ideally integrated by December. We have a waiting list of interested parties; all prospective sites can currently login to the portal and view dummy usage data. We are at varying stages of negotiation with the remaining NESLi2 publishers. We will not include publishers who do not provide COUNTER-compliant statistics or a SUSHI server; COUNTER compliance is a requirement in the NESLi2 licence, but implementation is at differing stages of completeness.

As new sites join, we will collect their monthly data routinely via SUSHI; we are actively working on SUSHI clients for the other publishers, with several of these ready to be rolled out. We will collect a limited amount of back data for each site, though we may gather additional years on request, subject to time constraints. We have also developed and are currently testing a JUSP SUSHI server. This enables institutions to access, gather, and download their usage data from JUSP via SUSHI rather than having to visit numerous publisher sites or develop numerous SUSHI clients themselves.

We continue to work on development of portal functionality, with a 'wish list' of new features to implement. These include a way to incorporate publisher price lists into reports – these are not generally available in a suitable format, so this will require discussions with the publishers. We are also keen to implement subject categories alongside journal titles – a key feature for libraries that wish to view usage of particular sets of titles or use by a particular demographic.

At the request of participating libraries, JUSP will also enable benchmarking. Each of your customers will be able to compare their own usage with that of other anonymized institutions in the same JISC band, or with a similar full-time equivalent number of students and academic staff. They will use this information to plan developments in their library systems and internal marketing that will aid resource discovery and increase usage. We do not have plans to implement a 'cost per download' model, however.

a crucial part of the development cycle is gathering user feedback and implementing suggestions

JUSP will also enable benchmarking

Finally, a key function of the portal will be to provide a basic analysis of usage for individual libraries, saving them the time and effort currently spent in this area. We hope that this will eventually inform decisions about purchasing, use of specific titles or deals, and aid annual statistical returns to the Society of College, National and University Libraries (SCONUL) and institutional reports. Further support will be made available through online tutorials, help guides and institutional visits where appropriate.

Summary

JUSP is working very closely with Oxford Journals and a range of other publishers to create a dynamic statistics portal resource, which is already being widely used in the academic community. A fundamental aspect of the service is the automatic gathering and processing of usage data directly from publishers, and to this end the development and implementation of a range of SUSHI clients is key.

The internal procedures and security in place for the Oxford Journals SUSHI service is similar to that established at many other publishers, although small differences remain. Our work with OUP has allowed us to develop software suitable for our needs and those of our institutions. The benefits are clear to both publishers and participating academic institutions – simplified gathering of data, automatic monthly ingests and updates, quality control and data checking, and a smoother and more transparent dia-

logue between libraries and publishers, facilitated by the JUSP team.

As JUSP continues to develop and expand throughout 2011, we expect that lessons learned from working with OUP can be successfully applied to our work with the remaining NESLi2 publishers – and hopefully wider as the service grows in future years.

For all questions or enquiries about JUSP, please send an email to: jusp@mimas.ac.uk.

References

1. <http://www.projectcounter.org/>, accessed 27 Jan 2011.
2. <http://jusp.mimas.ac.uk/>, accessed 27 Jan 2011.
3. <http://www.jisc-collections.ac.uk/nesli2/>, accessed 27 Jan 2011.
4. <http://www.jisc-collections.ac.uk/>, accessed 27 Jan 2011.
5. <http://mimas.ac.uk/>, accessed 27 Jan 2011.
6. <http://www.ebase.bcu.ac.uk/>, accessed 27 Jan 2011.
7. <http://www.cranfield.ac.uk/>, accessed 27 Jan 2011.
8. <http://www.niso.org/workrooms/sushi/>, accessed 27 Jan 2011.
9. <http://www.niso.org/workrooms/sushi/start/clients>, accessed 27 Jan 2011
10. <http://reports.oxfordjournals.org/pilot/hitlist/>, accessed 27 Jan 2011.
11. <http://www.marketwave.com/index.php/Product/producttext.html>, accessed 27 Jan 2011.

Chris CRADOCK

Oxford University Press

Email: christopher.cradock@oup.com

Paul MEEHAN

Mimas

Email: paul.meehan@manchester.ac.uk

Paul NEEDHAM

Cranfield University

Email: paul.needham11@btinternet.com

a fundamental aspect of the service is the automatic gathering and processing of usage data directly from publishers