The JISC MOSAIC Project

Making Our Scholarly Activity Information Count

Final Report - Appendices

January 2010

Project partners

Sero Consulting
Ken Chad Consulting
Mark van Harmelen, HedTek
Paul Miller, Cloud of Data
Dave Pattern, University of Huddersfield

List of Appendices

Appendix 1 – International Use Data Exemplars

In 2006 Lorcan Dempsey commented: 'Data is inert in our systems. We do not release the investment in structured data in engaging and interesting user experiences. We are seeing signs that this is changing, but it has taken some time. Nor do we make much use of 'intentional' data, data about choices, use and usage, to refine and manage services. Holdings data, circulation data, database usage data, resolution data: much more could be done to mine this for intelligence about how to develop services. Think of services like 'people who borrowed this, also borrowed that' for example '[1].

Looking specifically at recommender systems a European survey revealed 'that in scientific libraries recommender services are still not in wide use — despite the considerable benefits they offer for students and scientists' [2]. A few HE libraries use Google analytics[3]to track where their website users come from, what links the click on and the content they view and this information can be used to redesign library web sites. However the picture painted by Dempsey in 2006 remains essentially true. There are a few notable exceptions and some of these are described below. This is not an exhaustive list. It does illustrate a range of approaches using different methodologies and algorithms use by HE institutions and commercial vendors

California Digital Library[4]

The Melvyl Recommender project is significant as a US system that makes use of historical circulation data for generating recommendations. (It should be noted here that Melvyl at the time was *not* based on a commercial LMS system and *did* store historical transactions). The University of California System has since migrated to ExLibris and makes use of the bX recommender service. It is also important in being in a union catalogue setting. The Melvyl Recommender Project team explored two methods of generating recommendations. The first method used circulation data from the University of California, Los Angeles (UCLA) to determine linkages between items ("patrons who checked this out also checked out..."). A second, content-based, strategy used terms from the bibliographic records to develop queries for similar items ("more like this..."). The project showed strong evidence that University of California (UC) library users are interested in receiving recommendations to support both academic and personal information needs. The first attempt to produce

---

[1] 'The (Digital) Library Environment: Ten Years After.' By Lorcan Dempsey. Ariadne Issue 46. 8th February 2006.http://www.ariadne.ac.uk/issue46/dempsey/intro.html

[2] 'Recommender Services in Scientific Digital Libraries.' By Markus Franke, Andreas Geyer-Schulz and Andreas W. Neumann. In Studies in Computational Intelligence. [Pages 377-417]. Volume 120/2008. ISSN 1860-949X (Print) 1860-9503 (Online). Springer. DOI 10.1007/978-3-540-78502-6

[3]Eg Open University. See for Tony Hirst and Hassan Sheikh's presentation at the 2009 Internet Librarian International http://conferences.infotoday.com/documents/52/C102_Sheikh.ppt

[4] 'The Melvyl Recommender Project. Developing Library Recommendation Services.' By Colleen Whitney and Lisa Schiff. D-Lib Magazine. December 2006. Volume 12 Number 12. ISSN 1082-9873. http://www.dlib.org/dlib/december06/whitney/12whitney.html

recommendations using circulation data met with mixed results. Only about a third of the recommendations were helpful to participants in user testing. Nevertheless, participants were almost unanimous in their support for development of such services.

MESUR[5]

MESUR stands for MEtrics from Scholarly Usage of Resources. The MESUR database is very large scale. It contains 1 billion usage events obtained from 6 significant publishers, 4 large institutional consortia and 4 significant aggregators. The collected usage data spans more than 100,000 serials (including newspapers, magazines, etc.) and is related to journal citation data that spans about 10,000 journals and nearly 10 years (1996-2006). In addition it contains publisher-provided COUNTER usage reports that span nearly 2000 institutions worldwide. The data is being ingested into a combination of relational and semantic web databases.

MESUR is now producing large-scale, longitudinal maps of the scholarly community and a survey of more than 60 different metrics of scholarly impact.

ExLibris worked with the project and has adopted the Mesur algorithm in it bX recommender service (see below).

Ex Libris bX[6]

The ExLibris bX service is the result of collaborative research into scholarly recommender systems conducted by the bX team and leading researchers Johan Bollen and Herbert Van de Sompel from the Los Alamos National Laboratory who were responsible for the MESUR project. The bX service does not make use of MESUR *data* but it does make use of the algorithm employed in the project

Based on data captured through a large-scale aggregation of link-resolver usage logs, bX is an extension of the OpenURL framework. BX claims to provide highly granular recommendations that point to specific scholarly articles. The bX Recommender is offered as an on-demand service, hosted service meaning that there is nothing to install or maintain on the local library/university network and ExLibris say that the service can be up and running in one business day.

BibTip [7]

Between 2002 and 2007, several projects funded by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) were carried out in Karlsruhe in Germany, all of them concerning the development of recommender systems for libraries. Out of these BibTip emerged in 2007. Project partners in the development were Karlsruhe University Library and the Institute for Information Services and Electronic Markets.

---

[5] http://www.mesur.org/MESUR.html. There is a useful overview article 'MESUR: usage-based metrics of scholarly impact.' By Johan Bollen, Marko A. Rodriguez and Herbert Van de Sompel http://www.mesur.org/Documentation_files/JCDL07_bollen.pdf

[6] http://www.exlibrisgroup.com/category/bXOverview

[7] 'Adding Value to the Library Catalog by Implementing a Recommendation System'. By Dr. Michael Mönnich and Marcus Spiering. D-Lib Magazine. May/June 2008. Volume 14 Number 5/6. ISSN 1082-9873 http://www.dlib.org/dlib/may08/monnich/05monnich.html#2

The Institute developed the algorithms and the scientific basis for the Karlsruhe recommender system, and the Karlsruhe University Library was responsible for the system's integration into the library catalogue, the collection of statistical data and, most recently, the development and operation of BibTip as a commercial service.

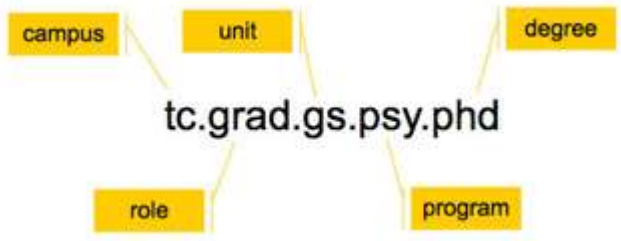
The BibTip recommendation system is based on the behavioural patterns of users interacting with a library catalogue (OPAC). This so called "implicit" recommendation service rests upon the observation of user patterns and the statistical evaluation of the usage data. All the data stored and processed are anonymous (identification numbers and session IDs).

On a technical level the BibTip architecture may be seen as an agent architecture involving three software agents: the OPAC Observation Agent, the Aggregation Agent, and the Recommendation Agent. The OPAC Observation Agent observes the selection of titles within defined OPAC sessions. These data are transferred to the Aggregation Agent, which then does computations on the statistical material to arrive at a list of recommendations. Lastly, the Recommendation Agent presents the list of recommendations to the user.

The University of Minnesota Libraries[8]

The University of Minnesota Libraries have created a 'MyLibrary' portal, with databases and e-journals targeted to users, based on their *affiliations*. It does this by means of what they term an affinity string. The University's enterprise authentication system provides this affinity string, which is used to deliver a personalised user experience. The affinity string describes a user's academic department and degree program or position at the University and so automates discovery of a user's relationship to the University. Affinity strings also provide the Libraries with an anonymized view of resource usage, allowing data collection that respects user privacy and lays the groundwork for automated recommendation of relevant resources based on the practices and habits of their peers

Affinity strings are created by the Office of Information Technology using data from the University's PeopleSoft system. Everybody at the University of Minnesota is assigned one or more affinity strings based on their area of study or work. For example, a PhD student in Psychology may receive the affinity string:



Every student, faculty member, and staff member has a unique Internet ID which is used to access a variety of systems, ranging from email to course registration to

[8] 'Affinity Strings: Enterprise Data for Resource Recommendations.' By Cody Hanson, Shane Nackerud, and Kristi Jensen. Code{4}lib Journal. Issue 5 15th December 2008. ISSN 1940-5758. http://journal.code4lib.org/articles/501

licensed library resources. Known as the Central Authentication Hub (CAH). Each time a user with an active CAH cookie accesses a library resource, we can capture his affinity string and associate it with the resource being requested. We thus gain valuable insight into which of our user populations are using (or not using) a given resource. By aggregating the activity of a number of users with the same affinity string, we can sufficiently anonymize our data to protect user privacy.

Appendix 2 - University of Huddersfield Implementation

Scope[9]

Huddersfield University has made extensive use of a variety if usage data to deliver a number of locally developed extended features to their OPAC. As part of the development of the UK version of (Dynix) Horizon back in the early 1990s, libraries requested that the company add code to log all circulation transactions. Up to 2009 Huddersfield had logged data for 3,157,111 transactions. Contrastingly, some US systems deliberately avoided tracking historical usage data, which was regarded as 'un-American'; consequently the Millennium (Innovative Interfaces) LMS still does not store historical circulation transaction data.

The work of Dave Pattern at Huddersfield illustrates that usage data can be exploited locally to provide a wide range of services to enhance a standard LMS OPAC. For example:

- "People who borrowed this, also borrowed…" suggestions. Dave Pattern commented in his blog that apart from the 'did you mean' spelling suggestions, 'this has been the most popular tweak we've made to our OPAC.'
- Once a user logs into the OPAC, the system can provide a personal suggestion by generating the suggestions for the books they've borrowed recently and then picking one of the titles that comes out near the top.
- A new book lists per course of study. Whenever a new book is added to our catalogue, a check is made to see which courses have previously borrowed heavily from that Dewey class and then add the book details to their feeds. The feeds are picked up by the University Portal, so students should see the new book list for their course and the titles should be relevant to their studies.
- 'Good' keywords to combine with current search terms. For example a user may start with a general search for "law" which brings back an unmanageable results set. In the background, the code searches through all of the previous keyword searches that contained law and pulls together the other keywords that are most commonly used in multi-keyword searches that included "law". With a couple of mouse clicks, the user can narrow the search down to a manageable 34 results for "criminal law statutes". In his blog Dave Pattern remarks ton two key attributes of this particular feature
  '1) I didn't have to ask our librarians to come up with the lists of good keywords to combine with other keywords — they've got much more important things to do with their time
  2) The service acts as a feedback loop — the more searches that are carried out, the better the suggestions become.'

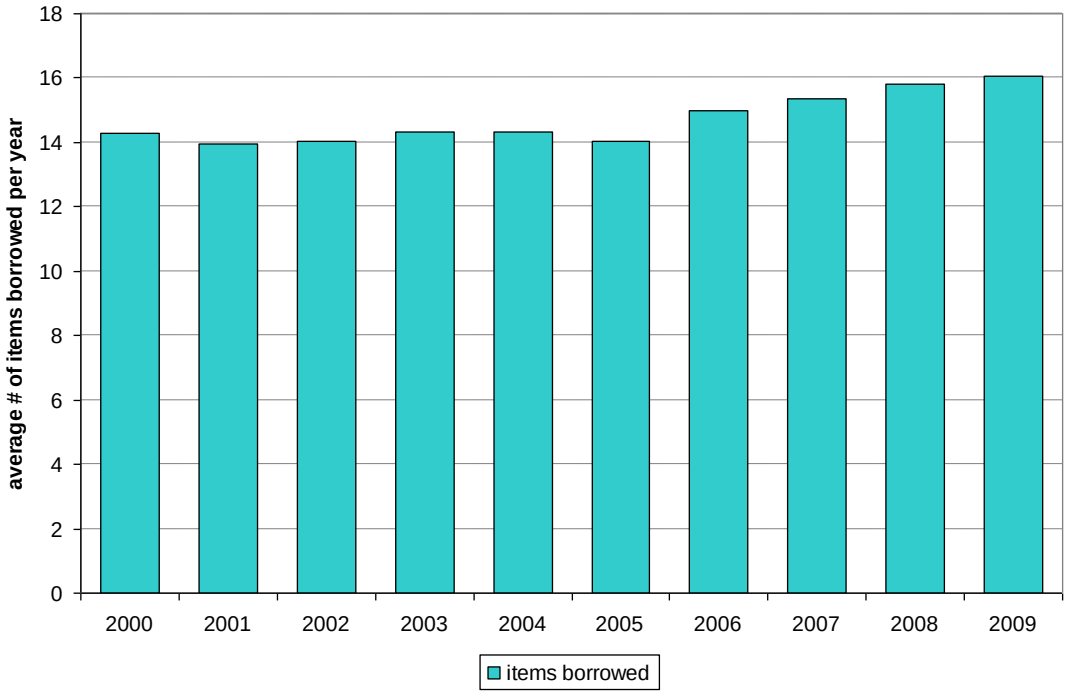
The blog makes an interesting observation about recommender services. 'One of the comments I frequently hear is that book recommendation services might create a

[9] A good summary of the Dave Patterns work at Huddersfield in s in his 2009 Internet Librarian International presentation blog entry http://www.daveyp.com/blog/archives/1317

"vicious circle" of borrowing, with only the most popular books being recommended. At Huddersfield, we've seen the opposite - since adding recommendations and suggestions, the range of stock being borrowed has started to widen'. He goes on to say. 'Not only are students borrowing more widely than before, they're also borrowing more books than before'

Evidence

Graph 1 - Average number of books borrowed per year by active borrowers

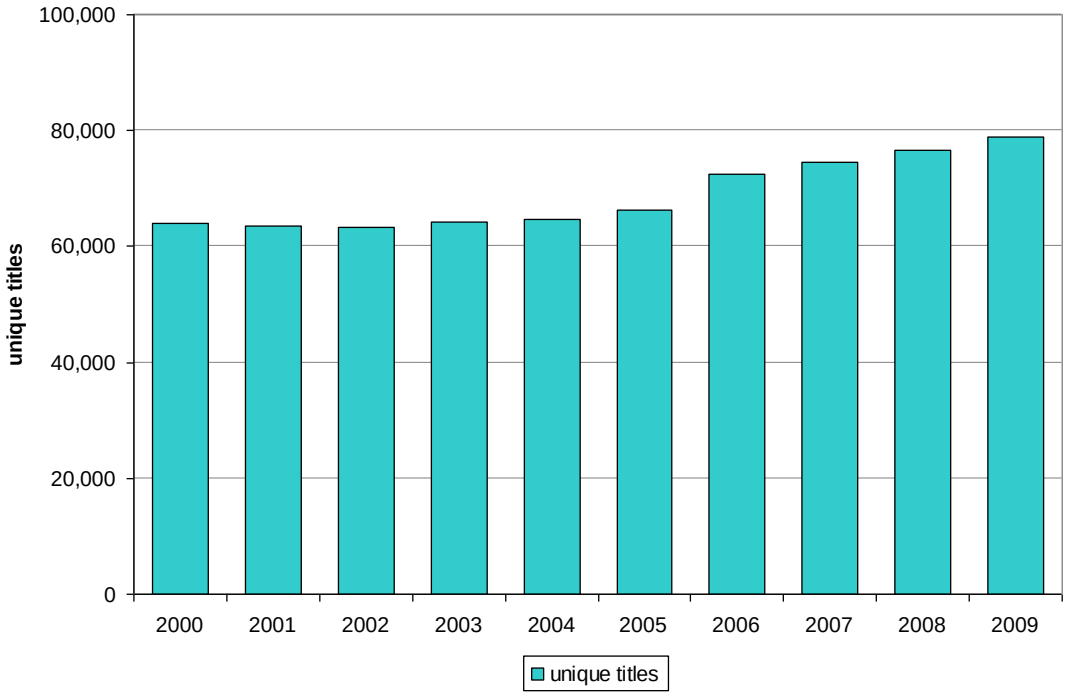


Comments

- Since the introduction of serendipity into the library catalogue at the start 2006, the average number of books borrowed by students has increased year-on-year.

- During the same period, the library has seen a decrease in the number of renewals.

Graph 2 - Range of unique titles borrowed per year



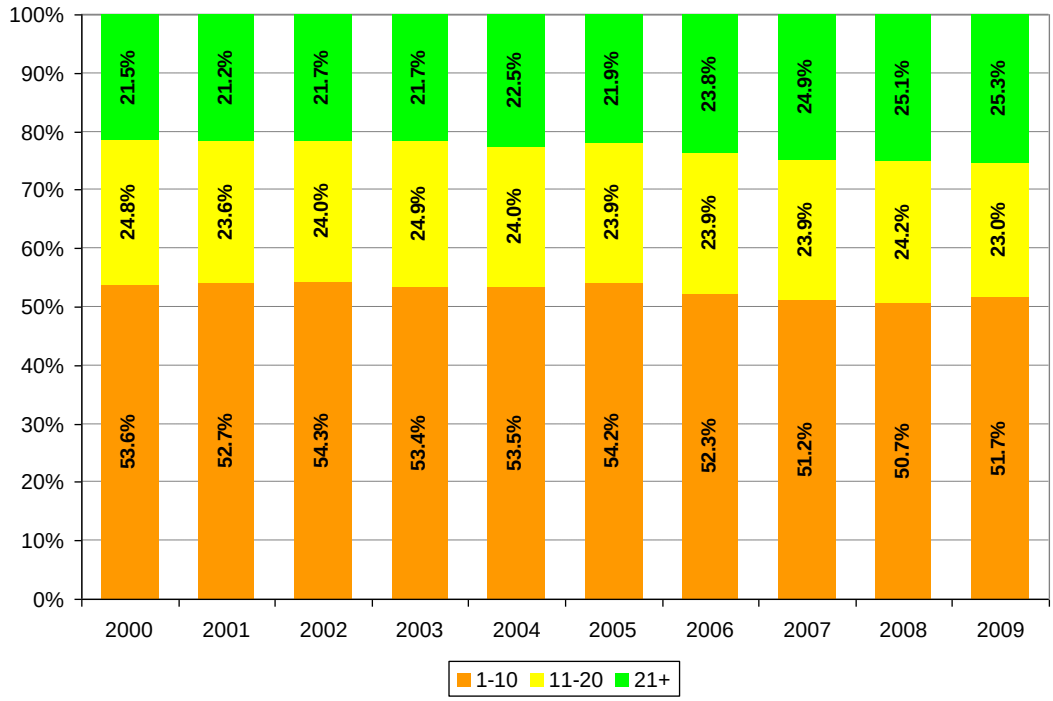
Comments

- The range of unique titles borrowed per year has increased since the introduction of suggestions and recommendations within the library catalogue at the start of 2006.

- During the period 2000 to 2005, approximately ¼ of the titles held by the library circulated per year.

- Since 2006, the range has increased year-on-year – in 2009, approximately ⅓ of the titles held by the library were borrowed. Approximately 2,000 additional titles are circulating each year.

Graph 3 - Number of books borrowed per year



Comments

- The above graph shows the percentage of students who fall into particular borrowing bands. The majority of students (around 50%) are low users of the library, borrowing 10 or fewer books per year.

- From 2000 to 2005, the borrowing profile of students at the University of Huddersfield was fairly consistent, with around 53% of students borrowing 10 or fewer books per year and around 21% borrowing more than 20 books per year.

- Since the introduction of recommendations and suggestions in the OPAC at the start of 2006, there has been an increasing trend towards higher borrowing, especially in the group who borrow more than 20 books per year.

- At a time when many HE Academic Libraries are reporting a general decline in book borrowing, the data seems to indicate that adding serendipity to the library catalogue helps to encourage students to borrow more books and to explore a wider range of titles.

Appendix 3 – MOSAIC data formats: A guide

Introduction

Sources

The dataset may be derived from a variety of sources:

- Reading lists
- Use records derived from library circulation/loan systems (LMSs)
- Use records derived from virtual learning environments (VLEs)
- OpenURL resolver / ERM derived information

In this appendix we discuss only the first two of these components.

Libraries should only supply data that is sufficiently anonymised so as not to require registration or other actions under the Data Protection Act (DPA) - more about this below.

Terminology - Courses and modules

In what follows we use **course** to refer to a course as an entity for which an institution grants a degrees or certificates. Two examples might be *BSc Computer Science at the University of Manchester*, or a *PhD in Geology at the University of Southampton*.

A **module** is an entity contributing to a course, so we might have *CS2034 Database Technology*, a second year Computer Science Module at the University of Manchester.

Desired data

We use symbols to indicate the importance of information to us in our current work.

▶ Fields that are absolutely essential
▷ Fields that we would really like to have
– Fields that are desirable if easily available

If you can only supply some data please discuss it with us, there is considerable flexibility, for example, we will be very happy with use records where the resource is only identified by an ISBN, ISSN, DOI etc. and lacks other resource information.

Reading lists

You may hold reading lists in Word or HTML formats, which are fine for trial use, but so are more structured forms - such as LMS, VLE or other structured records.

The data is

- The institution – e.g. *University of Manchester*
- The reading list's module ID – *CS3024*
- The reading list's module name – *Database Technology*
- The academic year the module is being offered in – *2008/2009*
- The progression level most students are at when they take the module – *UG2*
- The course(s) the module is offered for, as UCAS code [10] and course name pair(s) – *F1G1, BSc Chemistry with Mathematics*
- The reading list – a list of the recommended items

Some notes:

- If the entire reading list is published on the web please supply the content of the reading list, rather than the URL at which it is stored.
- If you are supplying catalogue entries, it is preferable to supply the catalogue entries in MODS rather than MARC format.
- Exports from Talis List, Talis Aspire or Ex Libris Reading Room are very welcome.

Loan based use records

An example use record might contain circulation information that *a first year Chemistry undergraduate at the University of Hull borrowed a book "Organic Chemistry: A Primer" in academic year 07/08*.

There could be other data in the record as discussed below.

Data levels for use records

We are expecting that library loans will be the most common kinds of data, so we illustrate their use records here. It is a small step to imagine use data for other sources, e.g. VLEs.

For library use records that represent loans, there are three levels of data. The level represents the amount of information in a record. All data supplied by the institution should be at the same level.

[10] UCAS code background information

http://www.ucas.ac.uk/he_staff/datamanagement/jacs/coursecodingprinciples

- Level 0 use data may be used to personalize search for resources like library items, promoting results for popular items according to various criteria.

- Level 1 use data may be used to personalize search for resources like library items, promoting results for popular items according to various criteria. It may also be used to implement search results faceted on any combination of institution, course, progression level and/or academic year associated with the loan.

- Level 2 Use Data may be used to provide more complex information about resource use, that is not derived from a single record; for example
  - *'borrowers who borrowed this also borrowed …'*
  - *'users of this went on to …'*

We describe the fields of interest in use records that represent library loans below.

Level 0 library loan use records

Basic data

- Institution identifier – the full name of your institution - without any preceding 'the' – *Oxford Brookes University, University of Oxford.*
- Academic year – formatted as yyyy/yyyy - *2007/2008*
- Record date - the date the record was extracted – formatted as yyyy.mm.dd – *2009.05.10*
- Item source – *LMS*

Note on Item sources: These are identified by a string, e.g., LMS for library management system, OUR for OpenURL Resolver, VLE for virtual learning environment, etc. We only consider LMS here, but it is our intention that the same or very similar format should work for all item sources.

Resource-related data

- Media type – e.g. *Book, e-Book, Journal, Article, Presentation*
- Author or editor – *A.N. Author*
- Title – *Global Pandemics in the Twentieth Century*
- Global ID type and its value, if the item has an ISBN, ISSN, SIKI, or DOI please supply it, it is more important than the author/editor and title – *ISBN 12345678*
- Reliable local ID – unique ID from your source system, such as internal catalogue reference which must is persistent for updating purposes - *0123456789*
- Persistent URL - for the resource or its catalogue entry – *http://www.lib.inst.ac.uk/cat/012345t789.html*
- Publisher – *John Wiley and Sons*
- Publication Year – *1978*
- Journal Issue - for a Journal related record volume, issue number and month – *volume 12, issue 3, July*
- acquisition date
- creator

Level 1 library loan use records

Basic data

- Institution identifier – the full name of your institution - without any preceding 'the' – *Oxford Brookes University*, *University of Oxford*.
- Academic year – formatted as yyyy/yyyy - *2007/2008*
- Record date - the date the record was extracted – formatted as yyyy.mm.dd – *2009.05.10*
- Item source – *LMS*

Note on Item sources: These are identified by a string, e.g., LMS for library management system, OUR for OpenURL Resolver, VLE for virtual learning environment, etc. We only consider LMS here, but it is our intention that the same or very similar format should work for all item sources.

Resource-related data

- Media type – e.g. *Book*, *e-Book*, *Journal*, *Article*, *Presentation*
- Author or editor – *A.N. Author*
- Title – *Global Pandemics in the Twentieth Century*
- Global ID type and its value, if the item has an ISBN, ISSN, SIKI, or DOI please supply it, it is more important than the author/editor and title – *ISBN 12345678*
- Reliable local ID – unique ID from your source system, such as internal catalogue reference which must is persistent for updating purposes - *0123456789*
- Persistent URL - for the resource or its catalogue entry – *http://www.lib.inst.ac.uk/cat/012345t789.html*
- Publisher – *John Wiley and Sons*
- Publication Year – *1978*
- Journal Issue - for a Journal related record volume, issue number and month – *volume 12, issue 3, July*
- acquisition date
- creator

User context related data

For students:

- global Student course UCAS code, please supply where available – *F1G1 - jacs*
- Course – the user's course name as a string – *BSc Chemistry with Mathematics*
- Progression level – *FD, UG1, UG2, UG3, UG4, MD, PhD1, PhD2, PhD3+*

For staff:

- Staff type, if this is available – *acStaff, resStaff, adminStaff* (or if not available, simply use the designator *staff*)

Level 2 use data

Level 2 use data contains a GUID (a large random number) to represent the user making the withdrawal. This should replace any traceable identifier such as the Institutional Student ID or Library User ID.

There will be many records with a single GUID to represent an anonymised user's use of many resources over time. For DPA purposes, the dates in all records for a particular GUID are replaced by a sequence number, and any 'singleton' or 'doubleton' records are removed.

DPA requirements

There is code to help libraries to meet the anonymisation requirements of the DPA.

For level 1 data this will strip out 'singleton' or 'doubleton' records.

For level 2 data this will strip out 'singleton' or 'doubleton' records, replace names with GUIDs, and replace dates of loans with sequence numbers (sequenced 1, 2, 3, … for each GUID).

Examples of use records

Three examples of the kinds of data that interesting in the context of the Mosaic project are tabulated below.

Institution Name	University of Huddersfield	University of Huddersfield	University of Huddersfield
Academic year	2008/2009	2008/2009	2008/2009
Record date	2009.04.06	2009.03.31	2009.02.01
Item Source	LMS	VLE	OUL
Title	The Waves	Woolf & the cult of modernism – lecture	Woolf & the cult of modernism
Media Type	Book	Presentation	e-Journal
Global ID	ISBN 98765		ISSN 98765
Reliable local ID	LMS123	VLE999	ERM987
Resource URL		www.hud.ac.uk/123.html	www.openurlhere
Author	Virginia Woolf	Prof. A. Smith	A.B.Smith
Publisher	Bloomsbury		Wolf Society Journal
Publication Year	1927		2005
Journal Issue			15.04
Course	English	English	English
UCAS code	1234	1234	1234
Progression level	UG1	UG1	UG3
Modules list		Female Authors	

Use record format

The following shows the data needed for a use record, e.g. a load of a book, the use of a VLE record, a reference to an item via an OpenURL resolver.

We propose two formats, as name value pairs, and as XML.

Use record name value format

Basic format details:

- One name=value pair per line, starting column 1, no blanks on either side of the = sign, the value terminated at the end of the line
  (i.e. no multi-line values, spaces allowed within the value, spaces after the last non-space character on the line ignored)

- The name value pairs that constitute a loan or use record appear on subsequent lines, and are terminated by a blank line, that may have no characters, or that may consist of one or more spaces.

Illustrative examples of name value pairs appear below

Basic data: Institution, year and dates

- Institution identifier – the full name of your institution - without any preceding 'the' – *Oxford Brookes University*, *University of Oxford*.
  `institution=Oxford Brookes University`

- Academic year – formatted as yyyy/yyyy
  `academicYear=2007/2008`

- Record date - the date the record was extracted – formatted as yyyy.mm.dd
  `extractedOn=2009.06.20`

- Item source
  `source=LMS`

  Note on Item sources: These are identified by a string, e.g., `LMS` for library management system, `OUR` for OpenURL Resolver, `VLE` for virtual learning environment, etc.

Resource data

- Resource global ID type and its value – if the item has an ISBN, ISSN, SICI, or DOI please supply it,; this is more important than the author/editor and title. Please use the capitalised global ID type `ISBN`, `ISSN`, `SICI` or `DOI`. Depending on how use records are constructed we can see one case where multiple ISBNs might be associated with a record, so if this is the case please supply them all; otherwise we only expect one global ID per use record. Some examples are

```
ISBN=0123456789
ISSN=0378-5955
SICI=0095-4403(199502/03)21:3<12:WATIIB>2.0.TX;2-J
DOI=10.1000/j.1365-313X.2008.03660.x
```

- Media type – choose from `book`, `ebook`, `journal`, `article`, `ppt`, `doc`, `jpg`, `wav`, `mpg` etc.
  `media=book`

- Author or editor
  `author=A.N. Author`
  `editor=A.N. Editor`

- Title
  `title=Global Pandemics in the Twentieth Century`

- Reliable local ID – unique ID from your source system, must is persistent for the item being used
  `localID=0123456789`

- Catalogue URL
  `catalogueURL=http://www.lib.inst.ac.uk/cat/012345.html`

- Resource URL - for an electronic resource; this might be a non-redirected URL, or a redirectable URL such as an OpenURL or a DOI redirection
  `resourceURL=http://www.lib.inst.ac.uk/res/something.html`

- Publisher
  `publisher=John Wiley and Sons`

- Publication Year
  `published=1978`

- Journal Issue - for a Journal related record volume, issue number and month – *volume 12, issue 3, July*
  `volume=12`
  `issue=3`
  `month=July`

User context data

For students and staff, level 2 data only, this must be anonymised

- User identity – the user name should be translated into a GUID
  `user=38a52be4-9352-453e-af97-5c3b448652f0`

- Date of use – this will be later transformed into a per-user sequence number
  `useDate=2008.04.27`

For students, in both level 1 and level 2 data:

- Course code – Where available please supply either or both of the UCAS code (followed by) comma-separated JACS codes in the 'letter three digits' format
  `UCAS=F1G1`

```
JACS=F102,G144,Z123
```

- Course name – the user's course name as a string
  `courseName=BSc Chemistry with Mathematics`

- Progression level – please choose from the following
  `F` (for foundation year), `UG1`, `UG2`, `UG3`, `UG4`, `M`, `PhD1`, `PhD2`, `PhD3+`
  `progression=UG2`

For staff, in both level 1 and level 2 data:

- Staff type – if this is available, please choose from
  `acStaff`, `resStaff`, `adminStaff` (or if not available, simply use `staff`)
  `progression=acStaff`

Use record XML format

We use a simple XML format: Names used in the name-value format are used as tags in the XML format, and values are the same as defined in the name-value format section above.

Start records

```
<useRecordCollection>
<useRecord>
```

Basic data: Institution, year and dates

```
<from>
```

- Institution identifier – the full name of your institution - without any preceding 'the' – *Oxford Brookes University*, *University of Oxford*.

```
<institution>
      Oxford Brookes University
</institution>
```

- Academic year – formatted as yyyy/yyyy

```
<academicYear>
      2007/2008
</academicYear>
```

- Record date - the date the record was extracted

```
<extractedOn>
      <year>2009</year>
      <month>6</month>
      <day>20</day>
</extractedOn>
```

- Item source – `LMS`, `VLE`, etc

```
<source>
      LMS
</source>
```

```
</from>
```

Resource data

```
<resource>
```

- Media type – choose from `book`, `ebook`, `journal`, `article`, `ppt`, `doc`, `jpg`, `wav`, `mpg` etc.

```
<media>
      book
</media>
```

- Global ID type and its value – if the item has an ISBN, ISSN, SICI, or DOI please supply it; this is more important than the author/editor and title. Please use the capitalised global ID type `ISBN`, `ISSN`, `SICI` or `DOI` as type attributes. Depending on how use records are constructed we can see one case where multiple ISBNs might be associated with a record, so if this is the case please supply them all; otherwise we only expect one global ID per use record.

```
<globalID type="ISBN">
      0123456789
</globaLID>
```

or

```
<globalIdCollection>
      <globalID type="ISBN">
            0123456789
      </globaLID>
      . . .
</globalIdCollection>
```

Any angle brackets in a global ID, such as may be found in a SICI, must be translated into < and >

```
<globalID type="SICI">
      0095-4403(199502/03)21:3&lt;12:WATIIB&gt;2.0.TX;2-J
</globalID>
```

- Author

```
<author>
      A.N. Author
</author>
```

or editor

```
<editor>
      A.N. Editor
</editor>
```

- Title

```
<title>
      Global Pandemics in the Twentieth Century
</title>
```

- Reliable local ID – unique persistent ID from your source system, such as an internal catalogue number

```
<localID>
      0123456789
</localID>
```

- Catalogue URL - for the resource or its catalogue entry

```
<catalogueURL>
      http://www.lib.inst.ac.uk/cat/012345.html
</catalogueURL>
```

- Resource URL - for an electronic resource; this might be a non-redirected URL, and OpenURL, or a DOI redirection, i.e. any URL that locates or redirects to the resource

```
<resourceURL>
      http://www.lib.inst.ac.uk/res/something.html
</resourceURL>
```

- Publisher

```
<publisher>
      John Wiley and Sons
</publisher>
```

- Publication Year

```
<published>
      <year>1978</year>
</published>
```

- Journal Issue - for a Journal related record volume, issue number and month – *volume 12, number 3, July*

```
<journalData>
      <volume>12</volume>
      <number>3</number>
      <month>6</month>
</journalData>
```

```
</resource>
```

User context data

```
<context>
```

- For students and staff, level 2 data only: User identity, the user name could be translated into a GUID, and the same GUID should be used to identify a record for each of the same user's loans in the same academic year.

```
<user>
    38a52be4-9352-453e-af97-5c3b448652f0
</user>
```

- Date of use e.g. that a library loan was made.

This will be transformed into a per user sequence number by our software

```
<useDate>
      <year>2009</year>
      <month>6</month>
      <day>20</day>
</useDate>
```

For students, in both level 1 and level 2 data:

- Course codes – Most HE courses have UCAS codes, and they are very important to the work we propose. We would appreciate as many use records with UCAS codes as possible. If available, we are also interested in a breakdown into letter and three digit format JACS codes

```
<courseCodeCollection>
      <courseCode type="ucas">F1G1</courseCode>
      <courseCode type="jacs">F102</courseCode>
      <courseCode type="jacs">G144</courseCode>
      <courseCode type="jacs">Z123</courseCode>
</courseCodeCollection>
```

If there is only one `courseCode` then simply use

```
<courseCode type="ucas">
      F1G1
</courseCode>
```

- Course name– the user's course name as a string

```
<courseName>
      BSc Chemistry with Mathematics
</courseName>
```

- Progression level, please choose from the following

`F`, `UG1`, `UG2`, `UG3`, `UG4`, `M`, `PhD1`, `PhD2`, `PhD3+` (`F` is for foundation year)

```
<progression>
      UG2
</progression>
```

For staff, in both level 1 and level 2 data:

- Staff type, if this is available, please choose from

`acStaff`, `resStaff`, `adminStaff` (or if not available, simply use `staff`)

```
<progression>
      acStaff
</progression>

</context>
```

End record, more records

```
</useRecord>
<!-- more useRecords here if need be -->
```

End all records

```
</useRecordCollection>
```

Reading List formats

The data is divided in three parts

- The institution and year the reading list is from
- The module that the reading list is from, together with any course and progression information that can be supplied
- The items that compose the reading list, these are called resources below

We would like to discuss formats with you, please, if you can only provide reading lists in formats that are not described below.

Reading list name=value format

Institution and year

As above.

Module

- The reading list's module ID
  moduleId=*CS3024*

- The module name
  moduleName=*Database Technology*

We do not know the extent to which institutions will be able to supply the course(s) for which the module is offered. If this information is available it is valuable. In the name value syntax, a reading list may only be associated with one module and course combination, and the reading list should be repeated for any other courses where the module is offered. If supplied this information is as in 0 p18, reproduced below.

- Course code – Where available please supply either or both of the UCAS code (followed by) comma-separated JACS codes in the 'letter three digits' format
  `UCAS=F1G1`
  `JACS=F102,G144,Z123`

- Course name – the user's course name as a string
  `courseName=BSc Chemistry with Mathematics`

- Progression level – please choose from the following
  `F` (for foundation year), `UG1`, `UG2`, `UG3`, `UG4`, `M`, `PhD1`, `PhD2`, `PhD3+`
  `progression=UG2`

Resources

- The reading list – a list of the recommended items

Ideally the format is

> Starting with an empty line,
> A list of resources in the format of 0, p17, separated by empty lines

However, other formats are possible, such as an associated word file.

Reading list XML format

```
<readingListCollection>

      <readingList>

          <from>
              <!-- as 0, p19 -->
          </from>

          <module>
              <moduleId>
                  CS3024
              </moduleId>
              <moduleName>
                  Database Technology
              </moduleName>
              <courseAndProgressionCollection>
                  <courseAndProgression>
                       <!-- as courseCodes, courseName, and
                       progression under 0, p21  -->
                       <courseCodeCollection>
                            <!-- ... -->
                        </couseCodeCollection>
                       <courseName>
                           BSc Chemistry with Mathematics
                       </courseName>
                       <progression>
                           UG3
                       </progression>
                  </courseAndProgression>
                  <!-- ... repeat courseProgression etc -->
              </courseAndProgressionCollection>
          </module>

          <resourceCollection>
              <resource>
                  <!-- as 0, p20 -->
              </resource>
              <!-- ... repeat resource as needed -->
          </resourceCollection>
```

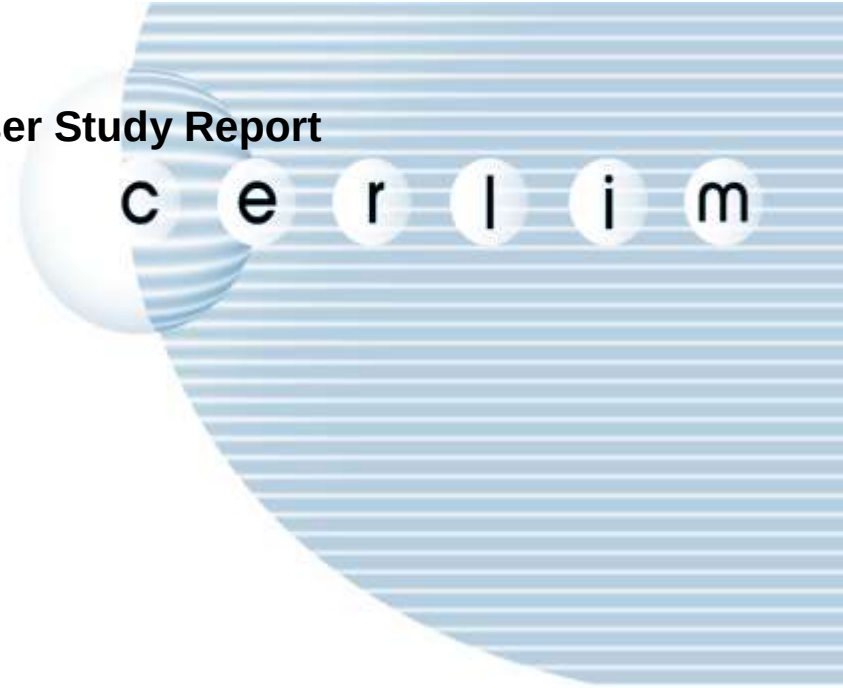
```
        </readingList>

        <!-- ... repeat readigList as needed -->
</readingListCollection>
```

- If there is only one course and progression then substitute element `courseAndProgression` for `courseAndProgressionCollection` above

- If there is only one course code then substitute the element `courseCode` for `courseCodeCollection` above

Appendix 4 – CERLIM User Study Report



MOSAIC Demonstrator Evaluation

User evaluation undertaken by CERLIM
Manchester Metropolitan University

November 2009

1. Introduction

The evaluation of the MOSAIC demonstrator was undertaken with 49 1st Year Undergraduates in the Department of Information and Communications. The evaluation took place over three 1 hour seminars which were a follow-up of a lecture for the Organising Information unit on tagging and recommender services, which provided a context in which to use the MOSAIC demonstrator. Although the participants were provided with a background to tagging and recommender services, they were not necessarily proficient users of such features.

The evaluation was undertaken using an online survey, covering the following areas:

- User profile
- Using libraries
- Using resources
- Using recommending services
- Trying out MOSAIC (the MOSAIC web address was provided with instructions for undertaking 2 tasks)
- MOSAIC evaluation

2. Key findings

Using the library

- Students are using the university library in relation to their studies
- Students find the university library catalogue useful/very useful
- Students use their local library less often.
- Those that use their local library showed a mixed response to whether it was useful for their studies, or other activities.

Using resources

- Students use a variety of resources for study, and also for leisure and work
- They make use of facets on shopping sites
- Students are interested in resources other students are using on the same course, a similar course or lecturers are using on the course
- They use reading lists provided by lecturers
- They think reading lists are important/very important

Resources other people are using

Students are mainly interested in resources used by

- People taking the same course
- Taking a similar course at another university
- Lecturers teaching on the same course

Reasons for interest are:

- To provide a bigger picture of what is available
- To aid retrieval of relevant resources
- To help with coursework

Using MOSAIC

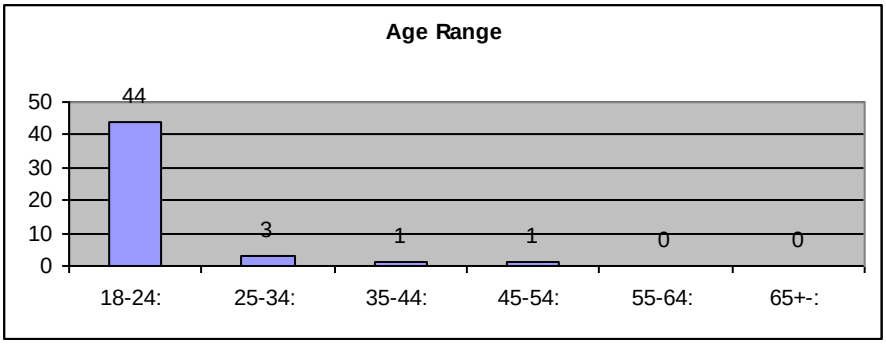
- Students mostly found both the iTunes and Classic interface user friendly/very user friendly
- In general, they showed no preference for the Classic or iTunes interface
- For studies, they showed a preference for Classic or liked Classic or iTunes equally
- Students mostly found the refining options useful/very useful
- They found the language/terms used understandable

3. Results

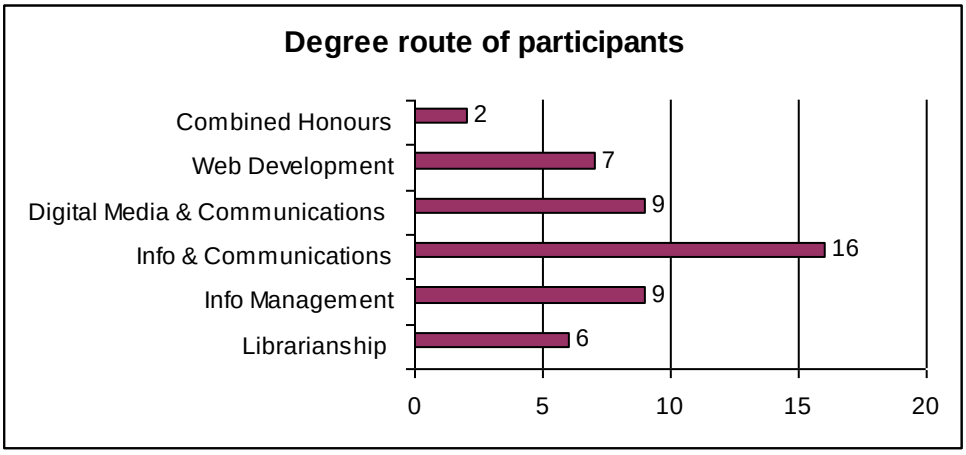
The results of the MOSAIC evaluation are as follows:

3.1 User Profile

Age of participants
Of the 49 participants, the age range shows that the majority were aged between 18 and 24 (44%).

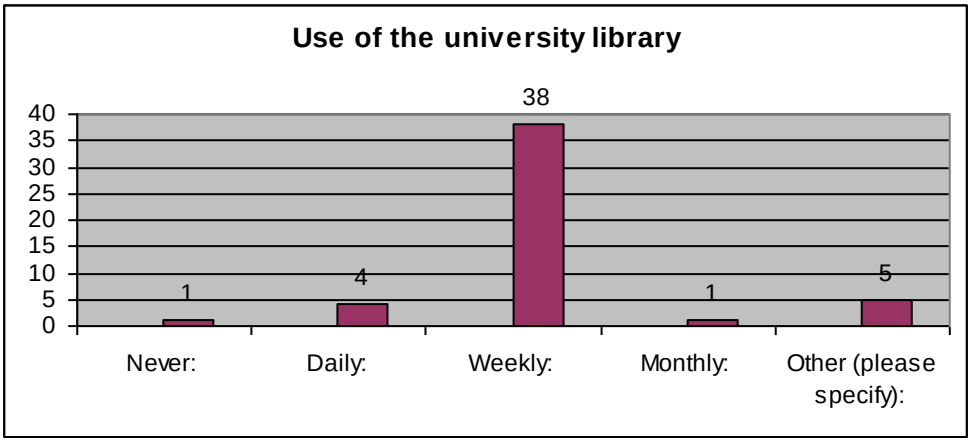


Degree route of participants
Participants were 1st year Undergraduates from the Department of Information and Communications. Degree routes taken include Information and Communications, Information Management, Digital Media, Web Development, Librarianship, and Combined Honours.

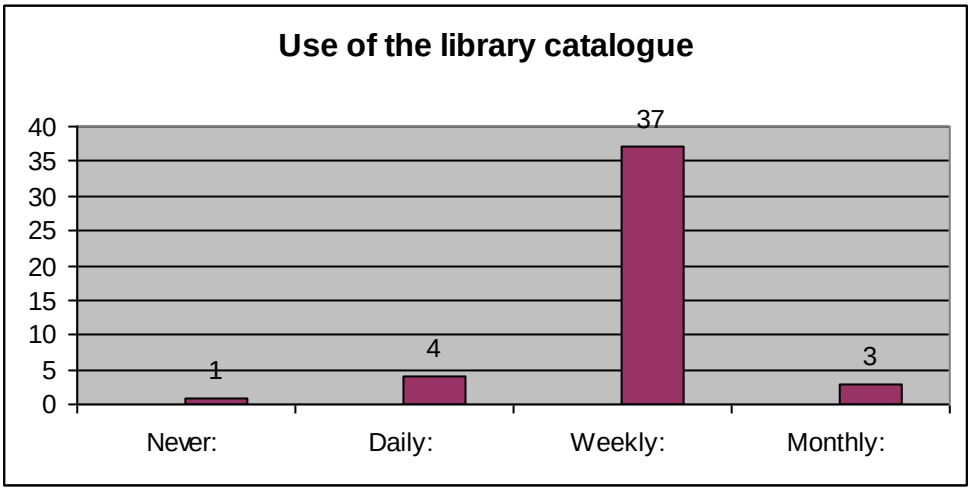


3.2 Using libraries

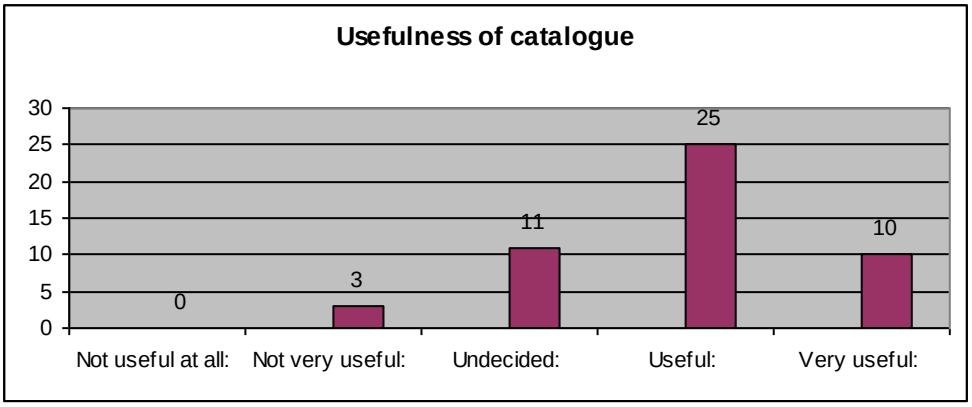
Use of the university library
The majority said they used the university library on a weekly basis. Those that selected 'Other' commented: '4-5 times weekly'; 'as often as necessary'; 'whenever needed'.



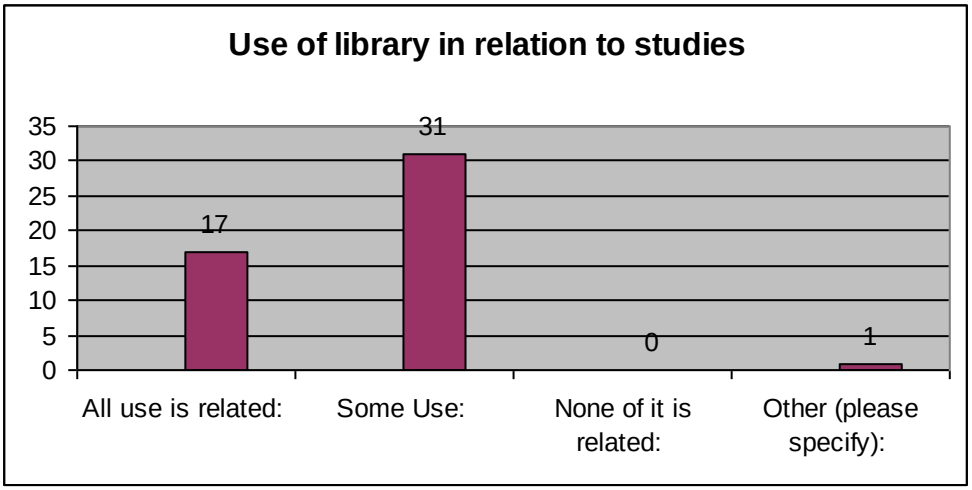
Use of the university library catalogue
The majority also said they used the library catalogue on a weekly basis. Those that replied 'Other' commented: '4-5 times weekly', 'a few times a week', 'Depends', and 'whenever needed'.



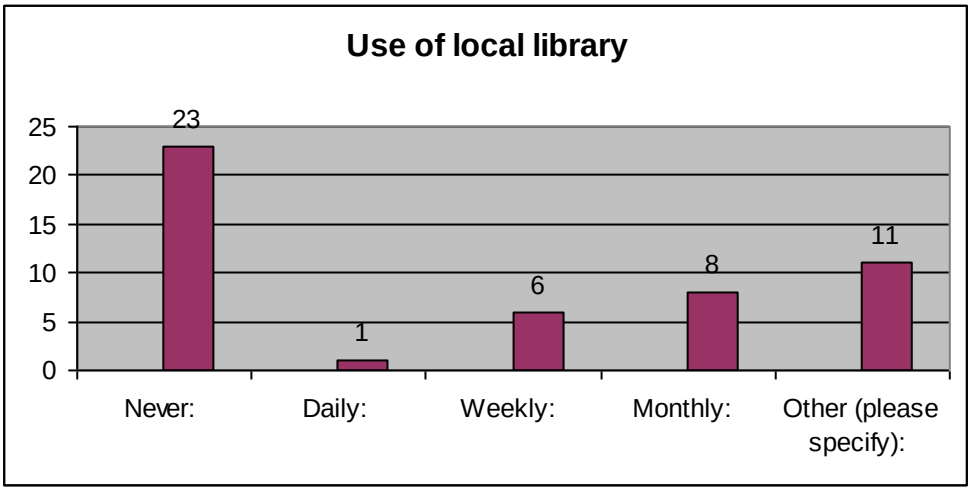
Usefulness of library catalogue
The majority of participants thought that the library catalogue was either useful or very useful, with 11 undecided and 3 who stated it was not very useful.



Use of library in relation to studies
All participants said they used the library in relation to some or all of their studies, the participant who selected 'Other' said 'most of it is related'.



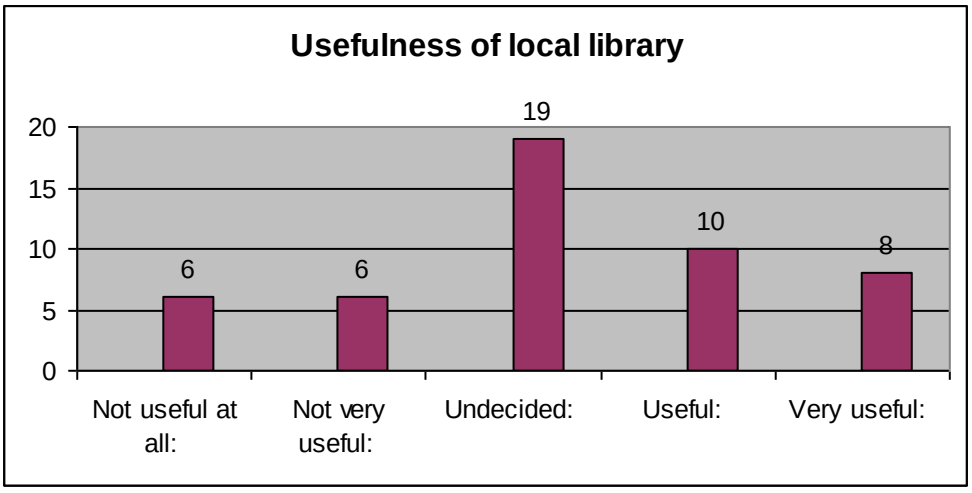
Use of local library
The majority did not use their local library (or the library nearest to where they currently live).



Comments from those who selected 'Other' included: 'Hardly ever', 'Rarely', 'Sometimes I go in there to use the photo copier or to print something', 'When I am home.'

Usefulness of the local library

A mixed response was given to the usefulness of the local library, with 18 stating that it was either useful or very useful, 19 undecided, and 12 stating that it was not very useful or not at all useful.

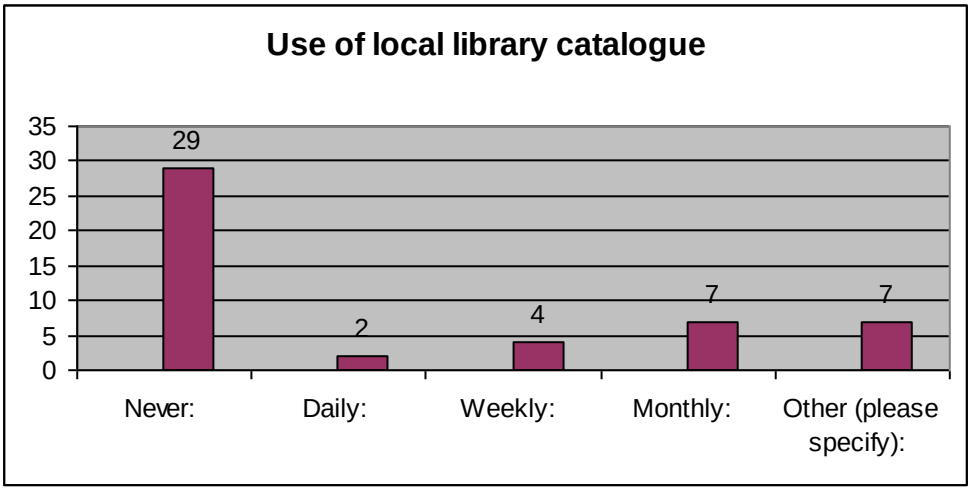


Comments included:

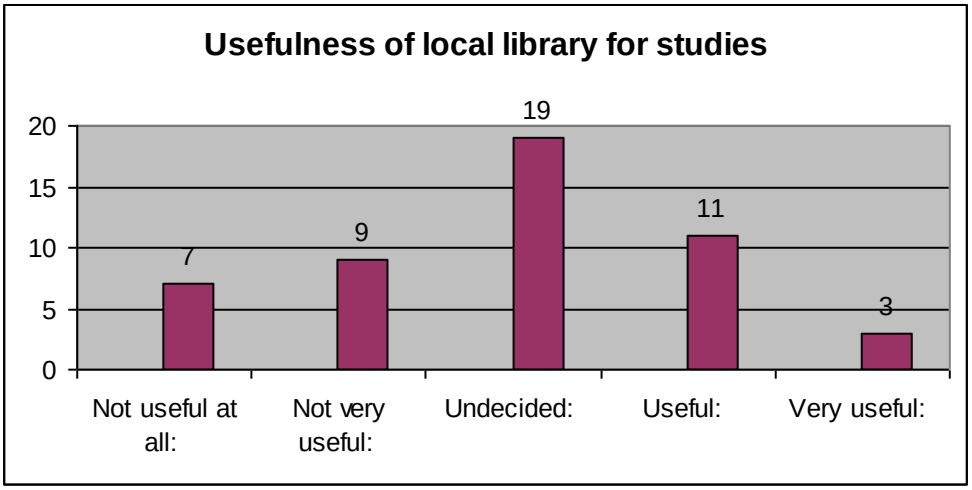
> *'Although the library in XXXX is not useful for university students, I found it is very helpful for local people, including children.'*
> *'I often find that being a local library and not a university or school library they don't have the field specific topics'*
> *'It is a small library with not many books related to my degree, therefore I prefer to use the University library more.'*
> *'It's mobile I cannot use it.'*
> *'The subject matter is very good fiction and non fiction as well as the online resources.*

Use of local library catalogue

The majority of participants (29) did not use the local library catalogue.



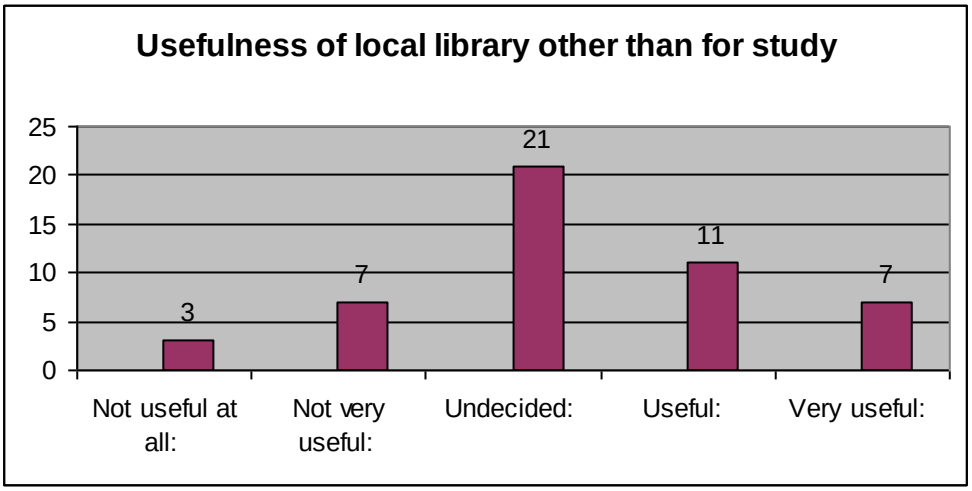
Usefulness of the local library for studies
A similar response to usefulness of the local library was given to the usefulness of the local library for studies.



Comments included:

'Don't have the books needed'
'Having not tried to look into the community library for my studies I cannot decide if it was useful for my studies.'
'If a book is not in stock and under a certain price the library will buy the book in for a small fee'
'They don't have many books on Librarianship, ironically.'

Usefulness of local library for activities other than study, such as leisure or hobbies.
A mixed response to whether the local library is useful for activities other than study, with 18 stating it was either useful or very useful, 21 undecided and 10 stating that it was either not very useful or not useful at all.



Comments revealed a number of activities either known or used:

'Different events, clubs etc. happen now and then, some on a regular basis.'
'My local library has a gym and swimming pool.'

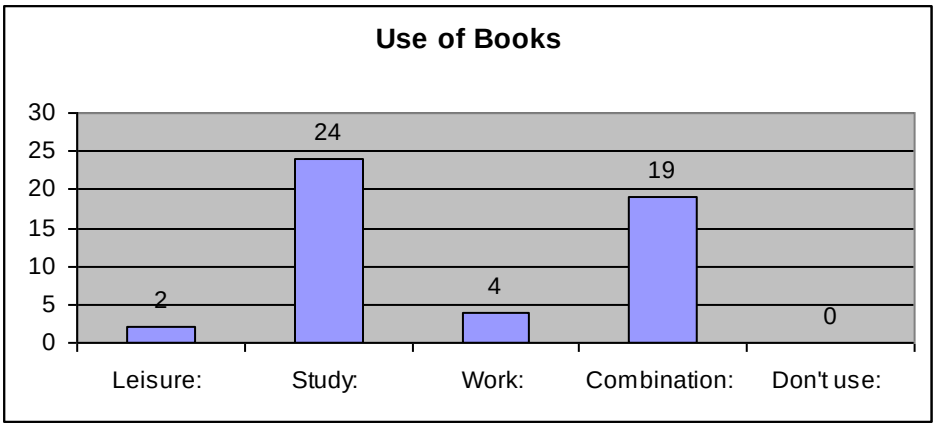
'There are events at the library as well as education and leisure courses available'
'They run clubs which I am not very interested in.'

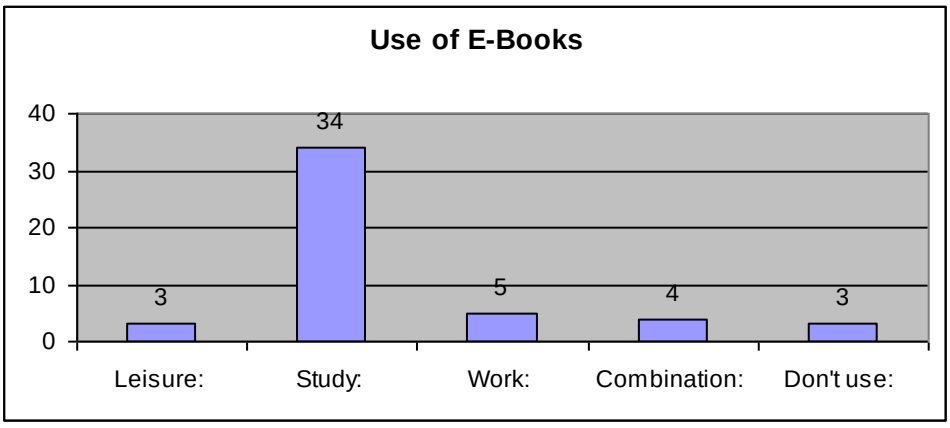
Other comments included:

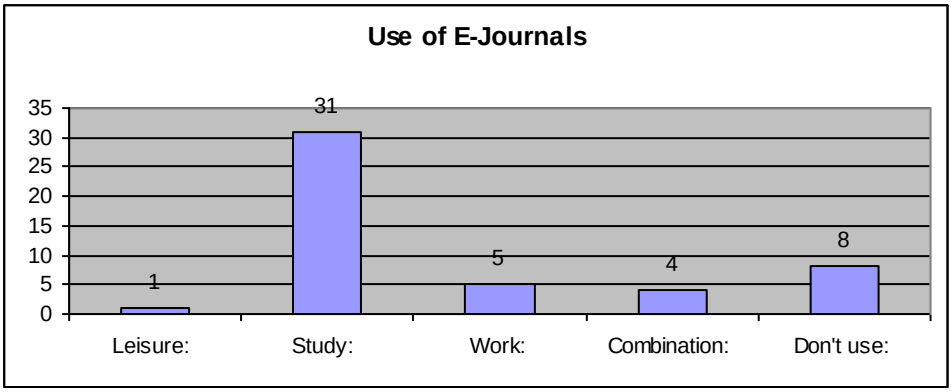
'I recently got internet access in my flat, so now I don't have to sit next to drunken people while I check my e-mails.'
'When I visit local libraries, I can learn how they are organising their libraries'.

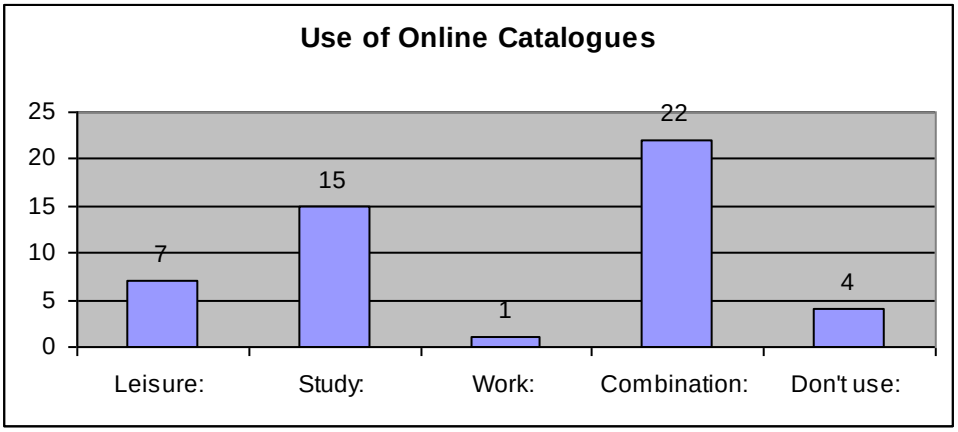
3.3 Using resources

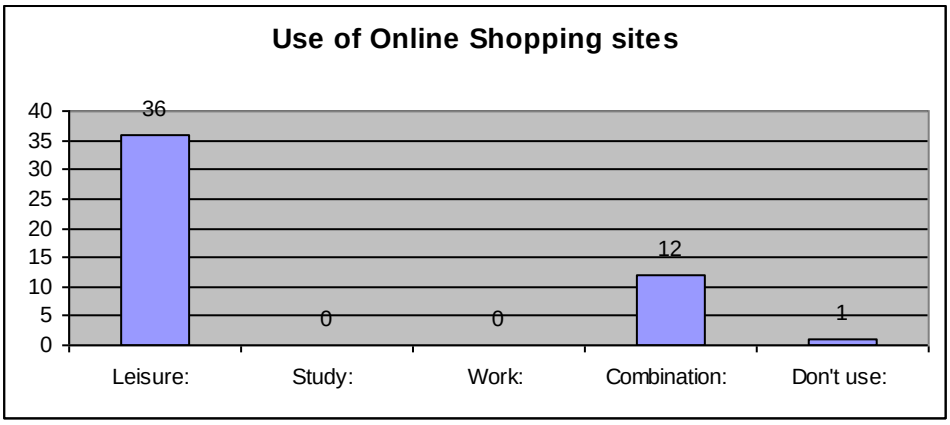
Apart from shopping, participants mostly use different resources for study or a combination of leisure, study and work. Some stated that they do not use specific resources (3 do not use e-books, 8 do not use e-journals, 4 do not use online catalogues and 1 does not use online shopping).





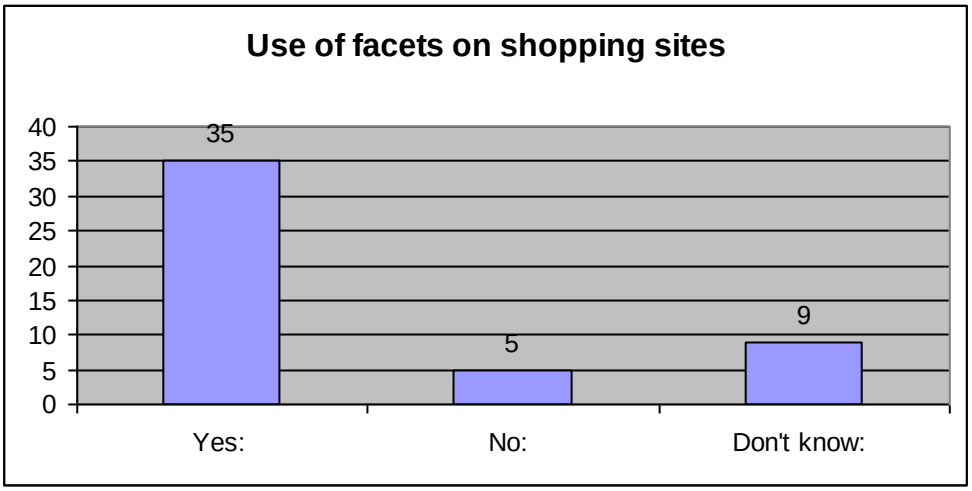

Use of E-Journals
35
30
25
20
15
10
5
0
1
31
5
4
8
Leisure:
Study:
Work:
Combination:
Don't use:


Use of Online Catalogues
25
20
15
10
5
0
7
15
1
22
4
Leisure:
Study:
Work:
Combination:
Don't use:


Use of Online Shopping sites
40
35
30
25
20
15
10
5
0
36
0
0
12
1
Leisure:
Study:
Work:
Combination:
Don't use:

3.4 Using recommender services

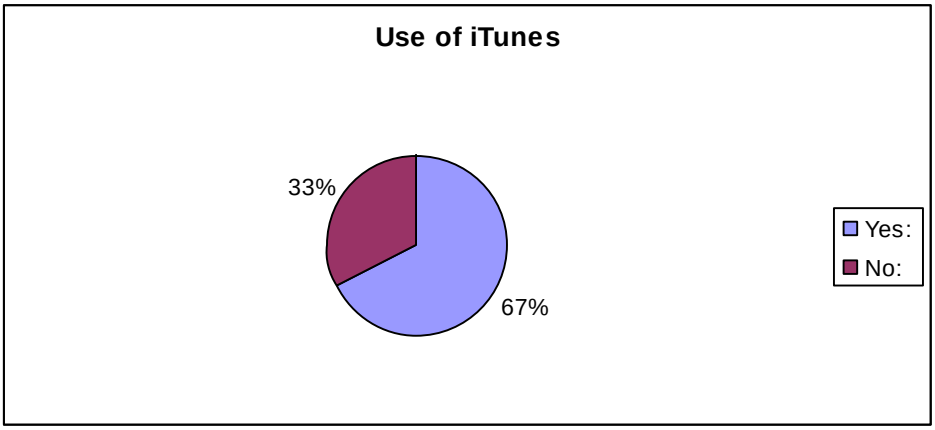
Use of facets on shopping sites
Most of the participants said they use facets (such as brand, size or price) when using online shopping sites.



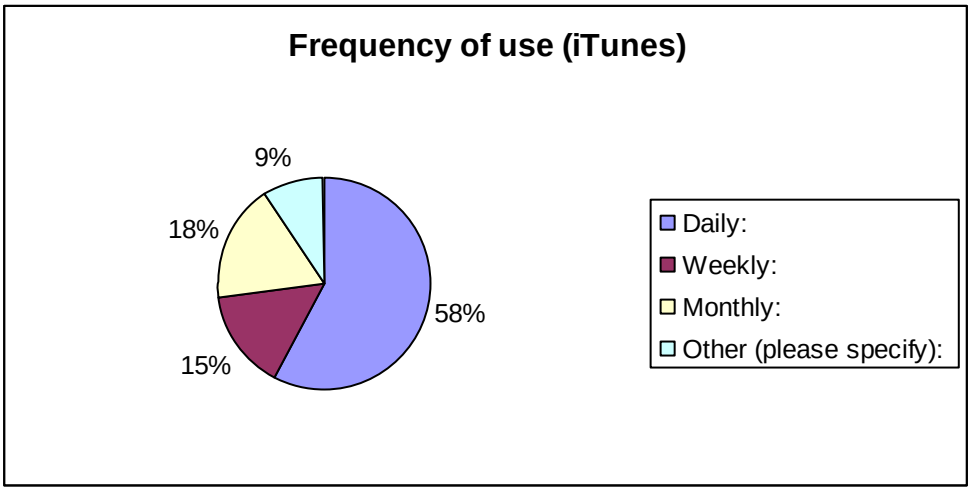
Some commented on the frequency of using facets; many respondents said they use facets 'all the time' or 'most of the time', with reasons such as 'it helps to narrow down the search', 'Every so often when shopping online, if the website uses that method of filtering', and ' I use it to see how many items are available, so I use it a lot'.

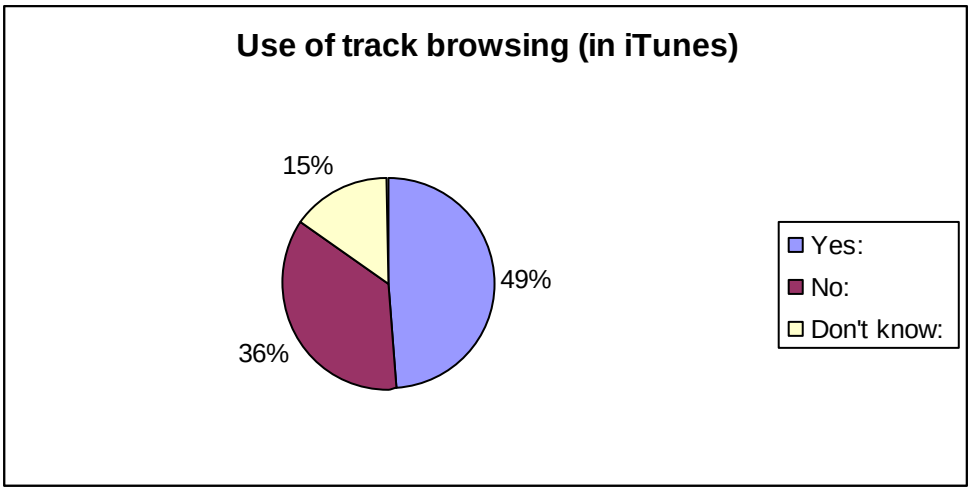
Other respondents said they use facets 'weekly' or 'monthly', with reasons such as 'Mostly when shopping for clothes', 'mainly during access to eBay and Amazon', 'Whenever I want to refine a search - in Amazon or Ebay for example' or 'whenever shopping online and I am looking for a certain brand or price range'. Only one respondent said they use facets 'infrequently'.

Use of iTunes
Over half the participants (67%) said they use iTunes.



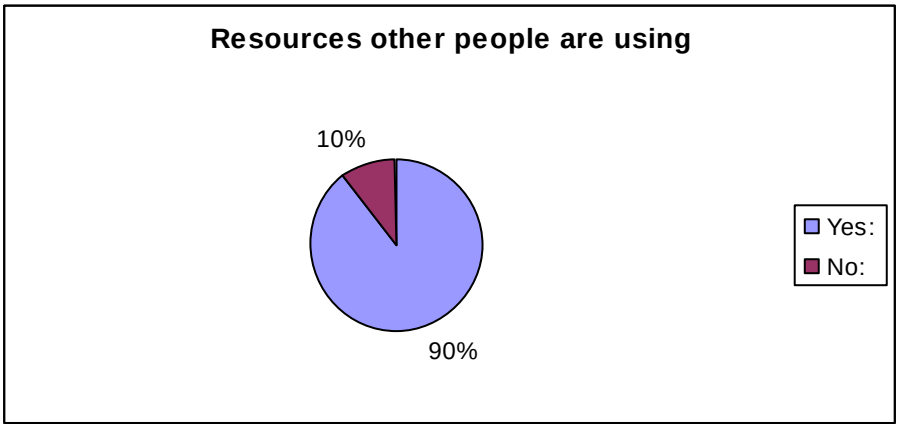
Of the 67%, over half (58%) use iTunes on a daily basis, and 49% use the track browsing facility.





3.5 Resources other people are using

The majority of participants (90%) said they would like to be able to find out what resources other students, lecturers or researchers are using.



Reasons given can be grouped into 3 main themes:

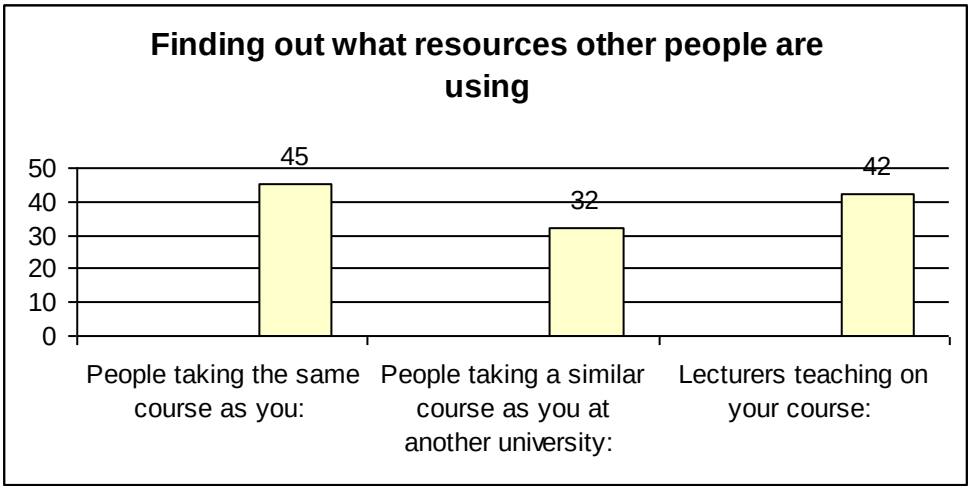
1. To provide a bigger picture of what is available
   'it would give me a universal picture of what I should be looking at, reading etc.'
   'I guess it would be a useful thing to see where other people are looking to give a guide on where to go.'
   'if there are any other resources that are useful; I would also like to use them'
   'To allow access to alternative resources to those I already use'
   'I would like to know new recourses and be kept up to date'
   'To widen my knowledge about where else … I could get resources from.'

2. To aid retrieval of relevant resources
   'I guess it would be a useful thing to see where other people are looking to give a guide on where to go.'
   'If it helps me find information.'
   'I think it would be helpful, so as to refine my own research process.'
   'It'd be handy to see what sources are most relevant.'
   'It would help filter out what information is useful to me or not depending on how many people have used the resource'

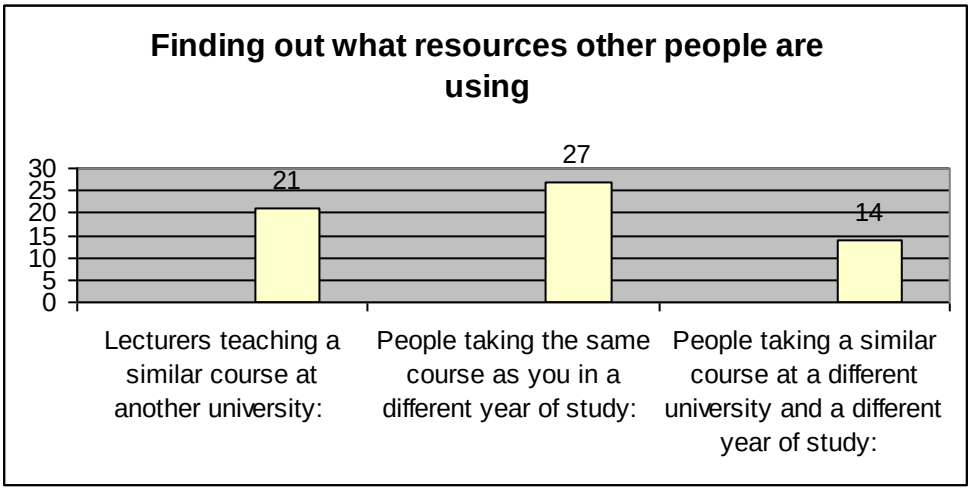
3. To help with coursework
   'It may help me to find the resources that they would use to get a better mark'
   'It will be helpful for my study and research.'
   'it would help me [with] my essay
   'this may help me with my studies'
   'It is always helpful to look at what other students are doing to further their studies in order to improve your own.'
   'would help with coursework'

Some gave reasons not related to reading lists, 'I prefer to chose my own music, not be subjected to look at what others like', 'it would increase the range of products and prices which I am looking for'.

Finding out what resources other people are using
Participants mostly want to find out about the resources used on their course or a similar course at another university.

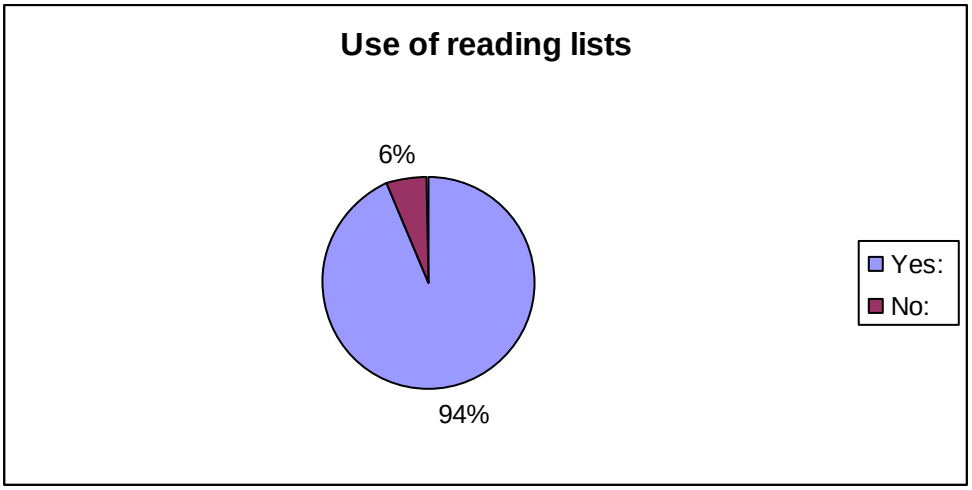


Participants are also interested in resources used by people taking the same course, but in a different year. Some are interested in finding out what lecturers teaching similar courses at other universities are using. Only 14 of the 49 participants thought they would like to see what people are using taking a similar course, but in a different year and at another university.

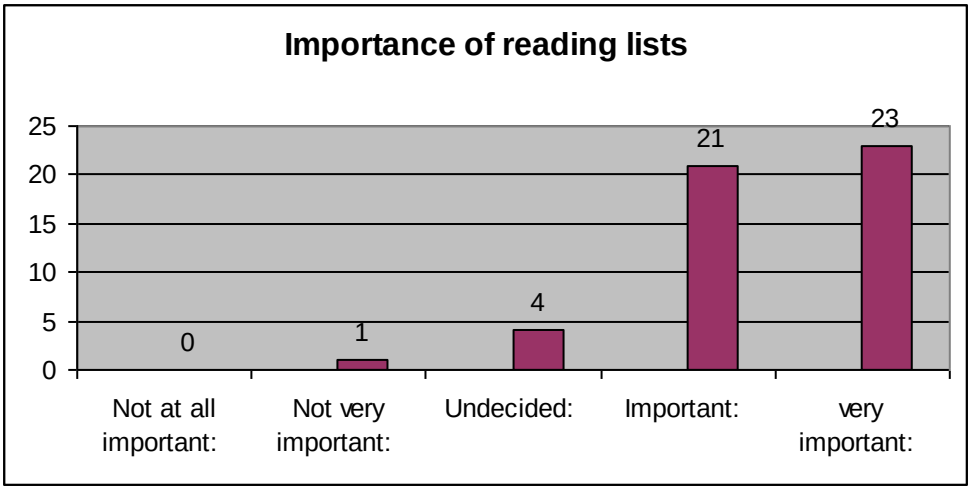


3.6 Use of reading lists

The majority (94%) of participants said they use reading lists given to them by lecturers.



Importance of reading lists
The majority thought that reading lists are either important or very important.



Comments can be grouped into 3 themes:

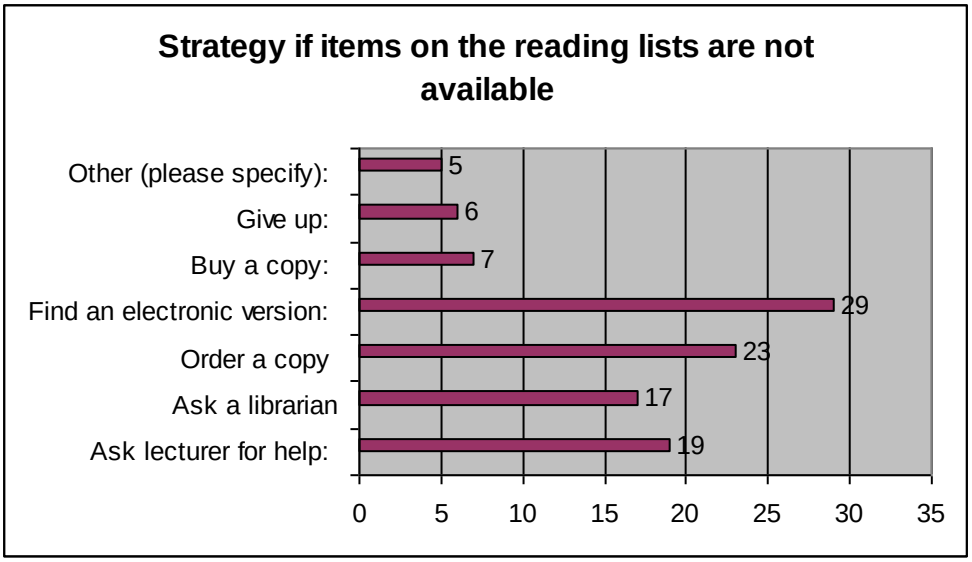
1. To provide a bigger picture of what is available
   'I find that background reading provides an important foundation for future learning in a particular subject.'
   'This helps us to broaden our knowledge and understand each aspect from our lectures point of view as he books from the reading lists are often the books that they have read themselves and now recommend us to read'.

2. To aid retrieval of relevant resources for studies
   'Gives explicit guidelines as a starting point'.
   'I find it useful as I know what books I need to read'
   'Essential for the information that you need'
   'If tutors are recommending these books or/and are basing their lectures on them, likely…. that they are amongst the most trusted and important sources to reference from'
   'I may find useful materials apart from particular key words.'
   'It shortens the amount of information that we need.'
   'It's not rocket science that the lecturer wants me to pass as much as I do, so they'd whatever they can to help me.'
   'It sets you off and gives you direction for your studies.'

3. To help with coursework
   'they give you really useful books that you can use for coursework and for looking things up related to the unit'
   'Important as they go towards to academic study and will help you throughout the years'.
   'points you in the right direction for coursework'
   'it dictates what is going to be used during the year and is useful if a student is struggling to understand'

Other comments include, 'Alternative books should also be listed.'

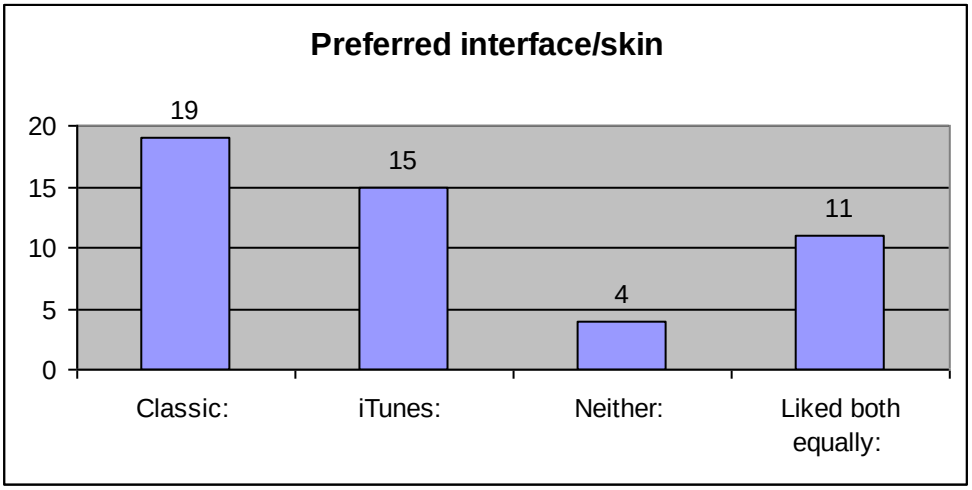
Strategies if items on the reading lists are not available
The most popular strategy (29 participants) is to try and find an electronic copy. This is followed by ordering a copy from another library (23), asking the lecturer for help (19), and asking the librarian (17 participants).



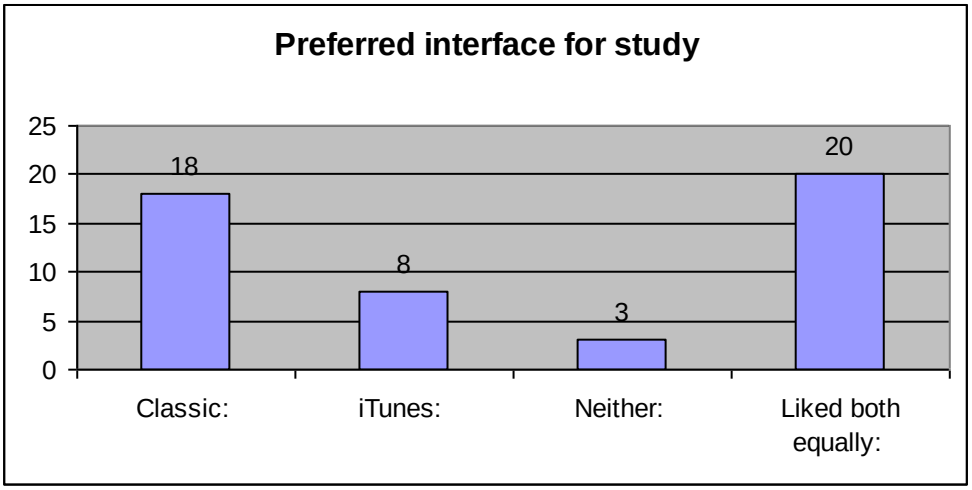
Other comments include ‘Find an alternative’ or ‘search for a related book’, ‘request it at the local library’ or ‘Search for recommended books'.

3.7 MOSAIC evaluation

Preferred interface/skin
Participants show no strong preference between the Classic and the iTunes interface, with 19 preferring the Classic interface, 15 preferring the iTunes interface, and 11 liking both equally.

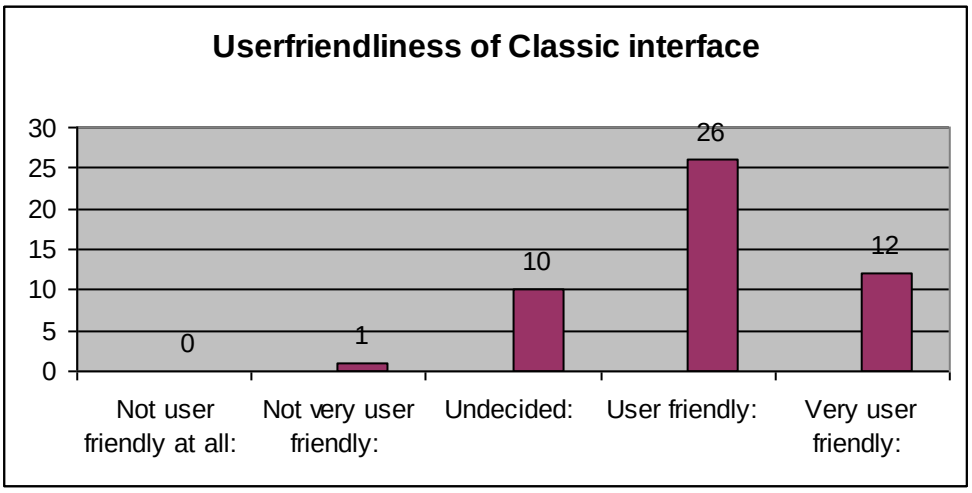


When asked about interface preferences for study purposes, 20 said they like both equally and 18 prefer the Classic interface, with only 8 stating a preference for the iTunes interface.



,

User friendliness of the Classic interface
The majority of participants said they thought the Classic interface was either user friendly, or very user friendly, with 10 undecided and just one finding it not very user friendly.



Comments can be grouped into 3 main themes:

1. Options required are all available on this interface
   'All the options are available for you so less thinking is required'
   'A good interface'

2. Like the options being displayed on the left hand side
   'Having the options on the left side is easier to navigate than having it on top.'
   'Similar to refining searches in Amazon and Ebay so I was familiar with it already.'
   'The advance search is on the left hand side makes it easier to search.'

3. Quick and easy to find things.
   'It is a very comprehensive way of narrowing down results.'
   'It is speedy.'
   'it was clear and understanding'
   'Simple and easy to navigate around.'

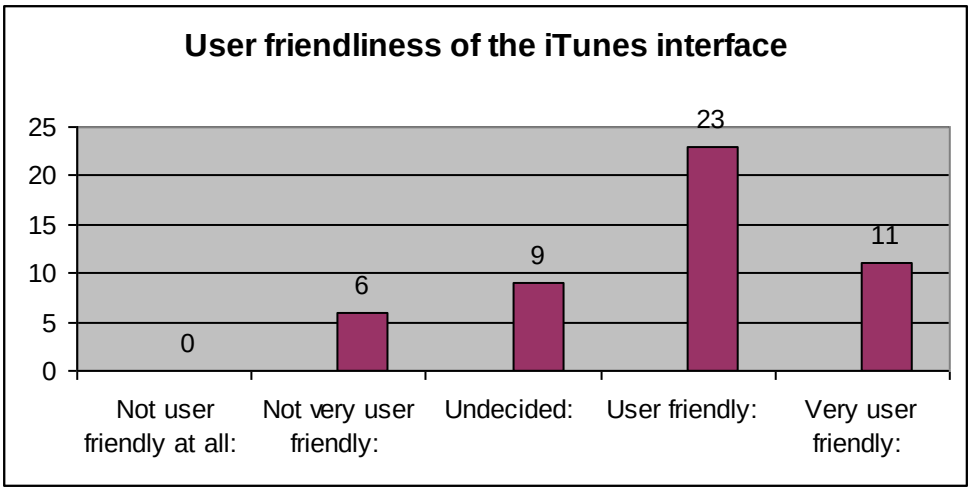
'Standardised processing order and retrieval.'

One participant had some criticisms:

'The aesthetics are overly bland, the results colour scheme is not in sync with the rest of the search page, the overall design is clean but not NICE. Look at this, http://www.autotrader.co.uk/search/form get some AJAX in your search engine'

User friendliness of the iTunes interface

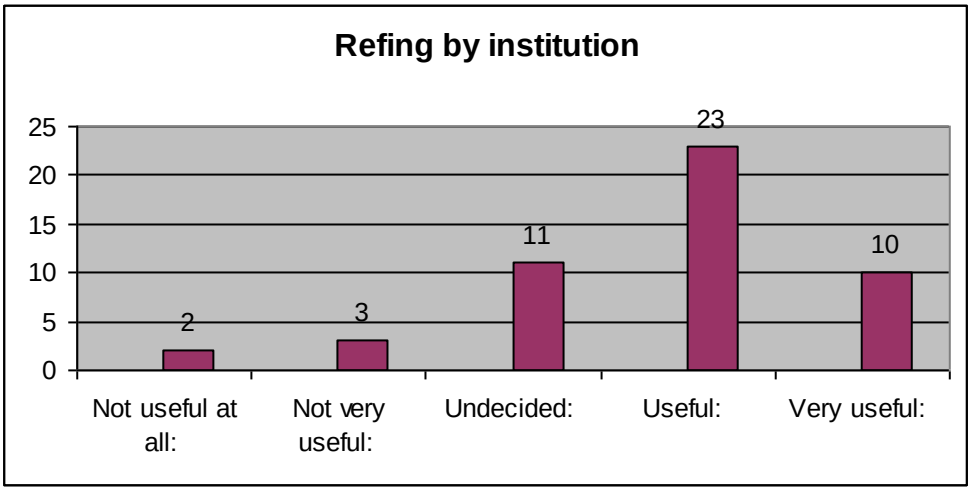
The majority of participants said they thought the Classic interface was either user friendly, or very user friendly, with 9 undecided and 6 finding it not very user friendly.

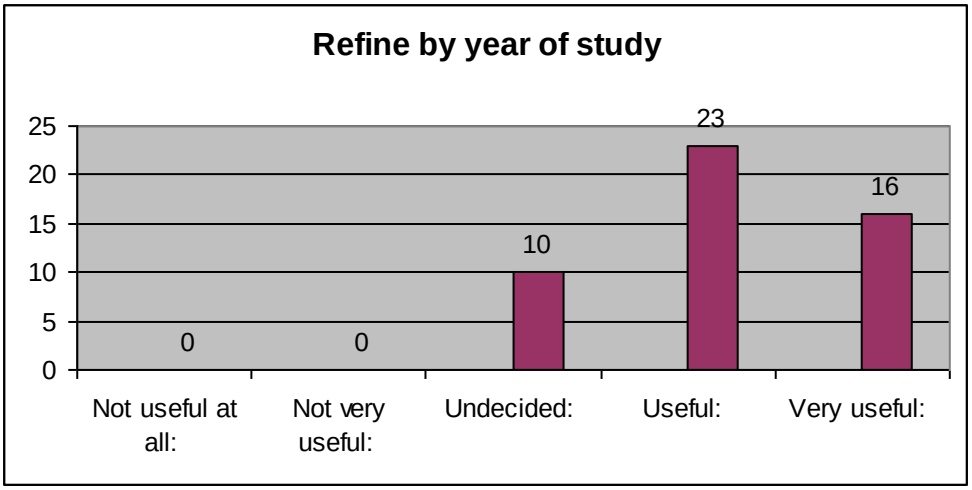


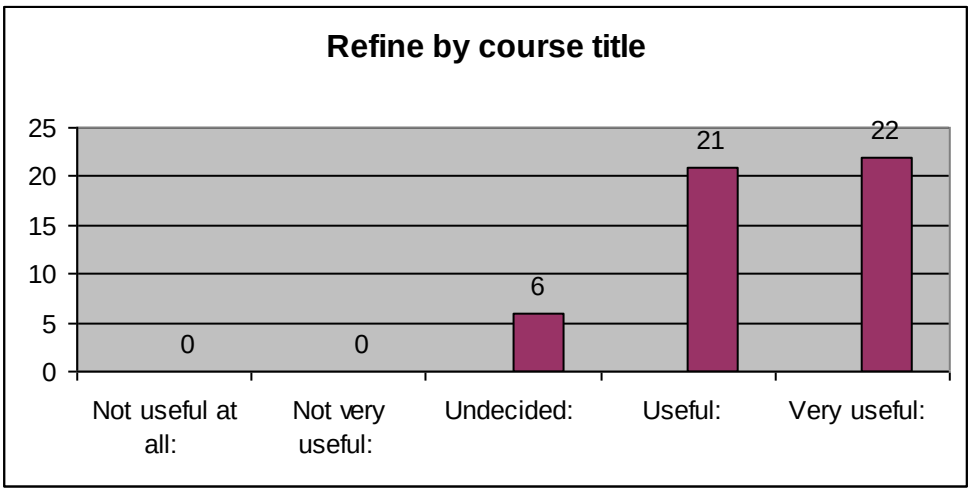
No further comments were provided.

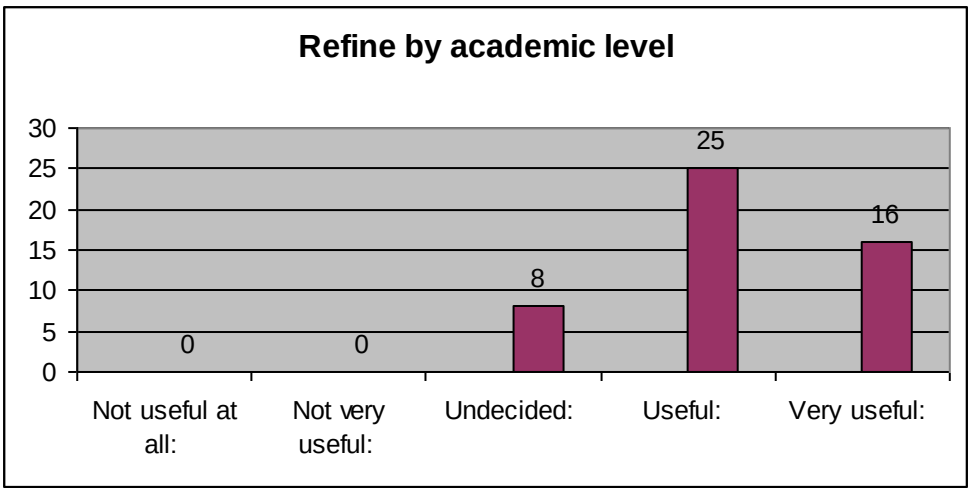
Usefulness of facets

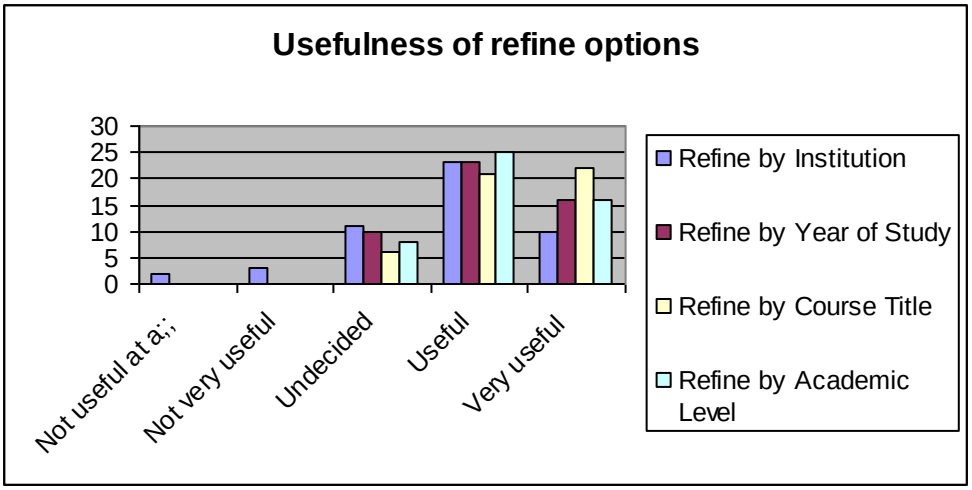
Participants were asked how useful it was to refine the resources, by institution, year of study, course title and academic level. Participants mostly thought the refine options to be useful or very useful (33). Some were undecided (11), and 5 found the refine by institution option not very useful or not useful at all.




Refine by year of study
25
20
15
10
5
0
0
0
10
23
16
Not useful at all:
Not very useful:
Undecided:
Useful:
Very useful:

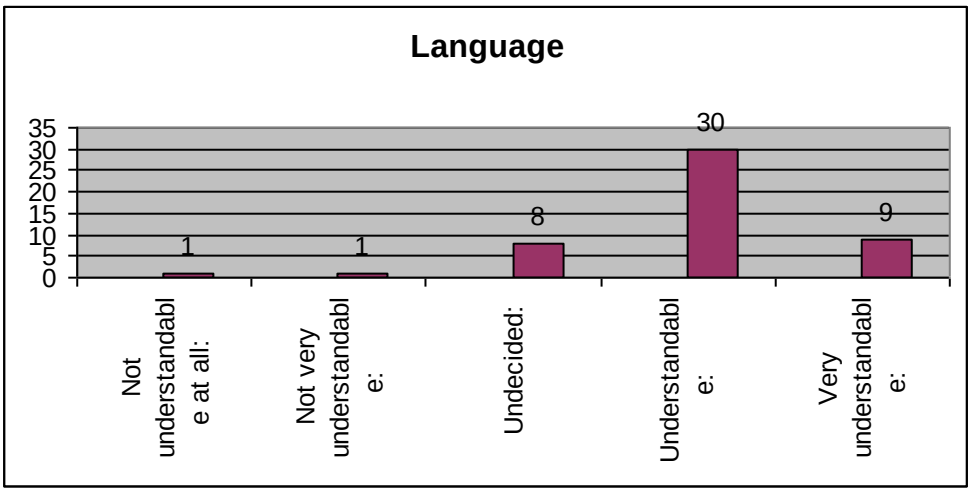

Refine by course title
25
20
15
10
5
0
0
0
6
21
22
Not useful at all:
Not very useful:
Undecided:
Useful:
Very useful:


Refine by academic level
30
25
20
15
10
5
0
0
0
8
25
16
Not useful at all:
Not very useful:
Undecided:
Useful:
Very useful:



Language used in MOSAIC

Participants were asked if any of the MOSAIC terminology was difficult to understand. Most thought that the language used is understandable (30) or very understandable (9), with 8 undecided and 2 who said they found the language difficult to understand.



Comments include: ‘it was all quite comprehensive’, ‘the abbreviations used for the academic level, need to be made with a key on the side or actually typed out in full’, ‘the PhD section should be joined by PhD1, PhD2 and PhD3’.

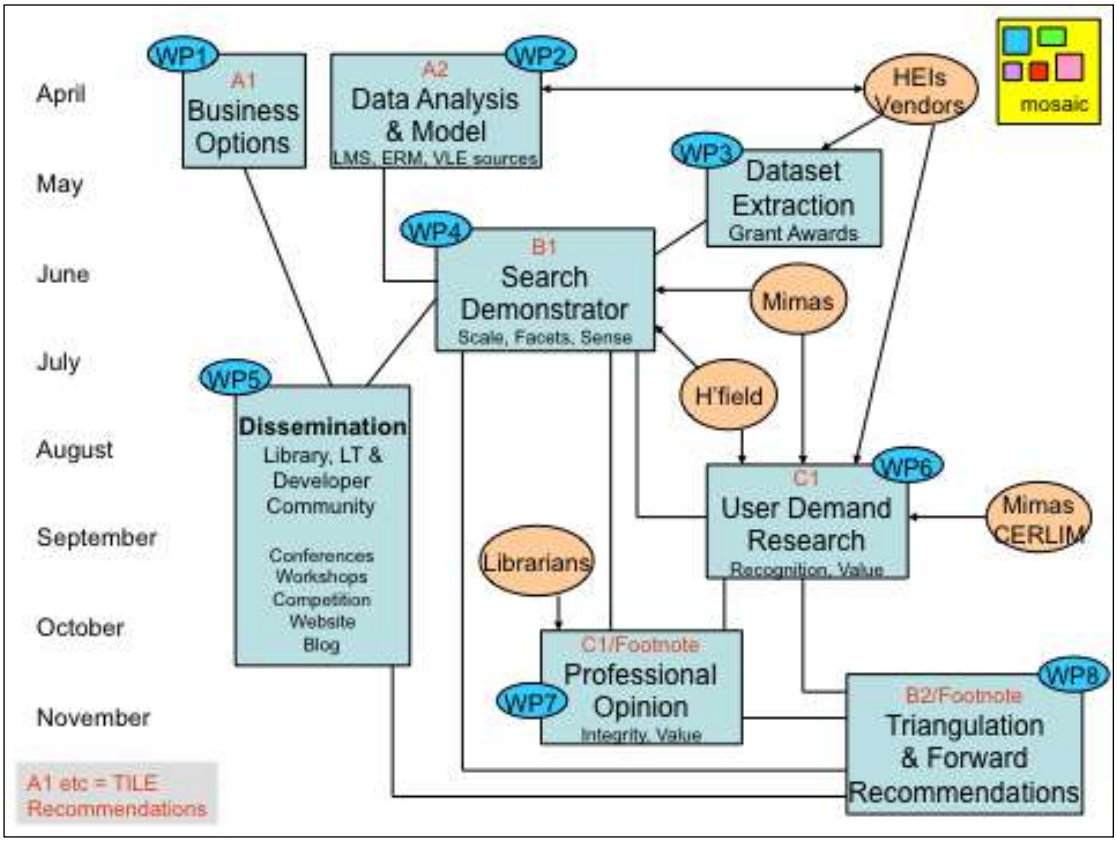
Appendix 5 – MOSAIC Project Work Plan

Work Plan

The MOSAIC project work plan was based on 8 work packages to be executed over a period of 8 months (April to November 2009) in order investigate the business case for use data services based on

- Appetite for services
- Access to data
- Technical feasibility

The work streams were designed to enable the team to triangulate evidence from those themes over the life of the project, as illustrated here:



Activities

The headline activities proposed for the individual work packages are as follows:

Work Package	Activity & Outputs
WP1 – Business Options	• Identify business options for development, implementation and operation of aggregated context / activity based services as specified by TILE • Review the Open Data approach to expose context data • Establish areas of cost, pain and risk for institutions • Consult with Mimas, Edina & Resource Discovery Task Force on the options
WP2 – Data Analysis	• Define 'first pass' data model based on the TILE & Huddersfield work • Specify immediately workable schemas for use in this project • Engage wider community, including LMS, ERM & VLE vendors • Possibly design/code a data collection add-in for one LMS & VLE • Undertake analytic 'desk' work to see if merge and search works in practice
WP3 – Data Extraction	• Engage institutions as contributors • Award Data Extraction Grants under Open Data Commons licence conditions • Provide developer / extraction advice & data QA • Possibly undertake RDF Conversion • Make raw data available on a platform hosted within '.ac.uk '
WP4 – Search Demonstrator	• Define SOLR data model • Bulk up data sets to test scale and import into SOLR • Tune model and undertake performance tests for import, indexing, search • Build rudimentary web forms to display results • Establish whether SOLR will support the aggregated approach • Identify links with EIE project
WP5 – Dissemination	• Focus: subject librarians, learning technologists, developers, system managers • Set up website presence that will facilitate dissemination and engagement • Ensure key LMS and service vendors are briefed and encouraged to engage • Link with key conferences and events • Organise up to 3 specialist workshops for selected communities • Run developer competition
WP6 – User Demand Research	• Determine target user groups • Define a range of Use Cases supported by 'screen shots' • Ascertain through focus group / survey whether users would be inclined to use such facilities • Identify scenarios that will motivate contribution of User Generated Content
WP7 – Professional Opinion	• Engage UK HE professional community • Make demonstrators available • Run workshop(s) focusing on integrity and value of search results and on service considerations
WP8 – Report	• Weigh up the possibilities for investment and interest of strategic partners in 'context and contribution services' from local up to national levels • Review Business Models • Develop a high level rollout plan for recommended scenario(s)

Evaluation against plan

Whilst the key objectives were achieved, as indicated in the final report, the self-evaluation conducted by the MOSAIC team indicates that the following activities did not work out as intended:

WP	Activity	Shortfall / Issues
WP2	Engage wider community, including LMS, ERM & VLE vendors	A significant number of universities engaged directly with the project; however, feasible activity on the ground was restricted to LMS data; in addition other JISC projects and also Ex Libris provided inputs regarding journal use data; engagement with the VLE community was not realistic in the timeframe.
WP3	Make raw data available on a platform hosted within '.ac.uk '	On account of the challenges faced by contributing libraries, it took right to the end of the project to gather the data together
WP5	Set up website presence that will facilitate dissemination and engagement	The bulk of dissemination was directly through project activities, notably the developer competition, the practitioner workshops, the student evaluation and the final workshop. In addition, the project presented at a number of events and conferences.
WP5	Ensure key LMS and service vendors are briefed and encouraged to engage	Whilst contact was established and maintained with vendors, the uncertainties faced by participating universities made it difficult to press forward directly with vendors. The project recommendations therefore present vendor cookbooks as a next step.