ARIADNE

Looking for the link between library usage and student attainment

Graham Stone, Bryony Ramsden and Dave Pattern introduce the JISC funded Library Impact Data Project

Introduction

In 2010, the University of Huddersfield shared results from its analysis of anonymised library usage data [1]. Data was analysed for over 700 courses over four years - 2005/6 – 2008/9; this included the number of e-resources accessed, the number of book loans and the number of accesses to the library. This investigation suggested a strong correlation between library usage and degree results, and also significant underuse of expensive library resources at both School and course level. At the time, it was highlighted that the correlation between library usage and grade had not yet been significance tested and that it was not known whether the Huddersfield findings were an anomaly or the norm [2].

As a result, a number of universities approached Huddersfield in order to benchmark against the data. Shortly afterwards the JISC released a call through the Activity Data programme [3], where potential bidders were asked to put forward a hypothesis as part of their project proposal.

In the light of the recent Comprehensive Public Spending Review and the Lord Browne's Review of Higher Education Funding and Student Finance [4], it was thought that if the Huddersfield experience was found to be of statistical significance across a broad range of universities there was potential for the results to be used as a factor to enhance student attainment. In parallel, there is a continuing focus on the student experience and a desire that all students should achieve their full potential whilst studying at University. Results could also be used by libraries to better target their resources where budgets are shrinking.

In February 2011 the University of Huddersfield along with 7 partners: University of Bradford, De Montfort University, University of Exeter, University of Lincoln, Liverpool John Moores University, University of Salford and Teesside University were awarded JISC funding to prove the hypothesis that:

'There is a statistically significant correlation across a number of universities between library activity data and student attainment'

Aims and objectives

The overall goal of the project is to encourage greater use of library resources thus leading to an increase in students' knowledge and understanding of their subject areas, and ultimately to ensure that student attainment is improved in areas of non/low use.

However, the project is keen to acknowledge that the relationship between the two variables is not a causal relationship and there will be other factors, which influence student attainment.

The project hopes that by proving the hypothesis there will also be tangible benefits to the wider higher education (HE) community. A successful conclusion to the project will help to create a greater understanding of the link between library activity data and student attainment. To this effect, the project will publish its methodology to allow other HE Institutions to benchmark their data.

The proof of the hypothesis will also be fed back into the original work from studies on non/low use in order to target improvement in these areas [5], including techniques such as course profiling, targeted promotion of information resources, raising tutor awareness of the link between use of resources and attainment and improved targeting of resources allocation.

Project plan

As part of the planning process, the project has been separated in 4 work packages

Project reports and outputs (months 1-6)

In guidance issued from Andy McGregor [6], the programme manager for the Activity Data strand, all projects are required to create a number of blog posts throughout the project:

- Post 1 - The Project Plan
- Post 2 - Hypothesis
- Post 3 - Users
- Post 4 - Benefits
- Post 5 - Technical and Standards
- Post 6 - Licensing & reuse of software and data
- Post 7 - Wins and fails (lessons along the way)
- Post 8 - Final post

The Library Impact Data Project (LIDP) has chosen to do this in a series of tagged blog posts from the project blog [7]; this will then be combined to create the final report.

An early blog post looked critically at the hypothesis, emphasising that there are other factors that impact on student attainment, and that library usage alone is not an indicator of how well a student will score in their assignments, and this will be reflected in our conclusions. There are also a number of other elements we will need to consider when we progress through analysis of our results. Due to the nature of some courses, borrowing and resource access may not necessarily link to a level of access viable for attainment. Recreational use of facilities, and the nature of some subject area materials being freely available outside of library subscriptions (for example primary sources necessary for history assignments), may lead to increased or reduced borrowing respectively. Some courses have information skills heavily embedded into teaching, which could result in increased access to electronic

resources, but may not reflect the skills of students accessing them. Distance learners and placement students may also skew results, and we may wish to consider their use in terms of electronic access rather than book borrowing. In terms of entry data, socialising and non-library facilities will need to be factored into the analysis.

The project will be presenting its findings at a number of conferences in Europe during the summer, including the 40th LIBER Conference in Barcelona, the SCONUL conference in Cardiff and the 9th Northumbria International Conference on Performance Measurement in Libraries and Information Services in York; further information on other seminars and conferences will be provided on the project blog.

In addition, the partners will seek to raise awareness of the project within their respective intuitions, e.g. at Huddersfield the team will submit a poster to the annual Poster Prom for academic staff and researchers.

The project is working closely with the Synthesis Project [8], which has been funded to support the Activity Data projects, the LIDP has already contributed to a series of recipes, which will enable other universities to collect and analyse their own data in order to benchmark themselves against the findings of the project.

Data collection (months 1-4)

The approach the project will take is to extract anonymised activity data from partners' systems and analyse the findings.

Potential partners were asked to provide activity data which would span at least one entire academic year (e.g. 2009/10), or ideally for multiple years if historic data was available as this would add robustness to the data. An ideal situation being that with multiple years of data it would be possible to analyse library usage by students in each year of their course. Data from the original study at Huddersfield suggests that library usage changes as a student progresses. This is in addition to evidence, which pointed to differences in library usage behaviour between departments; it would be interesting to know if both were replicated across other institutions. This builds on research from Kramer, which suggests certain disciplines have less need for books [9], extending this for the first time into the need for electronic resources.

For each student who graduated in a given year, the following data was required:

- Final grade achieved
- Number of books borrowed
- Number of times e-resources were accessed
- Number of times each student entered the library, e.g. via a turnstile system that requires identity card access
- School/Faculty

It should be noted that the number of times a student accessed e-resources through authentication such as Athens, Shibboleth, EZProzy or MetaLib is a crude but common measure of actual e-resource usage.

At the bidding stage it was anticipated that there may be problems in getting enough data to make the project viable, especially given the short 6 month timeframe the project had to work with. To this effect potential partners were asked to confirm that they could provide at least 2 of the 3 measures of usage as well as student grades.

Huddersfield has provided definitions on the data required and the form the data can be accepted in. At the time of writing data is now being successfully received from the partners and will be quality checked, collated and significance tested at Huddersfield. However, at this early stage of the project, some partners have already run into some issues with data collection, but it is felt that there is still enough information to prove the hypothesis one way or another. Therefore the project will be seen to have succeeded if the following measurable targets are achieved:

- Sufficient data is successfully captured from all partners
- Statistical significance is shown for the data
- The hypothesis that 'there is a statistically significant correlation across a number of universities between library activity data and student attainment' is either wholly or partly proved for each data type and partner

One of the partners has additionally submitted PC login data too; this is also being looked at by Huddersfield to see if there is a correlation between attainment and PC logins, or anticipating a lack of correlation between attainment and library entry, library entry and PC logins.

The project also hopes to make all anonymised data available for re-use under an Open Data Commons Licence, building on the recommendations of the MOSAIC project [10]. However, unanimous approval will have to be sought from partners about whether individual data will be released; otherwise the data will be further anonymised into one set data, thereby protecting the identity of each institution. This would probably have to be the case if the hypothesis was proved for all partners as a group, but not for an individual partner. It is thought that this could be construed as potentially damaging to that University despite the fact that this is not a casual relationship. Ideally, if all partners show a statistically significant correlation then the data sample from each will be large enough for the data to be released by University, thereby giving a snapshot of different types of University as well as type of use.

One of the major issues for the project so far has been to ensure that legal regulations and restrictions are being adhered to. The data we intend to utilise for our hypothesis is sensitive on a number of levels, and while we were already making efforts to ensure there is anonymisation, we have liaised with JISC Legal and both the University of Huddersfield's Legal Officer and Records Manager. We need to use identifiers to ensure we match the correct usage data to the degree result, but once the data is combined the identifier will be removed, and we are excluding data from small courses to ensure there is no traceable data back to student level. Student notification of data use is publicised via a revised version of a statement designed by the Using OpenURL Activity Data Project at EDINA [11] for its collaborators; it is intended to include the statement into the fair processing notice for future research use.

Analysis of data (months 1-6)

All partners have agreed to hold a number of focus groups in order to collect qualitative data from students on library usage. The initial project meeting agreed a set of questions that each partner would use in addition to guidelines for holding focus groups in the absence of any individual University guidelines on ethics. It is hoped that the data collected from the focus groups will complement the quantitative data and provide a more holistic picture of how students engage with library resources; this may form the basis for future project work.

As part of the project, Huddersfield recruited a research assistant to analyse the data using appropriate statistical methods and collate and theme the issues from the focus groups.

In the original project bid it was hoped to analyse National Student Survey data at course level, with a view to finding a correlation between satisfaction levels, library activity data and student attainment. However, as each University has a different Faculty/School set up, it may not be possible to release as part of the final report.

Evaluation and exit strategy (month 6)

The Library Impact Data Project has a finite scope and goal, that of proving or disproving the hypothesis.

With this in mind, some of the potential outcomes from analysing the data that does not specifically relate to the hypothesis may have to be left for future projects to investigate. One of the outcomes of the project, to release the data on an Open Data Commons Licence [12] will allow others to exploit the data further.

However, the hypothesis if proved or partly proved may have potentially far reaching implications and could be taken forward in a number of ways, which will be further explored in the final report:

- Liaison with other projects in the strand with a view to sharing data in the future, such as the STAR-Trak Project at Leeds Metropolitan University [13], which is looking to develop a student tracking and retention system
- Liaison with publishers about linking the results of the project with their usage data
- Use of the results of the hypotheses testing to develop generic and subject specific information skills sessions to increase library use over time in areas of non/low use and target promotion of resources at point of need
- Use of the methodology and experience from partners to develop a generic toolkit based on proven results
- To investigate the impact on Learning and Teaching strategy

Conclusion

The Library Impact Data Project is now at the halfway point; all indications show the project will have enough data to be able to prove that:

'There is a statistically significant correlation across a number of universities between library activity data and student attainment'

for at least some of the elements and that data collected from the partners is similar to that already collected at Huddersfield.

Further information on the project will be available in the final report at the end of July, with the project blog and Twitter (using the #lidp hashtag) being used to post regular developments and lessons learned. Details of all members of the project team can be found on the Library Impact Data Project blog.

References

1. White, S. and Stone, G. Maximising use of library resources at the University of Huddersfield. In: UKSG 33rd Annual Conference and Exhibition, 12-14 April 2010, Edinburgh International Conference Centre http://eprints.hud.ac.uk/7248/
2. Activity data: JISC: http://www.jisc.ac.uk/whatwedo/programmes/inf11/activitydata.aspx
3. White, S. and Stone, G. Maximising use of library resources at the University of Huddersfield. Serials, 2010, 23 (2). pp. 83-90 http://eprints.hud.ac.uk/7811/
4. Lord Browne's Review of Higher Education Funding and Student Finance: http://hereview.independent.gov.uk/hereview/
5. Goodall, D. and Pattern, D. Academic library non/low use and undergraduate student achievement: a preliminary report of research in progress. Library Management, 2011, 32 (3) http://eprints.hud.ac.uk/7940/
6. Evernote shared notebook: http://www.evernote.com/pub/andrewmcgregor/jiscad#
7. Library Impact Data Project Blog: http://library.hud.ac.uk/blogs/projects/lidp/
8. JISC Activity Data Synthesis Project Blog: http://blog.activitydata.org/
9. Kramer, L. and Kramer, M. The College Library and the Drop-Out, College and Research Libraries, 1968, 29 (4), pp.310-312
10. JISC MOSAIC: http://www.sero.co.uk/jisc-mosaic.html
11. Using OpenURL Activity Data: http://edina.ac.uk/projects/Using_OpenURL_Activity_data_summary.html
12. Open Data Commons: http://www.opendatacommons.org/
13. Leeds Met STAR-Trak Project: http://leedsmetstartrak.wordpress.com/

Author Details

Graham Stone
Electronic Resources Manager
University of Huddersfield

Email: g.stone@hud.ac.uk

Bryony Ramsden
Research Associate
University of Huddersfield

Email: b.j.ramsden@hud.ac.uk

Dave Pattern
Library Systems Manager
University of Huddersfield

Email: d.c.pattern@hud.ac.uk

Keywords: please list in order of importance the keywords associated with the content of your article (usually up to 8). These will not be made visible in the text of your article but may be added to the embedded metadata:

1. *Activity data*
2. *Library usage*
3. *Electronic resources*
4. *Books loans*
5. *Student attainment*
6.
7.
8.

Finally, it helps if you would kindly update the Word document properties before submission - in particular the Title and Author name(s). Thanks!