



University of HUDDERSFIELD

University of Huddersfield Repository

Showers, Ben and Stone, Graham

Safety in Numbers: Developing a Shared Analytics Services for Academic Libraries

Original Citation

Showers, Ben and Stone, Graham (2015) Safety in Numbers: Developing a Shared Analytics Services for Academic Libraries. In: Proceedings of the 10th northumbria international conference on performance measurement in libraries and information servicess. University of York, York, pp. 305-314. ISBN 978-0-901931-15-3

This version is available at <http://eprints.hud.ac.uk/id/eprint/18464/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

Safety in Numbers: Developing a shared analytics service for academic libraries

Ben Showers, Graham Stone

Abstract

Purpose

It is clear that libraries consider the use of data to inform decision making a top priority in the next five years. Jisc's considerable work on activity data has highlighted the lack of tools and services for libraries to exploit this data.

The purpose of this paper is to explore the potential of a shared analytics service for UK academic libraries and introduce the Jisc Library Analytics and Metrics Project (LAMP).

The project aims to help libraries effectively management collections and services as well as delivering pre-emptive indicators and 'actionable insights' to help identify new trends, personalise services and improve efficiencies, economies and effectiveness (student attainment and satisfaction and institutional reputation, for example) . The project builds on the Library Impact Data Project at the University of Huddersfield and the work of the Copac Activity Data and Collections Management tools.

The paper will deliver a case study of the project, its progress to date, the challenges of such an approach and the implications the service has for academic libraries.

Design, methodology or approach

The paper will be a case study of the project and its institutional partners and early adopters work to date and explore both the technical and cultural challenges of the work as well as its implications for the role of the library within the institution and the services it provides.

Specifically the case study will comprise of the following aspects:

1. A brief history of the work and the context of library analytics services in the UK (and internationally). A description of the approach adopted by the project, and the vision and goals of the project
2. Exploration of the challenges associated with the project. In particular the challenges around accessing and sharing the data, 'warehousing' and data infrastructure considerations and the design challenge of visualising the data sources in a useful and coherent way
3. Outline of the implications of the project and the resultant service. In particular the implications for benchmarking (within the UK and beyond), standards development for library statistics and impact (in particular the development of ISO 16439), service development, the role of the library within the wider institution and skills and expertise of librarians.

Findings

This paper will report on the initial findings of the project, which will run from January 2013 to December 2013. In particular it will consider the issues surfaced through the close engagement with the academic library community (through the projects community advisory and planning group) and the institutional early-adopters around data gathering and analysis.

Practical implications

Data accumulated in one context has the potential to inform decisions and interventions elsewhere. While there are a number of recognised and well understood use-cases for library analytics these tend to revolve around usage and collection management. Yet, the potential of a shared analytics service is in uncovering those links and indicators across diverse data sets.

The paper will consider a number of practical impacts:

- Performance: benchmarking, student attainment, research productivity
- Design: fine tuning services, personalised support
- Trends: research landscape, student marketplace, utilisation of resources.

The case study will explore these practical implications for libraries and what they mean for the future of the library within the academy.

Originality and value of the proposal

The paper will present a case study of a unique service that currently fills an important gap within the library analytics space. The paper will focus on the services potential to transform both the way the library works and how it is perceived by its users, as well as its role and relationship within the broader institution.

Uncontrolled keywords: Library usage, student attainment, low use, non-use, academic libraries, undergraduate students, achievement

Background

The 10th Northumbria International Conference on Performance Measurement in Libraries and Information Services is testament that libraries have been collecting and analysing data for many years. The variety and scope of this data has ranged from transactional data, such as catalogue searches, book issues and returns, access to electronic resources and entry through the library gates to the manual collection of statistics on space usage, student satisfaction and external visitors to the library and library spend.

Collection methods are as varied as the data collected, for example, quantitative methods such as the SCONUL (Society of College, National and University Libraries) annual statistics return (SCONUL, 2013) and internal spread sheets or qualitative methods such as LibQual (Association of Research Libraries Statistics and Assessment Program, 2013) or analysis of the National Student Survey (HEFC, 2013). For as long as libraries have been analysing statistics, they have also been benchmarking their data against others either through bespoke work and consultancy or through such services as LAMPOST (Libraries, Archives, Museums and Publishing Online Statistics Tables). (LISU, 2012)

The application of this data is also varied and can include library management functions, demonstration of impact and improvement of service and student experience.

Jisc has been working in this area through the Jisc Activity Data programme (Jisc, 2012), which has provided a significant body of evidence and practical guidance to make a library analytics service feasible. The Library Impact Data Project (LIDP), led by the University of Huddersfield, was one of these projects (University of Huddersfield, 2011). The first phase of LIDP looked at data from over 33,000 undergraduate students across eight universities. The results, announced at the 9th Northumbria conference (Stone et. al., 2012), supported the hypothesis that there was a statistical significance across a number of universities between library activity data, specifically book issues and electronic resource usage, and student attainment, although it is important to note that this relationship is not causal. (Stone and Ramsden, 2013) This work was supported by similar studies in Australia at Wollongong (Cox and Jantti, 2012 and Jantti and Cox, 2013) and the United States at Minnesota. (Nackerud et. al., 2013 and Soria et. al., 2013)

The second stage of LIDP looked at data from 3,000 full time undergraduates from the University of Huddersfield. This phase used continuous rather than categorical data, which allowed the project to do more with the data. The aim of this study was to dig deeper into the data and looked at a number of new data categories, such as demographic information (Stone and Collins, 2013), discipline, retention, on and off campus use, breadth and depth of e-resource usage and UCAS data. Using students' final grades as a percentage, rather than a class also allowed phase two to demonstrate a correlation in the phase one hypothesis in addition to the statistical significance found in phase one.

Library Analytics Survey

Towards the end of the second phase of LIDP, the project worked alongside the Copac Activity Data (Mimas, 2011) and Copac Collections Management projects (Mimas, 2013) at Mimas. The Copac activity data project analysed libraries' circulation data to provide detailed recommendations for researchers and students. Copac collections management uses Copac data to compare libraries' collections to see which items must be retained and which can be discarded.

Huddersfield and Mimas collaborated on a library analytics survey, which was held in the autumn of 2012. The survey tried to understand any potential demand for a data analytics service, which could enhance business intelligence at an institutional level to support strategic decision-making and whether there was appetite for a shared service approach to process the raw data and to provide analytics tools and data visualisations back to local institutions. The survey received 66 replies from library staff, including many library directors.

96% of those who replied confirmed that they would want automated provision of analytics demonstrating the relationship between student attainment and library usage within in their institution, with 94.6% wanting to benchmark their data with other institutions. Furthermore, 87.7 were interested in the richer data that was used as part of the second phase of LIDP, e.g. discipline, age, year, nationality and grade.

When asked if, in principle, their institution would be willing to contribute data that could be linked to anonymised individuals, those surveyed showed a significant appetite for an analytics service, however, there was more hesitation over sharing UCAS and student data than other forms of more traditional library focussed usage data. There was also a strong willingness to share a broad range of data, however only 47% wanted to be named. The majority (91%) preferred some kind of anonymisation, e.g. to be identified by JISC band.

Figure 1 shows the major concerns from participants that would prevent them from sharing library analytics data, these concerns are discussed below.

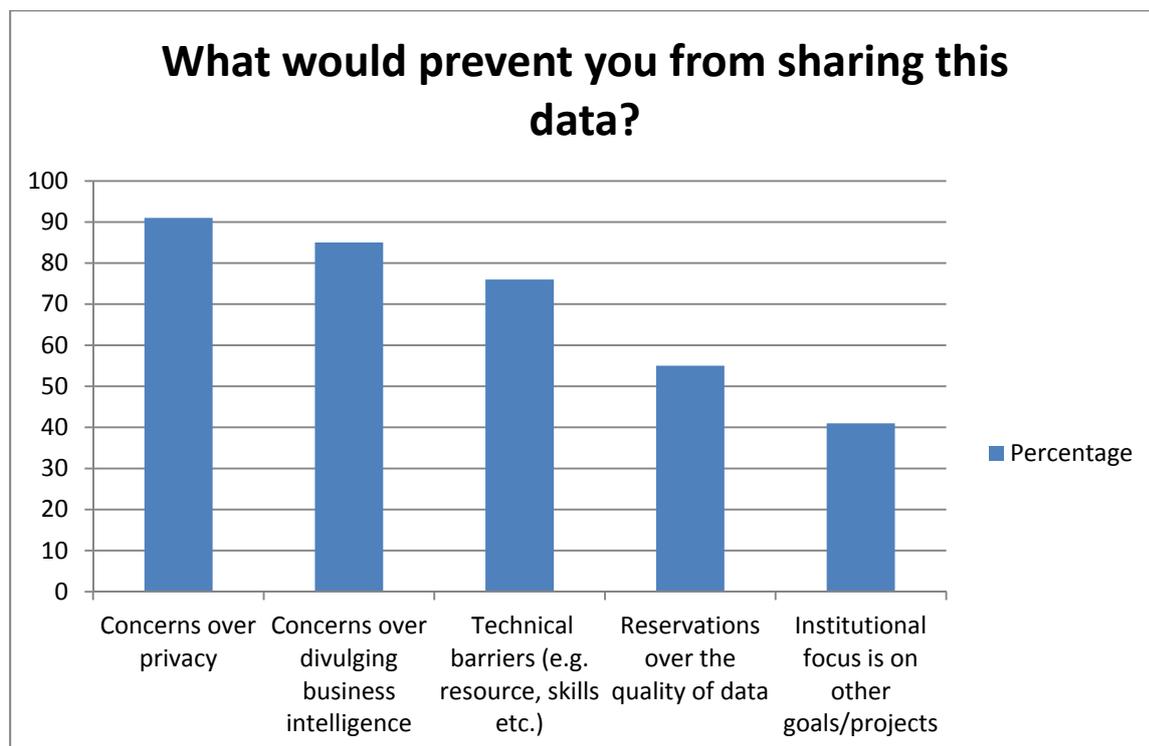


Figure 1 Concerns that would prevent the sharing of library analytics data

Finally, the survey asked specifically about the institutional focus regarding analytics. Figure 1 shows that 41% felt that focus was on other goals, however, figure 2 shows that many did see library analytics as a current strategic priority.

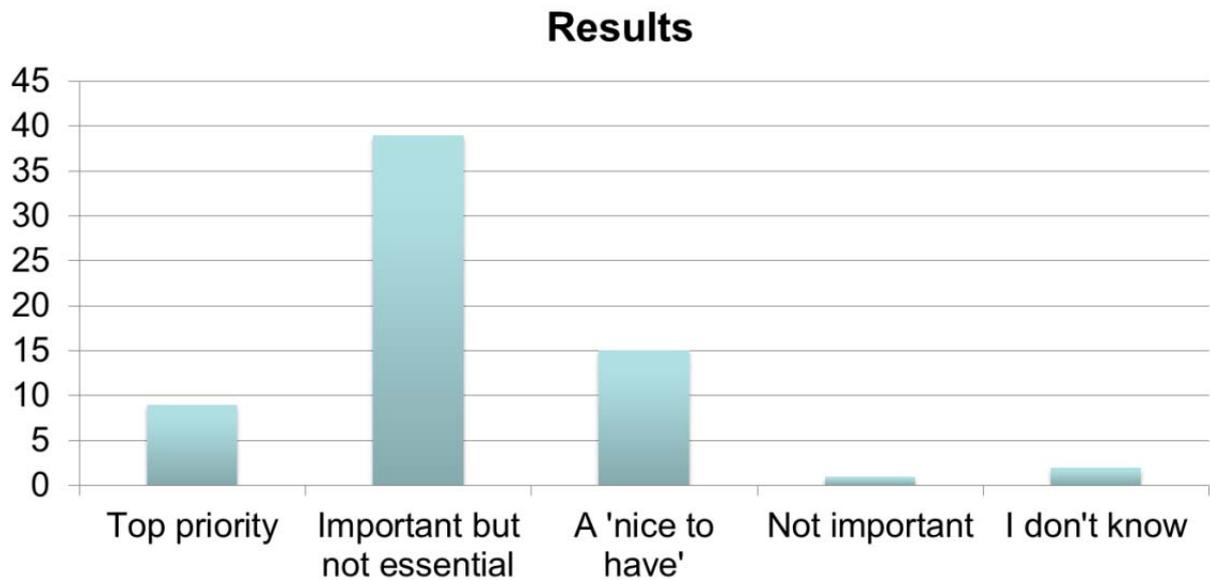


Figure 2 Library analytics as a current strategic priority

When asked if this was likely to become a top priority in the next five years, the evidence from the survey was clear (see Figure 3).

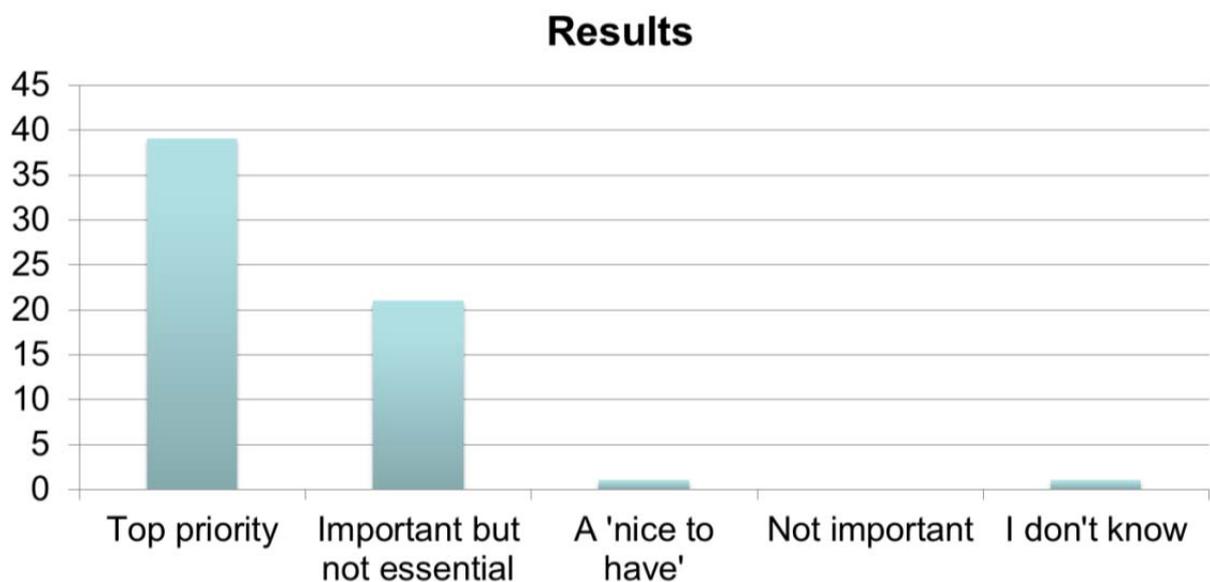


Figure 3. Library analytics as a strategic priority in the next 5 years

Appetite for a National Analytics Services

The key strategic drivers for the use of library analytics identified by the library survey were, perhaps unsurprisingly:

1. Enhancing the student experience
2. Demonstrating value for money
3. Support research excellence

A subsequent meeting of representatives from the LIDP and Copac projects, Jisc, SCOUNL and RLUK decided that there was sufficient evidence demonstrating the need, and desire, for a shared analytics service. This resulted in collaboration between Jisc, Mimas (at the University of Manchester) and the University of Huddersfield and funding for the Library Analytics and Metrics Project (LAMP) (Jisc, 2013).

Running from January to December 2013, LAMP aims to develop a prototype shared library analytics service for, and in collaboration with, UK academic libraries. The LAMP prototype will deliver a data dashboard enabling libraries to capitalise on the many types of data they capture in day-to-day activities, and will support the improvement and development of new services and demonstrate value and impact in new ways across an institution in line with the three strategic drivers highlighted above.

Technology: Architecture to Interface

In developing the LAMP service it is critical that it fits within the wider library, and institutional, systems ecosystem. Libraries already have a range of existing systems that are essential to the delivery of their services, so a new service must be able to happily co-exist within this wider systems environment; both locally in the library itself, and potentially with outside services and vendors to deliver data through existing systems and consume data from them.

The prototype LAMP architecture is therefore built around APIs (application programming interfaces). Adopting the approach described in figure 4 (Jisc, 2013), LAMP will use an API to deliver data to its own user interface (the dashboard), as well as consume external APIs from other data sources (such as JUSP, Copac or other third party services, for example). Such an approach would also allow LAMP users to get the results of the analysis on their own applications, if they preferred.

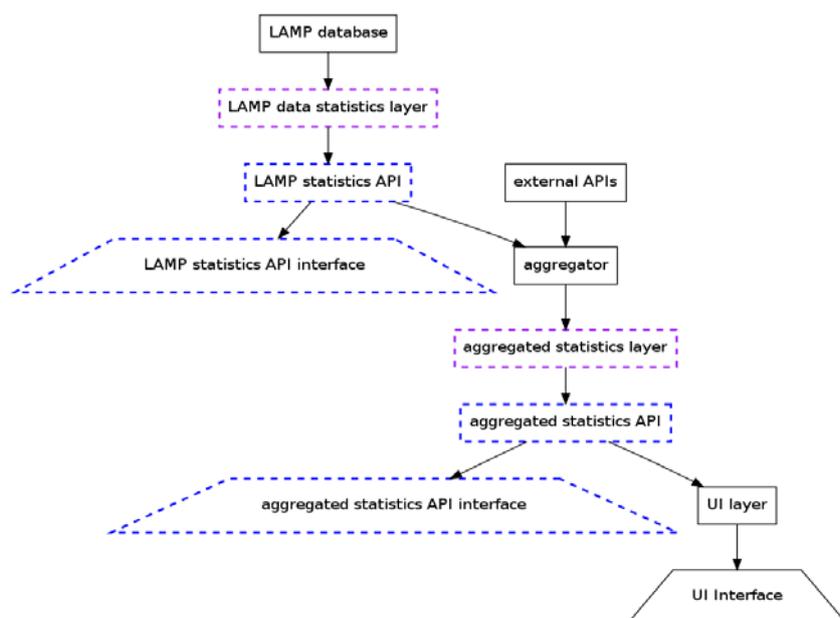


Figure 4. Example of possible structure for LAMP data analysis

In theory, the approach outlined would enable LAMP to analyse data both from the LAMP database and from third party sources and present this as a coherent view across these various data sets. In reality this may not provide any useful – or statistically valid – data, and the project is leaning towards implementing a processing layer and API on the LAMP data for now, and only including data from other APIs in the interface for comparison.

Further statistical analysis across various datasets could then be an option for a future release phase of LAMP.

Initially the LAMP prototype will be exploring the types of visualisations and analysis it can deliver. To begin with the visualisations may place a larger share of interpretation onto the user, but as the prototype is able to undertake more sophisticated analysis and visualisations, the aim is for the service to do the majority of the heavy lifting when it comes to analysis. But, from the outset, a guiding principle of the data visualisations will be that they should provide users with meaningful data, and provide the ability to adapt the types of visualisation depending on the users' requirements and/or needs.

Like the clichéd iceberg, beneath the LAMP user interface are the hidden technologies and architecture that ultimately deliver the data visualisations for the library. The ultimate aim of these visualisations is that the time and resources of the library can be spent on acting upon the data, not collecting it. This also means that, as far as is possible, analysis should be 'baked' into the visualisations being delivered; the service should already have done the majority of the analysis. In the current language of data analytics, the LAMP prototype should provide libraries with 'actionable insights'; it is then left to the library to interpret and act (or choose not to act) upon that data.

Data and Identifiers

For most users of the LAMP prototype it will be the dashboard that is the visible face of the service. Yet, it is the data circulating beneath the surface of this interface that will represent the heart of the

service. Without the data, or indeed with the wrong data or incorrect analysis, the service will fail to meet the requirements of its users.

To ensure the data is useful, and the prototype focuses on the right types of data and analysis, the project developed a number of primary use-cases for the service (for more details see Jisc, 2013). These use-cases were group under a number of headline areas:

- Demographics: Usage of resources by specific student or researcher groupings, such as age;
- Discipline: Usage of resources by discipline area;
- Student use: e-Resource usage against their final grade, for example;
- Staff use: e-Resource usage against H-Index, for example;
- Collection management: Benchmarking against other institutions;
- Outcomes: Library visits by students who did, and didn't drop out.

This is a high level overview of the use-cases, and these are in a state of constant refinement by the project and its community advisory group, but they ensure the eventual prototype is able to deliver two core objectives: Firstly, that the service is able to meet the needs of libraries in the delivery of their day-to-day services (such as collections management decisions), and secondly; provide the libraries with a way to uncover new narratives about their collections and the usage of their resources, and demonstrate new forms of impact.

In order to deliver these use-cases the project has identified the various data-sets it initially wants to use. For the first phase of the prototype these data sets can be roughly divided into two types:

1. Existing aggregations: Copac, JUSP, IRIS etc;
2. Institutional data: gate counts, circulation records, student records etc.

The first category of data is, to a large extent, easier to access (it is an existing aggregation) and helps enable the delivery of those essential services for libraries, such as collection management and resource usage. The second category is harder to access, and places a greater burden on the contributing institution (at least initially) in providing access to this data. But, it is this data that provides the project with an opportunity to deliver new types of insight and opportunities for libraries; from demonstrating impact through to improving student attainment and retention. It offers the potential for entirely new types of library services and how the library supports the wider mission of the institution.

The potential of the institutional data is significant, but it poses two main problems for the prototype LAMP service: The first is that a large proportion of the burden of the service is being shifted onto the institutions providing the data. They have to do the hard work in negotiating access for the project and ensuring LAMP is able to consume the data when it's been made available. While this is a problem for the project phase of the work and initial prototype, the burden on institutions will become less as the prototype service is able to develop ways of migrating that burden over to the service and make the institution's role as straight forward as possible.

More significantly, institutions will, currently, have to apply anonymous identifiers to the data so the project is able to track a student's data across a form. If the service cannot identify data relating to

the same individual that may be located in different parts of a form or spreadsheet, then the analysis of that data will be sub-optimal and provide little in the way of interesting information.

Initially the use of identifiers seems to imply the service will be able to identify data relating to specific individuals. This is not the case however: If the anonymisation process which generates the ID is robust, then the ID will tell the project nothing about the person to which it refers. It will simply provide the project with a way to join up separate tables of data. A further valid concern is that once data has been joined together using an anonymous ID, you could identify an individual through the unique data relating to her, for example: There is only one student who is on the Spanish history and computer science courses, originates from Norway and likes the novels of E.M Forster, for example.

This is a completely valid concern. However, this concern is removed due to the level of data the service will allow users to look at. Put simply, the LAMP prototype will not provide access to the data it holds at the granularity described above. The reasons for this are both the issues surrounding anonymity and, more importantly, ensuring the prototype is able to demonstrate statistically significant patterns for its users across the data. If the prototype provided analysis at too granular a level then it would no longer be highlighting patterns and correlations across the data; the analysis would effectively be meaningless.

The development of the LAMP prototype is, to a large extent, an exploration of these legal and ethical questions that the effective use and exploitation of data pose to libraries and institutions. LAMP poses a number of important questions to the libraries which engage with it, including how the service might help the library's role in the wider analytics mission of the institution, and the implications for the library (and institution) of not making use of this data for the benefit of its students and researchers.

Legal and Ethical Considerations

When considering the legal and ethical implications of a prototype service like LAMP it is easy to focus on the implications of exploiting the data. This is usually for a very good reason: There are specific legal obligations that the institution has regarding the data of its students, researchers and staff, and the implications of breaching those obligations are substantial. Equally, the institution, and its library, has certain ethical obligations towards its students and staff which cover how it uses the data about its students and researchers.

Yet, it is important to consider the flip side of this argument: The implications of **not** effectively using this data for the benefit of students and researchers. What might be the implications of a student who asks their institution why they didn't exploit the data they had about her to ensure she did better in her final exams, or didn't drop out? There are equally persuasive and compelling ethical imperatives for the effective exploitation of data for the institution and its students and researchers.

In effect the legal obligations and measures can be met by ensuring they are not contravened, but the ethical issues are less clear cut, nor so easily mitigated by specific actions or protocols (Kay et. al. 2012). With these various and complex legal and ethical challenges the LAMP project has adopted a three level approach to addressing the legal and ethical landscape within which it will be operating.

Firstly the project has developed a draft set of terms and conditions for participating institutions that will outline the legal obligations the service has, along with the legal status of the data centre

Mimas, which will host the data. These terms and conditions are currently drafted and available on the project website (Jisc, 2013) and provide the basis for a legal relationship between the contributing institution and the service. These terms and conditions will then be complimented with further legal advice and guidance for institutions to ensure institutions are clear about LAMP's legal and ethical commitments.

Secondly, the project has been developing a legal and ethical framework. This will provide the project with a clear understanding of the current legal and ethical requirements it has to its users, and also the best practice it should be adopting in delivering its service. The aim of this framework is to both provide the project with a clear understanding of its role and responsibilities, but also provide the library community with a clear picture of what the current legal and ethical landscape currently looks like for library data.

The final stage of the project's legal and ethical approach is its principles. These principles, made available on the project blog and web presence (Jisc, 2013), provide the 'rules' by which the project intends to undertake its work. These are not legally binding like the service's terms and conditions, nor as comprehensive as the framework, but they lay down the approach of the project and make it clear what it will and won't do. The principles include statements on data protection and confidentiality, licensing and standards, project development and how institutions can get involved.

The ability of the project to resolve these legal and ethical issues and to develop policies, guidance and support that enable institutions to trust and have confidence in the service and its use of institutional data, are critical to the ultimate success of the project.

In many ways LAMP is an experiment exploring the potential of analytics and measurement within libraries and finding out how far the data libraries hold can be used to deliver next generation services and provide enhanced student and researcher experiences. It is an attempt to work with the library community to see what it is possible to do with analytics.

Looking forward

In his opening keynote at the 10th Northumbria conference, Elliot Shore of the Association of Research Libraries (2013), called for libraries to move from a position of describing what's happening in their spaces and with their collections, to one where they are able to begin predicting need and usage: A shift away from the quantity of data libraries collect, towards one of quality and the ability to impact on student and researcher expectations and experience.

The aims of LAMP are neatly described by Shore's clarion call. The project wants to work with the library community to ensure it is able to deliver the analysis and information for librarians to be able to act upon what really matters to their students, researchers and users. Furthermore, after December 2013 when the prototype service is delivered, the project will work with libraries and institutions to ensure that library analytics are part of the wider enterprise of the institution and of UK academia more broadly.

Acknowledgements

Many thanks to Jisc, University of Huddersfield and Mimas (University of Manchester) and the LAMP team for their work on the project which includes: Joy Palmer (Mimas), Ross McIntyre (Mimas), Bethan Ruddick (Mimas), Lee Bayliss (Mimas), Siobhan Burke (Mimas), Ellen Collins (RIN) and David Kay (Sero Consulting).

References

Association of Research Libraries Statistics and Assessment Program (2013) *LibQUAL* [online] Available at: <http://www.libqual.org/home> [Accessed 6 September 2013].

Cox, B. and Jantti, M.H. (2012) 'Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement'. *Library and information science research*, 34 (4), pp.308-316. doi: <http://dx.doi.org/10.1016/j.lisr.2012.06.002>

HEFCE (2013) *National Student Survey* [online] Available at: <http://www.hefce.ac.uk/whatwedo/lt/publicinfo/nationalstudentsurvey/> [Accessed 6 September 2013].

Jantti, M. and Cox, B. (2013) 'Measuring the Value of Library Resources and Student Academic Performance through Relational Datasets'. *Evidence Based Library and Information Practice*, 8 (2), pp. 163-171 <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/19574/15217> [Accessed 6 September 2013].

Jisc (2011) *Digital Infrastructure: information and library infrastructure programme* [online] Available at: http://www.jisc.ac.uk/whatwedo/programmes/di_informationandlibraries.aspx [Accessed 6 September 2013].

Jisc (2012) *Activity Data: Delivering Benefits from the Data Deluge*. London, Jisc [online] Available at: <http://www.jisc.ac.uk/publications/reports/2012/activity-data-delivering-benefits.aspx> [Accessed 6 September 2013].

Jisc (2013) *Library Analytics and Metrics Project* [online] Available at: <http://jisclamp.mimas.ac.uk/>

Kay, D. et. al. (2012) *Legal, Risk and Ethical Aspects of Analytics in Higher Education* JISC CETIS Analytics Series No. 6 [online] Available at: <http://publications.cetis.ac.uk/2012/500> [Accessed 15th September 2013]

LISU (2012) *LAMPOST (Libraries, Archives, Museums and Publishing Online Statistics Tables)* [online] Available at: <http://www.lboro.ac.uk/microsites/infosci/lisu/lampost.html> [Accessed 6 September 2013].

Mimas (2011) *Copac Activity Data Project* [online] Available at: <http://copac.ac.uk/innovations/activity-data/> [Accessed 6 September 2013].

Mimas (2013) *Copac Collections Management* [online] Available at: <http://copac.ac.uk/innovations/collections-management/> [Accessed 6 September 2013].

Nackerud, S. et. al. (2013) 'Analyzing Demographics: Assessing Library Use Across the Institution'. *portal: Libraries and the Academy*. 13 (2), pp. 131-145. [online] Available at:

http://www.press.jhu.edu/journals/portal_libraries_and_the_academy/portal_pre_print/current/articles/13.2nackerud.pdf [Accessed 6 September 2013].

SCOUNL (2013) Statistics – performance portal [online] Available at:
<http://vamp.diglib.shrivenham.cranfield.ac.uk/statistics> [Accessed 6 September 2013].

Shore, E. (2013) 'The Role of the Library in the Transformative Higher Education Environment: Or Fitting Our Measures to Our Goals'. In: *10th Northumbria international conference on performance measurement in libraries and information services*. 22-25 July 2013, University of York, York,

Soria, K. et. al. (2013) 'Library Use and Undergraduate Student Outcomes: New Evidence for Students' Retention and Academic Success'. *portal: Libraries and the Academy*. 13 (2), pp.147-164. [online] Available at:
http://www.press.jhu.edu/journals/portal_libraries_and_the_academy/portal_pre_print/current/articles/13.2soria.pdf [Accessed 6 September 2013].

Stone, G. and Collins, E. (2013) 'Library usage and demographic characteristics of undergraduate students in a UK university'. *Performance Measurement and Metrics*, 14 (1). pp. 25-35. Doi:
<http://dx.doi.org/10.1108/14678041311316112>

Stone, G. and Ramsden, B. (2013) 'Library Impact Data Project: looking for the link between library usage and student attainment'. *College and Research Libraries*. Preprint [online] Available at:
<http://eprints.hud.ac.uk/15038/> [Accessed 6 September 2013].

Stone, G. et. al. (2012) 'Library Impact Data Project: hit, miss or maybe'. In: *Proving value in challenging times: proceedings of the 9th Northumbria international conference on performance measurement in libraries and information services*. University of York, York, pp. 385-390. [online] Available at: <http://eprints.hud.ac.uk/15563/> [Accessed 6 September 2013].

University of Huddersfield (2011) *Library Impact Data Project* [online] Available at:
<http://library.hud.ac.uk/blogs/projects/lidp/> [Accessed 6 September 2013].

Author Contact details

Ben Showers, Programme Manager, Digital Infrastructure, Jisc. b.showers@jisc.ac.uk

Graham Stone, Information Resources Manager/Senior Research Fellow, University of Huddersfield.
g.stone@hud.ac.uk