Can Enhanced Library Analytics Improve the Understanding of User Behaviours at City, University of London?

Christine Goodson

September 2018

Submitted in partial fulfillment of the requirements for the degree of MSc in Library Science

Supervisor: David Bawden
Abstract

Data that is automatically generated by users in an academic library as they engage with various library resources is often held in disparate systems. This prevents library staff from knowing how an individual uses these resources in combination and therefore opportunities to provide more targeted support or information may be missed.

This case study examining the linking of quantitative data sets held at the libraries of City, University of London uses data held in the following systems: the library management system, Sierra, for patron details and item borrowing; OpenAthens logins and WAM connections (web access management) for e-resource use; and LibApps for space bookings for study rooms, silent study spaces and specialist databases. The study is limited to students at the university and excludes staff, visitors and alumni.

Successful linking of the data sets allowed information to be combined and visualized to explore if low item borrowing correlates to high e-resource use, which was not found to be the case generally. Results also support previous studies’ findings regarding low use of library resources across computing and engineering undergraduates. The combined data has the capacity to provide in-depth understanding of what, how and when resources are utilized by students and provide a more targeted path for further qualitative investigations to answer questions of why.
Acknowledgements

My thanks to David Bawden, my supervisor for his guidance and advice. Thanks also to all the library staff at City whose interest, support and expertise provided me with insight and kept me buoyed through some of the more tedious parts of cleaning data and in particular those who took time to run reports or explain the detail of work already undertaken. Special thanks to Derek MacKenzie, Head of User Services and Lucy Clifford, Library Systems Manager for their time, expertise and smoothing the way for this project.
Table of contents

Abstract ................................................................................................................................................ 2
Acknowledgements ......................................................................................................................... 3
Abbreviations .................................................................................................................................... 6
List of tables ....................................................................................................................................... 7
List of figures ...................................................................................................................................... 7
Chapter 1. Introduction ................................................................................................................ 10
  1.1 Background ................................................................................................................................................... 10
  1.2 Research focus .............................................................................................................................................. 11
  1.3 Overall research aim and research objectives .............................................................................. 13
  1.4 Value of this research ............................................................................................................................... 13
  1.5 Disclosure ....................................................................................................................................................... 13
  1.6 Ethics ................................................................................................................................................................ 14
Chapter 2. Literature Review ..................................................................................................... 15
  2.1 Introduction .................................................................................................................................................. 15
  2.2 Defining analytics ....................................................................................................................................... 15
  2.3 Where does the data come from? ........................................................................................................ 16
  2.4 Library assessment, impact and value .............................................................................................. 17
  2.5 Measuring e-resource use ....................................................................................................................... 22
  2.6 Future directions ........................................................................................................................................ 23
  2.7 Personas .......................................................................................................................................................... 25
  2.8 Issues arising ................................................................................................................................................ 25
Chapter 3. Methodology ........................................................................................................... 28
  3.1 Introduction .................................................................................................................................................. 28
  3.2 Research Strategy ...................................................................................................................................... 28
  3.3 Data Collection ........................................................................................................................................... 29
    3.3.1 Item borrowing ........................................................................................................................................ 34
    3.3.2 WAM connections .......................................................................................................................... 35
    3.3.3 OpenAthens logins ........................................................................................................................... 36
    3.3.4 Space bookings ...................................................................................................................................... 36
    3.3.5 Assembling the data ......................................................................................................................... 37
  3.4 Framework for data analysis ........................................................................................................... 38
3.4.1 Library personas........................................................................................................................39
3.5 Limitations and potential problems ..............................................................................................43
  3.5.1 WAM connections information .........................................................................................43
  3.5.2 Room/Space Bookings .........................................................................................................44
  3.5.3 Item borrowing information ...............................................................................................45
  3.5.4 OpenAthens login information ............................................................................................45
  3.5.5 General limitations ................................................................................................................46

Chapter 4. Findings and discussion ..................................................................................................47
  4.1 Introduction..................................................................................................................................47
  4.2 Individual data sets .....................................................................................................................47
    4.2.1 Patron Record.......................................................................................................................47
    4.2.2 Item borrowing.....................................................................................................................49
    4.2.3 WAM connections................................................................................................................55
    4.2.4 OpenAthens logins................................................................................................................61
    4.2.5 Space Bookings...................................................................................................................55
  4.3 Combined data sets .....................................................................................................................66
    4.3.1 WAM downloads and item borrowing ..............................................................................66
    4.3.2 Combined resource engagement.........................................................................................74
  4.4 Discussion......................................................................................................................................83

Chapter 5. Conclusion ......................................................................................................................86

Chapter 6. References ......................................................................................................................89

Appendices .......................................................................................................................................95
  Appendix A: Supplementary data.....................................................................................................95
    Appendix A-1 Number of anomalous hours of WAM connections data.................................95
    Appendix A-2 Number of missing time stamps in WAM hourly use and their percentage of the total month.................................................................95
    Appendix A-3 Users of WAM connections who do not download items by student type ..................................................................................................95
    Appendix A-4 Comparison between persona 5 and all SHS undergraduates .......................96
  Appendix B: Original proposal and ethics checklist........................................................................98
  Appendix C: Reflection......................................................................................................................112
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALIEN</td>
<td>Automated Library Information Exchange Network</td>
</tr>
<tr>
<td>COUNTER</td>
<td>Counting Online Usage of Networked Electronic Resources</td>
</tr>
<tr>
<td>CPD</td>
<td>Continuing Professional Development</td>
</tr>
<tr>
<td>CPPD</td>
<td>Continuing Personal and Professional Development</td>
</tr>
<tr>
<td>Jisc</td>
<td>Previously stood for Joint Information Systems Committee, they are ‘the UK higher, further education and skills sectors’ not-for-profit organisation for digital services and solutions’ (Jisc, no date)</td>
</tr>
<tr>
<td>LIDP</td>
<td>Library Impact Data Project</td>
</tr>
<tr>
<td>PG</td>
<td>Postgraduate</td>
</tr>
<tr>
<td>SA&amp;SS</td>
<td>School of Arts and Social Sciences</td>
</tr>
<tr>
<td>SCONUL</td>
<td>Society of College, National and University Libraries</td>
</tr>
<tr>
<td>SHS</td>
<td>School of Health Sciences</td>
</tr>
<tr>
<td>SMCSE</td>
<td>School of Mathematics, Computer Science and Engineering</td>
</tr>
<tr>
<td>STEM</td>
<td>Science Technology Engineering and Mathematics</td>
</tr>
<tr>
<td>UG</td>
<td>Undergraduate</td>
</tr>
<tr>
<td>WAM</td>
<td>Web Access Management</td>
</tr>
</tbody>
</table>
List of tables

Table 1 Unique identifiers for each data set ................................................................. 33
Table 2 Library personas for City, University of London .................................................. 40
Table 3 Statistical summary of total length of item borrowing ....................................... 50
Table 4 Statistical summary of student WAM downloads ............................................. 56
Table 5 Statistical summary for hourly WAM use ......................................................... 58
Table 6 Comparison of WAM users that do not download ......................................... 60
Table 7 Statistical summary for OpenAthens logins ....................................................... 61
Table 8 Statistical summary for space bookings ............................................................ 64
Appendix A-1 Table 9 Number of anomalous hours of WAM connections data per day 95
Appendix A-2 Table 10 Number of rows without time stamps in WAM connections reports 96
Appendix A-3 Table 11 Users of WAM connections who did not download items split by student type 96

List of figures

Figure 1 Tables and fields for created for Access database with unique identifiers marked in red _ 38
Figure 2 Distribution of user records in Sierra by school and student type for August 2017 to June 2018 48
Figure 3 Percentage distribution of user records in Sierra by school for August 2017 to June 2018 49
Figure 4 Distribution of the total length of item borrowing by students from August 2017 to June 2018 ................................................................. 50
Figure 5 Distribution of total length of item borrowing by school (non-linear bin selection) 51
Figure 6 Total length of borrowing split by student type and School ............................. 52
Figure 7 Comparison of the number of students in each school who borrow items and those who do not for the period August 2017 to June 2018 ................................................................. 53
Figure 8 Comparison of borrowing and non-borrowing, with the later separated by the presence of a borrowing history in the Sierra patron record ................................................................. 54
Figure 9 Frequency of total WAM downloads for all students across August 2017 to June 2018 55
Figure 10 Number of downloads by school and student type ................................. 57
Figure 11 Comparison of student types and the distribution of download frequency ........................................................................................................ 57
Figure 12 Distribution of the number of WAM uses registered for a student across August 2017 to June 2018 using the hourly sampling ................................................................. 59
Figure 13 Distribution of WAM use by School ............................................................. 60
Figure 14 Distribution of OpenAthens logins per student across August 2017 to June 2018 62
Figure 15 Number of OpenAthens logins by school and student type .......................... 62
Figure 16 Distribution of the OpenAthens logins per student, separated by student type 63
Figure 17 Distribution of hours booked for library spaces per student across August 2017 to June 2018 64
Figure 18 Use of space types by school 66
Figure 19 Comparison between total days of book borrowing and WAM downloads for the five main schools 67
Figure 20 Comparison between total days of book borrowing and WAM downloads for undergraduates in the five main schools 68
Figure 21 Comparison between total days of book borrowing and WAM downloads for postgraduates in the five main schools 68
Figure 22 Item borrowing and WAM downloads for persona 1: part time MSc Finance and Investment, postgraduate at Cass Business School 69
Figure 23 Item borrowing and WAM downloads for all postgraduates in Cass Business School 69
Figure 24 Item borrowing and WAM downloads for persona 2: full time BA Business Management, undergraduate in Cass Business School 70
Figure 25 Item borrowing and WAM downloads for all undergraduates in Cass Business School 70
Figure 26 Item borrowing and WAM downloads for persona 3: full time BSc Computer Science, full time undergraduate in SMCSE 71
Figure 27 Item borrowing and WAM downloads for persona 4: full time BEng Civil Engineering, full time undergraduate in SMCSE 71
Figure 28 Item borrowing and WAM downloads for all undergraduates in SMCSE 71
Figure 29 Item borrowing and WAM downloads for persona 5: full time BSc Mental Health Nursing, undergraduate in SHS 72
Figure 30 Item borrowing and WAM downloads for all undergraduates in SHS 72
Figure 31 Item borrowing and WAM downloads for persona 6: full time MA Newspaper Journalism, full time postgraduate in SA&SS 73
Figure 32 Item borrowing and WAM downloads for all postgraduates in SA&SS 73
Figure 33 Item borrowing and WAM downloads for all undergraduates in City Law School 74
Figure 34 Resource engagement by undergraduates with courses expiring after 2019, shown as a percentage of all undergraduates within the given school 76
Figure 35 Resource engagement by all undergraduates in the five main schools, shown as a percentage of all undergraduates within the given school 76
Figure 36 Venn diagram showing resource engagement by students across August 2017 to June 2018 (not mathematically to scale) 78
Figure 37 Venn diagram showing resource engagement by Cass Business School postgraduates 79
Figure 38 Venn diagram showing resource engagement by Cass Business School undergraduates 79
Figure 39 Venn diagram showing resource engagement by SMCSE undergraduates 80
Figure 40 Venn diagram showing resource engagement by SHS undergraduates 81
Figure 41 Venn diagram showing resource engagement by SA&SS postgraduates 81
Figure 42 Venn diagram showing resource engagement by City Law School undergraduates 82
Figure 43 Infographic highlighting indicating resource engagement within the five main schools at City, University of London
Appendix A-4 Figure 44 Item borrowing and WAM downloads for persona 5: full time BSc Mental Health Nursing, undergraduate in SHS
Appendix A-4 Figure 45 Item borrowing and WAM downloads for all undergraduates in SHS
Chapter 1. Introduction

1.1 Background

Libraries have a long history of recording information about themselves: counting the number of books in the collection; the number of registered users and how many seats in its reading rooms are a few examples. And while it is essential to have a clear view of a library’s inventory, counting things does not demonstrate use. Thus, a more detailed view using such measures as circulation statistics to assess which books are being borrowed and for how long, head counts or card swipes indicating the degree to which physical space is being used are more illuminating. It is still the case that a borrowed book is not necessarily a read book or its content understood by the borrower but it does at least signify some degree of intent.

Academic libraries are accustomed to this type of assessment; questioning and investigating how well they are serving the needs of its users, employing a range of techniques including quantitative analysis of data recorded in the different library systems alongside qualitative analysis of such items as feedback forms or live messaging support. There is a great deal of literature available regarding suitable methods of library assessment, numerous books and a biennial conference (Library Assessment Conference), increasingly though there has been a marked shift toward the concept of library value and impact assessment i.e. the contribution the library makes toward supporting its home institution’s strategic objectives (Madsen and Hurst, 2018). This shift emphasises the library’s requirement not just to be good at what it does as measured by such standards as Customer Service Excellence accreditation (SCONUL, 2018) but also to align its activities toward its home institution’s goals.

Within an academic library setting, users will automatically create a trail of data as they interact with the different resources available. This includes use of swipe cards for entry gates, computer logins, library catalogue searches, accessing e-
resources, printing, enrolment of library instruction classes, room bookings, book borrowing, returns and renewals, inter-library loan requests, queries raised using library chat applications and this list is by no means exhaustive. Library analytics affords the opportunity for those responsible for managing library services to develop a greater understanding of its users: who is using what resources and how they are navigating them; who is not using the services and are there any common factors; where the peaks and troughs in demand occur and how they fluctuate from year to year. This kind of detailed understanding means support can be more efficiently targeted whether through instruction, staffing levels at help points or provide the basis for further investigation particularly with regard to low or non-users of the library.

1.2 Research focus

This project is focused on data automatically collected or generated at the libraries of City, University of London, which has four library locations:

- Northampton Square
- Cass Learning Resource Centre
- City Law School Library (Gray’s Inn Place)
- City Law School Library (Northampton Square, Innovation Centre)

(Note that City Law School Library has since moved location but this did not occur until after the study had concluded)

The library staff at City currently use analytics to generate and visualise information based on a variety of data sets including gate swipes, a number of different website analyses, room bookings and circulation statistics to name a few. Additionally service desk and library chat enquiries are logged and categorised to indicate the types of query asked and whether they were resolved directly or passed to another service. There are also a number of library feedback initiatives across the academic year that generates qualitative
information regarding how users see the library. Some of these data sets have been used to create library personas, a fictional character used to represent a group of users. Their primary role is to assist in user experience investigations and problem solving as they help ‘keep staff focused on the needs of actual patrons, by putting a human face on a lot of disparate data’ (Borg and Reidsma, 2016, p. 46).

Many of the data sets do not contain any information that could identify the user, in keeping with the Code of Professional Practice and the Ethical Principles that library, information and knowledge professionals adhere to (CILIP, 2017). However, those data sets that do contain user identification are often separate from one another as they are held in different administrative systems and consequently it has not been possible to combine them to produce a more integrated picture of how individuals utilise the available resources.

These information silos mean it is not possible to definitively connect borrowing of physical items from the library with the use of electronic resources at an individual level or assess if heavy book borrowing correlates to regular visits to the library. Thus, this research will focus on combining data sets at the user level as this has the potential to unlock insights into user behaviours not previously accessible at this university’s libraries.
1.3 Overall research aim and research objectives

The overall aim of this project is to refine the understanding of student user behaviour of library resources at City, University of London using pre-existing, automatically generated data sets and combining them at the user level to generate a more detailed data set that can answer a greater range of questions.

The objectives are:

1. Explore what data sets are available and what information they contain.
2. Establish if the data sets can be successfully combined.
3. Examine the kind of questions that the combined data can answer.
4. Evaluate whether the combined data provides evidence to support the underlying assumptions regarding the pre-existing work on library personas.
5. Assess whether the combined data adds further understanding to user behaviour.

1.4 Value of this research

This work will provide evidence of whether student users at City conform to behaviours seen in previous studies conducted at other institutions, detail some of the issues arising from the individual data sets and provide potential paths for further investigation.

1.5 Disclosure

The author is employed by City, University of London within the library as a part-time service assistant. This investigation does not from any part of that paid employment and there has been no interference or influence regarding editorial content except where staff opinion has been specifically requested and has consequently been recorded as such.
1.6 Ethics

With regard to data management and evaluating potential privacy issues, the Head of Users Services and the Library Systems Manager were consulted and kept informed of the various investigations and the data being extracted. All access to data was conducted on site via library staff computers and no data was accessed, analysed or held on the author's personal equipment or in any portable memory format until it had been fully anonymized and aggregated. This prevented any accidental loss of data.
Chapter 2. Literature Review

2.1 Introduction

This Literature Review will begin by providing a working definition for library analytics and review the reasons for its rise in use and key findings. The overall focus will be on academic libraries in the US, UK and Australia though contributions to this field are spread over a far wider number of territories. The issue of user privacy will not be covered here, but it should be noted that all the library research that has linked library data with individual user information e.g. course completion, have had to conduct their own assessment of how privacy was managed and a summary of considerations is given in Showers (2015, pp. 153-167). A short section regarding personas is also included as this forms part of the research objectives. This chapter will set the context for this investigation and provide a point of triangulation for the case study undertaken.

2.2 Defining analytics

It is useful to have a clear view on what is meant by analytics and despite its prevalence in a variety of guises such as web analytics, business analytics and more recently learning analytics, it appears to lack a precise or absolute definition. In the introduction to his book on library analytics, Showers defines analytics as being:

about analysing data to uncover information and knowledge (discovery) and using these insights to make recommendations (communication) for specific actions or interventions. (Showers, 2015, p. xxx).

They also note that communicating the insight is often done through the use of visualisations; where complex data is represented in a visual and meaningful way. Such an example can be seen in work undertaken by Khoo et al. (2014) where heat maps are used to visualise seat occupancy in an academic library.
Austin Booth and Hendrix suggest that:

*Analytics is not just data, it is a process to use data to answer strategic questions and make strategic decisions in order to produce strategic results.* (Austin Booth and Hendrix, 2015, p. 695).

Barneveld et al. acknowledge the term has different meanings in different contexts and propose the following definition for analytics:

*An overarching concept that is defined as data-driven decision making* (Barneveld et al., 2012, p. 12).

It has also been described by Campbell et al. as follows:

*Analytics marries large data sets, statistical techniques, and predictive modeling. It could be thought of as the practice of mining institutional data to produce "actionable intelligence."* (Campbell et al., 2007, p. 42).

Though the definitions vary, there is a general theme of analysing data and acting on the results. For purposes of this investigation, the view taken by Showers (2015, p. xxx) will be utilized. Though the term itself was not commonly used in the research reviewed for this investigation, it nevertheless conforms to the definition as it refers to analysis of data collected within the library setting and the actions or interventions that are recommended from the insights gained.

### 2.3 Where does the data come from?

Farmer and Safer (2016, pp. 20-21) list a large number of points where data is generated along with a breakdown of the types of information that could be extracted and utilized for analysis which are summarised below:

- Acquisitions: tasks and processes.
- Collection holdings: bibliographic data, missing and lost items, additions and withdrawals.
- Circulation: borrowing records, reservations and holds, borrower demographics.
- Computer hardware and software: quantity and specification, access by users.
• Internet: technical data such as connectivity or speed.
• Other technologies: includes printers, copiers and assistive technology support items. Access and usage data.
• Library website: usage, updating.
• Library publications: dissemination, quantity or hits if online, costs.
• Reference: query types, time frame, length of interaction.
• Instruction: quantity, time, number in group.
• Internal communication: type, function.
• Staffing: quantity, organisational structure.
• Staff training and development: purpose, cost, frequency.
• Facilities: usage patterns, area, location.

The data generated by many of these items are routinely analysed in libraries with regard to ensuring cost effective resource provision but may not always be used to analyse user behaviours.

**2.4 Library assessment, impact and value**

In 2010, the ACRL (Association of College and Research Libraries) released a prescient report, the Value of Academic Libraries (Oakleaf, 2010). Its content highlighted the need for academic libraries to demonstrate their value to the higher education system rather than passively assume it was intrinsic, the report went on to provide research methods and metrics that could be used to achieve this. It was a direct call for action, which has inspired a significant body of scholarly literature. Oakleaf has subsequently followed this up with a further work refining and updating the concepts laid out in the original report (see Oakleaf and Kyrillidou, 2016; Oakleaf 2017).

Research published by practitioners since the ACRL report has served a number of purposes; its primary function is most often to provide evidence of the library’s value to its institution by tying various library metrics to some form of student success measurement. However, there are additional aspects that may
not always be explicitly acknowledged; as the body of research builds and increases, it begins to create more general themes that have the potential to be applied to and used by libraries that are not necessarily in a position to conduct some of the types of research for themselves. This is seen with findings that indicate use of e-resources and book borrowing have a statistically significant correlation with student grade attainment (White and Stone, 2010), which is detailed later in this chapter. Thirdly, the publication of this type of research provides a body of knowledge to facilitate other libraries conducting their own research; clear research questions and methodology are a feature in many of papers reviewed, and clarity regarding data limitations or problems encountered assist other prospective researchers in refining their own proposals to maximise their efficacy. Finally, where no correlation is found, it provides insight regarding those aspects less likely to be fruitful for other institutions though it may still be worth considering inclusion if it does not add undue work to the investigation.

As identified by Schweider and Hinchcliffe (2018), the range of investigations tend to fall into one of three main types study which are summarised below:

- **Small-group studies**: typically investigate some form of library service offered to students for example through single instructional sessions, online modules or supplying LibGuides. Often the studies relate to specific courses but sessions on information literacy skills were not necessarily discipline specific. Evaluation is conducted to assess the effectiveness of the instruction. Examples include work conducted by Horn et al. (2013) who examined the effect of embedded practice on students from low socioeconomic status backgrounds. Booth et al. (2015) targeted first year students to assess their levels of information literacy and compared it against library instruction.

- **Single-institution studies**: examines a larger group of students at a single academic institution usually focusing on statistically analysing many cases to correlate individual library experiences with student outcomes such as grade achievement or retention rates. See White and Stone
Christine Goodson

(2010), Cox and Jantti (2012) and Soria, Fransen and Nackerud (2013) for early examples.

- Multi-institution studies: while not as common, they often undertake to show the same kinds of correlations as for single institute studies. Crawford’s (2015) work in the United States found that amongst a range of expenses, library expenditure per full-time equivalent student had strongest correlation with retention and graduation. Womack (2016) used advanced analysis on 99 US university libraries to assess their impacts on grant funding, and concludes that no effects were found.

Schweider and Hinchliffe (2018) also go to describe the benefits and drawbacks of the different types of studies. Small-group studies with sound methodologies can assist in confirming the most suitable methods of instruction or indeed if instruction in a particular aspect is of any benefit at all. However, given the effort involved in evaluating such offerings with suitable rigour, they require careful targeting and alignment with the institution’s goals if they are not to be considered too niche by institutional stakeholders. Single-institution studies of library value usually have the benefit of being large enough to resonate with key stakeholders and commonly utilize data that is already being collected such as computer logins or book borrowing. Concerns over privacy often need to be addressed but once these are satisfied library data can be joined to student information such as grades, re-enrollment (retention) and graduation. Certain library factors are difficult to isolate in single-institute studies, these can include expenditure per student or collection size, as they tend not to vary to a large extent from one year to the next. Multi-institution studies have the potential to reinforce or dispute findings made in single-institution studies and are also able to compare institutional library characteristics such as expenditure, size and hours. Difficulties can arise when seeking to obtain standardized data across different institutions when there is no formal standardization for a particular measure and can significantly limit the factors that can be investigated on a large scale.
A substantive portion of recent, practice-based literature regarding the use of academic library resources by students has related it to student attainment. This has been undertaken in part to indicate how such libraries support their institution’s goals and mission and thereby establishing its value to the institution, but it also allows for opportunities to further improve the service to the users. One of a number of key studies was undertaken at the University of Huddersfield; staff in Computing and Library Services published a paper investigating low/non-use of library resources and whether there was a correlation between library use and final degree results (White and Stone, 2010). The project selected book loans, access to e-resources, and visits to the library using gate entry data as its indicators and the initial findings indicated some positive correlations.

This work led to the Jisc funded Library Impact Data Project (LIDP), a multi-phased research project managed by the team at Huddersfield. Phase 1 investigated the hypothesis ‘that there is a statistically significant correlation across a number of universities between library activity data and student attainment’ (Jisc, 2011). The project used the same three indicators as the 2010 investigation although due to timescales and data capture issues some of the partner universities were only able to supply one or two of the indicators (Stone and Ramsden, 2013). Additionally, qualitative data was sought through the use of focus groups and potential bias was acknowledged with an expectation that low-use students would be less likely to be motivated to attend focus sessions or meetings with library staff. The reporting of the research is comprehensive and shows a strong understanding of the limitations of the data being used, including the shortcomings of measuring e-resource use and was open about what could and could not be achieved in the time frame that was available for the project. It concludes that there is a statistically significant relationship between student attainment and e-resource use and book borrowing but not for library gate entry. It was also stressed in the article, and has been in subsequent work carried out in other institutions, that this is not a causal relationship and that other factors affect student outcomes.
Findings from this work was supported by concurrent research undertaken in Australia (Cox and Jantti, 2012), where they created a dataset that joined the University of Wollongong Library resources with student data held by the university to create a tool called the ‘library cube’. The paper also contained helpful information regarding the issues encountered, specifically how e-resource use was measured and the decision to register use for each 10 minute period a student's ID was registered in.

The second phase of the LIDP looked at additional demographic data of Huddersfield undergraduates to investigate if there were causal factors for library use and drilled down further into the data to examine age, gender, ethnicity and country of domicile (Stone and Collins, 2013) and also school and course level (Collins and Stone, 2014). The first of these two investigations indicated that there is a relationship between library usage and demographic factor, which was also the case in Cox and Jantti’s research (2012). The second of the investigations found the social science group were the highest users of library resources, whereas students in computing, engineering and the arts were the lowest.

The third phase of the LIDP used the data and findings from phases 1 and 2 to explore ways to increase library usage amongst those groups identified as being low users (Stone et al., 2015). Several approaches were taken which involved engaging with teaching staff as well as students, and while informative in providing a framework for other practitioners and disclosed the level of uptake on various initiatives, the timeframe did not allow for further analysis of usage data to indicate whether it did in fact increase use of resources.

Jisc also published an article utilising results of the LIDP (Jisc, 2012) and went on to commission the Library Analytics and Metrics project (JiscLAMP) and in 2013 produced a ‘prototype shared analytics service for UK academic libraries’ (Ruddock, 2013). This early attempt to allow UK academic libraries to supply information and subsequently benchmark themselves against similar institutions
has subsequently evolved into Jisc's Learning Analytics, ‘using data and analytics to support students’ (Jisc, 2018).

2.5 Measuring e-resource use

Use of e-resources is a popular metric in the literature and has been used in a variety ways but what at first appears to be a straightforward method of measurement is not without its problems. A student is usually required to login to which confirms their authority to use a particular publication and it should be possible to capture that transaction in a way that does not unduly infringe on the users privacy. This is most commonly done by an authentication service of some sort, which allows the institution to know what publisher a user has accessed but not the precise content. The publisher simultaneously knows the precise details of the publication what was selected, such as the article or book chapter viewed or downloaded. However it only knows the institution the user a user has been authenticated by, not the individual identity. Thus if, if a publisher considers there may have been a misuse of their service, sometimes caused by a compromised user account, they are required to contact the institution to investigate further as they have no capacity to identify the individual. This can sometimes result in the halting of service until the matter has been investigated (Guajardo, Hamparian and Katz, 2018). Electronic publications supply institutions with usage reports and many subscribe to the Counting Online Usage of Networked Electronic Resources (COUNTER) code of practice. In use since 2002, it lays out guidelines for publishers and vendors on the provision of usage statistics (Bull and Beh, 2018).

COUNTER reports have the benefit of following a set of standardised measures meaning it is possible to compare use of different publishers as metric in a multi-institutional study however, it does not contain individual identifiers and prevents it from being tied directly to a user’s record. Research into understanding users’ needs and behaviours was conducted by Gustafson (2018)
and used data from COUNTER reports, Summon (a discovery service) analytics and information from reference librarians to answer its research question: ‘how can the vast usage data collected by electronic resources services empower reference to better serve our users?’ Although only an overview of its methodology was given, and little indication of what contribution the COUNTER reports made, its findings did result in changes being made for the benefit of the user. These included changing instruction from Internet Explorer to Chrome and Firefox to reflect the browsers being used on Summon. Another opportunity for changes with instruction arose from more subtle inferences from the statistics. Gustafson found that as the school year progressed, Summon data indicated an increase in visits and searches but the average search per visit decreased. It was shown that users were learning to refine their searches by using faceted searching, but it remained that had they articulated their search better this would not be necessary.

The measurement of e-resource use with a user identifier has been imperfect. The work done at the University of Huddersfield (Stone and Ramsden, 2013) acknowledged different institutions collected different data and, at the time many institutions did not collect it at all. As mentioned earlier, Cox and Jantti (2012) discussed the problems of measuring using Ezproxy logs (the system used to access licensed materials), specifically noting that simply counting entries was not appropriate as these varied depending on the resource being accessed. Their decision to register a single use every 10 minutes resolved this. Nevertheless it remains a popular measure of library use, see Cherry et al. (2013); Haddow (2013) and Thorpe et al. (2016) in addition to the previously discussed examples.

2.6 Future directions

With the increasing quantity of research being conducted on students’ use of library resources for the purpose of providing evidence as to the library’s contribution to the institution’s goals there are also a number of articles
examining the possibility of predictive analytics. Smith (2012, p. 25) defined predictive analytics as ‘a process that uses a variety of analysis and modeling techniques to discover patterns and relationships in existing data – then using the insight to make accurate predictions’. In one of the first articles regarding the use of predictive analytics in the library context, Massis (2012) suggests that the transactional nature of many core library services lends itself to the kinds of forecasting seen in the business world using predictive analytics. At Texas Tech University Libraries, Litsey and Mauldin (2018) describe their use of ‘the first library-developed machine learning/predictive analytic tool’, the Automated Library Information Exchange Network (ALIEN). Still in development, ALIEN was used to address interlibrary loans, specifically predicting the number of requests for future semesters. Results were mixed however, the system was able to recognise a change of library behaviour. While this work is still at an early stage it could assist in providing libraries with the kinds of information that will allow them to tailor services to demand rather than reacting retrospectively.

As touched on earlier when mentioning Jisc’s evolution of the LIDP, learning analytics is still in the early stages of development (Sclater, Peasgood and Mullan, 2016) but is gaining popularity amongst higher education institutions and has been defined by Long et al. as:

> the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs. (Long et al., 2011).

Learning analytics utilizes a range of metrics aimed at gauging student engagement and there is the potential for data held in library systems to be utilized as part of these metrics. This presents opportunities for assessing the library’s impact on student outcomes but naturally raises further concerns for user privacy. Oakleaf discusses these issues along with practical obstacles such as data unavailability or imprecision in a number of articles (see Oakleaf, 2016; Oakleaf et al., 2017 and Oakleaf, 2018)
2.7 Personas

A persona is a fictitious individual that represents a discrete group of users with shared characteristics (Fox, 2014; Guenther 2006). Although they are fabricated characters, they are based on information from real users’ behaviours, needs and goals and serve as a proxy for a wider group. Libraries are beginning to use them for a range of situations such as library service changes (Al-Shboul and Abrizah, 2014) and their use within the library setting is aimed at ensuring the service works for the user rather than those who work there. The use of questions such as, ‘How would X react to that change?’ or ‘How would Y access this service?’ assist in making more informed strategic decisions (Zaugg and Rackman, 2016). Grand Valley State University Library have put the three personas they use on their website to show visitors how the library is trying to build services around the user’s needs (Grand Valley State University, 2018) and a number of libraries have published their methodology for persona creation and its use (see Ward, 2010 and Tempelman-Kluit and Pearce, 2014).

2.8 Issues arising

With academic libraries needing to do more to measure their effectiveness than calculate the size of their collection (Schonfeld and Malpas, 2017) and instead be able to provide robust information on how their service assists with supporting the institution’s goals, a number of issues arise that need to be addressed.

As the body of research grows, with a wide variety of measures, methodologies and analyses, it is becoming increasingly necessary to formulate a more structured framework to evaluate the library service. Oakleaf (2010) provided a range of starting points for evaluation and has refined this further with guidance on composing inquiry questions (Oakleaf and Kyrillidou, 2016). Elsewhere, various literature reviews are examining the research through different lenses to provide overviews (Oliveira, 2018), identify emerging themes (Blummer and Kenton, 2018) and consider frameworks for impact assessment (Urquhart,
The emergence of a need to provide suitable frameworks, which are flexible enough to suit different types of academic libraries, is an indicator that this type of assessment is maturing. Existing library assessment frameworks as discussed by Atkinson and Walton (2017) in their editorial, tend to focus on the quality of the service and whether it meets a set of expectations, either the customer’s or an external requirement. However, Madsen and Hurst’s (2018) investigation into library assessment concludes that there is a shift away from this type of ‘service quality improvement’ and suggest that the move towards impact assessment will continue.

It is important to acknowledge that the adage known as Goodhart’s Law, ‘When a measure becomes a target, it ceases to be a good measure’ (Strathern, 1997). Libraries need to be mindful that measurement is for the purpose of understanding and improvement, an issue Oakleaf and Kyrillidou (2016) address with reference to the problems using student-learning surrogates.

With much of the data libraries need to undertake these types of assessment sitting in different systems, in different states of completeness and containing varying levels of detail it is not an easy task to join this to individual students. When this is combined with insufficient resources such as lack of time, skills or financial support it pushes the research towards what is easiest or most straightforward. When selecting the measures and methods to assess the library’s impact Showers (2015, p. xxvi) warns against the streetlight effect, a form of observational bias where one searches for something only where it is easiest to look. This bias runs the risk of undermining the results obtained or failing to draw useful conclusions.

As this area of research matures, frameworks for evaluation are likely to emerge and a degree of standardization with regard to the measures seems probable. Hurst et al. (2016) noted that many library leaders were seeking answers to a common set of questions including wanting to know who is or is not using the library, and there was also a desire to be able to compare data with other institutions. This shared ambition will assist in driving forward an integrated
framework for assessment that allows libraries to serve its users well by increasing the understanding of how its resources are being used and, at the same time, provide evidence of how the library is supporting the institution's teaching and research goals.
Chapter 3. Methodology

3.1 Introduction

The methodology will provide an explanation of the research strategy used to conduct this investigation followed by overview of the available data sets and why the selected sets were chosen. The extraction and cleaning process for the data will be described and the relational database which forms the framework for analysis is outlined. A section providing an overview of the different library personas which have been developed by a project group within the library staff is also included to provide clarity regarding the sections of analysis relating to these. The chapter concludes with the limitations and potential problems relating to the study.

3.2 Research Strategy

This case study will explore a number of data sets that are automatically generated by different systems within the library of City, University of London and combine them to form an improved understanding of the usage of library resources by students. A mixed methods approach (Bawden and Robinson, 2012, p.305) was utilized to enable triangulation. Yin states that a case study ‘relies on multiple sources of evidence’ (2014, p.17) and that these sources should converge. This was achieved through the following methods:

- Desk research for the literature review.
- Analysing and combining five different data sets from three different systems, which contained different biases in terms type of resource being used.
- Informal discussions with library staff who were responsible for or were primary users of the data sets.
- Comparison of findings with the results of previous work conducted by library staff to produce the library personas.
The analysis of pre-existing, externally created material necessarily raises the issue of ‘content rigidity’ (Webb, 2000, p. 30) where the researcher is restrained by the types of data that have been collected. Pickard (2013, p. 258) further explains that the risk of using statistics because of their availability rather than their suitability is the research could become ‘inappropriately shifted simply to accommodate the means.’ This investigation is an exploratory case study rather than a descriptive or explanatory one (Biggam, 2011, p. 120) and does not seek to provide a full breakdown of all the available sources nor combine them all as this would not be practicable within the time constraints for this project. This project is investigating the possibility of repurposing data, which already exists, towards a straightforward method of understanding user behaviours with regard to the resources provided by the library. The mix of research methods employed serves as a solid basis from which to draw initial conclusions and highlight future possibilities.

3.3 Data Collection

The investigation focuses on being able to combine the data sets by converting them to a common format and linking them using the various unique student identifiers to produce a more multi-dimensional view on the use of library resources. Quantitative analysis on the individual and combined data was then conducted to analyse the use of resources and gain insight into student behaviour regarding their usage.

Throughout the process, a number of library staff were informally interviewed by the author to understand the available data sets and identify the following points:

- What information they contained.
- The data sets that were currently extracted and what were they used for.
- The voracity and robustness of the data - whether they considered the method of recording was a true reflection of the behaviour of the user.
• What questions they would like to answer but are unable (either due to constraints in the data or other limitations such as time pressures).
• Whether it was possible to extract data with a unique identifier.

In order to be able to complete the collection, cleaning and assembling of the data within the timeframe allocated it was necessary to assess what was available and what would be most suitable.

Selecting data sets that related to as broad a range of students as possible would mean that insights gained and any conclusions drawn would be relevant to a wide range of students. Using this criterion, data sets relating to core activities that were not discipline specific were considered for use. This included WAM connections, OpenAthens logins, item borrowing information, space/room bookings, email enquiries and security gate swipes. Although it may be the case that some disciplines could be represented to a greater or lesser degree within the dataset depending on the nature of the course and its study requirements, all these resources or activities are anticipated to be used by a range of students.

A considerable number of the data sets did not contain student identifiers, as would be expected of a library where user privacy is a priority. Such examples include searches on the library catalogue and enquiries at the service desks or via Libchat, the online live chat function. Records are kept of the nature of the interactions e.g. desk enquiries are split into categories such as circulation, IT and access to name a few and other details such as the site are also recorded. This information assists the management of the library in the efficient allocation of support and highlights whether regularly requested information can be answered in a different format such as an FAQ entry on the library’s website. Equally, reports are run which analyse the use of the library website, the various guides within the site and Summon, the library catalogue’s resource discovery tool. Nevertheless such interactions are out of scope for this investigation as the lack of an identifying tag means it cannot be joined with other data sets.
The final selection of data sets was chosen to represent the use of electronic resources, traditional item borrowing and use of the on-site library facilities. This allowed for comparisons between a diverse range of resources; on site versus off site and analogue versus digital and consequently had the potential to answer questions not previously possible within this particular institution's libraries. Questions such as ‘are users who do not borrow books using e-resources to a greater extent than those who do borrow?’ are difficult to answer when the data remains in separate systems.

The data sets used for this investigation are as follows:

- Sierra patron record
- WAM connections information
- OpenAthens login information
- Room/space bookings
- Item borrowing

Sierra is the library management system used at City and the patron record holds the range of unique identifiers that are needed to access the different resources selected. It also contains information on total item borrowing and number of inter-library loans.

WAM and OpenAthens information are used in this project to indicate a student’s use of online resources, which includes electronic journals, databases and e-books.

Room/space bookings show student use of group study rooms, individual study spaces and dedicated database terminals. They were selected to indicate physical use of the library.

Item borrowing by students is also an indicator of the use of the physical library.

Other data sets that were given serious consideration but not selected include:

- **Gate entry data**: this is a large and rich data set but also has a number of flaws for use in this project because not all the library sites have security barriers that operate exclusively for the library. Additionally, students
could be using resources around the entrance level of the library at Northampton Square where no swipe access is required.

- **Subject Librarian 1-2-1 appointments**: joining this data with the sets already selected has the potential to unlock some detailed user behaviour, however it represents a small percentage of the overall student body and there would also be a conflict with user privacy and a potential risk of re-identification.

The data was narrowed down to include all undergraduate, postgraduate and short course students at City, and also included affiliate and validated courses. Out of scope were all university staff, visiting staff, alumni users and student visitors (using the SCONUL access scheme or under the University of London access agreement). It is acknowledged that library is required to support the needs of academic staff who will also be undertaking research but as the majority are likely to be well versed in the skills required to research their chosen field, their user behaviour ran the risk of skewing the results or generating unnecessary outliers in the spread of results.

A time range for the data also needed to be considered. It would have been possible to use more data sets if the time span was kept short however it increased the risk that any observations made regarding the data would not necessarily be reflective of behaviour across a longer time period. Library statistics, which are already collected for use internally as well as supplying to external bodies such as SCONUL for the annual libraries statistics return, are assessed across an academic year. This is August to July and this time frame tends to be used across a range of analyses already undertaken.

In order to examine a full academic year, it was initially proposed that August 2016 to July 2017 would be selected because the current academic year would not be fully available during the data collection phase. However, there was an issue of a large amount of information unavailable in the WAM connections data set (see 3.5 Limitations and potential problems, later in this chapter) and a
decision was made to use the current academic year. Thus the period investigated was August 2017 to June 2018 inclusive.

The patron record in Sierra holds the various unique identifiers used across the range of library resources (see Table 1) as well as a quantity of other details to ensure efficient support for users, as not all these details were pertinent or necessary for this investigation they were excluded from the data extraction. The following fields were extracted:

- Patron record number
- Patron barcode
- Patron IT username
- City email address
- Course type (indicates part time or full time)
- Total checkouts
- Total renewals
- Inter-library loan requests
- Patron home library (which of the four libraries the user assigned to)
- Patron type (indicates level of study e.g. PG/UG/research/CPD)
- School
- Department

Personal information such as names were avoided, however it was necessary to extract the university email addresses for students as this was the unique identifier used for room bookings. As the addresses contain the names of students all subsequent use of the table when creating queries was done without showing (i.e. as a hidden field) the email address to maintain as much user privacy as possible.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Unique identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patron Record</td>
<td>Holds all the unique identifiers</td>
</tr>
<tr>
<td>Item borrowing</td>
<td>Patron record number</td>
</tr>
<tr>
<td>WAM connections</td>
<td>Patron record number</td>
</tr>
<tr>
<td>OpenAthens logins</td>
<td>IT user name</td>
</tr>
<tr>
<td>Space bookings</td>
<td>City email address</td>
</tr>
</tbody>
</table>
3.3.1 Item borrowing

The patron record does not hold the borrowing history for an individual, the record only contains what an individual has currently on loan and any reservations that are waiting to be fulfilled or picked up. Historic information regarding what a user has borrowed is only available to the individual themselves via their personal library account accessed with their IT username and password. However, each item in the library management system has a borrowing history, including the patron record number, when it was borrowed and when it was returned.

The library already run SQL queries to extract information from Sierra for check outs and returns to monitor circulation levels and item usage and includes the following range of fields:

**Checkout SQL**
- Title
- Transaction (time and date stamp GMT)
- Item location (which library the item is held)
- Patron type (includes study level and school)
- Patron home library
- Item number
- Status [type of study level e.g. undergraduate]
- School
- Department

**Checkin SQL**
- Title
- Transaction (time and date stamp GMT)
- Item location (which library the item is held)
- Item number

The queries were adapted by the library systems manager to include the patron record number and re-run for August 2017 to June 2018 for the purposes of this investigation.
3.3.2 WAM connections

WAM (web access management) is a process of identity authentication to allow access to specific resources or secure servers. In this context, it allows students to use a single username and password, supplied to them at the start of their studies, to access a large array of online journals, databases, e-books and a number of other online resources without the need to create separate accounts with each vendor used. Sierra has software that creates reports which collate the activity of users accessing resources via this authentication route. The reports used are ‘connections summary reports’ and included the following information:

- IP address
- Patron record number
- Date and time
- Database
- Category name (for type of download e.g. PDF)
- Patron type and School

The reports contain a lot of information, which for the purposes of this investigation were considered noise. When a user connects to a resource by logging in a variable number of ‘packets’ of information are returned such as page header information, the abstract or citation information and each of these potentially generate a line of data. Different resources create different quantities of data rows so it is not possible to be clear how many rows of data relate to a specific item particularly if the user is browsing a number of things in a single resource, such as different papers in the same journal. The reports also indicate when an item has been downloaded and this provides a reasonably robust measure of e-resource usage by students. However, it was a concern that by restricting the data to exclude everything except downloads there may be users who access e-resources but choose not to download items. To ensure that this potential risk was minimised it was decided that another table should be created with the data that showed hourly use. This meant that a user was sampled once per clock hour regardless of how many rows of data they generated. If their
connections data straddled a clock hour they would be sampled twice and if they worked across many hours they would be represented once for each hour. Despite the imperfections of this method it allows for a general picture to be drawn of e-resource usage that accommodates potentially different styles of information behaviour.

3.3.3 OpenAthens logins

OpenAthens is an identity and access management service that allows users to access specific resources and servers. Students use their IT username and password to access resources via this platform. The monthly reports that were available for this data set had the IT user name as the unique identifier and detailed the resource which had been logged into and how many times the user had logged into it over a month.

3.3.4 Space bookings

This data set is held and managed in Springshare, a library-based software application which, among other features, contains a booking calendar for the different spaces within the various library locations. The different spaces are made of assistive technology rooms, group study rooms some of which contain white boards or PCs and presentation screens, individual study spaces that are available to book during high use periods in the library (January-February and March-June) and terminals, which allow access to specific financial resources such as Bloomberg or Morningstar. The majority of library spaces can be booked for a maximum of two hours per day. The data fields extracted were the user's City email address; name and type of the space; date, time and length of the booking and the date the booking was made.
3.3.5 Assembling the data

The data was extracted as CSV files and converted into Microsoft Excel xlsx files. All the data sets required various degrees of data cleaning to get the information into a consistent format and functions within Excel were utilized to perform the cleaning. The data sets contained users out of scope for this investigation and these were removed along with any fields that were not considered pertinent to indicating user behaviour.

The WAM connections data required the most processing to get in to a stage where it could be exported to the database. This included cleaning the patron record numbers so they could be joined with the other tables, removing rows without patron numbers and using a range of Excel formulas to create the hourly use information.

The number of columns generated by including all the resources via OpenAthens was unwieldy so another table was created with the monthly totals and a VLOOKUP calculation was added to generate the Patron Record Number prior to uploading the table to the database.

Item borrowing was made up of two sets of files for each month; items checked out and items checked in. The data from each needed to be combined into a single file for all months across the total eleven-month period. Using MS Excel sorting functions and calculations it was possible to generate the length of borrowing and number of instances for users across this period. In order to maintain as much privacy as possible the subsequent table had the item titles cleared and only held the identity numbers for each item. The data was time-stamped so students who borrowed books for very short periods would still accrue borrowing time.
3.4 Framework for data analysis

A basic relational database was created using Microsoft Access 2016. Figure 1 shows the tables and field headings that were created and fields highlighted in red indicate where the tables were linked. The query function within Access was then used to interrogate the database and create new tables. These queries were exported to Excel for statistical analysis and visual representations. As the author had no prior experience of building a relational database, reference was made to MacDonald (2010) for guidance on construction, field properties and appropriate style and technique. The book refers to an older version of Access than was used for the project but the principles of construction remained the same.

<table>
<thead>
<tr>
<th>Patron data 2017 2018</th>
<th>Sierra item borrowing</th>
<th>Space bookings</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>ID</td>
<td>ID</td>
</tr>
<tr>
<td>PatronRecordNumber</td>
<td>PatronRecordNumber</td>
<td>CityEmailAddress</td>
</tr>
<tr>
<td>PatronBarcode</td>
<td>ItemNumber</td>
<td>Location</td>
</tr>
<tr>
<td>Username</td>
<td>CheckOut</td>
<td>Category</td>
</tr>
<tr>
<td>CityEmailAddress</td>
<td>CheckIn</td>
<td>SpaceName</td>
</tr>
<tr>
<td>ExpiryDate</td>
<td>DaysBorrowed</td>
<td>DateOfSpaceBooking</td>
</tr>
<tr>
<td>CourseType</td>
<td>HomeLibrary</td>
<td>SpaceBookingStartTime</td>
</tr>
<tr>
<td>TotalCheckouts</td>
<td>PatronStatus</td>
<td>SpaceBookingEndTime</td>
</tr>
<tr>
<td>TotalRenewals</td>
<td></td>
<td>SpaceBookingDuration</td>
</tr>
<tr>
<td>ILLRequests</td>
<td></td>
<td>DateBookingCreated</td>
</tr>
<tr>
<td>HomeLibrary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PatronType</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WAM hourly use</th>
<th>WAM downloads</th>
<th>OpenAthens Monthly totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>ID</td>
<td>ID</td>
</tr>
<tr>
<td>TransactionDate</td>
<td>TransactionDate</td>
<td>Date</td>
</tr>
<tr>
<td>TransactionTime</td>
<td>TransactionTime</td>
<td>Username</td>
</tr>
<tr>
<td>DatabaseName</td>
<td>DatabaseName</td>
<td>PatronRecordNumber</td>
</tr>
<tr>
<td>EResourceCategory</td>
<td>EResourceCategory</td>
<td>TotalAccesses</td>
</tr>
<tr>
<td>PatronType</td>
<td>PatronType</td>
<td></td>
</tr>
<tr>
<td>Department</td>
<td>Department</td>
<td></td>
</tr>
<tr>
<td>PatronRecordNumber</td>
<td>PatronRecordNumber</td>
<td></td>
</tr>
<tr>
<td>ClockHour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 Tables and fields for created for Access database with unique identifiers marked in red

The initial assessment of the final data tables included frequency distributions to gain a basic understanding of the information they contained (Pickard, 2013, p.
and assist in developing the queries that would be most useful to run. The mean, median and mode were also calculated along with the standard deviation to further develop an understanding of the data range. The data was then broken down by school and study level to assist in providing insight and allow alignment with previous work undertaken on library personas.

### 3.4.1 Library personas

A persona is used to describe a group of users that display a number of characteristics in common. Within an academic library, developing a range of personas facilitates the refinement of services to meet user needs (Zaugg and Rackham, 2016). A library project group has recently formulated seven personas, some of the details of which are summarised in Table 2. A number of the data sources that were selected for this project, were used to develop the personas and as such the combined data has the potential to provide additional detail.
<table>
<thead>
<tr>
<th>Persona 1</th>
<th>Cass postgraduate, part time (1st year MSc Finance and Investment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivations:</td>
<td>• Teaching self skills (wants to be self-sufficient).</td>
</tr>
<tr>
<td></td>
<td>• Highly proactive, will follow up on issues/problems, etc.</td>
</tr>
<tr>
<td></td>
<td>• Technically adept, familiar with online tools.</td>
</tr>
<tr>
<td></td>
<td>• Appreciates value of human interactions when requesting help.</td>
</tr>
<tr>
<td>Frustrations:</td>
<td>• Not enough Bloomberg terminals. Problems accessing online resources (and not enough database support).</td>
</tr>
<tr>
<td></td>
<td>• Library not opened 24/7 all year round.</td>
</tr>
<tr>
<td></td>
<td>• Cannot access everything required via Moodle.</td>
</tr>
<tr>
<td></td>
<td>• Long waiting times for core (print) textbooks.</td>
</tr>
<tr>
<td>Key brands:</td>
<td>• Dow Jones Factiva</td>
</tr>
<tr>
<td></td>
<td>• Business Source Complete</td>
</tr>
<tr>
<td>Bio notes:</td>
<td>Well organised, but sometimes sacrifices her personal life to fulfill work or study commitments.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Persona 2</th>
<th>Cass undergraduate, full time (1st year BA Business Management)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivations:</td>
<td>• Keeping up to date (technology/current affairs).</td>
</tr>
<tr>
<td></td>
<td>• Increase employability.</td>
</tr>
<tr>
<td></td>
<td>• They want to make the most out of their time and finish tasks quickly.</td>
</tr>
<tr>
<td>Frustrations:</td>
<td>• Not enough core text books.</td>
</tr>
<tr>
<td></td>
<td>• Not enough quiet study space.</td>
</tr>
<tr>
<td></td>
<td>• Things not working.</td>
</tr>
<tr>
<td>Key brands:</td>
<td>• Statistica</td>
</tr>
<tr>
<td></td>
<td>• Orbis</td>
</tr>
<tr>
<td></td>
<td>• CB Insights</td>
</tr>
<tr>
<td></td>
<td>• Zephyr</td>
</tr>
<tr>
<td></td>
<td>• ProQuest</td>
</tr>
<tr>
<td>Bio notes:</td>
<td>Would prefer to work in the library if there was more space.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Persona 3</th>
<th>SMCSE Computing undergraduate, full time (1st year BSc Computer Science)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivations</td>
<td>• Keeping up to date (technology/software/current affairs).</td>
</tr>
<tr>
<td></td>
<td>• Increase employability.</td>
</tr>
<tr>
<td></td>
<td>• Need to fit work around family/work commitments.</td>
</tr>
<tr>
<td>Frustrations</td>
<td>• Library isn’t open 24/7 365.</td>
</tr>
<tr>
<td></td>
<td>• Library is often overcrowded.</td>
</tr>
<tr>
<td></td>
<td>• Advanced software not available in the Library.</td>
</tr>
<tr>
<td></td>
<td>• Hot food isn’t allowed in the Library.</td>
</tr>
<tr>
<td>Key brands</td>
<td>• Cambridge Core</td>
</tr>
<tr>
<td></td>
<td>• ProQuest</td>
</tr>
<tr>
<td></td>
<td>• ScienceDirect</td>
</tr>
<tr>
<td>Bio notes:</td>
<td>Uses the library to work and socialise on evenings and weekends when he’s not at work.</td>
</tr>
<tr>
<td>Persona 4</td>
<td>SMCSE Engineering undergraduate, full time (2nd year BEng Civil Engineering)</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Motivations</strong></td>
<td><strong>Frustrations</strong></td>
</tr>
<tr>
<td>To work with colleagues on the 3rd Floor.</td>
<td>Would be happier with more study space.</td>
</tr>
<tr>
<td>Keen on study space, she is one of the biggest room bookers.</td>
<td>Wants more specialist software.</td>
</tr>
<tr>
<td>An independent worker, she does not often seek help.</td>
<td>Would like more journals and conference reports.</td>
</tr>
<tr>
<td>A user more of space than physical resources.</td>
<td>Frustrations are likely to lead to non-completion of course.</td>
</tr>
</tbody>
</table>

**Bio notes:**
Enjoys using the Library's shared study spaces to meet with friends, socialise and work on her course.

<table>
<thead>
<tr>
<th>Persona 5</th>
<th>SHS undergraduate, full time (3rd year of BSc Mental Health Nursing)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivations</strong></td>
<td><strong>Frustrations</strong></td>
</tr>
<tr>
<td>Obtain a work placement.</td>
<td>Has Dyslexia; difficulty writing essays.</td>
</tr>
<tr>
<td>Pass the exams.</td>
<td>Lack of confidence when using both physical and online library resources.</td>
</tr>
<tr>
<td>Career change.</td>
<td>Not enough spaces for group study.</td>
</tr>
<tr>
<td></td>
<td>Not being able to remember her username and password. Keeps getting locked out of her account.</td>
</tr>
<tr>
<td></td>
<td>Not enough time with the Subject Librarian.</td>
</tr>
<tr>
<td></td>
<td>Getting to the library, childcare commitments.</td>
</tr>
</tbody>
</table>

**Bio notes:**
Literature searches are challenging, and lacks the confidence to use databases.

<table>
<thead>
<tr>
<th>Persona 6</th>
<th>SASS postgraduate, full time (MA Newspaper Journalism)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivations</strong></td>
<td><strong>Frustrations</strong></td>
</tr>
<tr>
<td>Keep up with changes in media industry.</td>
<td>Having to wait for more diverse titles to arrive in the library (More Books).</td>
</tr>
<tr>
<td>Access to latest news media.</td>
<td>Takes a while/ has strict deadlines.</td>
</tr>
<tr>
<td>Keep an eye out for developing stories.</td>
<td>Finds it difficult to bring books back on time, and is annoyed by fines accruing.</td>
</tr>
<tr>
<td>Course is expensive - I want to get the most out of my time at City.</td>
<td></td>
</tr>
</tbody>
</table>

**Bio notes:**
Uses the library for quiet study space.
<table>
<thead>
<tr>
<th><strong>Persona 7</strong></th>
<th><strong>City Law School undergraduate, full time (final year GELLB)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivations:</td>
<td>Frustrations:</td>
</tr>
<tr>
<td>• High standard of achievement.</td>
<td>• Lack of silent study space.</td>
</tr>
<tr>
<td>• Professionally focused.</td>
<td>• Lack of Group Study Space at Law libraries / Separate spaces - use NSq spaces.</td>
</tr>
<tr>
<td>• Accuracy and thoroughness is important to them.</td>
<td>• Lack of core textbooks - asked to buy own copies of certain core texts.</td>
</tr>
<tr>
<td>• High standards for service provision.</td>
<td>• E-books not available or inaccessible with assistive software.</td>
</tr>
<tr>
<td>• Developing in depth research skills for task completion - low feedback from law firms, poor research skills of employees.</td>
<td>• Evening / weekend opening hours and lack of 24/7.</td>
</tr>
<tr>
<td>• Workshops and 1-2-1 support.</td>
<td>• Noise.</td>
</tr>
<tr>
<td>• Engagement with resources based on need (e.g. More Books).</td>
<td>• Things not working (e.g. equipment / resources) no specialist IT support at the Law Library Gip/IC: pressure on Library staff / staff sensitivity.</td>
</tr>
<tr>
<td>• Maintaining knowledge of current affairs.</td>
<td>Key brands:</td>
</tr>
<tr>
<td>• Time pressured: intensive programmes</td>
<td>• Thomson Reuters</td>
</tr>
<tr>
<td></td>
<td>• Westlaw</td>
</tr>
<tr>
<td></td>
<td>• LexisNexis</td>
</tr>
<tr>
<td></td>
<td>• HeinOnline</td>
</tr>
<tr>
<td>Bio notes: Structured, organised and disciplined. Makes the most of library facilities. Registered blind and needs extra time to read material.</td>
<td></td>
</tr>
</tbody>
</table>
3.5 Limitations and potential problems

CAVEAT

All data regarding a student’s patron type and School affiliation refers to the Sierra patron record and not official University student records as supplied to HESA (Higher Education Statistics Agency). Any reference to numbers of students in specific schools are not official figures as the library records are not updated in the same way as the university student record system and do not hold information to indicate a student has left a course before its conclusion. This means library records may include students who have left courses or have taken time off from their course and therefore, it is likely that the library management system over-estimates the current number of students in each school.

Each data set came with its own particular shortcomings and issues, which are described below.

3.5.1 WAM connections information

There were reservations on the absolute accuracy of the information supplied. The programme which collects the server data is acknowledged in user forums as having bugs. Missing data for the academic year 2016/2017 meant that the original period for investigation was changed to the current academic period of August 2017 to June 2018. This does mean that the full academic year was not available at the time of data collection although July is a relatively low usage month for students as indicated by the Summon statistics reports by month, which are already collated and analysed by library staff.

Further issues were discovered when running the connections reports. When the monthly reports were reviewed it was noted that most months appeared to contain days where no connections were indicated and some days contained
Christine Goodson

many hours where connections were absent. Running reports on these days individually revealed the missing connections for almost all days but undermined the validity of using monthly reports for this investigation. Consequently it was decided to run an individual connections report for each day of the period under investigation. There were still two days of missing data, namely the 5th and 6th May 2018, and a number of clock hours were also registering zero connections at times when this seemed improbable. Appendix A-1 shows a brief summary the number of hours with missing or anomalously low connections (usually adjacent to hours with zero connections) and the date this occurred. These missing days and compromised days represent less than 2% of the sample period so their absence should have only a minor effect.

Another issue arising from this data set was that not all connections had a time stamp. To exclude this information from the investigation ran the risk of skewing the results as the cause for the missing information was not clear and could be due to issues arising from the student’s point of access or simply another bug within the programme. Therefore the decision was made that all connections without a time stamp would be allocated the time 00:00. The quantity of connections without a time stamp was less than 2% per month and details are given in Appendix A-2

3.5.2 Room/Space Bookings

Group study rooms are booked by a single student using their City IT account and consequently do not provide any other details of the other potential users of the room. A booking does not necessarily mean the student(s) used the facility but it is reasonable to assume that it does indicate intent. An individual can only book one room for a maximum of two hours once a day, which gives rise to an aspect of room booking behaviour, which has been observed by the author in their staff role and acknowledged by other library staff, whereby different members of the group book the same room for consecutive slots.
3.5.3 Item borrowing information

Only items that were borrowed and returned within the time frame were included in the data set; meaning that items borrowed prior to August 2017 but returned before June 2018 were excluded. Similarly items borrowed during the investigation period but returned after June 2018 were also excluded. This was to keep a strict sense of only assessing behaviour during the period in scope but also to reduce adding inaccurate information e.g. if an item were returned on 1st August 2017 it would have a borrowing period of one day within the scope regardless of whether it had been on loan for many months and would not accurately reflect the behaviour of the user. Using this restriction of only including items loaned and returned during the scope period will mean some information is excluded but maintain the validity of what remains.

As the data was time stamped and borrowing could be calculated to minutes and seconds, students who checked books out but returned them immediately either because they had made a selection error or changed their mind would have registered in the data collected. It was decided to keep this data rather than exclude it as it indicates an interaction with library resources but the borrowing length would be very small.

3.5.4 OpenAthens login information

The OpenAthens login information indicates when a student has been prompted to supply their IT username and login details in order to access content. It does not necessarily indicate successful retrieval of information. Therefore it can be considered a sign of activity relating to e-resource use but caution should be used if relying on the detail of the reports.

OpenAthens provides access to, amongst other things, a large body of health related resources. A number of students, particularly within the School of Health Sciences, are employed by the NHS (National Health Service) some of whom are granted access to this resource using an NHS account in addition to their
university access. This means their use of this resource, as part of their studies, may not be captured.

3.5.5 General limitations

As with students employed by the NHS, other students may have access to relevant e-resources via their employer and use different login credentials to their student account. As such they could appear to be a non-user whereas it would be more accurate to infer they are simply not using library resources.

More broadly it needs to be remembered that these data sets indicate use of library resources and that an individual’s apparent lack of use of these resources may simply mean they have access to the resource beyond the library’s purview. Thus information gleaned from the data analysed reflects usage of library resources not a student’s overall information behaviour relating to their studies, which is likely to operate beyond the bounds of library. This is particularly true where students are using Google or Google Scholar to source information.
Chapter 4. Findings and discussion

4.1 Introduction

This chapter will explore the individual data sets described in Chapter 3.0 Methodology, and then describe the results of a series of queries undertaken on successfully combined data sets which intend to discover additional information about use of library resources previously unavailable due to the separate nature of the information they held. The discussion that follows draws on what has been uncovered in this research to provide some further areas for investigation.

4.2 Individual data sets

4.2.1 Patron Record

There are 33,263 patron records in Sierra that are in scope for this investigation. Figure 2 shows the distribution of student numbers as identified in the Sierra patron record. They have been separated into the different schools and the makeup of each school in terms of undergraduate, postgraduate and other is also shown. The term ‘other’ covers students on courses run by City that are not defined by postgraduate or undergraduate and are usually either short courses or alternatively, Continuing Personal or Professional Development (CPPD or CPD). Additionally, City has a number of affiliations with other institutions and students at these institutions may be included within the patron records, however full details of their course and study level are not necessarily held and will consequently be categorized as ‘other’. Students who are not part of the universities five main schools (Cass Business School, School of Health Sciences, School of Mathematics, Computer Science and Engineering, School of Arts and Social Sciences and City Law School) have restricted access to a number of the resources available via the library, for example students on a short course have access to the library itself and may borrow a limited number of books however,
in most instances they do not have access to e-resources in the same way as other students.

The limitations of resource access by students who are not part of the five main Schools of the university means that for a number of the results reporting they are not included in the tables and graphs. Figure 3 provides a simple percentage breakdown of the different Schools’ contribution to the total student record for the period of time under investigation.

Note the following abbreviations will be used:

SHS – School of Health Sciences
SMCSE – School of Mathematics, Computer Science and Engineering
SA&SS – School of Arts and Social Sciences

![Figure 2 Distribution of user records in Sierra by school and student type for August 2017 to June 2018](image_url)
4.2.2 Item borrowing

The total length of item borrowing for each student was calculated i.e. if a student borrowed 3 books for 2 days the total borrowing length would be 6 days. The distribution of total number of days of item borrowing by each student is shown in Figure 4. The total number of students borrowing in this time period (according the limitations set out in Chapter 3) was 8308 and Table 3 shows the arithmetic mean, median, and standard distribution to indicate the variance in the data.
Figure 4 Distribution of the total length of item borrowing by students from August 2017 to June 2018

Table 3 Statistical summary of total length of item borrowing

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Undergraduate</th>
<th>Postgraduate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of borrowers</td>
<td>8308</td>
<td>4544</td>
<td>3992</td>
<td>72</td>
</tr>
<tr>
<td>Mean</td>
<td>270.1</td>
<td>272.2</td>
<td>268.9</td>
<td>200.9</td>
</tr>
<tr>
<td>Median</td>
<td>107.5</td>
<td>101.9</td>
<td>112.6</td>
<td>77.8</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>436.4</td>
<td>462.6</td>
<td>403.8</td>
<td>77.8</td>
</tr>
</tbody>
</table>

It can be seen from the distribution that the majority of students’ total borrowing is for the shorter period of time but there is also a long tail on the distribution with a few outliers in the longest borrowing durations.

Figure 5 shows the distribution of total borrowing length by school. The spread of borrowing lengths along the horizontal axis is non-linear and has been selected to provide more detail for shorter borrowing lengths as this was where the majority of activity occurred. It can be noted that City Law School is significantly over-represented in the ‘up to one week’ of total borrowing, this range was also the most populated class for Cass Business School and SMCSE, although longer borrowing times were well represented for Cass Business School. Longer total borrowing periods were favoured for students of SHS and SA&SS with 60-100 weeks and 20-30 weeks being the most populated classes respectively.
Figure 6 illustrates the total length of borrowing for postgraduates, undergraduates and other according to each School. Cass Business School shows an almost equal split of borrowing length between undergraduates and postgraduates which when compared to the student make up, where undergraduates account for 39% of the total school, indicates a disproportionate split of item borrowing towards undergraduates. The over-representation of undergraduates is also seen with SA&SS and SHS. A smaller skew towards undergraduates was observed for City Law School and may be partly due to the school supplying students with some core textbooks, thereby reducing the requirement to borrow them from the library. Opposing this pattern, SMCSE shows a small under-representation of undergraduates. The school is made up of 59% undergraduates but they account for 55% of the item borrowing.
In addition to assessing the information on borrowing, it is helpful to look at non-borrowers across the period (see Figure 7). Non-borrowers constitute 75% of the total number of student records in Sierra and if only the five main schools are considered, they constitute 68% of the students. The high proportion of non-borrowers for SHS relates to the high number of CPD courses and although many of these courses confer borrowing privileges they are not utilized, possibly because the courses do not require book borrowing. If only undergraduates and postgraduates are considered for SHS, then the percentage of non-borrowers for the school drops from 75% to 65% and the figure across the five schools moves down to 62%. SMCSE shows the highest proportion of non-borrowers within a school at 81%.
Figure 7 Comparison of the number of students in each school who borrow items and those who do not for the period August 2017 to June 2018

It needs to be noted that non-borrowing could arise from a number of situations:

- The student uses books within the library but does not check them out.
- Borrowing occurred but started or finished outside of the timeframe under investigation (see Chapter 3.5.3).
  - This is more likely for courses finishing in the early part of the timeframe.
- The course supplies the core textbooks reducing the need for borrowing.
- The student has left the course but their library record does not indicate this.
- The student does not use or borrow books from the library.

The patron record in Sierra does show the total checkouts and renewals for each individual. It was therefore useful to cross reference non-borrowers as defined in this investigation with total checkouts to remove some of the possibilities above and give a clearer view on the numbers of students who do not borrow books from the library as part of their information behaviour.
Figure 8 shows the percentage split between borrowers and non-borrowers along with the percentage of users with no borrowing history on their patron record i.e. zero checkouts. It indicates that the highest proportion of users who did not borrow items from the library were SMCSE students.

There are some considerations to be borne in mind when making observations regarding the data collected, particularly when considering the levels of non-use or non-engagement with resources. As stated in the methodology (Chapter 3.5), the caveat that the student records are drawn from Sierra rather than the definitive university records means it is likely the student body has been overstated and it is also possible this overstatement may not be evenly distributed across the schools or student types. Thus it could be that the apparent lack of resource usage by SMCSE students has been disproportionately represented. However, the percentage of non-borrowing with a history of checkouts (23%) is approximately in line with that seen in the other schools so it is reasonable to assume that the over-estimation of students is not excessively biased toward SMCSE.
4.2.3 WAM connections

WAM connections were considered in two ways to account for the possibility of students using the resources it provided access to without downloading items. The distribution of the number of downloads per student was reviewed and is shown in Figure 9. Over half the students that downloaded items did so fewer than 21 times across the eleven-month period and the remainder were spread over a long range. This long tail on the range is reflected in the statistical breakdown (Table 4), which shows how outliers at the upper end of the number of downloads are skewing the mean when compared to the value for the median. These outliers (a student with more than 1,000 downloads) occurred within each of the five main schools.

A large portion of the students downloaded at least one item using a WAM connection, with a skew towards postgraduates.

Figure 9 Frequency of total WAM downloads for all students across August 2017 to June 2018
Table 4 Statistical summary of student WAM downloads

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Undergraduate</th>
<th>Postgraduate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of WAM downloads</td>
<td>825,050</td>
<td>337,257</td>
<td>453,615</td>
<td>34,178</td>
</tr>
<tr>
<td>Mean</td>
<td>57.6</td>
<td>59.4</td>
<td>59.1</td>
<td>35.0</td>
</tr>
<tr>
<td>Median</td>
<td>19.0</td>
<td>18.0</td>
<td>22.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Mode</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>124.2</td>
<td>137.0</td>
<td>117.3</td>
<td>92.3</td>
</tr>
<tr>
<td>Number of students with</td>
<td>14,329</td>
<td>5,680</td>
<td>7,673</td>
<td>976</td>
</tr>
<tr>
<td>WAM downloads</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of total students</td>
<td>43.1%</td>
<td>17.1%</td>
<td>23.1%</td>
<td>2.9%</td>
</tr>
</tbody>
</table>

The total number of downloads was also reviewed by school and student type (Figure 10) and the distribution of downloads per student was broken down to student type (Figure 11). It was anticipated that there would be a skew towards postgraduates for downloads – given the general expectations of an increased level of in-depth research and independent study. In total this was the case however, when each school was looked individually there were pronounced differences in the overall number of downloads as well as the ratio between undergraduate and postgraduate.

Both City Law School and SMCSE showed high proportion of postgraduate downloads compared to undergraduates, 80%/20% and 79%/21% respectively. By contrast, SHS and SA&SS showed an almost equal split. Cass Business School was somewhere between, with a 66%/34% split. When the relative size of the postgraduate body within each school is taken into account, all schools show a skew towards higher numbers of downloads by postgraduates, and although the balance in City Law School and SMCSE remains high it is more reflective of a relatively low number of downloads by undergraduates.

In terms of total downloads by school, SA&SS produced a high level relative its size and given the school’s strong level of item borrowing, it indicates that SA&SS students are making substantial use of physical and electronic resources.
The postgraduate skew was further investigated to see if there was a difference in the distribution between the different levels of study; Figure 11 indicates that the relative balance between undergraduate and postgraduate downloads is reflected through the main segments of the distribution. Postgraduates overall tend to download more items but a particularly large outlier in the undergraduate data resulted in a very similar arithmetic mean for both, but as Table 4 indicates, the median is higher for postgraduates. There were a total of 13 undergraduates downloading in excess of 1000 items over the eleven months whereas there were 21 postgraduates doing the same.
The high download outliers were scrutinised; it was possible that the figures could have been the result of an error in the data or an abuse of the system either by a student or an unauthorized access (i.e. cybercrime). In the instances investigated the data did indicate either of these possibilities although relevant library staff did acknowledge it was unusual behaviour.

The possibility that students may be using resources via WAM connections but not downloading items was true. Table 5 shows the statistical breakdown for the hourly WAM use and shows an extra 1,901 students used the WAM connections that had not downloaded items. There were more undergraduates contributing to this number than postgraduates, 983 compared to 713 respectively. This method of sampling the resource usage had the benefit of diminishing the effect of large numbers of downloads over a short space of time and focused more on how often or long the resource was used. As with the previous distribution there were still a number of students who had high levels of hourly use, but the range was considerably reduced overall (Figure 12).

Table 5 Statistical summary for hourly WAM use

<table>
<thead>
<tr>
<th>Total number of WAM uses sampled</th>
<th>All students</th>
<th>Undergraduate</th>
<th>Postgraduate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>9.4</td>
<td>9.3</td>
<td>10.1</td>
<td>5.1</td>
</tr>
<tr>
<td>Median</td>
<td>5.0</td>
<td>6.0</td>
<td>6.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Mode</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>10.1</td>
<td>9.8</td>
<td>10.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Number of students with hourly use</td>
<td>16,230</td>
<td>6,663</td>
<td>8,386</td>
<td>1,181</td>
</tr>
<tr>
<td>Percentage of total students</td>
<td>48.8%</td>
<td>20.0%</td>
<td>25.2%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>
The same median values for undergraduates and postgraduates for hourly use could indicate that while both types of student commonly access resources a similar number of times, the latter is more likely to download. Though beyond the scope of this project, this information raises a number of further questions, such as:

- Is the information seeking behaviour of postgraduates more attuned to finding what they want and this indicates a higher success rate?
- Are the information requirements different for the two student types e.g. an undergraduate could be downloading a book as opposed to a postgraduate downloading a number of journal articles.

The level of hourly use separated by school was also assessed and is illustrated in Figure 13. Relative to the size of its student body, SA&SS showed a strong level of use and SMCSE the least.
The percentage of users that registered on the hourly use data set but did not download was also separated by school (Table 6) and showed that this behaviour was most prevalent amongst Law students and least common amongst the Arts and Social Science students. When the difference between the types of study level was reviewed it was seen that 15% of undergraduates and 8.5% of postgraduates who used WAM based resources did not download (see Appendix A-3).

**Table 6 Comparison of WAM users that do not download**

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Cass Business School</th>
<th>SHS</th>
<th>SMCSE</th>
<th>SA&amp;SS</th>
<th>City Law School</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of WAM users not downloading</td>
<td>1901</td>
<td>538</td>
<td>342</td>
<td>297</td>
<td>212</td>
<td>459</td>
<td>53</td>
</tr>
<tr>
<td>Number of WAM users</td>
<td>16230</td>
<td>4353</td>
<td>3594</td>
<td>2197</td>
<td>3641</td>
<td>2106</td>
<td>339</td>
</tr>
<tr>
<td>Percentage using WAM but not downloading</td>
<td>11.7%</td>
<td>12.4%</td>
<td>9.5%</td>
<td>13.5%</td>
<td>5.8%</td>
<td>21.8%</td>
<td>15.6%</td>
</tr>
</tbody>
</table>

The differences between schools with regard to WAM users who do not download could be a result of different information needs, different information seeking behaviours or simply reflective of the resources they are accessing. As a simple example, using the OED online (Oxford English Dictionary) does not produce a download but may result in a successful information search for the individual. It is also worth noting that SA&SS covers a diverse range of courses
including economics, sociology, history and music to name a few, and as such there would be an expectation that a wide range of information requirements and behaviours would be exhibited reflecting the diversity of subjects covered. Therefore it is interesting that, as a school, its resource requirements and usage has not sat in the middle of the range but has in fact shown higher levels of item borrowing, WAM usage and downloading. This is supported by findings in Collins and Stone (2014) where the social science group was found to be the highest users of library resources and students in computing and engineering the lowest.

4.2.4 OpenAthens logins

The OpenAthens logins indicate usage of e-resources not available using WAM connections. From the statistical overview, Table 7, it can be seen that this resource is used by fewer students and less often per student compared to WAM downloads. It is reasonable to assume that this is not an active or deliberate choice, as when a student selects an e-resource from the online catalogue there is no obvious indication which service will used. This more simply reflects the type and quantity of resources available through OpenAthens.

As observed in the previous distributions of usage, the majority of students use the resource a few times across the eleven-month period, in this case over half logged in between one to six times (Figure 14). The distribution also showed a long tail with a number of outliers at the far end, which caused a large standard deviation.

Table 7 Statistical summary for OpenAthens logins

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Undergraduate</th>
<th>Postgraduate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of OpenAthens logins</td>
<td>91,070</td>
<td>43,108</td>
<td>46,759</td>
<td>1,203</td>
</tr>
<tr>
<td>Mean</td>
<td>21.1</td>
<td>27.5</td>
<td>18.2</td>
<td>7.0</td>
</tr>
<tr>
<td>Median</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Mode</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>233.6</td>
<td>327.1</td>
<td>161.7</td>
<td>16.2</td>
</tr>
<tr>
<td>Number of students using OpenAthens</td>
<td>4,308</td>
<td>1,569</td>
<td>2,568</td>
<td>171</td>
</tr>
<tr>
<td>Percentage of total students</td>
<td>13.0%</td>
<td>4.7%</td>
<td>7.7%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
Figure 14 Distribution of OpenAthens logins per student across August 2017 to June 2018

Figure 15 shows the number of logins by school and SHS is by far the largest user. This was anticipated based on information from library staff specialising in e-resources, as many of the publishers available through the service are health related. Figure 16 shows the distribution of the number of logins per student type and indicates the relationship between undergraduate and postgraduate use of OpenAthens was similar to that observed with WAM downloads (Figure 11).
Information from one of the subject librarians supporting Health Sciences students revealed that some courses require students to conduct a systematic review as part of their assessment, and an element of the subject librarian’s role was providing instruction for using the resources to undertake this kind of research. This information on course requirements sheds light on observed behaviour and could be a reason for why the proportion of logins is much higher for undergraduates in this school compared to postgraduates. Although another explanation could be that larger numbers of postgraduates are using OpenAthens via a non-university login.

The second largest user of OpenAthens was City Law School; it provides access to a number of law related resources such as LexisLibrary however, it was noted that postgraduates conducted the vast majority of use. This leads onto a further question as to whether this is simply down to course requirements for undergraduates not requiring the resources available on OpenAthens or, whether there is some other factor creating this imbalance. This was not a question that was able to be answered by this project.
4.2.5 Space Bookings

Space bookings are an indicator of the use of physical library resources and from Table 8 it can be seen that almost 12% of students made bookings for the various types of spaces available and the distribution of the number of hours booked per student is shown in Figure 17. The lack of extreme outliers in this data set is indicated by the substantially smaller difference in magnitude between the mean and standard deviation than compared to the number of WAM downloads or OpenAthens logins.

Table 8 Statistical summary for space bookings

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Undergraduate</th>
<th>Postgraduate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of space bookings made</td>
<td>16,544</td>
<td>10,368</td>
<td>6,154</td>
<td>22</td>
</tr>
<tr>
<td>Total number of hours booked</td>
<td>30,464.5</td>
<td>18,875.5</td>
<td>11,551.5</td>
<td>37.5</td>
</tr>
<tr>
<td>Mean no. of hours booked per student</td>
<td>7.7</td>
<td>7.9</td>
<td>7.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Median</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Mode</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>12.2</td>
<td>12.6</td>
<td>11.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Number of students booking spaces</td>
<td>3,931</td>
<td>2,399</td>
<td>1,521</td>
<td>11</td>
</tr>
<tr>
<td>Percentage of total students</td>
<td>11.8%</td>
<td>7.2%</td>
<td>4.6%</td>
<td>0.03%</td>
</tr>
</tbody>
</table>

Figure 17 Distribution of hours booked for library spaces per student across August 2017 to June 2018
To look a little closer into who is using what type of space, the figures were further broken down by school and the different spaces grouped according to their type. Assistive technology and group study rooms were already identified in the data and grouped accordingly. Specialist database computer terminals in the Cass Learning Resource Centre and Northampton Square Library were collated as ‘terminal’ and finally the individual study spaces available during library peak use months and the two ‘silence sound centre chairs’ were grouped as individual study.

From the breakdown of space bookings by school (Figure 18) it can be seen that, with the exception of City Law School, booked hours for group study rooms are similar. This does not track with the size of the student body in each school and could indicate that group study space is a larger requirement for SMCSE and SA&SS compared to Cass Business School. The high use of specialist terminals by students in Cass Business School was expected; they provide access to databases required by students for their courses and the majority of the terminals are based at the library situated within the school’s main building. Students of this school were also the highest users of individual study spaces, the vast majority of which were the temporary space allocations during peak library times (655 hours). By comparison, the remaining schools showed a largely similar number of hours booked for individual study spaces. The high level of booking for individual spaces by Cass Business School students, relative to the other schools, aligns with one of the frustrations listed for persona 2 (Cass Business School undergraduate) that states ‘not enough quiet study space’.
The creation of a relational database allowed the data sets to be combined and sliced in a variety of ways. In order to provide useful insight, queries which focused on large numbers of students and provided information regarding their use or non-use of resources was selected rather than analysing the habits of outliers or the more extreme ends of the data distribution.

### 4.3.1 WAM downloads and item borrowing

One area of behaviour which had been previously been too difficult to assess due to the separate nature of the data sets was the comparison or relationship between e-resource use and item borrowing. Figure 19 shows the total days of item borrowing along with the total number of WAM downloads for each student that used either or both resources; students who used neither resource are not represented. A value is shown on the WAM downloads data point where it intersects with item borrowing reaching zero. It can be seen that there are outliers in terms of WAM downloads throughout the range of item borrowing.
and that lower levels of item borrowing does not appear to correlate to increased use of e-resources.

![Figure 19 Comparison between total days of book borrowing and WAM downloads for the five main schools]

A further look at undergraduates and postgraduates separately was conducted to see if a pattern could be discerned, particularly as postgraduates had earlier been shown to display shorter total borrowing but higher e-resource use. Figures 20 and 21 both show that the highest outliers occur in the longer borrowing lengths but that there is little in the way of an obvious pattern between the two resource types.
This was further broken down to reflect the library personas (as detailed in Table 2, Chapter 3.4.1). The selection of students was to department level rather than purely individual course as course details were not selected as one of the fields extracted from the patron record.

Figures 22-33, show the library persona followed by a comparison to the School and student type to assess whether they are approximately reflective of the broader student body they aim to represent.

Looking at the Finance postgraduates (Figure 22), it does indicate that heavier borrowing occurs with reduced WAM downloads and vice versa, although when the larger school is graphed it appears much like the overall postgraduate
population seen in Figure 21. This behaviour was not observed with the other personas.

Figure 22 Item borrowing and WAM downloads for persona 1: part time MSc Finance and Investment, postgraduate at Cass Business School

Comparing the Cass Business School undergraduate persona to the broader school population (Figures 24 and 25) and to the total undergraduate population it is noticeable that the point where item borrowing reaches zero is much later; 65% of the students borrowed items compared to 58% in the broader school. This could support one of the frustrations listed in the persona information of ‘not enough core text books’. Unlike postgraduate persona 1, there is a sizable quantity of WAM downloads throughout the item borrowing section above zero. The mean of the WAM downloads occurring where item borrowing is present is 35.0 compared the previous persona which was 14.4.
Figures 26-27 highlight the sparse use of WAM downloads for SMCSE undergraduates, which concurs with the evidence of WAM usage by school (Figure 12). The percentage of students with item borrowing but no WAM downloads was 56% for persona 3, this ranked highest of all the personas. Persona 4 came second with 38% of borrowers having no WAM downloads. This may be due to the nature of the courses undertaken in this school; that compared to other schools the requirement for information available via WAM connections is not as large, it could also be that students are getting the information they need outside of library resources e.g. using Google to locate information sources. By comparing the personas to all undergraduates at the school (Figure 28), it can
be seen that they are reflective of the broader body of students they are representing in this case.

Figure 26 Item borrowing and WAM downloads for persona 3: full time BSc Computer Science, full time undergraduate in SMCSE

Figure 27 Item borrowing and WAM downloads for persona 4: full time BEng Civil Engineering, full time undergraduate in SMCSE

Figure 28 Item borrowing and WAM downloads for all undergraduates in SMCSE
A cursory comparison between persona 5 (Figure 29) and all SHS undergraduates (Figure 30) would lead to the assumption that there is little similarity between the two. However the presence of an anomalous WAM downloads outlier of 5198 in Figure 30 has the effect of shrinking the majority of WAM usage due to scaling requirements. When the secondary axes are scaled identically (see Appendix A-4) there is a high level of similarity between the two. Persona 5 showed the smallest percentage of zero downloads where item borrowing had occurred at 2% and this compares well with the 7% measured for all SHS undergraduates.

Figure 29 Item borrowing and WAM downloads for persona 5: full time BSc Mental Health Nursing, undergraduate in SHS

Figure 30 Item borrowing and WAM downloads for all undergraduates in SHS
Persona 6 (Figure 31) follows a similar pattern to persona 5 with regard to these two measures. A small percentage (5%) of the persona borrowers registered zero downloads whilst all SA&SS postgraduates registered 3%. Given the wide range of subjects covered by this school it is useful to note that the persona is reasonably reflective of the broader population it represents in this aspect.

![Figure 31 Item borrowing and WAM downloads for persona 6: full time MA Newspaper Journalism, full time postgraduate in SA&SS](image1)

For persona 7, the full time undergraduate studying Law it was decided to use the whole undergraduate body for the school as the difference between the two was the addition of two part-time students. When the graphs were compared side by side there was no discernable difference.

The percentage of borrowers who did not download anything via WAM was 26%, which was in the middle of the range seen for other personas. The graph

![Figure 32 Item borrowing and WAM downloads for all postgraduates in SA&SS](image2)
(Figure 33) shows a number of high downloads spread through the borrowing range but nothing over 130 downloads once the non-borrowing segment was reached.

![Figure 33 Item borrowing and WAM downloads for all undergraduates in City Law School](image)

There were some points in common across the graphs comparing item borrowing and WAM downloads, namely the presence of a number of outliers for WAM downloads as well as the presence of item borrowing but no WAM downloads and vice versa. The personas were also largely reflective of the broader student body they represented in this aspect. There were also distinct differences between the personas such as the balance between borrowing and downloading and non-use.

### 4.3.2 Combined resource engagement

With regard to resource use by SMCSE students, there is a significant quantity of literature discussing the high dropout rates for some STEM subjects and how it can be addressed (see Chen and Soldner, 2013; Hardy and Aruguete, 2014; Johnson and O'Keeffe, 2016; Stone et al., 2015 and Times Higher Education Supplement, 2001) and, as covered in Chapter 2, there has been acknowledged a positive correlation between use of library resources and student outcomes. A question was raised to the author regarding how reflective the 1st year Computer Science persona is with respect to the choices of electronic resources. This
prompted another line of investigation, namely an examination of into the
general use of library resources for first year SMCSE undergraduates compared
to all SMCSE undergraduates. The Sierra patron record holds the start dates for
students however this was not included in the fields selected for this project as it
was not anticipated to pursue this route of enquiry. As an alternative measure,
undergraduates whose access was expected to expire in 2020 or later were
compared to all undergraduates. Though an imprecise measure, with obvious
flaws, it would provide an approximation for first year students. Each school was
assessed this way and WAM hourly use was combined with OpenAthens use to
provide a general indicator of e-resource usage.

Figure 34 shows the percentage of first year (as defined above) undergraduates
that used a resource at least once. It does not aim to indicate how much a
resource was used but simply that it has been engaged with. The balance
between item borrowing and e-resource varies from school to school, however
the use of library resources by SMCSE first years is significantly lower that seen
in the other schools, in particular the use of e-resources. When all
undergraduates are viewed (Figure 35) the use of e-resources improves
considerably indicating that these are used predominantly by second and/or
third year students. Elsewhere, the elevated use of item borrowing ahead of e-
resources seen in Cass Business School first years reverses when all
undergraduates are viewed and moves them in line with the trend seen in the
other schools. Item borrowing also drops away for SHS undergraduates as they
progress.
The potential flaws with the library management system overstating student populations are a risk when considering this particular comparison. However, given that the use of space bookings is of a similar magnitude to the other schools it would be reasonable to assume that any disproportionality is not to an excessive degree and thereby reinforces the view that SMCSE first year undergraduates are showing a relatively small amount of engagement with library resources and in particular e-resources. It could also be considered that if there is a disproportionate skew for over-estimation of the first year population of SMCSE, then this group is still exhibiting some unusual behaviour compared to
the other schools with regard to space booking being used to a greater extent than other the resources investigated.

To further assist in developing an approach to view the data regarding the combined use of resources, a series of Venn diagrams were drawn to indicate the way resources overlap with one another. The personas were used as guidance but in this instance the broader school was selected e.g. Cass Business School postgraduates. For clarity, it should be noted that the diagrams are not mathematically to scale. As with Figures 34 and 35, this perspective is looking at whether a student has used a particular resource rather than considering how much a resource has been used.

Figure 36 shows the split of resource engagement across all the patron records, the high level of non-use of resources will therefore include patrons who do not have permission to some of the resources in addition to the previously mentioned records for patrons are who are no longer attending courses. The diagram illustrates that the majority of resource usage relates to e-resources and while a sizeable proportion of engagements were just for this resource the student body registered relatively few ‘borrowing only’ or ‘spaces only’ users.
The subsequent Venn diagrams (Figures 37-42) for the most part follow this pattern. All the types show approximately similar levels of non-users ranging between 25-31% except for Figure 39, representing SMCSE undergraduates, where non-users are 56% of the available records. This was to be expected given the analysis so far.

The two types of postgraduates shown, Cass Business School and SA&SS (Figures 37 and 41) both displayed similar levels of engagement with borrowing and e-resources but the later had a lower use of spaces, this is most likely due to a reduced requirement to access the specialist terminals needed for many of the Business School courses.

Of the selection of student types illustrated, Cass Business School undergraduates (Figure 38) also showed the largest proportion of users who had used all three resources. When this is combined with the assumption that they also exhibit the lowest level of non-use it indicates that the majority of these
students are aware of some of the key resources available to them and how to access them.

The use of resources by SMCSE students has been highlighted at several points as being significantly different compared to other types of students. Figure 39 shows very clearly this contrast, and even if the number of Sierra records

Figure 37 Venn diagram showing resource engagement by Cass Business School postgraduates

Figure 38 Venn diagram showing resource engagement by Cass Business School undergraduates

Figure 39 Venn diagram showing resource engagement by SMCSE students
significantly overstates the student body, the balance of resource use shows a
greater emphasis on space usage. Also, whereas the other types of students’ use
of spaces has had a considerable degree of overlap with at least one other
resource, SMCSE undergraduates have exhibited a greater level of space
bookings and no other resource.

Figure 39 Venn diagram showing resource engagement by SMCSE undergraduates

The advantage of using this type of assessment is that even if the number of non-
users is discounted from the view, the relationships and balance between the
remaining numbers and their intersections can be compared and contrasted to
highlight significant differences or similarities between the student types. With
this in mind, it can be seen that in addition to the use of space bookings, the
overlap of borrowing and e-resource is very different for SMCSE undergraduates
compared to the other student types, which show a similar degree of overlap to
each other.
Figure 40 Venn diagram showing resource engagement by SHS undergraduates

Figure 41 Venn diagram showing resource engagement by SA&SS postgraduates
From the Venn diagrams it can be seen that where students have only engaged with a single resource it was most likely to be an e-resource as opposed to space booking or item borrowing.

Finally, a simple infographic was created using a free graphics website, Canva (no date), to provide a succinct summary for resource engagement by the five main schools (Figure 43). Unlike the Venn diagrams, the data used for the infographic kept WAM hourly usage and OpenAthens logins as separate resources. The diagram not only highlights that 32% of students who did not use one of the resources investigated but also separates how many had zero checkouts on their borrowing history. Though it is entirely feasible that a student undertaking a course within one of the schools may simply not have engaged with the library resources investigated for this project it is reasonable to assume that a portion of this 20% are records where the student is no longer registered with or undertaking their course.

The use of infographics is helpful to provide a very general overview of the complex interlinking of resource engagement. Whilst unable to show the depth of information that is held by the data explored, for this project it serves as a
simple summary of what has been examined as well as assisting in the formulation of the next level of questions to be explored in the data.

![68% of students in the five main schools used one or more resources*](image)

30% of students used one of the resources investigated
23% of students used two of the resources investigated
15% of students used three or more of the resources investigated
32% of students used none of the resources investigated
20% of patron records used none of the resources and had no borrowing history

*Resources as defined by the limitations of the investigation

Figure 43 Infographic highlighting indicating resource engagement within the five main schools at City, University of London

Figure 43 created using Canva (no date)

4.4 Discussion

The data collated in this project has been successfully combined and has demonstrated it can test assumptions about resource usage, answer questions regarding the types of students that use the various resources and in what combination, as well as highlighting anomalies and unusual behaviour. The previous work undertaken on the personas has, in some instances been supported by the data.

Combining item borrowing with WAM download data has shown that generally, students who download larger numbers of items are more likely to have also borrowed books. It does not support the proposition that if a student is not
borrowing books then perhaps they are heavier users of e-resources, the implication being that they may be accessing e-books rather than using the physical copies. It may be the case that some students are using e-books in preference to physical items but this behaviour cannot be accurately uncovered by the data collected.

That the findings in this investigation concur with those undertaken elsewhere (Collins and Stone, 2014) regarding social science groups and computing and engineering groups implies the student users of City’s libraries are typical and thus other research undertaken in this respect may also apply. This could be helpful when considering solutions to increasing engagement in low user groups.

The study indicates that postgraduates do not borrow as much as undergraduates but use WAM resources to a greater extent and download more items than undergraduates, this accords with the general assumption that the nature of postgraduate courses requires an increased level of research compared to undergraduate courses. It should also be noted that the postgraduate group comprises of taught and research degrees.

The investigation for this project examined a broad perspective of the data, looking across the whole year in order to gain an overview of how the different resources were used by different sections of the student body. Time constraints meant that a deeper dive into the data was not possible but potential avenues of investigation are suggested below.

WAM connections reports contain IP address information and a cross check for library terminals could be conducted to compare patterns of use versus students using their own computers: is their behaviour different; are there common traits; do some schools use library terminals more than others. This information could also be compared to the item borrowing and space use to examine whether there are any patterns of behaviour observable.
There are normal peaks and troughs for resource use through the academic year and these are reflected in the provision of bookable individual study spaces and extended opening hours. It may be worth analysing whether all schools share similar rates of resource use across the year or whether there are differences. It may also be useful to compare this against when library instruction or outreach events are conducted to assess whether these programmes have any observable effect on resource use.

Analysis of low and non-users is an area that warrants further investigation to gain an understanding of the factors in play, specifically the apparent low use of resources by SMCSE undergraduates. The data collected is capable of being broken down to department level, which may allow for more subtle differences to reveal themselves and it may be useful to examine other data sets to assess the rates of use are consistent with these findings.

Low borrowing for CPD students could be due to the course content not requiring items to be borrowed, students may have access to materials elsewhere or use the library for reference purposes. Liaison with relevant teaching staff could shed light on this observation and further initiatives undertaken if appropriate. Similarly, further investigation of City Law School undergraduate behaviour regarding the use of e-resources may be beneficial. Some of the significant differences seen between undergraduate and postgraduate use compared to other schools could simply reflect course demands and expectations but if this were not the case, then again additional support may be required.

It is also worth considering a more detailed analysis of the data, to department level as well as undergraduate/postgraduate split, to assess whether heavy resource use in one department is masking low resource use in another. This type of breakdown may be especially useful for SA&SS, where the diverse nature of courses could reveal very different levels and combinations of resource use.
Chapter 5. Conclusion

The overall aim of this research was to uncover a more detailed understanding of how students at City, University of London use a selection of library resources by linking existing, automatically generated data sets.

The specific objectives were:

1. Explore what data sets are available and what information they contain.
2. Establish if the data sets can be successfully combined.
3. Examine the kind of questions that the combined data can answer.
4. Evaluate whether the combined data provides evidence to support the underlying assumptions regarding the pre-existing work on library personas.
5. Assess whether the combined data adds further understanding to user behaviour.

The exploration of data sets held and analysed by the library indicated many that did not contain unique identifiers in accordance with upholding the ethical principles of privacy for users.

The data sets selected represented three types of library use, namely item borrowing, e-resource use and bookings for designated spaces or rooms and were successfully linked using the different student identifiers each set contained.

The technique allows a more detailed examination of the information than may be gained through qualitative methods and has the particular advantage of capturing information regarding low use. As such SMCSE students were identified as having an unusual balance between the types of resources engaged with: low use of e-resources and item borrowing but not for space booking. However, the insight gained by using this method does not answer questions of
why these patterns of use occur but does facilitate a more targeted approach for further investigation possibly using focus groups or interviews for a qualitative analysis.

The examination of the data provided some supporting evidence for the library personas and importantly did not indicate and contradictory information.

The findings of this report could be used as the basis to increase the level of collaboration between particular schools and the library especially where low use has been identified as a potential concern. Establishing whether there is an expectation that students should be using library resources to a greater extent than is currently the case could be a first step and then working as partners to identify opportunities to increase engagement and examine the barriers or causes of the current low use.

In an ideal world, this type of linked data could be expanded to include more datasets to provide a more holistic view of library usage and, if data could be added to on a continuous basis as it became available, it could be used to monitor changes over time rather than the single snapshot this study undertook. This would require some significant obstacles to be addressed, not least a careful assessment of how user privacy would be impacted and managed as this would be different to single-shot use. A straightforward method of data collection, cleaning and verification would also be required as some data sets required a substantive amount work to get them into a useful format. As has been suggested by others when discussing library impact (Hurst et al., 2016; Madsen and Hurst, 2018), a dashboard interface with appropriate visualization tools would be the most productive way to incorporate analytics into normal working procedures. This would allow the data to be sliced according to the needs of the investigator and harness the full potential of automatically generated data. Such a tool could allow provide the information necessary to increase the precision with which library instruction and outreach events could target users and supply them with the knowledge of the full range of resources the library has to offer and skills they need to navigate them. It could also become the basis of a more predictive
approach to resource provision and assist in modeling resource demand when faculty departments expand or contract their education offering.

Anomalies, relationships, correlations and patterns of behaviour relating to library use are discoverable in data which is automatically created by the systems students interact with as they use the various library resources. This information can provide a roadmap for avenues requiring further investigation either through mining deeper into the data or by contextualizing it with qualitative research.
Chapter 6. References


Jisc (no date) Who We Are and What We Do. Available at: https://www.jisc.ac.uk/about/who-we-are-and-what-we-do (Accessed: 24 September 2018).


Appendices

Appendix A: Supplementary data

Appendix A-1 Number of anomalous hours of WAM connections data

Appendix A-1 Table 9 Number of anomalous hours of WAM connections data per day

<table>
<thead>
<tr>
<th>Date</th>
<th>Number of anomalous hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-Aug</td>
<td>9</td>
</tr>
<tr>
<td>13-Oct</td>
<td>4</td>
</tr>
<tr>
<td>19-Mar</td>
<td>3</td>
</tr>
<tr>
<td>03-Apr</td>
<td>1</td>
</tr>
<tr>
<td>04-Apr</td>
<td>3</td>
</tr>
<tr>
<td>05-Apr</td>
<td>4</td>
</tr>
<tr>
<td>09-Apr</td>
<td>6</td>
</tr>
<tr>
<td>10-Apr</td>
<td>4</td>
</tr>
<tr>
<td>11-Apr</td>
<td>4</td>
</tr>
<tr>
<td>16-Apr</td>
<td>3</td>
</tr>
<tr>
<td>19-Apr</td>
<td>1</td>
</tr>
<tr>
<td>23-Apr</td>
<td>8</td>
</tr>
<tr>
<td>24-Apr</td>
<td>3</td>
</tr>
<tr>
<td>25-Apr</td>
<td>2</td>
</tr>
<tr>
<td>27-Apr</td>
<td>3</td>
</tr>
<tr>
<td>28-Apr</td>
<td>5</td>
</tr>
<tr>
<td>29-Apr</td>
<td>4</td>
</tr>
<tr>
<td>01-May</td>
<td>9</td>
</tr>
<tr>
<td>05-May</td>
<td>24</td>
</tr>
<tr>
<td>06-May</td>
<td>24</td>
</tr>
<tr>
<td>08-May</td>
<td>10</td>
</tr>
</tbody>
</table>
Appendix A-2 Number of missing time stamps in WAM hourly use and their percentage of the total month

Appendix A-2 Table 10 Number of rows without time stamps in WAM connections reports

<table>
<thead>
<tr>
<th>Month</th>
<th>number of missing time stamps</th>
<th>number of rows in total</th>
<th>Percentage of missing time stamps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug-17</td>
<td>77</td>
<td>5559</td>
<td>1.4%</td>
</tr>
<tr>
<td>Sep-17</td>
<td>57</td>
<td>5800</td>
<td>1.0%</td>
</tr>
<tr>
<td>Oct-17</td>
<td>99</td>
<td>16951</td>
<td>0.6%</td>
</tr>
<tr>
<td>Nov-17</td>
<td>151</td>
<td>19471</td>
<td>0.8%</td>
</tr>
<tr>
<td>Dec-17</td>
<td>127</td>
<td>10800</td>
<td>1.2%</td>
</tr>
<tr>
<td>Jan-18</td>
<td>176</td>
<td>14353</td>
<td>1.2%</td>
</tr>
<tr>
<td>Feb-18</td>
<td>119</td>
<td>15573</td>
<td>0.8%</td>
</tr>
<tr>
<td>Mar-18</td>
<td>198</td>
<td>19832</td>
<td>1.0%</td>
</tr>
<tr>
<td>Apr-18</td>
<td>352</td>
<td>23573</td>
<td>1.5%</td>
</tr>
<tr>
<td>May-18</td>
<td>198</td>
<td>13248</td>
<td>1.5%</td>
</tr>
<tr>
<td>Jun-18</td>
<td>102</td>
<td>7562</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

Appendix A-3 Users of WAM connections who do not download items by student type

Appendix A-3 Table 11 Users of WAM connections who did not download items split by student type

<table>
<thead>
<tr>
<th></th>
<th>total</th>
<th>Undergraduate</th>
<th>Postgraduate</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of WAM users not</td>
<td>1901</td>
<td>983</td>
<td>713</td>
<td>205</td>
</tr>
<tr>
<td>downloading number of WAM</td>
<td>16230</td>
<td>6663</td>
<td>8386</td>
<td>1181</td>
</tr>
<tr>
<td>users</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>percentage of students using</td>
<td>11.7%</td>
<td>14.8%</td>
<td>8.5%</td>
<td>17.4%</td>
</tr>
</tbody>
</table>
Appendix A-4 Comparison between persona 5 and all SHS undergraduates

Secondary axis in Figure 45 has been adjusted to match that shown on Figure 44.

Appendix A-4 Figure 44 Item borrowing and WAM downloads for persona 5: full time BSC Mental Health Nursing, undergraduate in SHS

Appendix A-4 Figure 45 Item borrowing and WAM downloads for all undergraduates in SHS
Appendix B: Original proposal and ethics checklist

Dissertation proposal by Christine Goodson

Working title:

Can enhanced library analytics improve the understanding of user behaviours at City, University of London?

Introduction

This project hopes to contribute to an improved understanding of how students use various resources at the libraries of City, University of London.

A library user automatically generates data by at a number of different points as they utilise the available resources; this includes the use of swipe cards by staff, students and visitors to the library which provides details on which users enter the library with a date and time, along with what course they are studying if they are a student. The library management system, Sierra, holds data on all the physical stock the library has as well as records for each user, what they have borrowed, how long items are borrowed and information on interlibrary loans. The Library guides system contains data sets that include booking of group study rooms. Additionally, access to e-books via the Athens portal requires a user to log in.

The libraries at City University currently use analytics to generate information on usage of the various resources however, to date, the different data sets that each system produces have not been combined to produce a more integrated picture of how individuals utilise the available resources.

Each of the systems are managed separately and combining the data is not straightforward so it is not possible to definitively correlate borrowing of physical items from the library with the use of electronic resources at an individual level or assess if heavy book borrowing correlates to regular visits to the library. Being able to combine the data sets has the potential to unlock insights into user behaviours.
Aims and objectives

The overall aim of this project is to combine data sets acquired from different administration sources within the university to gain a better understanding of student user behaviour of the library’s resources.

The objectives are:

1. What data sets are available and what information do they contain?
2. Can the data sets be successfully combined?
3. What kind of questions can the combined data answer?
4. Does the combined data support the underlying assumptions regarding the pre-existing work on ‘library personas’?
5. Does the combined data add further understanding to user behaviour?

Scope and definition

The initial part of the project will be to identify the data sets that are available and only those that contain information relating to the use of library resources will be considered for this project.

- City University has four library locations:
  - Northampton Square
  - Cass Learning Resource Centre
  - City Law School Library (Gray's Inn Place)
  - City Law School Library (Northampton Square, Innovation Centre)

Each student will be allocated a one of these sites as their ‘home’ or main library according to their course.

Some data sets will have site specific information such as where certain resources are located other data sets such as accessing e-books will not contain this information.

The data held in Sierra includes the details of borrowing of books, CDs, DVDs, specialised assessments and tests, audio equipment and iPads by individual user. It also contains the information on which site the resource is located, although users can request books from another site to pick up at their home library.
It is possible to use the swipe data for entry to the library to record who visits and how often but there is no requirement to swipe out so the data will not indicate the duration of the visit. It should also be noted that Northampton Square library has study spaces, computer terminals and printers on its entry level (floor 2) where there is no requirement to swipe in, therefore it is feasible that a student could be visiting the library regularly to use electronic resources but that this would not be recorded in the swipe data. Although other sites require swipe access to enter the building, it is not for the exclusive use of the library so would not be included in this project.

There are some limitations regarding certain types of data; for example, group study rooms are booked by an individual, which is recorded by the system, but there is no ability to capture who else will be using the room.

The source data available may include university students, staff, alumni and visitors, however the data used for this project will focus solely on City University students. It is anticipated that the academic year October 2016 to September 2017 will be the period of time analysed but the number of data sets selected for combining could vary depending on the amount of work required to clean and merge them.

In line with ethical practices for a library, the browsing history and search details of individual users are not recorded so this aspect of information behaviour will be out of scope.

**Research context and literature review**

The area of library analytics is practice orientated and there are a number of on-going projects that allow higher education libraries compare and contrast themselves with others, which enables benchmarking against similar institutions. An example includes the Society of College, National and University Libraries (SCONUL) Annual library statistics return (SCONUL, 2018).

The Library Impact Data Project (LIDP), led by a team at Huddersfield University, investigated the hypothesis ‘that there is a statistically significant correlation across a number of universities between library activity data and student attainment’ (Jisc, 2011). The project looked at whether library usage data that was routinely generated could be used ‘to understand patterns of library use among students from different
disciplines at the University of Huddersfield’ (Stone and Ramsden, 2013; Collins and Stone, 2014, pp. 51). Jisc also published an article utilising results of the LIDP (Jisc, 2012) and went on to commission the Library Analytics and Metrics project (JiscLAMP) and in 2013 produced a ‘prototype shared analytics service for UK academic libraries’ (Ruddock, 2013). This early attempt to allow UK academic libraries to supply information and subsequently benchmark themselves against similar institutions has subsequently evolved into Jisc’s Learning Analytics, ‘using data and analytics to support students’ (Jisc, 2018).

Other investigations undertaken by the LIDP included a review of whether demographic characteristics affected library usage (Stone and Collins, 2013) and although there were statistically significant differences, it was not possible to draw definitive conclusions as to the reason why this was the case.

Following a brief search, the number of books on the subject appears limited; one example, *Library Analytics and Metrics, using data to drive decisions and services* (Showers, 2015), contains a number case studies and the contributors may provide further routes to investigate as part of a more detailed literature review.

Many of the previous investigations have focused on combining library usage with student academic performance such as drop-out rates or degree class awarded. This project will be focused on finding methods to refine and combine the data available at City University that will enable improved understanding into current behaviours regarding the use of resources and provide the means for this to be repeated over time to gain insight into the way this changes.

**Methodology**

The project will be a quantitative analysis of the usage of City University library resources by students at the institution. It will utilise data that is routinely generated by various IT systems and servers.

The review of what data can be collected will be done by consulting with the library staff who are responsible for managing that particular area and reviewing what information
can be exported from the various systems. As part of the review there will be informal discussion with relevant members of staff as to whether there are areas of enquiry that are not currently pursued that would provide useful insight.

Once there is a clear picture of the available data sets, then a portion of the various sets will be extracted into an Excel spread sheet for examination with regard to the requirements for cleaning the data and deciding which parts of the data will be useful.

At this stage it is uncertain whether the data cleaning would require some coding using a simple programming language such as Python or if the functionality within excel would suffice. However it is anticipated that some work will be required before the data can be successfully combined.

Once there is a unifying individual identifier (either the City IT login or student number) then the data tables can be loaded into Libinsight. At this point it will be possible to review the output and the types of questions that can be asked.

At this preliminary stage it is difficult to ascertain the number of data sets that can be used or combined for this project. However, it would be preferable to combine a smaller amount of data but to do so across a whole academic year rather than a broad spectrum across a few months. This would allow assessment of seasonal variation and requirements to be highlighted when looked at separated by month. The period under review will be September 2016 to August 2017.

With regard to data management, Derek MacKenzie, Head of User Services and Lucy Clifford, Library Systems Manager will be responsible for allowing access to the data and will be consulted regarding the exact detail of the data being extracted to ensure confidentiality is maintained. All access to data will be via a library staff computer and no data will be accessed, analysed or held on the author’s personal equipment or in any portable memory format. This prevents any accidental loss of data. The results of analyses will held on the author's personal computer equipment as all data will be aggregated and anonymized by that point.

There will be a hand-written journal/notebook to record notes from various discussions with staff, record of methods used, issues encountered and the approaches considered to resolve them during the course of the investigation. Permission to take notes during
Christine Goodson

meetings will be sought and no personal or confidential material will be included in the notebook.

**Dissemination**

It is anticipated that the dissertation could be deposited in the CityLIS area on the Humanities Commons repository subject to a satisfactory grade. Additionally, it is expected that the insights gained would be shared with colleagues at the library. Further possibilities may present themselves subject to the nature of the findings.

**Work plan**

<table>
<thead>
<tr>
<th>MAY</th>
<th>JUNE</th>
<th>JULY</th>
<th>AUGUST</th>
<th>SEPTEMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature search</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and review</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>available data sets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and options for</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cleaning and linking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean and</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>combine data sets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>introduction and</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>literature review</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyse and</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interrogate the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>combined data set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write up methodology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write up results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and conclusions of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finalise report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and submission</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
May

- Literature search.
  - As this area is practice driven it will include searches for information from conferences and blog posts as well as the traditional routes of articles and books.
- Discussions with members of library staff regarding the detail of what data is held.
- Discussions with members of library staff regarding the library personas they have identified.
- Sample of data extraction to assess the quality, any shortcomings that may be identified and how much cleaning would be required.

June

- Full extraction of selected data sets.
- Clean data in Excel spread sheets.
- Export cleaned data into Libinsight.
- Write introduction and literature review.

July

- Analyse the data in Libinsight by considering different combinations and what questions could be answered.
- Review whether the data can provide additional information regarding library personas.
- Write up methodology.

August

- Write up analysis and conclusions.
- Discuss analysis and conclusions with Head of User Services.

September

- Checking of dissertation, reference material, proof reading.
- Submission of dissertation.

Resources

This project requires the use of a library staff computer and the accompanying software, including Microsoft Excel and Springshare Libinsight. This has already been discussed with Derek MacKenzie who did not envisage any difficulties with this.
The additional travel costs associated with the requirement to be at the library have been accounted for and will be met by the author.
Ethics

In addition to the ethics check list, a note regarding the data that will be used:

The project will have access to some information regarding individual students however it is not expected to have any access to sensitive information i.e. information which falls into the following categories:

- health and medical conditions
- certain criminal convictions
- racial or ethnic origin
- religion or similar beliefs
- sexual orientation

In order to safeguard privacy, data from the various sources will be stripped of as much identifying information as possible. It is the intention that data extracts will only identify individuals by their student number or City IT login.

Disclosure

The author is employed by City University in the library as a part-time evening and weekend service assistant. All work carried out on the project will not be paid for and it is not expected that the library will direct research goals or editorial decisions.

The author will be using resources provided by the library including a computer terminal and software licensed for library staff use.

Confidentiality

As stated earlier, there are senior staff at the library who will be ensuring the data accessed is appropriate and measures in place to prevent a breach of privacy.

All the individual data sets and combined data will remain the property of City University and permission would be required from City University to release it. All published data within the report will be aggregated and anonymized to ensure there is no possibility of re-identification.
References for proposal


**Ethics Review Form: LIS Masters projects**

In order to ensure that proper consideration is given to ethical issues, all students undertaking the LIS dissertation project must complete this form and attach it to their dissertation proposal. Consult your supervisor if anything in this form is unclear or problematic. There are two parts:

*Part A: Ethics Checklist.* All students must complete this part. The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval. Students who answer 'yes' to any of questions 1-18 should consult their supervisor, as they may need approval from the ethics committee.

*Part B: Ethics Proportionate Review Form.* This part is an application for ethical approval of low-risk research. Students who have answered “no” to questions 1 – 18 and “yes” to question 19 in the checklist must complete this part; students who have answered ‘no’ to all the questions 1-19 may ignore this part. The supervisor has authority to approve this application.

### If your answer to any of the following questions (1 – 3) is YES, you must apply to an appropriate external ethics committee for approval:

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Does your project require approval from the National Research Ethics Service (NRES)? (E.g. because you are recruiting current NHS patients or staff? If you are unsure, please check at <a href="http://www.hra.nhs.uk/research-community/before-you-apply/determine-which-review-body-approvals-are-required/">http://www.hra.nhs.uk/research-community/before-you-apply/determine-which-review-body-approvals-are-required/</a>)</td>
<td>Yes/No</td>
</tr>
<tr>
<td>2. Will you recruit any participants who fall under the auspices of the Mental Capacity Act? (Such research needs to be approved by an external ethics committee such as NRES or the Social Care Research Ethics Committee <a href="http://www.scie.org.uk/research/ethics-committee/">http://www.scie.org.uk/research/ethics-committee/</a>)</td>
<td>Yes/No</td>
</tr>
<tr>
<td>3. Will you recruit any participants who are currently under the auspices of the Criminal Justice System, for example, but not limited to, people on remand, prisoners and those on probation? (Such research needs to be authorised by the ethics approval system of the National Offender Management Service.)</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>

### If your answer to any of the following questions (4 – 11) is YES, you must apply to the Senate Research Ethics Committee for approval (unless you are applying to an external ethics committee):

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Does your project involve participants who are unable to give informed consent, for example, but not limited to, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf?</td>
<td>Yes/No</td>
</tr>
<tr>
<td></td>
<td>Question</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>5</td>
<td>Is there a risk that your project might lead to disclosures from participants concerning their involvement in illegal activities?</td>
</tr>
<tr>
<td>6</td>
<td>Is there a risk that obscene and or illegal material may need to be accessed for your project (including online content and other material)?</td>
</tr>
<tr>
<td>7</td>
<td>Does your project involve participants disclosing information about sensitive subjects?</td>
</tr>
<tr>
<td>8</td>
<td>Does your project involve you travelling to another country outside of the UK, where the Foreign &amp; Commonwealth Office has issued a travel warning? (<a href="http://www.fco.gov.uk/en/">http://www.fco.gov.uk/en/</a>)</td>
</tr>
<tr>
<td>9</td>
<td>Does your project involve invasive or intrusive procedures? For example, these may include, but are not limited to, electrical stimulation, heat, cold or bruising.</td>
</tr>
<tr>
<td>10</td>
<td>Does your project involve animals?</td>
</tr>
<tr>
<td>11</td>
<td>Does your project involve the administration of drugs, placebos or other substances to study participants?</td>
</tr>
</tbody>
</table>

If your answer to any of the following questions (12 – 18) is YES, you should consult your supervisor, as you may need to apply to an ethics committee for approval.

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Answer Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Does your project involve participants who are under the age of 18?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>13</td>
<td>Does your project involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)? This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people.</td>
<td>Yes/No</td>
</tr>
<tr>
<td>14</td>
<td>Does your project involve participants who are recruited because they are staff or students of City University London? For example, students studying on a particular course or module. (If yes, approval is also required from the Project Tutor.)</td>
<td>Yes/No</td>
</tr>
<tr>
<td>15</td>
<td>Does your project involve intentional deception of participants?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>16</td>
<td>Does your project involve identifiable participants taking part without their informed consent?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>17</td>
<td>Does your project pose a risk to participants or other individuals greater than that in normal working life?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>18</td>
<td>Does your project pose a risk to you, the researcher, greater than that in normal working life?</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>
If your answer to the following question (19) is YES and your answer to all questions 1 – 18 is NO, you must complete part B of this form.

<table>
<thead>
<tr>
<th>19.</th>
<th>Does your project involve human participants? For example, as interviewees, respondents to a questionnaire or participants in evaluation or testing.</th>
<th>Yes/No</th>
</tr>
</thead>
</table>

**Part B: Ethics Proportionate Review Form**

If you answered YES to question 19 and NO to all questions 1 – 18, you may use this part of the form to submit an application for a proportionate ethics review of your project. Your dissertation project supervisor will review and approve this application.

<table>
<thead>
<tr>
<th>The following questions (20 – 24) must be answered fully.</th>
<th>Delete as appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. Will you ensure that participants taking part in your project are fully informed about the purpose of the research?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>21. Will you ensure that participants taking part in your project are fully informed about the procedures affecting them or affecting any information collected about them, including information about how the data will be used, to whom it will be disclosed, and how long it will be kept?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>22. When people agree to participate in your project, will it be made clear to them that they may withdraw (i.e. not participate) at any time without any penalty?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>23. Will consent be obtained from the participants in your project, if necessary? Consent from participants will only be necessary if you plan to gather personal data. “Personal data” means data relating to an identifiable living person, e.g. data you collect using questionnaires, observations, interviews, computer logs. The person might be identifiable if you record their name, username, student id, DNA, fingerprint, etc. If YES, attach the participant information sheet(s) and consent request form(s) that you will use. You must retain these for subsequent inspection. Failure to provide the filled consent request forms will automatically result in withdrawal of any earlier ethical approval of your project.</td>
<td>Yes/No</td>
</tr>
<tr>
<td>24. Have you made arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential? Provide details:</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>

If the answer to the following question (25) is YES, you must provide details

<table>
<thead>
<tr>
<th>If the answer to the following question (25) is YES, you must provide details</th>
<th>Delete as appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>25. Will the research involving participants be conducted in the participant’s home or other non-University location?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>If YES, provide details of how your safety will be ensured:</td>
<td></td>
</tr>
</tbody>
</table>

**Attachments (these must be provided if applicable):**

<table>
<thead>
<tr>
<th>Attachments (these must be provided if applicable):</th>
<th>Delete as appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant information sheet(s)</td>
<td>Yes / No / Not applicable</td>
</tr>
</tbody>
</table>
**If these items are not available or not applicable at the time of submitting your project proposal, preliminary approval through proportionate review can still be given. This will be subject to you submitting the items to your supervisor for approval at a later date. Approval must be obtained prior to the research commencing.**

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes / No / Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consent form(s)</td>
<td><strong>Yes / No / Not applicable</strong></td>
</tr>
<tr>
<td>Questionnaire(s)**</td>
<td><strong>Yes / No / Not applicable</strong></td>
</tr>
<tr>
<td>Topic guide(s) for interviews and focus groups**</td>
<td><strong>Yes / No / Not applicable</strong></td>
</tr>
<tr>
<td>Permission from external organisations (e.g. for recruitment of participants)**</td>
<td><strong>Yes / No / Not applicable</strong></td>
</tr>
</tbody>
</table>
Appendix C: Reflection

My interest in library analytics came from a lecture given by the Head of User Services at City's library as part of a Digital Libraries module. I had not worked in a library prior to the start of this course and so had not fully realised the breadth of data academic libraries held or had access to regarding their services and how they were used. Becoming aware of some of the issues faced within the library regarding the analysis of its own data felt like knowing there was a three dimensional world but only being able to see it in two dimensions. I was curious to answer questions that so far were not easily answered.

Through the dissertation process I have learnt a great deal about the methodical approach needed to undertake such a project. Decisions taken early in the process regarding the fields to extract or retain in the various data sets were done so from the perspective of only holding information that was pertinent to the investigation, rather than keeping everything on the basis of 'maybe it will be useful'. Consequently it meant isolating first year SMCSE undergraduates was less precise than it could have been. This reflects the kind of decision-making that occurs when constructing data warehouse systems where a large number of data fields and therefore reporting flexibility has to be balanced with the cost of data processing and storage requirements. Course expiry dates were needed to exclude students out of scope but I found it difficult to consider what I might lose by not selecting start dates until after the decision had been made.

Despite a change to the timeframe to be investigated, the project did not change a great deal from the original proposal although having to repeat sections of work due to unavailability of data for 2016/2017 did slow my progress somewhat. I learnt a valuable lesson about treating the data with a degree of circumspection before any further work was undertaken, and put in place some cross checks to prevent a repetition.

Once all the data had been assembled and was ready for analysis, getting an understanding of the dimensions of each set helped to give some initial structure.
However, when combining the different data sets, it was tempting to be distracted by wanting to answer questions that served my own curiosity about understanding the exceptions to the norm rather than focusing on developing an understanding of a broad range of users. I also found the large array of investigative options made it difficult to settle on a route through the information that would prove helpful or illuminating with regard to resource usage. If the time frame had been longer, I would have liked to drill down further into the data particularly with regard to low use by SMCSE undergraduates.

I also found selecting the way to display the information in a format that allowed for patterns or general dimensions stretched me. I had originally anticipated that I would use the Springshare LibApps package, which the library already uses for a number of data sets and also features a dashboard style interface for combining different data sets. This approach was eventually disbanded as the system was not intuitive for a new user, importing the data was time consuming and required reformatting, and the visual outputs were not a significant improvement on those found in Microsoft Excel, just easier to use once the data was in there. Being aware that good choices of data visualization can make a big difference in how unwieldy data sets can be understood, it was disappointing not to have found a better way to interpret and present the information obtained but it has made me appreciate how important this aspect is in complex data sets and how difficult it is to get right.

Overall I have enjoyed the rather immersive experience of conducting my own research, particularly the exploration of what the combined data could be used for. I was also struck by the enthusiasm shown by other members of staff regarding the work I was undertaking. Their willingness to give up time from their own schedules to provide insight, assistance as well as running reports helped a great deal not just with the mechanics of the research but also in keeping me motivated through some of the more troublesome parts. Furthermore, the experience has left me wanting to explore this area of library assessment further and I hope to be able to incorporate it in a future role.