

Usage Statistics Portal Scoping Study Phase 2: Phase ii Technical Design and Prototyping

**Summary report
January 2010**

**Evidence Base
Research and Evaluation Services
Birmingham City University**

And

**MIMAS
University of Manchester**

with

Cranfield University

Report authors:

**Dr Angela Conyers
Pete Dalton
Ross MacIntyre
Paul Needham**



Acknowledgments

The project team would like to thank the libraries and publishers who took part in the project as well as the support of the Journals Working Group.

Disclaimer

The study team is not liable for the accuracy of any information gathered to compile this document. All data collected is to be construed as contributions towards meeting the aims of the study. The study team accepts no liability for errors or omissions in this document and accepts no responsibility for loss or injury which may occur as a result of reliance placed on any part of its contents.

Contents

1. Introduction	6
2. Approach and achievements.....	7
2.1 Libraries.....	7
2.2. Content providers	7
2.3 Setting up the database.....	8
2.4 User requirement development	9
2.5 User requirements - achievements.....	10
3. Issues raised.....	15
4. Next steps.....	20

Executive summary

Evidence Base and Mimas were commissioned by JISC Collections to produce a technical design and prototype for a JISC Usage Statistics Portal to assist and support libraries in the analysis of NESLi2 usage statistics and the management of their e-journals collections.

A prototype has been developed, taking in usage data (JR1 and JR1A reports) from five libraries in respect of three NESLi2 publisher deals.

A simple database and data processing code has been developed to verify and load the usage data files obtained via institutions. A number of scripts have been developed, presented in 'dashboard' style user interface to query the database.

At a basic level, we have demonstrated that the portal can provide a basic 'one-stop shop' where libraries can go to view and download their own usage reports from NESLi2 publishers. This in itself will be welcomed by libraries who currently have to go into each publisher's passworded site separately. They will also get a truer picture of usage from the addition where necessary of gateway/aggregator statistics, which some libraries use as their main route to e-journal articles. This has been a major achievement of the project. As well as looking at usage through gateways/aggregators, libraries are able to view their use of titles in a current deal separately from use of titles in a backfile or archive collection. This is another feature important to libraries. It will also be possible for libraries to compare their usage of different NESLi2 publisher deals, to look at trends over time and to complete their annual SCONUL return in respect of usage of titles from NESLi2 publishers.

This level of service alone responds to many of the essential requirements set for this project, and includes the desirable requirement of adding gateways/aggregators. It provides a response to many of the issues Evidence Base has encountered in its work with libraries and which are frequently the subject of debate on mailing lists.

The prototype has also considered a number of 'added value' services that a usage statistics portal could offer:

- It has demonstrated how those titles included in usage statistics but **not currently available within the deal** could be identified. This will help particularly in exploring the reasons for nil use of certain titles and will help track journals that change publisher.
- It has also explored how **prices** could be added to journal titles to allow libraries to compare usage of high and low priced titles, and to calculate the yield value of the deal.
- Adding **subject categories** has also been considered, to help libraries analyse deals by department or faculty, especially useful for those who do not have central budgeting.

It was apparent that while these added value services could be offered, much depended on how individual publishers presented the information. Further exploration is needed of how other developments, eg ONIX, British Library subject classification of journals can provide this information in a standard format. The portal then has the potential to hold this additional information against journal titles within its database and greatly increase its usefulness to the community.

There is potential and value in the development of a Journal Name authority file, assisting in the identification of titles, (e-)ISSN, etc and tracking journal name changes over time. However the first candidate, using a CrossRef file proved problematic.

The prototype has also demonstrated that it will be possible to provide for JISC Collections and its negotiating agents regular reports on usage of the NESLi2 deals. This will be a very useful service in future negotiations, and to illustrate the value of the NESLi2 deals to the JISC community. Libraries themselves are also very interested in being able to benchmark their usage against others. The extent to which such benchmarking can be delivered will depend on JISC Collections' negotiations with publishers and libraries, but the portal will be able to offer whatever methods (JISC bands, type of university etc) are agreed.

The project has not been able to test the SUSHI protocol, though has kept in close touch with SUSHI developments and publishers' plans. To date, the three publishers in the project Elsevier, Springer and OUP have not implemented SUSHI in production, though, as it is a requirement for COUNTER release 3 from 31 August 2009, they are expected to do so shortly. The study team have made a number of recommendations as issues have arisen during the project. They are confident that a case can be made for continuing the work which the prototype has started on, adding to the number of publishers and libraries, exploring SUSHI and the addition of more features to the portal.

Future development of the usage portal

With the SUSHI protocol now a COUNTER requirement, there is no doubt that a NESLi2 usage statistics portal can be developed to build on the prototype and provide a range of services to the HE community. Once developed, there is further potential to develop the portal to include non-NESLi2 journal publishers, databases and e-books, and to offer the service beyond the UK HE community, in a way that is responsive to changing needs. The feedback gathered to date on the portal has suggested that there is strong support for it to be developed further to benefit JISC and the community. For example one publisher commented:

“This is indeed really a milestone on its way. Congratulations with the project so far.”

Whilst the focus of the project has been on exploring the feasibility of developing a usage statistics portal, the study team were keen to gather comments from participants as well as the JISC Journals Working Group and SCONUL. At the point of writing, feedback from project participants has been extremely positive as the following demonstrates:

“It's exactly what all stats librarians need and a very positive step in the right direction for data collection.”

1. Introduction

In September 2008 JISC Collections commissioned Evidence Base and Mimas to undertake a project to develop a prototype NESLi2 Usage statistics portal. The need for such a portal had been identified in previous work undertaken by Evidence Base as well as a phase 1 NESLi2 Usage Portal Scoping Study carried out by Content Complete.

Based on an initial set of user requirements the study team set out to refine the user requirements and develop a small scale prototype working with data from three NESLi2 publishers and five information services in UK HEIs.

The project set out to:

- identify what would be feasible to include in an automated usage portal
- identify issues that needed addressing in the development of such a portal
- get feedback from participating libraries on the portal developed
- provide suggestions about what would be required to scale the portal into a full service.

The project was undertaken by a team from Evidence Base Research and Evaluation Services at Birmingham City University and Mimas at the University of Manchester. Work was undertaken between September 2008 and September 2009 following an extension to the original timescale. Following staff illness, Cranfield University was sub-contracted to progress the technical development aspects of the project. The core team comprised: Dr Angela Conyers (Evidence Base), Pete Dalton (Evidence Base), Ross McIntyre (Mimas), Sean Dunne (Mimas), Paul Needham (Cranfield University). Pete Dalton at Evidence Base was the project leader. Involved initially were Andrew Weeks and Lisa Charnock (Mimas).

2. Approach and achievements

The following section describes the approach taken and what has been achieved by the project.

2.1 Libraries

The five libraries that participated in the project are presented in Table 1 below:

Library	JISC band
University of Birmingham	A
Cranfield University	E
University of Glasgow	A
University of Liverpool	B
University of Westminster	C

Table 1. Participating libraries

For speed and consistency, JR1 and JR1A reports for 2007 and 2008 from the three participating publishers were obtained directly from the participating libraries. They provided details of the particular deals they had, and comments on their use of gateways and host services. As a result of their comments, and the team's prior experience of working with libraries in analysing usage statistics, it was decided at an early stage that the portal should allow for the addition of gateway and host statistics.

Separate usernames and passwords were set up for each of the participating libraries, so that they could view only data from their own institution. These were supplied to the libraries in July 2009, so that they could view and comment on their own data following the presentation to the Journals Working Group on 16th July 2009.

2.2. Content providers

Three NESLi2 publishers (Elsevier, OUP and Springer) were approached and agreed to take part in the project. Publishing Technology (previously Ingenta) were also approached and agreed to take part. All publishers agreed to our obtaining usage data directly from the participating libraries, though one also supplied reports directly.

Implementation of SUSHI protocol

An approach was made to the contacts provided by each of the publishers, asking them in particular how they planned to implement SUSHI. All were aware that they needed to be SUSHI compatible by the end of August 2009 in order to comply with COUNTER release 3, and were at various stages of development during the project. Because of this, it proved impractical to test obtaining publisher usage reports via SUSHI during the project and as SUSHI had been identified as the only way that publishers could readily supply reports to the portal, other means were not investigated.

Electronic Resource Management (ERM)

In February 2009 a conference call was held with senior staff at ExLibris: UK Managing Director (Robert Bley), US-based Verde Product Manager (Nettie Legace) and external consultant (Jenny Walker) and they expressed enthusiasm for the project and confirmed their willingness to be involved as appropriate in any follow on activities to test how data from the portal could be sent to it.

2.3 Setting up the database

The database was set up in a way that was aimed at meeting the following essential requirements:

- ***will be NISO Standardized Usage Statistics Harvesting Initiative (SUSHI) compatible. This will enable the system to query publishers' sites and be authorised to collect data relating to UK universities that are subscribers to individual publishers' titles.***

The study team has kept in touch with SUSHI developments, however due to the timing of Counter Sushi compliance for publishers being mandatory from 31st August 2009 it has only been possible to test SUSHI ingest on a very small scale. It will be essential to test SUSHI compatibility if the prototype is to be developed further.

Though not part of this project, Mimas have recently developed support for SUSHI for MDL' CrossFire service usage statistics. This was a straight-forward piece of code development, completed within 2 days. For CrossFire the DB3 COUNTER XML Reports are approx. 14 times the size of the corresponding csv files (7 times for DB1).

- ***will provide usage statistics that are deliverable to other vendors for integration with electronic management systems and databases***

A small scale test was planned to investigate incorporating Cranfield's SUSHI XML data from the portal into Ex Libris' Verde. This has not been possible within the timescales of the project.

- ***will adopt open source software standards and compliance with the JISC Information Environment standards¹***

No proprietary software was used in the development of the portal. It has been built as a 'standard' web service, using MySQL database and PHP scripting. Authentication has been implemented as username and password, though some discussion has taken place regarding UK Access Management Federation, refer to Section 3 'Authentication'.

We have presented data in both HTML for onscreen viewing and CSV format as an open standard for downloading. We investigated also offering XLS format, but decided against this because of time and versioning issues.

COUNTER XML format will be investigated in connection with SUSHI.

¹ <http://www.jisc.ac.uk/fundingopportunities/projectmanagement/working/standards.aspx>

2.4 User requirement development

The following describes the user requirements which informed the development of the portal. The user requirements were iterated over two stages which are described below.

Stage 1

An initial set of ten requirements were drawn up and are listed below. These were designed to meet the following essential and desirable requirements set out in the ITT:

Essential

- *cover all of the NESLi2 usage deals*
- *provide access to the usage data and reports to JISC Collections, the NESLi2 negotiating agent and NESLi2 participating university libraries*
- *provide a benchmarking facility so that individual libraries could benchmark against other libraries in the same JISC band (anonymised).*
- *provide bi annual benchmarking reports to JISC collections*
- *provide interactive functionality so that libraries can obtain reports when required and tailor them to their needs as well as provide graphical representation of usage statistics.*
- *provide JISC Collections and its NESLi2 managing agents with monthly consortia usage reports breaking down usage on a monthly basis by individual publisher.*
- *will be Counter Compliant with JR1 reporting standards (JR2 will be discussed with JISC further to ascertain its relevance)*

Desirable

- *Be able to account for usage via subscription agent gateways and aggregators and to combine such statistics where necessary with the usage from publisher platforms.*

The ten initial user requirements designed to meet the above essential and desirable requirements are given in summary form below. These formed the basis on which the portal was developed.

1. Single point of access to all JR1 and JR1A usage statistics as currently downloaded individually from publisher websites
2. Addition of aggregator/gateway JR1 statistics where relevant
3. Excluding usage of backfile collections
4. Summary tables to show trends over time, compare publishers etc

5. Summary table to show use of aggregators
6. Summary table to show use of backfiles
7. Some more figures (average, median and maximum number of requests)
8. What titles have the highest use?
9. Tables and graphs (monthly or annual usage over time)
10. Benchmarking (comparing usage by JISC band)

Stage 2

As the project developed, a further 10 user requirements were drawn up. These were intended to explore the 'added value' elements included in the team's proposal and covered three areas: adding price information; adding deal information; and providing a download area for libraries to add their own data (costs, subscribed titles etc) and do further analysis. These requirements were:

11. Getting price information for journals
12. Adding price information for journal lists
13. What titles are in the deal?
14. Adding deal information to journal lists
15. Showing usage/non-usage of titles listed in the deal and titles not listed
16. Summary table showing usage/non-usage of titles listed in the deal and not listed in the deal.
17. Summary table showing average and median use of titles listed in the deal and titles not listed
18. Download area 1. Cost per request
19. Download area 2. Usage of subscribed titles
20. Download area 3. Charts and graphs

2.5 User requirements - achievements

More detail is given below on progress made against each of the user requirements identified in the ITT.

Essential requirements

Cover all of the NESLi2 usage deals

The prototype was designed to work with three NESLi2 publisher deals and a selection of aggregator data and demonstrated that providing a single point of access

for NESLi2 usage deals was possible. This represents a considerable advance for libraries, who currently have to go to separate passworded sites to download usage statistics. It also means that usage statistics can be retained in the portal for an indefinite period to allow for later analysis of trends. Usage reports on publisher web sites are available only for a limited period, generally up to two calendar years.

Provide access to the usage data and reports to JISC Collections, the NESLi2 negotiating agent and NESLi2 participating university libraries

The prototype has through the use of database queries and password authentication enabled full access to the data to JISC Collections and the NESLi2 Negotiating Agent and for individual libraries access to their own data.

The monthly usage data is presented both by calendar year and by academic year (August-July). Libraries can also download monthly statistics to fit their own subscription patterns. The academic year is particularly useful for libraries when completing their SCONUL return for the question 'number of successful full-text article requests' as it will provide in one place a total for all article requests from NESLi2 publishers. A special table has been included in the prototype portal for this purpose.

Provide a benchmarking facility so that individual libraries could benchmark against other libraries in the same JISC band (anonymised).

The prototype has demonstrated that anonymised benchmarking by JISC Band is achievable. In addition, technically it has been shown that benchmarking by any conceivable criteria is also possible, but this will depend on the agreement of the publishers and the libraries.

Provide bi annual benchmarking reports to JISC collections

The prototype has demonstrated that benchmarking reports can be provided for any time period required.

Provide interactive functionality so that libraries can obtain reports when required and tailor them to their needs as well as provide graphical representation of usage statistics.

The prototype has demonstrated that graphical representation of data can be produced and has provided some examples of this. This data can be exported by JISC and libraries into other packages. Data can be downloaded as CSV or HTML for use locally for analysis and manipulation.

The download area (user requirements 18-20) represents an attempt to build on this initial functionality and provide a means for libraries to download data from the portal and combine it with their own data (eg costs and details of subscribed titles). This has the potential to allow libraries to do simple analysis eg cost per request compared across deals and for those who wish to do further analysis, the potential to develop further tables, charts and graphs relating to their own usage.

Provide JISC Collections and its NESLi2 managing agents with monthly consortia usage reports breaking down usage on a monthly basis by individual publisher.

The portal has demonstrated that it is possible to deliver this.

Will be Counter Compliant with JR1 reporting standards (JR2 will be discussed with JISC further to ascertain its relevance)

The portal is compliant with Counter JR1 and JR1A reporting requirements. JR2 turnaways has not been included in the portal as a result of discussions with libraries and JISC who considered this to not be an important measure.

By including JR1A the portal has been able to demonstrate how a report can be produced on current usage by subtracting JR1A usage from JR1. This is a requirement that was considered important by the participating libraries, who currently have to do their own subtraction after downloading reports to look separately at their usage of backfiles and current deals.

JR1A was an optional report for COUNTER release 2 and some publishers did not supply information on backfile use in the COUNTER compliant form. From Release 3 (31 August 2009) publishers are required to report on backfile usage using either the JR1A report or the new JR5 report (number of successful article requests by year and journal). It will therefore be necessary also for the portal in future to be COUNTER compliant with JR5 in order to ensure all backfile use can be recorded.

Desirable requirements

Be able to account for usage via subscription agent gateways and aggregators and to combine such statistics where necessary with the usage from publisher platforms.

The portal has demonstrated that such data sets can be combined where appropriate. It is considered that this is essential in order to provide a more accurate picture of usage especially for those libraries that use these services as a main source.

To do this, it was necessary to draw up business rules for inclusion, to establish for which publishers gateway and aggregator JR1 usage statistics had to be added and for which publishers these were already included. This is a major achievement of the project and one that is welcomed by libraries who are often unaware of the need to add gateway and aggregator statistics. .

Additional value added elements

Be able to provide data in a format that libraries could use locally to enhance and undertake additional analysis and data manipulation.

A number of possible additional data sets were suggested and explored. These included:

Linking usage statistics to publisher list prices of journals in order to demonstrate value for money.

The prototype explored adding publisher list prices to journal titles. Due to inconsistencies in the ways that price list data is presented by the publishers it did not prove possible to incorporate price list data in the prototype. There were inconsistencies in the way the price lists were presented, some being in PDF, others in Excel.

It is possible that the ONIX Serials–Subscription, Products and Subscriptions (SPS)² will provide a solution here, and this needs further exploration. EDItEUR have recently published a dedicated serials price message on behalf of ICEDIS. This message is a member of the ONIX for Serials family and utilises the same constructs, vocabulary and code lists. Wiley are already offering this format to agents (Summer 2009) and other publishers, including Taylor & Francis, Elsevier and others, plan to do so over the next 12 months.

While not an essential requirement for the portal, the possibility of being able to add list prices to the journals in the database gives added value, by allowing libraries to more easily assess the value of a deal. If this information can be supplied in a standard format then the possibility of adding it should be further explored.

Linking usage statistics to detailed subject breakdown of titles.

Where publishers supply information on subject categories, this is not supplied in a consistent form, and is often linked to price lists where inconsistencies have already been noted. Subject categories vary between publishers, and even between different lists provided by the same publisher. This means that publisher subject categories are of limited value. Subject classification provided by the British Library or by Cross Ref has the potential to be useful, and the portal database itself could hold subject information if that were available from a reliable and consistent source. Further exploration is needed.

As with price information, subject categories if supplied in a standard format would be a valuable addition to the portal, allowing libraries to assess both content and usage of a deal in relation to its subject coverage. This is especially valuable for those libraries whose budgets are devolved to departments or faculties. It has been noted that Dewey will be used for subject classification as part of UKSG's Journal Usage Factor project³.

Identifying those nil or low use titles within each deal that are not fully available within the deal (eg changed, transferred, ceased or new titles).

The prototype has provided a breakdown of titles listed in one of the publisher deals and those titles appearing in the library usage statistics that are not shown in the deal. This work has shown that in order to accurately identify which nil use related to titles not fully available much more work would be required. In addition, not all publishers make this information readily available or do not present it in a standard format. This type of information is considered extremely valuable to libraries to provide them with an accurate impression of the volume of nil and low use in a deal.

Enabling libraries to add local data in order to carry out more detailed analysis, in particular cost per request. Examples include:

- analysis by cost of deal (including both subscribed titles and additional titles within the deal) to show value for money and 'yield per £' value. Useful both to show trends over time and also for benchmarking with other libraries in same JISC band and for comparing different deals to which the library subscribes.

² <http://www.editeur.org/18/Current-Releases/#SPS>

³ <http://www.uksg.org/usagefactors>

- analysis by subscribed and unsubscribed titles. While maintenance of existing subscriptions continues to be a major cost of the deal, analysing how much use is being made of both subscribed titles and additional titles within the deal can demonstrate value of subscribing to the deal rather than relying just on individual subscriptions.

The prototype has investigated ways that the portal data can be used for more detailed analysis locally by individual institutions at a local level. The portal has provided a few examples of how this can be done using additional templates into which the data can be imported.

2.6. Feedback on the portal to date

Whilst the focus of the project has been on exploring the feasibility of developing a usage statistics portal, the study team were keen to gather comments from participants as well as the JISC Journals Working Group and SCONUL.

In September 2009, all the participating content providers were provided with access to the portal with permission to view usage data they have created and surveyed over their immediate plans. At the point of writing, formal comments have been received from two content providers. These have been overwhelmingly positive.

The portal has been well received by the libraries, who welcomed the opportunity to view all their usage statistics from one central point and also to see them reported in a format that they could use when completing the annual SCONUL return.

It's exactly what all stats librarians need and a very positive step in the right direction for data collection.

It is great to have access to a site with our usage data on. The functionality looks really good and easy to use.

As expected, a number of suggestions and comments were also made for the future development of the portal.

A demonstration of the usage portal was made to the JISC Journals Working Group in July 2009. The reception was extremely positive.

As a key organisation representing libraries and information services in UK Higher Education institutions the view of SCONUL is important to consider in assessing potential demand and value attributed to such a portal by the community of potential users. As a result SCONUL was approached to give a view on behalf of its members about the desirability of further development of the portal. Dr Angela Conyers gave a demonstration of the portal to the SCONUL Working Group on Performance Improvement on 24th September 2009. The Committee were very impressed with the portal and felt it had great potential as a service to SCONUL members. The SCONUL Executive Board has given formal support for the development of the portal 'which promises to save work, and to help formulate informed judgements, in a very large number of libraries'.

3. Issues raised

Developing the portal has highlighted a number of issues which need to be considered and addressed in order to develop an effective service. These are described below and recommendations made about actions that should be taken to address these issues.

Obtaining usage statistics-SUSHI

For the prototype, libraries were asked to provide their own usage data. For a scaled up working version to be effective, data ingest must be automated and come directly from the publishers. The SUSHI protocol provides a potential means of harvesting publisher data, but we have not been able within the timescale to fully test the viability of SUSHI for ingest. Under the Counter Code of Practice Release 3, SUSHI compliance became mandatory on 31st August 2009, though a number of major publishers have yet to implement it

Recommendation

More work is needed to test SUSHI as a means of data harvesting from providers, it is vital that SUSHI be fully validated in a production environment.

Reporting on backfile use

Feedback from libraries showed that they would like separate reports on backfile use and current use. However, some publishers do not currently supply the relevant JR1A report. It is noted, however, that either JR1A or a new report JR5 will be compulsory in COUNTER release 3 that came into effect on 31 August 2009.

In addition, for the prototype portal we have supplied data on the usage of the current deal by subtracting JR1A usage from JR1. While this has been one way of splitting the data, if this could be done as separate reports by the publishers it would improve the situation.

Recommendation

COUNTER to be asked to consider a journal report just for current titles ie JR1 minus JR1A

Consistency of data – COUNTER compliance

In a scaled up working version the data from publishers which is harvested needs to be fully COUNTER Compliant. We have encountered some examples where data supplied by institutions was not strictly COUNTER compliant.

There are COUNTER 'rules' about ISSNs and the COUNTER audit process should ensure these are followed. However, the study identified some cases where wrong ISSNs were assigned. The checks carried out by the portal should throw up any ISSN errors, and thus have a positive effect in helping to fix these. However, this implies that human intervention is needed and the data cannot just be loaded without this.

Recommendation

COUNTER to be informed of inconsistencies.

Level of complexity

The project further demonstrates the level of complexity involved in this area. This needs to be borne in mind when considering how to scale up and sustain the portal as a service. For example, will costing models change? Will article level usage be introduced to better reflect value of deals? Are there other changes that are currently unforeseen that will impact on the area?-

The portal should initially be based on 'simple' functionality. It should not be regarded as a replacement for the detailed analysis performed by institutions.

Recommendation

Research and development may need to underpin a live service to ensure that it reflects any changes.

Preservation

Libraries downloading JR1 reports for this project found that in some cases two full years of data were no longer available. While publishers are required by the COUNTER code to keep data available for the last calendar year and this calendar year's to date, usage statistics will be available for an indefinite period in the portal, making it possible to do longitudinal studies.

Recommendation

The portal should retain all usage data associated with NESLi2 deals indefinitely whilst a preservation policy is developed and agreed.

Nil use

An important use of the portal is to identify nil usage of titles. At present nil use can come from lack of actual usage, but also from titles not being fully available through a publisher. Whilst it may be possible to construct a system that identifies these cases, it is a complex undertaking which would be unnecessary if the publisher could provide current details of title ownership in a standard format.

Recommendation

Publishers should be expected to fully identify where titles listed in the deal are actually no longer available.

Scope and coverage.

The current portal was developed around data from three NESLi2 publishers. For the prototype we have collected JR1 and JR1A reports directly from the libraries. Some of these have different versions of the NESLi2 deal. As JR1 reports include all potential titles available, we believe it will be extremely difficult (and of less use to the community) to single out just those who have NESLi2 deals.

If data from libraries without a particular NESLi2 deal was included, libraries would be able to use the portal to compare their usage of NESLi2 and other approaches (eg subscribed titles only) and hopefully with benchmarking be able to better assess the value of the NESLi2 deal.

Recommendation

We would therefore propose that the portal deals initially only with publishers who offer NESLi2 deals (it has the potential to go beyond that once set up) but that we ask these publishers to supply usage data for all UK HEIs who take any of their products.

Benchmarking

We believe that technically benchmarking against any criteria is possible. Libraries are interested in being able to compare results with others of similar types. The SCONUL Statistics on the Web has been cited as a good example of how libraries would like to benchmark. Originally the intention for the portal was to benchmark just by anonymised JISC band, although other anonymised banding would also be possible eg RLUK, Russell Group, 1994 group, post 1992 universities. However it is clear that libraries would like to be able to undertake more tailored benchmarking. We are unclear at this point what the contractual implications are for publishers in allowing this level of benchmarking. In addition, we expect that in a scaled up service should detailed benchmarking be permitted, libraries would have to actively consent to their data being used for these purposes.

If the portal is to include libraries who do not take a particular NESLi2 deal from the publisher it will be necessary to identify these libraries when showing total usage through NESLi2 deals so that these libraries can be considered separately. Technically it would appear straightforward to do this within the portal, however it would require the correct information on which library has which deal to be provided.

Recommendations

JISC lawyers to fully explore the rights and wishes of the publishers in this matter

JISC Collections in collaboration with the legal team to consider what form agreements with libraries would take

JISC Collections to agree precisely what level of benchmarking to permit

Once benchmarking has been agreed, a clause in the NESLi2 model licence should be included.

The NESLi2 Negotiating Agent may be best placed to supply information about which libraries have which NESLi2 deals. Further investigation of the best way to provide this data to the portal is required.

Journal usage through aggregators

The study has shown that some libraries are making considerable use of aggregators and gateways and that it is important to include this usage where it is not recorded by the publishers.

Recommendation

In order to get a full picture of the usage of the NESLi2 deals, it is essential that the portal includes both statistics from the publisher and from all aggregators/gateways appropriate for the deal.

Journal usage through databases

The prototype has not undertaken work to incorporate journal usage through databases e.g. Business Source Premier, Ebsco Host. While this usage is outside the NESLi2 deals, in order to get a fully accurate picture of title usage per se this data would need to be included.

Recommendation

Decisions need to be made about the scope of the portal and whether to incorporate this data in the future.

Working with other publishers and aggregators

The prototype has used the data relating to five libraries from three NESLi2 publishers and has demonstrated the feasibility of providing a usage statistics portal. In order to scale up the system further testing of data from other publishers and aggregators/gateways is needed. The portal has been carefully designed to give a basic functionality as its starting point, with the ability to drill down to more advanced views. This provides a way forward for a staged expansion.

Recommendation

Before moving to a full service more testing of data from more publishers, aggregators and libraries is needed.

Inconsistency in presentation of publishers prices lists

Through testing the incorporation of additional data sets to add value to the portal it became apparent that while publisher price lists are presented in different formats and file types it would be difficult to incorporate them into an automated service. The ability to add in publisher price lists was seen as of considerable value by participating libraries in order to do further analysis on value for money of the deal.

Recommendation

Publishers should provide price lists in a consistent format. Industry standards – ONIX SPS.

Community should be encouraged to define appropriate standards especially for machine consumption. Eg ONIX, NISO

Authentication

User names and passwords have been issued to JISC allowing access to all usage data and to participating libraries allowing access only to their own usage. For the future, access mechanisms will need to be addressed within context of the UK Access Management Federation. This was discussed with Nicole Harris, JISC, who subsequently initiated a JISCmail discussion and found:

1. Institutions fall broadly in to three groups:

those that are able to populate entitlements and have IDM processes to request and revoke;
those that are able to populate but do not have well defined IDM processes to request and revoke;
those that are not capable of managing this process at the moment.

2. JISC Collections should define an entitlement for NESLi2-reps in a URL form. The URL should resolve to a full explanation of that entitlement and its usage.

3. Recommendations on uses of entitlement for groups should clearly state that they should not be used if policies for requesting and revocating such entitlements are not in place.

4. The Identity Management Toolkit plus the upcoming JISC AIM Programme should carry out more work to help advise institutions on managing entitlements / managing groups.

5. Other areas within the scope of this investigation that might be resolved by the use of entitlements are the JISC Collections 'Additional Authorised User Licenses' for UK Partnerships, Overseas Partnerships and Commercial Purposes.

These findings were considered and Nicole issued the recommendation below.

Recommendation

JISC Collections should define an eduPersonEntitlement, this should be in the form of a URL, and the URL should resolve to a page describing the scope of the entitlement. So something like:

<https://www.jisc-collections.ac.uk/groups/representatives>

with text explaining that this entitlement should be used by JISC Collections Members and Affiliate Members and applied to the accounts of users within the organisation that are authorised to sign JISC Collections sub-license forms on behalf of the organisation. The Member organisations are responsible for effectively assigning and revocating this Entitlement to appropriate members of staff

4. Next steps

The following section outlines some of the key considerations and suggested actions for future activity to move towards developing a full usage portal service.

A staged approach is recommended for expansion:

JUSP				
Participants	Pilot	Year 1	Year 2	Year 3
Institutions	5	20	~50	132+
Publishers	3	6	~10	20+
Intermediaries	1	3	~5	5

Note: 132 is the total number of institutions and 20 the total number of publishers participating in NESLi2. These figures may increase.

The roles & tasks envisaged are summarised below, assuming the existing partners are involved. A detailed project plan, would be developed should the project be taken further.

Project Management (Mimas)

- Project planning and reporting
Including: project and work plans; setting-up of appropriate communication channels with all partners; meetings, regular reporting and other project documentation in accordance with JISC reporting requirements; progress monitoring; overall QA and risks monitoring and maintenance of the project web site.
- Consortium agreement
A Consortium Agreement will be drawn up, incorporating JISC generic terms and conditions and any specific conditions relating to this project. will be defined.
- Financial management
Mimas, will administer all financial aspects, including payments to partners and reporting within the project and to JISC.

Application Development (Cranfield)

- SUSHI harvesting code
Development of code to harvest usage data from designated sites as and when Publishers & Intermediaries become compliant.
- Authentication code
Redevelopment of authentication layer to support UK Access Management Federation.
- Inclusion of supplementary data, eg price
Develop mechanism to allow for upload of institutional data.
- Existing additional requirements
A number of requirements have been identified, some taken to specification stage, but not developed or implemented.
- Future requirements and extension
Additional requirements will be identified during the further development of the portal.

Application Support (Mimas)

- Institutional contact and liaison
Agreeing and arranging how each institution's usage data will be obtained.
- Data transfer management, including ftp of CSV and harvesting of SUSHI XML

Over time, the use of SUSHI will reduce the set-up and maintenance effort associated with new institutions, publishers and intermediaries being included. This will allow the rate of expansion to accelerate year on year.

- Execution and support for data processing suite.
All aspects for data processing functions, including parsing, loading and exception handling.

Database Administration (Mimas)

- Creation of development and production environments
The existing portal is hosted on a small development server housed in Cranfield library, which will be unsuitable for scaling up to full-fledged system, with appropriate levels of monitoring, back-up etc.
- Database instance creation and configuration
Both live and development. The database could continue to be MySQL, or utilise Mimas' Oracle enterprise licence.
- Database migration
The existing database and data will be migrated to newly created environments at Mimas.
- Creation of (Shibboleth SP) instance for Authentication/Authorisation
Standard technical infrastructure implementation will be required to provide Shibboleth support.

Community Engagement (Evidence Base)

- Community engagement, including recruiting additional participants
A significant expansion is planned for the portal and targeted approaches will need to be made to Libraries, Publishers and Intermediaries to recruit additional participants, subsequent liaison and to feedback collection. Also liaison with SCONUL and other user groups will be required.
- Requirements gathering, analysis and specification
Capturing, refining and specifying additional requirements and enhancement requests, including analysis of publisher deal format and data.
- Dissemination and reporting
Attend and make presentations at relevant conferences and seminars, and will write articles for publication in media relevant to the various stakeholder groups.
- Examination of related initiatives, eg ONIX-SPS
There are a number of initiatives that may or should benefit the target user community. During the pilot, the inconsistency of price information was problematic. Since July 2009, some publishers are aiming to provide standardised price information/data using the ONIX-SPS format.

The Consortium Lead will be Mimas, who will define the consortium agreement and their Finance Officer will administer all financial aspects, including payments to partners and reporting within the project and to JISC.

Practical day to day management of the project will be led by a Project Manager based at Mimas, who will ensure the cohesion of the overall project. The Project Manager will work in close collaboration with a Project Management Team, including representatives from each of the primary partners, which will meet face-to-face or by conference call on a monthly basis and interact by email or the telephone as and when necessary.

Evidence Base will chair the Steering Committee, whose strategic role will include advising the Project Management Team, representing the interests of the project partners, endorsing any proposed major changes to the original work plan (subject to

approval by JISC) and contributing to maintaining a high level of visibility for the project across the UK and abroad.

The Steering Committee is expected to meet regularly throughout the life of the project, and will include a manager from each partner, ie Evidence Base (Chair), Mimas (Project Manager), Cranfield (Application Developer) and appropriate representatives including JISC and their nominee(s). A draft budget for developing a full usage portal service has been drawn up..