UKSG

Developing a shared analytics service for academic libraries

Based on a breakout session presented at the 37th UKSG Annual Conference, Harrogate, April 2014 by Joy Palmer (Mimas, University of Manchester), Graham Stone (University of Huddersfield) and Ben Showers (Jisc)

Universities and colleges have a rich seam of data, created as students, researchers and staff interact with the systems and services across their campuses. The potential for this data to transform the student and researcher experience, and improve the efficiency and effectiveness of the institution, is currently locked away in the diverse and complex data sets and systems employed by each institution. Even where these data sets are available, making sense of them can be difficult and resource intensive.

This article describes the efforts of the Library Analytics and Metrics Project (LAMP) to develop a shared, national service to enable institutions and libraries to make sense of their disparate and diverse data sets and enable them to spend their precious time and resources acting upon it to improve the student and researcher experience.

Introduction

Libraries are no strangers to the potential of data. Indeed, it is data that drives the systems and services that libraries deliver to their users: from the metadata that enables discovery and access to content, to the management and transactional data that helps ensure the library collections and services meet the requirements of the students and researchers and demonstrate the value the library delivers to the host institution.

Yet, identifying, collecting and analysing data that will help drive service development and demonstrate impact is not straightforward. Indeed, three key factors conspire to make this task increasingly difficult:

BEN SHOWERS

Head of Scholarly and Library Futures
Jisc

1. *Cost versus value*: collecting the data can be resource intensive, often requiring far more time to identify and collect the data than is spent acting upon it.

2. *Increasing volumes*: the sheer amount of data and data sources is increasing. As the business of the library becomes increasingly digital and more processes are automated, the library is confronted with trying to make sense of voluminous amounts of (often unstructured) data.

3. *Data beyond the library*: to develop effective services and systems, and to demonstrate value to the wider institution, the data libraries are interested in is not just located in their local systems, but also across the campus and in external aggregations.

The resource implications, as well as the sheer amount of data, make the task of analysing and acting on the data potentially more difficult; especially when library budgets and resources are already pushed to the limit.

These challenges, however, have not inhibited the increasing strategic desire of libraries to be able to make use of this data to both inform the development of existing and new services and to explore ways to demonstrate the library's value and impact to the institution and its students and researchers.

An appetite for analytics

Libraries are actively engaged in exploring the potential of analytics. This is visibly the case in the increasingly rich landscape of both UK and international activity aimed at exploiting and maximizing the many possible uses for library and institutional data.

The Library Impact Data Project (LIDP)[1], led by the University of Huddersfield, looked at data from over 33,000 undergraduate students from across eight universities to explore the validity of 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[2].

> "Libraries are actively engaged in exploring the potential of analytics."

The project was successful in supporting the hypothesis and demonstrating the (non-causal) relationship between library usage and student success. This hypothesis has also been further validated by work in both Australia at the University of Wollongong[3] and the US at the University of Minnesota[4].

As part of the Jisc Activity Data programme[5], the work of Huddersfield exemplifies some of the institutional-level activity around analytics. This work includes both a focus on library impact and value as well as an interest in personalization and recommendation services by projects at the Open University[6] and Manchester Information and Associated Services (Mimas) at the University of Manchester[7]. This local, institutional activity is also complemented by the national, shared services that have emerged to provide easy access to some of this transactional and usage data, including, for example: the Journal Usage Statistics Portal (JUSP)[8]; the Institutional Repository Usage Statistics (IRUS)[9], and Copac Collections Management (CCM)[10].

Building on this fertile environment of experimentation and service development in library data, Jisc, the University of Huddersfield and Mimas wanted to explore the sector's appetite for analytics in general, and the potential demand for a shared data analytics service in particular. So, in the autumn of 2012, a survey was distributed to all UK academic library directors, resulting in 66 responses.

Of those who replied, 96% confirmed that they would want automated provision of analytics demonstrating the relationship between student attainment and library usage within 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.

There was also a strong willingness to share a broad range of data, though only 47% wanted to be named. The majority (91%) preferred some kind of anonymization, e.g. to be identified by Jisc band.

When asked if this was likely to become a top priority in the next five years, the evidence from the survey was clear, with almost all respondents saying it was either a top priority (39) or important but not essential (21)[11].

> "… a strong willingness to share a broad range of data …"

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

- enhancing the student experience

- demonstrating value for money

- support research excellence.

With a set of clear, near-term strategic drivers for institutions, a rich library data/analytics landscape and the support of the Society of College, National and University Libraries (SCONUL) and Research Libraries UK (RLUK), it was agreed there was compelling evidence of the need, and desire, for a shared library analytics service.

Library Analytics and Metrics Project (LAMP)

Running from January 2013 to March 2015, the Library Analytics and Metrics Project (LAMP) aims to develop a prototype shared library analytics service for, and in collaboration with, UK academic libraries. By March 2015 the LAMP prototype will deliver a data dashboard enabling libraries to capitalize 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.

As LAMP begins its second phase of work, it is clear that there are three core components of the work so far, and of the work to develop the prototype service by March 2015. These are:

- data – the disparate institutional data sets that will be ingested and analysed by the service

- analysis – how the data is meaningfully displayed to users

- community – the role of the library community in helping develop, shape and deliver LAMP.

"LAMP will use the opportunities of scale to access a much larger number of data sets …"

Data

LAMP will use the opportunities of scale to access a much larger number of data sets, analysable at both the local, institutional level and at the shared, above-campus level. In both cases the hope is that new insights can be gained, such as national usage patterns, for example, as well as enhanced services and functionality for institutions, such as benchmarking and personalization.

Specifically, this means that LAMP will ingest and normalize various institutional data sets, including: UCAS data (which is data from the Universities and Colleges Admissions Service in the UK and is LAMP's primary source for individual student data such as age, gender, course and so forth), library loans data, e-resource logins, library gate counts, examination data (student attainment/records). Even with the relatively limited amount of data LAMP aims to collect, there are still a number of cultural and technical challenges, as these data sets are often in different systems and are owned by different departments. This conspires to make their collection and aggregation challenging. While the project will require a minimum number of core data sets to provide a certain level of service, it is ultimately a case of providing back to institutions a view (or rather views) on the data sets they were able to supply. LAMP, as a service, must recognize that institutional access to different data types and sets is not uniform, and will vary widely between institutions.

In the longer term, however, it will become increasingly important for institutions to be able to track the 'journey' of the student from prospect to alumnus if they are to be able to offer the kinds of experience that will quickly become the norm in education. This will make accessing more data sets across the campus critical. Also, while at the prototype stage LAMP is able to utilize one-off anonymized data sets, the need to track over time will mean that the project (or the institution) will need to find ways of using anonymized identifiers that can track the same student over the three or more years of their course. As the prototype develops into a service, this will become a critical challenge to address[12].

"… it will become increasingly important for institutions to be able to track the 'journey' of the student …"

Working with six institutional data providers[13] is enabling the project to explore exactly these kinds of challenge. At present much of the burden of contributing data sits with the institution and library but, as LAMP develops from a prototype to a service, this burden will need to shift onto the service.

One area LAMP is actively exploring to reduce the institutional burden of submitting data is the ingestion of other national data sets, some of which may include the local data sets

required by LAMP. The prototype LAMP architecture is therefore built around application programming interfaces (APIs). 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 an approach would also allow LAMP users to get the results of the analysis on their own applications, if they preferred.

Those other services and aggregations include: JUSP, IRUS, the Higher Education Statistics Agency (HESA) and SCONUL, for example. These provide a potential route to automating data consumption, as well as potentially enabling the sharing of appropriate data with other services and integrating processes where possible. Similarly, LAMP will itself become a national data set which will enable development of applications such as benchmarking and performance measurement. Already, for example, the prototype can tell whether the output described is statistically significant or not – not something easily achieved at the individual level.

> "… LAMP will itself become a national data set …"

The ability to bring together more data sets from more sources opens up the possibility of being able to answer difficult questions that institutions want to know. Indeed, it offers the possibility of asking entirely new questions.

But so much of this potential functionality depends on the ability of the LAMP service to be able to offer *meaningful* answers in response to the kinds of questions that librarians and others will want to ask of it.

Analysis

For LAMP, much of the first year of development has been about addressing the challenges of obtaining and 'cleaning' the necessary institutional data sets[14]. Much of this was behind-the-scenes development, and in close collaboration with the contributing libraries. But, while data represents the bedrock of the LAMP prototype, it is the service's ability to provide 'a view' or analysis of that data that is critical. It is the view on the data which will provide the library with the insights that will help inform their decisions and actions.

The project is now at the stage of exploring how the data is presented to the user and understanding the division between how much the service can provide, and how much needs to be analysis brought by the user. Ultimately, the service will be delivering graphs and analysis to the user; much of the work of interpretation and understanding will still be in the hands of the user. While the service can make access and analysis of the data much easier, the data literacy of the user is still a critical factor in how that data is used.

In order to explore these questions, the project has developed, in conjunction with our community advisory group, a number of user and job stories to get a clear sense of how libraries might want to use the data, the kinds of things they would want it to tell them and what they would want to do with it[15]. But, maybe more importantly, the project has also developed an 'ugly prototype' (as it has become affectionately known) to test with potential users the kinds of analysis they would want to be able to do on the data (and indeed to show the kinds of data LAMP can provide).

> "… data literacy of the user is still a critical factor in how that data is used."

An example of the kinds of data LAMP can provide at the moment are shown in Figure 1. This is a simple pie chart demonstrating loans activity among students studying for a bachelor's degree in psychology, broken down by gender.

At first glance it would appear women borrow a lot more than men – or do they? Are there simply more women in that discipline? Already, even the most straightforward graph begins to throw up questions for anyone viewing the data. Indeed, it may be that there are other, more important or interesting questions we should be asking with this particular set of data, such as what are the differences between full-time and part-time students and book borrowing, and how do they break down between genders, for example.
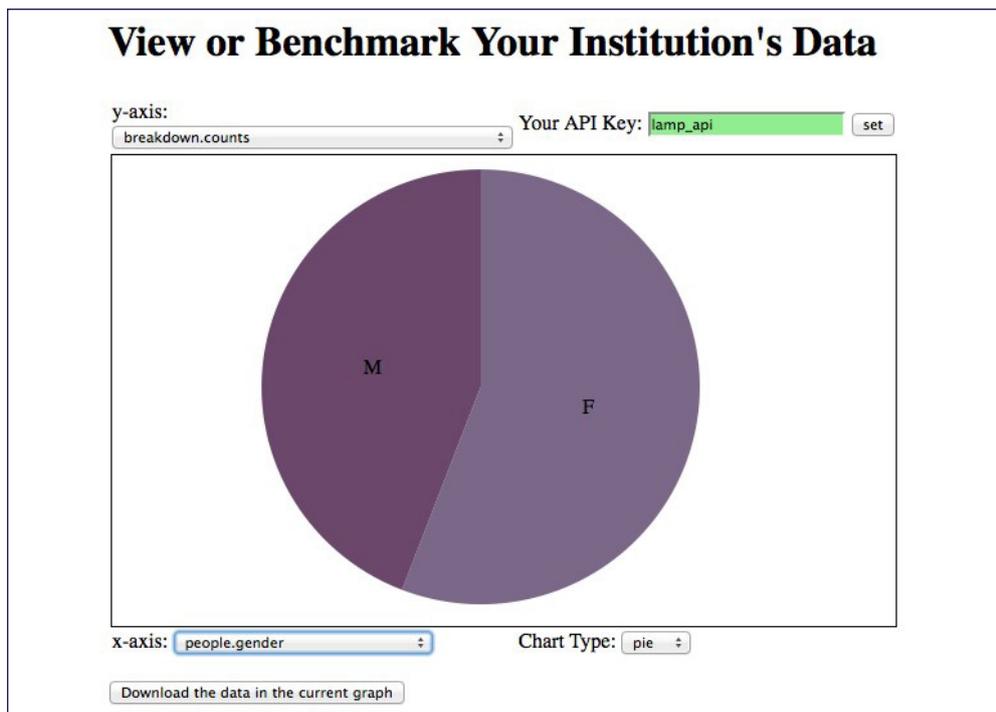
Figure 1. An example of the kind of data delivered by the LAMP 'ugly prototype': loans activity for a specific subject broken down by gender.

Is it enough to provide just that particular graph? The question for LAMP is what story is this data actually telling?

The ability to misinterpret or take the data out of context is a potentially important consideration, and it is beholden upon LAMP to ensure it does all it can to mitigate the risks inherent in the complexities of data.

The ugly prototype and collection of job stories have provided the project with a starting point from which to begin exploring these kinds of questions and to understand the requirements for both the visualizations the service provides and for the data literacy of the users who are interacting with the service.

Over the next 10-12 months, the project will be explicitly exploring these issues and developing the prototype with librarians to refine the way the data is described via the interface[16]. As with the data, the analysis aspects of LAMP require a close relationship with the library community to enable the project to test and refine the prototype over the coming months.

Community

The academic community and, specifically, the academic library community, is the critical partner in ensuring the success of the eventual LAMP prototype and service.

From accessing the institutional data through to developing, testing and refining the user and job stories, the project has relied on its Community Advisory and Planning group (CAP)[17]. This group of engaged experts has helped shape much of the work so far, and is helping ensure that the LAMP prototype being developed will meet the needs of libraries and institutions.

Importantly, this group and the sub-set of data contributors is willingly taking upon itself a significant amount of burden in order to help the project access the relevant data and build the prototype interface. As described already, accessing the data is no simple feat, and negotiating across the institution is a considerable task.

"The ability to misinterpret or take the data out of context is a potentially important consideration ..."

Ultimately, LAMP needs to work with the library community to ensure it is able to deliver the analysis and information required for institutions to be able to act upon what really matters to their students, researchers and users. Engaging with the library community becomes more critical as the project begins to develop its prototype service and looks to involve more institutions in both testing the interface and providing data.

A data-driven future?

At one of the early CAP group meetings there was a discussion about the legal and ethical implications of a service like LAMP. The discussion had turned to what might happen if students and researchers begin questioning an institution's use of this data. At this point one of the members argued that the focus of discussion is changing from one where students are asking about the use of data, to one where they are questioning why institutions are *not* using the data.

> "There is a real opportunity for libraries …"

What if students complained that the university had not used the personal data it had collected about them, in order to prevent their foreseeable failure? This is increasingly the reality for universities and colleges as students and researchers become accustomed to the personalized and data-driven experiences of the wider digital, and indeed physical, environment. This, of course does not mean academic institutions should be complacent, but that of all the organizations students will be interacting with, the university or college library is one of the most trusted. There is a real opportunity for libraries to both take a strategic lead on campus in the data and analytics area, and to use this data and expertise to develop new services to improve the student experience.

But, as the discussions around data analysis make clear, LAMP is just one part of a wider analytics conversation. There are considerations around data literacy across the academic sector and for users of services like LAMP in particular. Importantly, LAMP also enables libraries to consider analytics more broadly, to explore various approaches to gathering data required for improving and developing services, from quantitative to qualitative approaches.

LAMP is part of a much wider dialogue institutions and libraries are beginning to have with their students and staff.

Getting involved

LAMP itself is only really just beginning. So far the project has been working with a small group of libraries to develop the database and begin prototyping the interface (unfortunately losing the ugly prototype!). Over the next 12 months, the project will be engaging more libraries and looking to access more data. It will also explore some of the challenges and opportunities described above, including:

> "… LAMP is just one part of a wider analytics conversation."

- developing a user interface to deliver meaningful and compelling data visualizations

- exploring usage data as a way to 'profile' individuals for use cases such as the REF or intervention purposes (and the legal/ethical implications)

- item-level usage for e-resources

- looking at the possibility of integration with other data sets like SCONUL and National Student Survey (NSS)

- enabling benchmarking across libraries

- data literacy: what can be automated/what needs to be part of the user skillset?

If you would like to be involved in this next phase of the LAMP work, there are a number of ways you can engage:

1. Follow and comment on our blog: **http://jisclamp.Mimas.ac.uk**

2. Attend a LAMP workshop. These will be taking place throughout 2014 and early 2015 so keep an eye on the blog for further information.

3. Become a data contributor. Contact us via the information at the end of this article to see how you can become one of the early partners on the project.

References

1. Library Impact Data Project:
   https://library3.hud.ac.uk/blogs/lidp (accessed 14 May 2014).

2. Stone, G and Ramsden, B, Library Impact Data Project: looking for the link between library usage and student attainment, *College and Research Libraries*, 2013, 74(6), 546–559.

3. Jantti, M and Cox, B, Measuring the Value of Library Resources and Student Academic Performance through Relational Datasets, *Evidence Based Library and Information Practice*, 2013 8(2), 163–171.

4. Nackerud, S et al, Analyzing Demographics: Assessing Library Use Across the Institution'. *Portal: Libraries and the Academy,* 2013, 13(2), 131–145.

5. Activity Data Programme Synthesis:
   http://www.activitydata.org (accessed 14 May 2014).

6. The RISE project:
   http://www.activitydata.org/RISE.html#RISE (accessed 14 May 2014).

7. The SALt project:
   http://www.activitydata.org/SALT.html#SALT (accessed 14 May 2014).

8. JUSP:
   http://www.jusp.Mimas.ac.uk/ (accessed 14 May 2014).

9. IRUS:
   http://www.irus.Mimas.ac.uk/ (accessed 14 May 2014).

10. CCM:
    http://copac.ac.uk/innovations/collections-management/ (accessed 14 May 2014).

11. Further detail on the survey questions and the range of responses can be found on the LAMP blog:
    http://jisclamp.Mimas.ac.uk/2013/04/library-analytics-community-survey-results (accessed 14 May 2014).

12. Further discussion on the issue of anonymous identifiers can be found on the LAMP blog:
    http://jisclamp.Mimas.ac.uk/2013/06/unique-identifiers-which-dont-identify-anyone (accessed 14 May 2014).

13. Further information on the partners can be found on the LAMP blog:
    http://jisclamp.Mimas.ac.uk/team-and-partners (accessed 14 May 2014).

14. See, for example:
    http://jisclamp.Mimas.ac.uk/2013/08/creating-the-lamp-database (accessed 14 May 2014).

15. Further detail on the job and user stories can be found on the LAMP blog:
    http://jisclamp.Mimas.ac.uk/2013/11/personas-job-stories-and-simple-planes-wireframing-a-lamp-interface (accessed 14 May 2014).

16. An update on the work to develop the LAMP interface can be found here:
    http://jisclamp.Mimas.ac.uk/2014/03/screen-flows-wireframes-prototypes-and-guts (accessed 14 May 2014).

17. More information on the CAP group can be found here:
    http://jisclamp.Mimas.ac.uk/community-advisory-and-planning-group (accessed 14 May 2014).

Ben Showers
Head of Scholarly and Library Futures, Jisc
Brettenham House, 5 Lancaster place, London WC2E 7EN
Tel: +44 (0)203 006 6015; Mob: 07891 470735 | E-mail: b.showers@jisc.ac.uk | Skype: ben.showers | Twitter: @benshowers

ORCID iD: http://orcid.org/0000-0002-2619-4400