Exploring online research methods/TRI-ORM Website usage report May 2009

This report aims to provide an overview of usage of the Exploring online research methods / TRI-ORM website between 1st November 2006 and 30th April 2009. Accurate figures of website usage are notoriously difficult to gather and the report will begin by outlining the mechanisms used to gather and analyse the website usage information (section 1). It will then attempt to analyse overall usage in terms of the 'quality' of usage as well as the raw quantitative information (section 2). This will be followed by an attempt to track changes in patterns of usage over the course of the period (section 3) as well as providing information on origin of traffic to the site in an attempt to establish information on the site's geographic reach (section 4) and patterns of use (section 5). Finally, it will provide an overview of the relative popularity of different areas of the site (section 6) before attempting to draw conclusions on overall usage patterns over the period (section 7).

1. Analysis

The mechanism for obtaining this data is the Google Analytics service (http://www.google.com/analytics/). This is a cookie-based tracking service which offers a relatively accurate and extensive range of data. However, it should be noted that the figures used can only be considered accurate as 'minimum figures'. Most browsers provide users with options to disable cookies, JavaScript and images (including in some cases the option to disable images that are requested from domains other than the current page being viewed). Any of these actions would prevent a visit from the site from being recorded. There is thus a possibility that the figures will be skewed against more technically 'savvy' users with an interest in controlling their browser and security settings and the knowledge to do so. Despite this, the service was chosen as it offers a range of useful measures which can provide an indication of 'quality of visits' as well as number. These include information on 'returning users', length of time spent on the site before navigating away, and number of pages viewed. Accurate information on geographic location of site visitors and referring page information indicating the source of traffic to the site also provide useful information on site reach and utility. With the exception of uses who have disabled the options above in their browsers, the cookie-based system is highly accurate. If a user's return visit to the site calls upon the 'cache' rather than the server, it will still be recorded by the service, and automated hits from 'spiders' or 'bots', such as those used by search engines to index web pages will not be included.

Alongside the information provided from Google Analytics, reference will also be made to the Analog log file software which has also been used to track usage of the site throughout the period. The service is not influenced by the user's browser settings, but is based on IP + User agent methods of collecting data, which use log file analysis to record calls for pages on the website. This service provides less data on 'quality' of visit and is likely to be less accurate in terms of overall figures as return visits are not be recorded if they draw upon the users 'cache' rather than the server, and traffic originating from automated 'spiders' or 'bots' tends to be included. However, as there is consistency within these inaccuracies, it does offer useful trend data on overall numbers and source of visits to supplement that provided by Google Anlaytics. This is especially useful in that the Analog service was established before the Google Analytics one and offers a longer period for analysis of trends in usage change over time. The timescales for usage data gathering on the site are shown below:

April 2006 – Analog log file analysis service begins collecting usage data.
May 2006 – Exploring online research methods website launched.

November 2006 – Google analytics service established.
May 2007 – TRI-ORM materials first added to the Exploring online research methods website.
May 2009 – TRI-ORM project drawing to a close.

2. Overall usage

The Google Analytics figures show that over the period 1st November 2006 to 30th April 2009, the website was accessed 62,656 times (an average of 69 visits per day), from a total of 51,321 different computers. In analysing these figures, however, it is vitally important to consider 'quality' of visits as well as quantity. The service offers a number of useful indicators of this quality, including the number of return visitors, the number of pages viewed, and the length of time spent on the site. Obviously, the higher these figures, the more likely that the site is being used actively and that its contents and facilities are proving to be useful to the research community on an on-going basis.

An indication of the number of return visitors is provided by the service and this shows that overall 18% of users returned to the site during the period. This could be seen as an underestimate as anyone returning to the site via a different computer will be recorded as a new visitor. However, within these figures it is possible to obtain a useful overview of the number of return visits made, as follows:

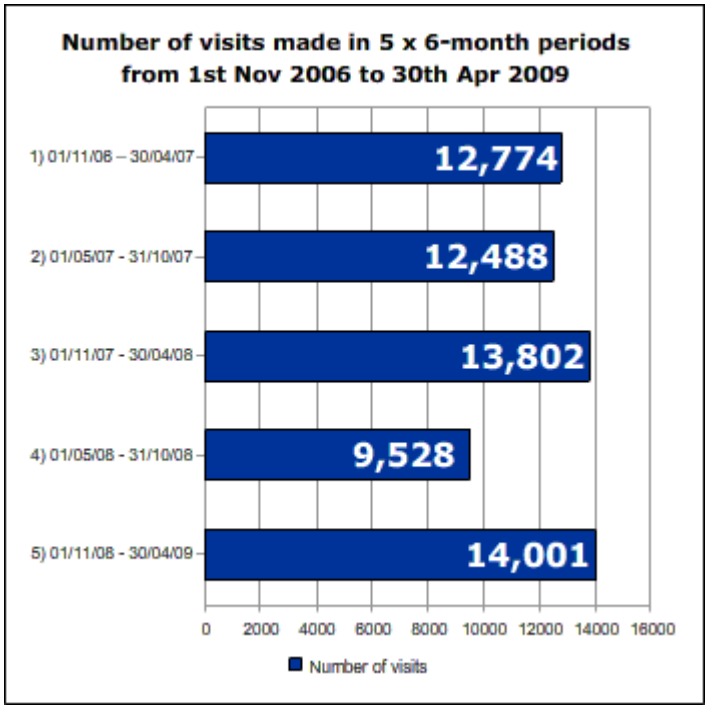
Number of return visits made	Number of returners	% of returners
2-4	7838	69.17
5-8	1422	12.55
9-14	639	5.64
15-25	502	4.43
26-50	459	4.05
51-100	310	2.74
101-200	152	1.34
200+	9	0.08

This shows that well over 10,000 users have visited the site repeatedly, and around 1000 users have clearly made it part of their researcher toolkit, returning to the resource on an average of a monthly basis or more.

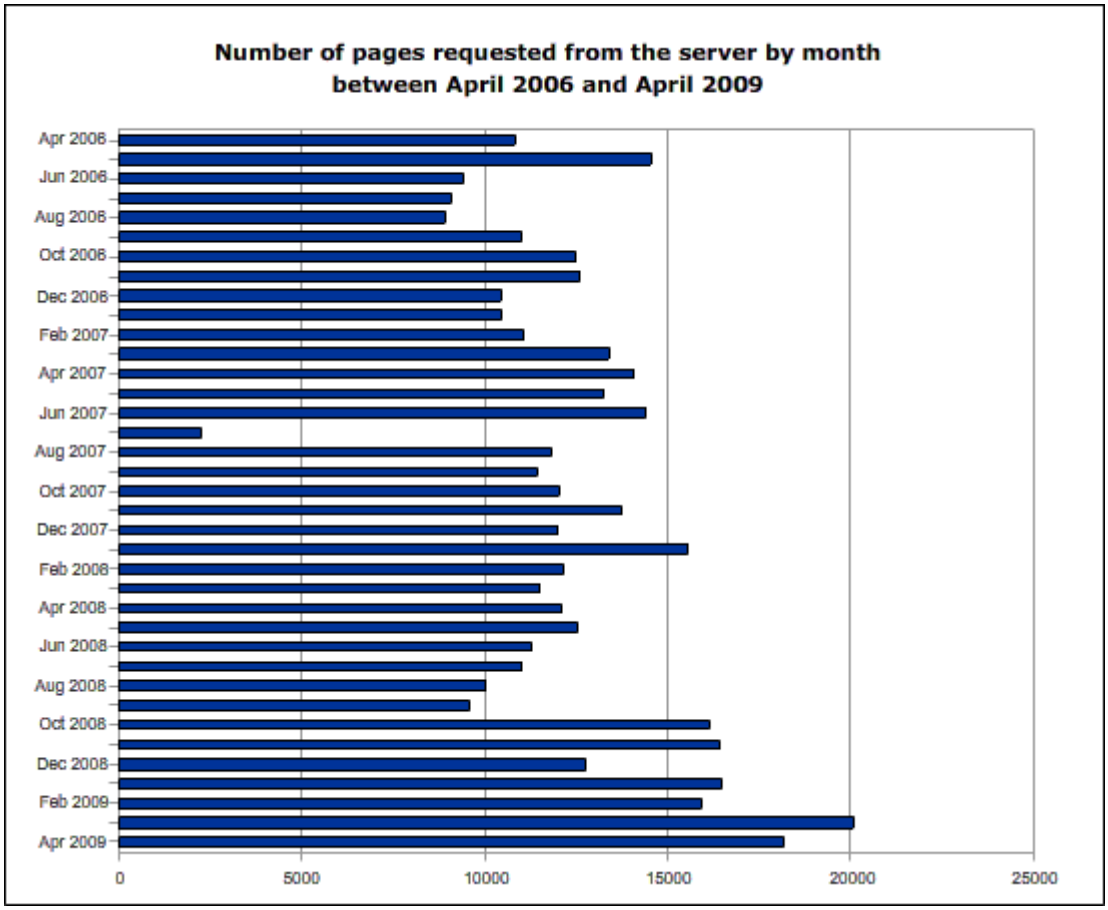
There is also evidence for a sustained and focused usage pattern in the statistics around length of visit and number of pages viewed. Almost 5,000 visitors viewed the site for 10 minutes or more during the period, while 5,500 viewed more than 10 pages in a single visit. These figures suggest that he site has become a valuable resource and reference point for a large number of users. We would recommend the establishment of a common site analysis procedure to provide benchmark figures for comparison and would welcome the use of these figures in this way.

3. Changes in usage over time

A comparison of five equal 6-month periods for overall site usage reveals high levels of usage throughout the period of the TRI-ORM project, with the average number of visits daily ranging from 52 to 77. The general pattern is one of increased usage over the period, with a peak of over 14,000 visits in the period 01/11/08-30/04/09. The total number of visits in each period is shown in the chart below.



The overall indication of growth in site usage during the period is corroborated by the Analog figures which provide an accurate overview of usage trends and changes over time. As shown in the chart below, the pattern of daily requests for pages from the server is general one of consistent increase, with the six busiest months falling in the period October 2008 to May 2009.

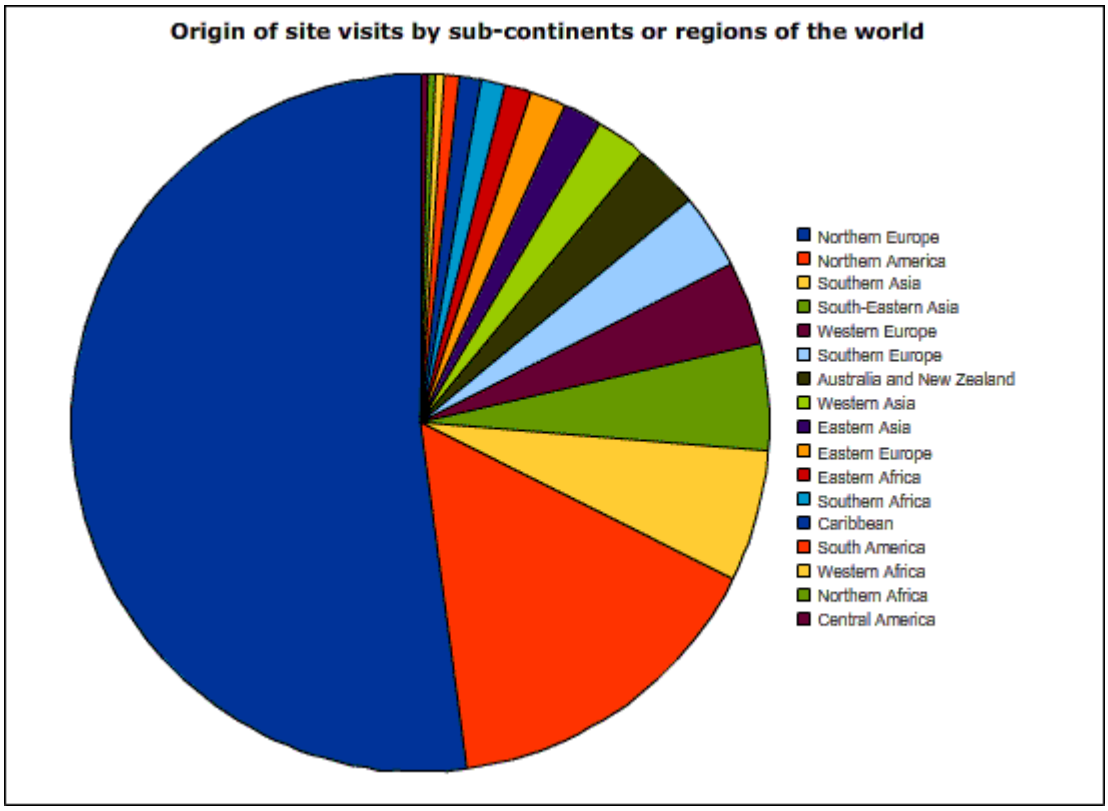


Thus, overall it is clear that the site is continuing to have robust and consistent levels of usage with increases through the TRI-ORM period.

4. Geographical origin of visitors

Between 1st November 2006 and 5th February 2009 visits were made to the site from 179 countries/territories. 49.61% of traffic during this period came from the UK, followed by 12.92% from the US, 4.84% from India, 2.52% from Canada, 2.41% from Australia, 1.75% from Malaysia, 1.53% from Germany, 1.32% from the Philippines, 1.2% from Slovenia and 1.04% from Ireland. These are the top 10 countries that visits to the site originated from, and a total of 79.14% of traffic came from these countries. The remaining 21% came from 169 other countries across the globe, each making up less that 1% of the total.

To give an overview of the number of visitors from different areas of the world, the following chart shows the proportion of visitors from sub-continents or regions of the world from which 150 or more visits were made:



The largest number of visits came from Northern Europe (28,749 or 52%), followed by Northern America (8,536 or 15%) and Southern Asia (3, 368 or 6%). While the vast majority of users originated from 'The West', the figures also reveal that the site has global reach with over 10,000 visits from Asia and Africa during the period (18.35% of all visits). Analysis of changes in the geographical origin of visitors throughout the period also shows substantial increases in visits from Asia, Eastern Europe, Africa and South America.

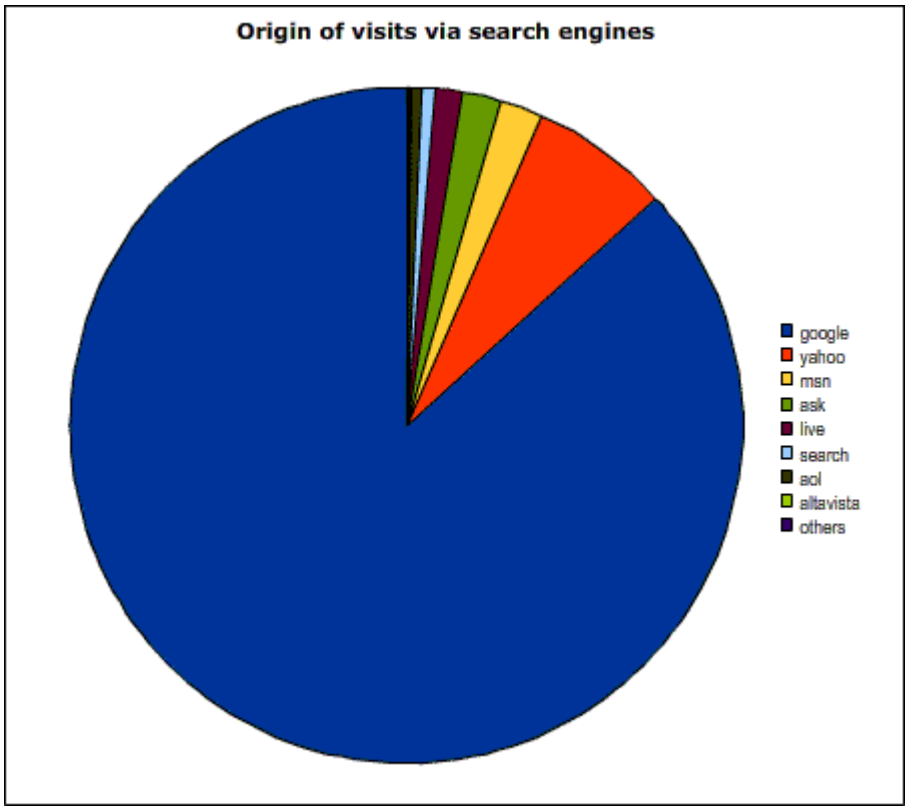
The overall picture is clearly one of high and increasing levels of geographical reach.

5. Traffic origin

During the period 1st November 2006 to 30th April 2009, the majority of traffic to the website came via search engines (63.61%). 19.46% of traffic came via links within other websites and 16.91% was through direct traffic.

5.1 Search engines

The following chart shows the origin of search engine traffic to the site:



Of the search engines used, Google was by far the most common (34,648 visits or 87%). Analysis of search times can provide a useful indicator of the purpose of quality of visits via search engines. The top 20 search terms used to find the site were:

Search term	Number of times used	Pages viewed per visit	Average time spent on site
1. online research methods	610	6.61	4 mins 47 secs
2. advantages of questionnaires	522	1.56	1 min 21 secs
3. online research	498	5.88	4 mins 47 secs
4. advantages of interviews	443	1.36	32 secs
5. advantages and disadvantages of questionnaires	407	1.63	1 min 43 secs
6. types of questionnaires	359	2.80	1 min 32 secs
7. advantages and disadvantages of interviews	332	1.61	1 min 2 secs

8. online interviews	297	4.92	4 mins 7 secs
9. research methods	234	4.59	1 min 38 secs
10. online questionnaires	228	3.88	2 mins 27 secs
11. advantages and disadvantages of questionnaire	200	1.68	1 min 56 secs
12. disadvantages of questionnaires	196	1.52	1 min 5 secs
13. advantages and disadvantages of interview	187	1.47	1 min 15 secs
14. methods of training	166	2.18	18 secs
15. advantages of face to face interviews	157	1.67	1 min 13 secs
16. benefits of questionnaires	139	1.57	1 min 49 secs
17. disadvantages of interviews	122	1.66	1 min 6 secs
18. advantage of questionnaires	107	1.31	1 min 34 secs
19. online questionnaire software	101	3.46	2 mins 9 secs
20. disadvantages of online surveys	97	1.87	2 mins 13 secs

These search terms generally find the website on the first page of results returned by Google. The most popular search term which brought a high degree of 'quality' usage 'online research methods' returns the website in first position. The site is clearly useful for users and this is reflected in the ranking within google.

In addition to these general search terms, some very specific searches have brought some extremely high quality visits to the website. Sorting search terms by number of pages viewed returns a range of terms such as the following suggesting that the depth and detail of content in the site is providing for a range of advanced uses:

Search term	Number of times used	Pages viewed per visit	Average time spent on site
european union internet research ethics	1	69	11 mins 12 secs
upload online questionnaire	1	58	23 mins 2 secs
training courses for online research	3	38.4	21 mins 7 secs
how to design online questionare	1	35	11 mins 17 secs
informed consent internet health research	1	33	62 mins 5 secs
ethical issues for online questionnaires	1	28	11 mins 33 secs
layout of online questionnaire softwares	1	27	54 mins 41 secs
internet sampling with examples	1	24	32 mins 51 secs
reponse rates questionaires online	1	23	47 mins 37 secs
consent forms for online interviews	1	21	9 mins 7 secs

Sorting by average time spent on the site also returns similar results, showing a range of 'quality' visits focused on a variety of technical, methodological and ethical issues.

Search term	Number of times used	Pages viewed per visit	Average time spent on site

php open and hide contents of page by checkbox	1	12	136 mins 48 secs
advantages to online chat room interviews and applications	1	1	54 mins 13 secs
returning questionnaires via email	1	6	52 mins 6 secs
informed consent confidentiality ethics	1	16	40 mins 53 secs
online research + ethics	2	11.5	38 mins 6 secs
how to check whether textbox is empty or not in javascript	1	5	35 mins 12 secs
incentives for online interviewing	1	9	32 mins 35 secs
research ethic training self study	1	15	30 mins 50 secs
asynchronous focus group	2	19.5	30 mins 20 secs
"client side and server side processing"	1	12	25 mins 17 secs

These examples provide a sample of over 20,000 search terms that were used to reach the site. Site access via search engines is clearly robust. The figures suggest that the site is adequately designed to be suitable for Google searching and its position in the Google rankings suggests that it has continued relevance for users.

5.2 Direct traffic

The figures for direct traffic to the site show that this method of reaching the site tends to bring traffic of a relatively high quality. 9,387 Visits (16.94% of the total number) arrived direct. These visitors spent, on average, 65.94% longer on the site (4 mins 4 secs compared to 2 mins 27 secs), and to view 54.97% more pages (5.42 pages per visit, compared to the site average of 3.5). This suggests that there has been a tendency to bookmark the site and return to it direct which adds weight to suggestions that it has become an important part of the research toolkit for a number of users. Spikes in the direct traffic figures coinciding with and following delivery of the project workshops and online module also provide some evidence that the site is well integrated with the other strands of training provided through TRI-ORM.

5.3 Referral from other websites

A total of 10, 652 visits to the website came about as a result of links from other websites, accounting for 19.22% of total traffic. A wide range of websites include links to the site and the top 30 are shown below. The majority of these are institutional websites such as academic institutions, research organisations and subject academies, reflecting the academic quality of the site. Institutional Virtual Learning Environments or portal websites also make up a number of referring sites, including the University of Leicester which again suggests a close synergy between the website and the other strands of the TRI-ORM training offering. Others still are 'web 2.0' services such as blogs, wikis, social bookmarking websites and personal homepage sites.

Referring site	Number of referrals
www.le.ac.uk	1314
gsociology.icaap.org	1259
en.wikipedia.org	630
gesis.org	363

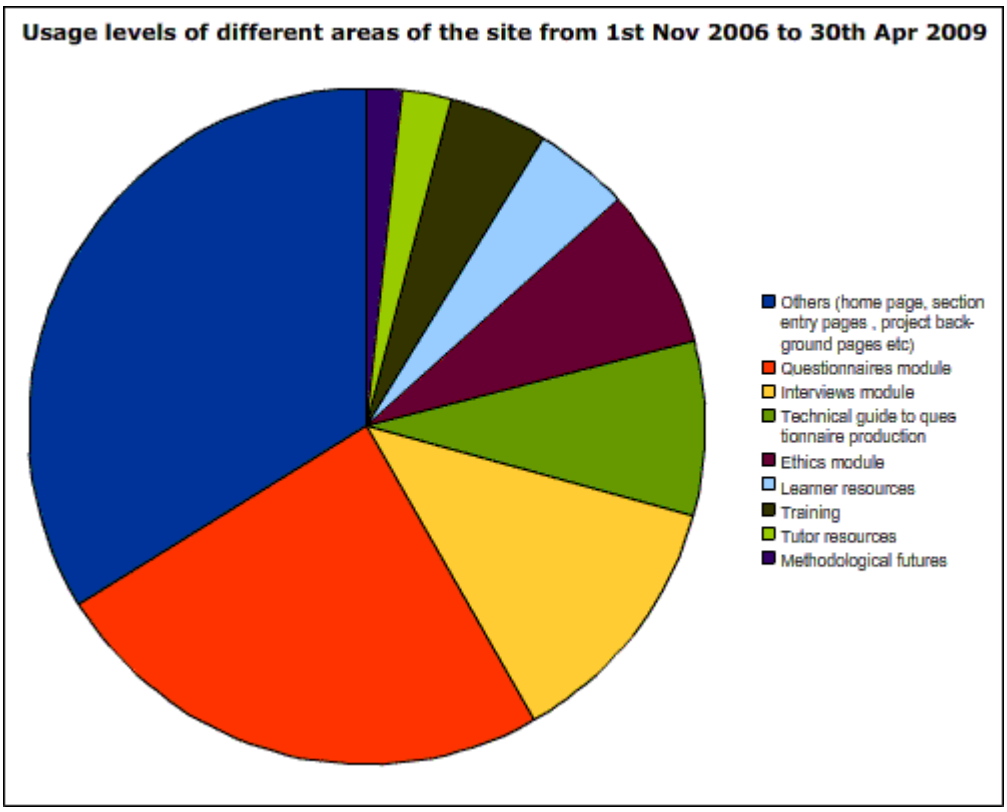
images.google.com	326
scie-socialcareonline.org.uk	301
websm.org	293
images.google.co.uk	274
york.ac.uk	266
ccsr.ac.uk	206
business.heacademy.ac.uk	189
rdi.ac.uk	172
aoir.org	171
ncrm.ac.uk	168
researchmethodsarena.com	168
nottingham.ac.uk	167
blogs.usyd.edu.au	164
data-archive.ac.uk	164
ilod.cot.org.uk	139
learn.open.ac.uk	93
del.icio.us	89
msxml.webcrawler.com	83
www2.warwick.ac.uk	80
esds.ac.uk	75
google.com	71
ljmu.ac.uk	61
answers.com	60
www2.rcn.org.uk	57
netvibes.com	56
moodle.univ-ab.pt	53

The site has clearly become embedded as a resource which is referenced in institutional information sources and educational courses, as well as in web services created and used by individuals and organisations within the 'web2.0' environment.

Overall, the range of sources of traffic to the website suggest it continues to be a highly relevant and useful resource accessible in a wide range of ways and with a generally high level of visibility reflected in search engine usage, links from a range of web sites and services, and in individual and course-based usage direct.

6. Contents

The approximate breakdown of usage levels of different sections of the site between 1st November 2006 and 5th February 2009 is shown in the following chart:



The site self-study materials were the most well used sections accounting for 55% of total traffic, with the questionnaires module being the most popular (24% of all traffic). Of the remainder, 12% was made up of learner and tutor resources and the training section. 33% consisted of introductory pages, site information and project background pages.

Content usage for 1st November 2008 to 30th April 2009 show that access patterns have remained largely consistent. However, the learner and tutor resources sections show an overall increase, as does the ethics module at the expense of the technical guide. The methodological futures section also shows a relatively substantial increase. It is likely that these 'newer' areas of the site will increase in popularity following the official release of the TRI-ORM version of the website.

7. Conclusion

While there are clearly difficulties in establishing usage figures with complete precision, a range of measures are available to allow the quality as well as the quantity of site usage to be estimated. These measures suggest that the site continues to show robust and growing usage patterns. The site has a wide range of users who access the site in a variety of ways from a vast range of geographical locations. It clearly has continuing relevance.