

Library Impact Data Project

Summary

This project aims to prove a statistically significant correlation between library usage and student attainment. Using activity data from three separate systems and matching these against student records which are held in a fourth system, this project will build on in-house research previously undertaken at the University of Huddersfield. By identifying subject areas or courses which exhibit low usage of library resources, service improvements can be targeted. Those subject areas or courses which exhibit high usage of library resources can be used as models of good practice.

The partner Universities represent a cross-section of size and mission and will provide a rich data set on which to work.

Hypothesis

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

Project Partners

- University of Huddersfield
- University of Bradford
- De Montfort University
- University of Exeter
- University of Lincoln
- Liverpool John Moores University
- University of Salford
- Teesside University

Appropriateness and Fit to Programme Objectives and Overall Value to JISC Community

Background

The current economic climate, the recent Comprehensive Public Spending Review and the Lord Browne's Review of Higher Education Funding and Student Finance¹ are all putting pressure on Universities to maximise use of their resources and ensure value for money. 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. The Library Impact Data Project (LIDP) addresses all of the above.

LIDP relates to the User Activity Data category within the Activity Data strand, by analysing users' actions with regards to library usage and then linking those to final degree award. By identifying subject areas or courses which exhibit low usage of library resources, service improvements can be targeted. Those subject areas or courses which exhibit high usage of library resources can be used as models of good practice.

Internal investigation at the University of Huddersfield suggests a strong correlation between library usage and degree results, and also significant under usage of expensive library resources. Three indicators of library usage were used:

- Access to e-resources (using logins to *MetaLib* as a measure)
- Book loans (using statistics from the *Horizon* Library Management System)
- Access to the library (using gate entry statistics from *Sentry*)

Data was analysed for over 700 courses over four years - 2005/6 – 2008/9. It was then matched against the student record system (SITS:Vision) and anonymised. The correlation between library usage and final degree was particularly strong in relation to access to e-resources and book loans, and exists at School and course level.

Course level charts were produced for each School showing non/low use for all students per year group over the four year period. Academic and library staff are currently exploring 'hotspots' of apparently low or no use for further in-depth work with a purposively selected sample of around 12-16 courses. Incorporating qualitative non-library data about student experience is also possible within the School of Human and Health Sciences by working alongside the School's in-house researcher for student engagement and student retention.

In April 2010 the initial findings were presented at the UKSG Conference^{2,3}. It was highlighted that the correlation between library usage and grade had not yet been significance tested and that it was not known whether Huddersfield was an anomaly or the norm. To this extent other academic institutions were invited to benchmark against the findings. As a result a number of universities from across the HE sector expressed an interest in becoming partners in this project.

Aims and Objectives

The main aim of the project is to prove the hypothesis that there is a statistically significant correlation across a number of universities between library activity data and student attainment. The project acknowledges however, that the relationship between the two variables is not a causal relationship and there will be other factors which influence student attainment.

Benefits to the wider community

- Greater understanding of the link between library activity data and student attainment
- Publication of methodology to allow other HE Institutions to benchmark their data
- The proof of the hypothesis can be fed back into earlier work from studies on non/low use in order to target improvement in these areas. Techniques would include:
 - Course profiling
 - Targeted promotion of information resources
 - Raising tutor awareness of the link between use of resources and attainment
 - Improved targeting of resources allocation
- Demonstration that activity data from a number of separate systems can be 'joined up' to provide valuable management information
- National Student Survey data can be matched to attainment and used to see if there is a further correlation

Overall Approach

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

The required activity data should span at least one entire academic year (e.g. 2009/10), or ideally for multiple years if historic data is available. For each student who graduated in that year, the following data would be required:

- Final grade achieved
- Number of books borrowed
- Number of times e-resources were accessed , e.g. the number of times the student accessed e-resources through authentication such as Athens, Shibboleth, EZProxy or MetaLib (with the caveat that it is a crude but common measure of actual e-resource usage)

- Number of times each student entered the library, e.g. via a turnstile system that requires identity card access
- School/Faculty. At Huddersfield, differences in library usage behaviour have been noted between departments and it would be interesting to know if this was replicated across other institutions

With multiple years of data it would be possible to analyse library usage by students in each year of their course. Huddersfield data suggests that library usage changes as student's progress in addition to differences noted between departments. This builds on research from the 1960s⁴ that suggests that certain disciplines have less need for books, extending this for the first time into the need for electronic resources. Huddersfield will collate the data, normalise it, and then analyse it from month 3 of the project onwards.

In addition all partners would be asked to hold a number of focus groups in order to secure qualitative data from students on library usage. This would complement the quantitative data and provide a more holistic picture of how students engage with library resources.

Finally, analysis of National Student Survey data will be conducted at course level, with a view to finding a correlation between satisfaction levels, library activity data and student attainment.

Project Outputs

The primary outputs will be

- the final project report and project blog
- dissemination events including seminars/conferences
- to make all anonymised data available for re-use under an Open Data Commons Licence (the project aims to build on the recommendations of the MOSIAC project⁵)

The project will work with the synthesis project and other projects in the strand with a view to sharing common data as the project progresses, e.g. activity data from this project could be matched against the proposed SALT (Surfacing the Academic Long Tail) project and vice versa.

Project Outcomes

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

The project will be a success if the following measurable targets are achieved:

- Sufficient data is successfully captured from all partners
- Statistical significance is proved for all 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

The project will examine whether different levels of library use by subject areas are replicated across institutions. Focus groups will explore these issues in more details, and may form the basis for future project work.

Quality of Proposal and Robustness of Workplan

Workpackages

Work package 1: JISC Reporting Schedule	
Project Plan, Work Plan, Budget	Month 1
JISC website project page and project blog	Month 1
5 days time dedicated to working with the synthesis project	Months 1-6
Interim reports to be delivered as appropriate	Months 2-6
Final Report/ Completion Report	Month 6
Project Web site and associated reports and data to be archived in the appropriate University Repositories	Month 6
Work package 2: Data collection	
Supply partners with details of activity data required	Month 1
Seek advice from JISC Legal regarding open data and data protection	Month 1
Partners to supply activity data for collation	Month 2
Release of data under Open Data Commons Licence	Month 3-4
Work package 3: Analysis of data	
Recruit Research Assistant	Month 1-2
Analysis of data	Month 3-6
Analysis of National Student Survey Data	Month 3-4
Focus groups	Month 3-4
Collation of focus group data	Month 5
Dissemination of results	Month 6
Workpackage 4: Sustainability Planning – working with future contributors	
Identify contributors/ other projects in the strand, e.g. SALT (Surfacing the Academic Long Tail)	Month 6
Participate in meeting/workshop to collect feedback	Month 6
Issues and recommendations report	Month 6

Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Inability to recruit a suitable and competent research assistant	2	2	4	Source alternative staff
Retention of staff	2	2	4	Motivation via good working environment and personal development
Failure to get all parties to share same understanding of purpose (project team)	1	5	5	Use sound project methodology
Partners unable to provide sufficient data	2	5	10	Agree data to be submitted at project start Use data from other projects if applicable
Internal technical issues	2	4	8	Include appropriate staff in project team Provide workarounds
Failure to observe data protection issues	2	5	10	Ensure correct permission received at start of project; Consult with JISC Legal

				Follow guidelines from Activity Data Q&A with Professor Charles Oppenheim and Naomi Korn ⁶
Other internal projects that might have an impact	3	3	9	Include key staff in project and steering groups Schedule key staff time according to work packages
Organisational changes that might take place during the project	2	4	8	Include key staff in project and steering groups
Data is not statistically significant	3	2	6	It is expected that there may be some issues with some of the data Whatever, the outcome, the hypothesis will either be proved or disproved

Standards

The project will follow the guidelines and recommendations laid out in the Mosaic Final report⁵.

Intellectual Property Rights

All IPR generated in the course of the project will be owned by the partner universities under terms negotiated between them. The universities are committed, nevertheless, to full public disclosure and dissemination of the results of the project within the sector.

Project Management

The project will be managed in accordance with the Universities' standard project management methodology and the JISC's best practice guidelines. A project team will be established consisting of members drawn from all partners and will report to senior University managers and the JISC as appropriate. Support will be provided by the JISC Programme Manager in the first instance, including notification of relevant events or developments. The project will set aside up to 5 days to work with the synthesis project.

Exit and Sustainability Plans

The work on linking library activity data to student attainment in the scope of this project has a finite goal. However, the hypothesis if proved or partly proved will have potentially far reaching implications. Planned outputs of the project will be taken further as follows:

- Release of the data on an Open Data Commons Licence for others to exploit further
- Release of the types of activity data required and methodology for others to test their data
- Liaise with other projects in the strand with a view to sharing data in the future
- Liaise with publishers about linking the results of the project with their usage data
- Use the results of the hypotheses testing to develop generic and subject specific information literacy sessions to increase library use over time in areas of non/low use and target promotion of resources at point of need
- Use the methodology and experience from partners to develop a generic toolkit based on proven results
- Work with partners to look at the impact on Learning and Teaching strategy

Engagement with the Community

Stakeholder Analysis

Stakeholder	Interest / stake	Importance
JISC	Funding body	High
Vice Chancellors	Universities Strategic Plan	High
Pro Vice Chancellors – Learning and Teaching Teaching staff Undergraduates	Utilisation of results to enhance student attainment	High
Wider community	Release of results and methodology	Medium
Library	Better targeting of resources	High

Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Months 1-2	Definition of activity data required sent to partners	Can the partners provide the data?	Observation	All data supplied
Months 3-5	Data analysis	Is the data in an acceptable form?	Statistical significance testing	Successful manipulation of data
Months 3-4	Focus Groups	Did the groups provide useful quantitative data?	Analysis of responses	Relevant quantitative data
Months 6	Hypothesis testing	Is the hypothesis proved	Statistical significance testing	Positive outcome

Quality Plan

- Quality of data. Each partner's data to be checked.
- Quality of analysis. All data to be tested for statistical significance. Further assistance to be sought from JISC as appropriate.

Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
Feb-Jul 11	University newsletters	University Staff	Raise Awareness	News item to outline project
Feb – Jul 11	Project Web pages JISC Web Pages	University Staff JISC Community	Raise Awareness	Project plan and report to outline key aims and objectives
Jun-Jul 11	Internal University Learning and Teaching Conferences	University Staff	Engage/Promote	Workshop to show hypothesis
Jun-Jul 11	Conference papers, e.g. interest from Business Librarians' Association Conference 2011	Academic Community, JISC Community	Inform	Inform about lessons learned
Jul 11	Contributors workshop in collaborative with the proposed SALT projected	JISC Community and future projects	Collaborate/Inform	Data sharing

Jul 11	Publish findings via green/gold open access	International community	Inform	Inform about lessons learned
Jul 11	Final Report to the JISC	JISC and future projects	Inform	Inform about lessons learned

References

1. Lord Browne's Review of Higher Education Funding and Student Finance: <http://hereview.independent.gov.uk/hereview/>
2. White, S and Stone, G (2010) *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/>)
3. White, S and Stone, G (2010) *Maximising use of library resources at the University of Huddersfield*. *Serials*, 23 (2). pp. 83-90 (<http://eprints.hud.ac.uk/7811/>)
4. Kramer, L. and Kramer, M. (1968) The College Library and the Drop-Out, *College and Research Libraries*, 29 (4), pp.310-312.
5. The JISC MOSAIC Project, Making Our Scholarly Activity Information Count. *Final Report*, January 2010. (http://www.sero.co.uk/mosaic/100322_MOSAIC_Final_Report_v7_FINAL.pdf)
6. Oppenheim, C and Korn, N, *Questions and answers on data protection issues associated with the use of activity data*, Gaining Business Intelligence from User Activity Data, 14 July 2010, The Hatton Conference Centre, London.