



M3.1.2 – Europeana Log Analysis Report 1

Europeana: an evaluation of users, usage and information seeking behaviour derived from web-server log-files (October 2009—April 2011).



co-funded by the European Union

The project is co-funded by the European Union, through the **eContentplus** programme

<http://ec.europa.eu/econtentplus>

Distribution

Version	Date of sending	Name	Role in project
0.7	2011-05-29	CIBER-research.eu	for UCL Consultants Ltd
0.99	2011-06-17	CIBER-research.eu	For Harry Verwayen and VeronikaPraendl-Zika
0.995	2011-07-15	CIBER-research.eu	with review comments noted.
1.0	2011-08-21	PUBLIC	

Approval

Version	Date of approval	Name	Role in project
0.9 Final draft			
1.0	2011-08-21	D J Clark	CIBER-research.eu / UCL Consultants Ltd.

Revisions

Version	Status	Author	Date	Changes
0.5	Draft	DJ Clark	2011-04-15	
0.6	Draft	DJ Clark	2011-05-11	with notes and track changes
0.9	Final draft	DJ Clark	2011-06-17	Revisions
0.995	Revised	D J Clark	2011-07-15	Review comments noted.
1.0	PUBLIC	D J Clark	2011-08-15	Notes and Revision History Removed



Table of contents

Table of contents	3
Summary of key findings	4
Introduction	5
[0] Context	6
Europeana	6
Data processing	6
[1] Search Engine Optimisation (SEO)	9
Browse-all	9
Most frequent referring sites	12
As noted, Google is the most frequent referring site by some margin, accounting for 57% of all externally-generated traffic to Europeana.	12
User navigation within Europeana.....	12
[2] Temporal patterns:	14
Case study 1: Visit (session) timing.....	18
[3] Local patterns	21
Click-through: popular providers and collections	21
Case study 2: The French question	23
[4] Media	25
[5] Future analysis of Europeana users and information seeking.....	27
Appendices	28

Summary of key findings

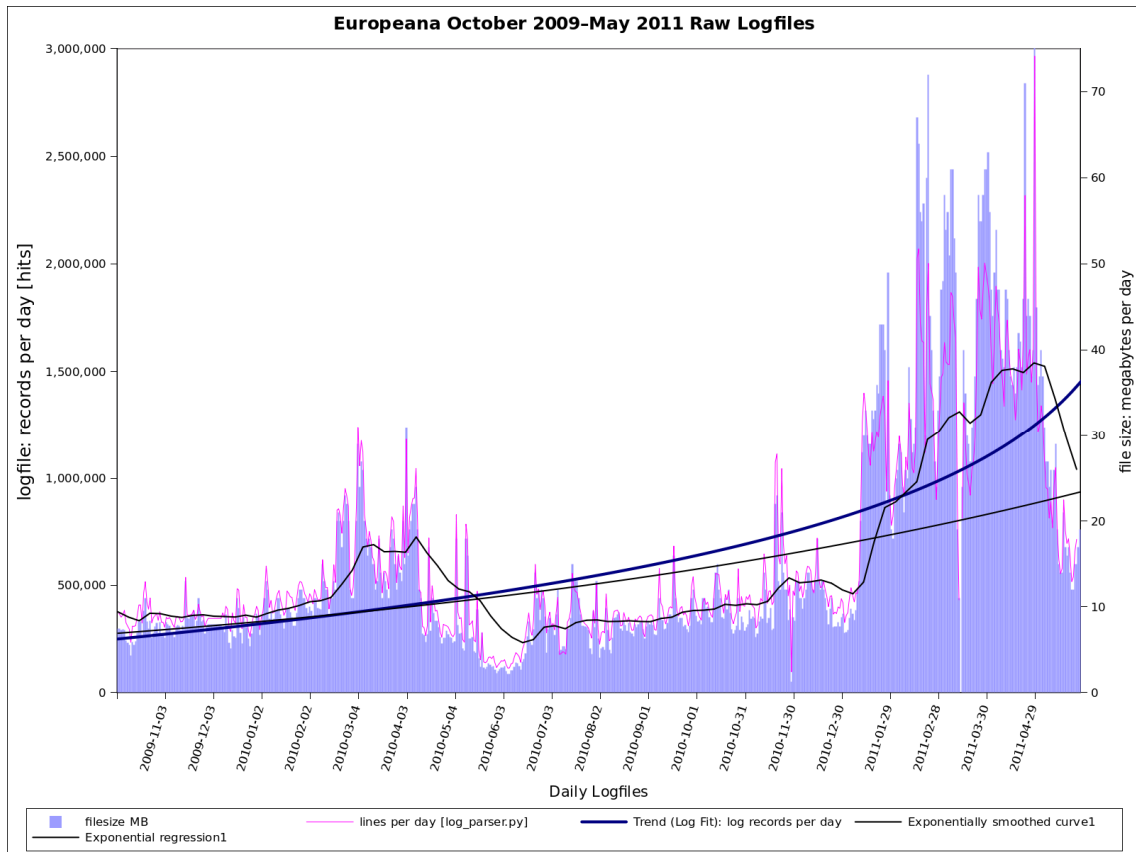
- Search engine optimisation (SEO) has been a huge success. Introduced in the latter half of 2010, optimisation has led to a five-fold increase in traffic to Europeana, measured as volumes of log transaction data.
- Since January 2011 there has been a huge increase in referrals from Google, the only search engine that counts, and as a result, Europeana content is now more widely available and used than ever before. Google was responsible for 57% of all visits between January and April 2011, up from 23% for the equivalent period in 2010.
- These factors are reflected in the very rapid growth in the number of unique visitors to Europeana, and the underlying growth trend is estimated at around 146% per annum. In April 2011, Europeana saw 620,000 unique visitors and CIBER predicts that this will more than double by December 2011, when we expect 1.3 million visitors to use Europeana content.
- As well as creating a wider audience for Europeana content, search engine optimisation has changed the way the site is used. New, Google-directed, users go direct from the search engine to content (record page) rather than navigating from the home page. We are seeing more users, but many more of them are bouncers. This phenomenon is particularly notable in the case of users located in France.
- Now that, thanks to Google, many users are arriving direct at a 'record' rather than navigating their way around Europeana, it might be a good moment to pause and think about the design of this page. Could more be done to encourage browsing and to make the record page more interesting and 'stickier'?
- Mobile devices are proving to be very popular. Although still a relatively small component of overall Europeana use the growth rates for this platform are very impressive, growing by more than 190% per annum. This growth is being led by the popularity of the iPhone and iPad, which in April 2011 accounted for two-thirds of all mobile page views.
- France is the largest single user of Europeana, accounting for 16% of all visits. The next highest visiting countries are Germany (14%), the USA (10%), Poland (7%) and Spain (7%). These top five countries account for over half of all (54.1%) of Europeana usage over the period January to April 2011.
- Use of Europeana tends to be nationally-focused rather than truly pan-European: the heaviest relative use of national collections tends to be by users based within the same country.
- Data on the duration of visits or sessions as recorded by commercial web analytics software, such as Google Analytics, are seriously misleading. The typical pattern of usage is very skewed when measured by time: most visits are very short and a few are very long. As a result the arithmetic mean used by GA and others overstates the duration of a 'typical' visit. CIBER uses the median for visit lengths and these have reduced dramatically since SEO: from 1 minute 24 seconds (January-April 2010) to just 54 seconds (January-April 2011) as the effect of increased bouncing rates attributable to Google indexing have become more prominent.
- Blogging sites are beginning to make a very significant impact, accounting for around one referred visit in ten between January and April 2011, up tenfold from the equivalent period in 2010. In both absolute and relative terms, the incidence of visits referred to Europeana from other, non-blogging, social media sites remains small albeit, with signs of growth.

Introduction

In our last report on the Europeana Prototype (M3.1.4) we noted the high 'noise levels' in the log data; the difficulty of pointing to reliable and useful conclusions when significant levels of logged activity may represent internal use by developers and irregular bursts of testing by partners. The situation changed in 2011 when the effects of search-engine optimisation (SEO) introduced in the later half of 2010 led to a dramatic five-fold increase in traffic to Europeana.

Consequently this report has been put back two months so as to present a more representative analysis. It is still a little early to make a year-on-year evaluation of the transformation; in late May 2011 there appears to be the onset of the seasonal lull in usage but it is not yet certain how close the comparison with 2010 should be. So we will concentrate on the first four months of 2011 and for context compare with the same period of the previous year (Jan–Apr 2010 v Jan–Apr 2011).

The graphic below clearly shows the impact of search optimisation, with traffic to Europeana following a pattern of exponential growth.



[0] Context

Europeana

"a multilingual point of access, a network and a channel for digital content distribution."

Europeana, the European digital library, originated with a 2005 proposal supported by six European heads of state (France, Poland, Germany, Italy, Spain, and Hungary): the Digital Libraries Initiative. It is a project to "to make all Europe's cultural resources and scientific records: books, journals, films, maps, photographs, music, etc., accessible to all, and preserve it for future generations". Europeana is conceived as a single access point for all these digital materials: the wandering scholar no longer has to travel the length and breadth of Europe seeking the original, digital copies are accessible online. It is also intended to provide stimulus to a 'digital economy', content creation and to 'democratise access to culture and knowledge'.

The Europeana.eu website was launched in November 2008 as a "multimedia online library". Analysis of the server log-files is part of the Europeana Connect project which commenced in May 2009. After an initial assessment of sample files in the summer of 2009 arrangements were made to transfer the server logs on a daily basis to the research team at CIBER-research.eu. This automated transfer of the complete files has been in operation since October 2009. Thus now, in May 2011, we are able to present a report covering over 18 months of stable operation of the europeana.eu web-site. A full-year's data enables observation of seasonal patterns. In the latter half of 2010 a major upgrade of the site ('Rhine release') and a programme of search-engine optimisation (SEO) led to a fivefold increase in traffic to the site in the first four months of 2011. This has required some recalibration of our log-file analysis to account for both remodelled features of the site and a genuine 'step-change' in usage.

Data processing

Since January 2011 there has been a significant increase in the size of log-files and the peak was reached on 29 April when three million hits were recorded in a single day. 'Hits', each of which generates a log-file record, do not translate directly into web pages viewed or counts of unique visitors: a web-page is composed of many components, some visible such as images, others unseen by the user such as style-sheets and javascript. We also remove records of error pages; server hits that do not result in data presented to the user. The result is a set of records of pages viewed, from which we extract relevant information about the page, its content and the viewer. The result is a very large database table: 150 million page-views, 4 million visitors, since October 2009. For each of those rows we can identify a multitude of attributes, but only a few hundred, among several thousand, occur with sufficient regularity and with the stable range of values that permit effective data mining. The aim of data-mining is not just to summarise these records in convenient tables, it must also find the hidden patterns and connections, cell to cell, within the whole table.

Page-view

Logs are analysed from a user perspective; the fundamental unit is the 'page view': what new display results from clicking on a link or typing in a URL. By new display we mean a complete page refresh: thus changes to the display such as pop-ups on mouse-over or the suggestions displayed when typing in a search box are not considered a new page.

For Europeana.eu a canonical sequence of page views would be: the [Home page](#) , a [search result](#) displayed as a set of thumbnail images], a [detailed record](#), and a '[click through](#)' to a provider site. This last item opens a [new window on another site](#) ; strictly speaking this is not therefore a 'page view' of the europeana.eu site (and would not count as such for advertiser

oriented 'page impressions' counters such as Google Analytics), however we are able to record these and they are included in our page view counts as 'redirect'. Additional analysis of 'shownAt' (the link text 'View in original context') and 'shownBy' (the link on the main image on the record page) is used to discover the popularity of providers and content. Clicking on the Picture (shownBy) is far more popular: the popularity varies with type of collection and image but 60-80% of redirects are from clicking on a picture.

Visits

Many Europeana visitors are 'bouncers', they view only one page. Of these a high proportion may be promiscuous singletons, making one visit to one page, never to return. On the other hand there are a few visitors who view thousands of pages. We group page-views into 'visits' (sometimes called sessions), a sequence of page-views we can ascribe to one user and displaying the Aristotelian unities of action, time and place. That is a visit has one actor, begins with a referral from another site, follows a chain of links from one Europeana page to another, and lasts no longer than one day.

Robots, outliers and real users

A singleton user may have visited Europeana in error, following a search result that did not satisfy, a reference that was not relevant, or perhaps quite the opposite: going direct to the answer and thus having no need to explore further. Either way, the singleton or one-shot user is a particularly interesting category, we have an opportunity to study a binary result: success or not? Those users who view thousands of pages pose a quite different question: are they really users at all?

The distinction between 'user' and 'robot' is a separation of interactive use from automated data gathering. For the most part this is clear cut: much robot activity can be clearly identified from the User-agent alone and is obviously distinct from a person seeking to find specific information or generally browsing. Such personal users are also clearly identified by their use of common browsers such as Internet Explorer or Firefox. The user/robot partition is thus reliable but nonetheless there are some ambiguities, particularly when our intent is to study how well a website serves the needs of normal conscious interactive Users.

Services such as "Google Web Preview" this will show a thumbnail image of the site's page alongside the search result presented by Google. In such cases the end-user has not visited our site but the Google agent has. The user has had a preview of the site: should that count as something akin to visiting the site or is it just another robot?

Some apparent users are in fact robots with malicious intent (e.g. DTS which gathers email addresses), others are benign (e.g. ECIS/Documentum Federated Search Services) but by acting a proxy for a search will misrepresent the user's activity. We also need to consider that sometimes there is considerable activity logged by developers and others close to the project. A careful study of user behaviour therefore requires more than the binary distinction of user/robot. In addition to the Robot/User classification which largely relies on identification by user-agent, we use the location and institutional information that can be derived from the network address together with patterns in the timing and frequency of access to sub-set the 'User' category.

For Europeana we have settled on four classes of 'real user' and two classes of 'robot'

The **One-Shot** class has been described above. Because of rapid growth over the past year, a distinct page-format and implied information need we treat **Mobile** users as a special category. **Heavy users** are probably based in institutions and may well be associated with the Europeana project as a developer or provider. The remainder, accounting for the majority of pageviews and over half of all visits we simply class **Users**.

Of the **Robots** there are the obvious, overt robots: Googlebot, Yahoo, Yandex, msnbot. But there is another class, not declaring themselves as robot user agents but consuming resources on a scale far in excess of the heaviest of users. This category, the '**Outliers**', have been particularly noticeable since January 2011.

[1] Search Engine Optimisation (SEO)

In the later part of 2010 changes were made to the structure of European.eu that made the indexing of the site by robots, in particular Googlebot, far more effective. As a result from January 2011 there have been far more users of the Europeana site. And it has changed the way the site is used: these new users go direct from search engine to content (record page) rather than to the home page. We see far more users but they are bouncers.

There were two components to this change. Firstly a change in the format of the URLs made individual items on the site (records) much more visible to search engines.

Old Style: full-doc

```
http://www.europeana.eu/portal/full-doc.html?query=Scotland&qf=YEAR:1900&tab=&start=4&startPage=1&uri=http://www.europeana.eu/resolve/record/00401/248AE744BF9E19BA04FCDBDCF9EE9ACADA5AEB78&view=table&pageId=bd
```

The 'query string' component of the URL, everything after the '?', is usually ignored by search engine robots, consequently the full-doc page appears to be one page with ever changing content.

New Style: record

```
http://www.europeana.eu/portal/record/09405a/EDDB8C4DD7DF41EE9F5A5569C03F70E15AF2199F.html?query=modes+OR+fashion&qf=TYPE:IMAGE&start=13&startPage=13&view=table&pageId=brd
```

The record identifier is now part of the resource id, each record is seen by the robot as a distinct page and is indexed.

A similar problem with the search result thumbnails

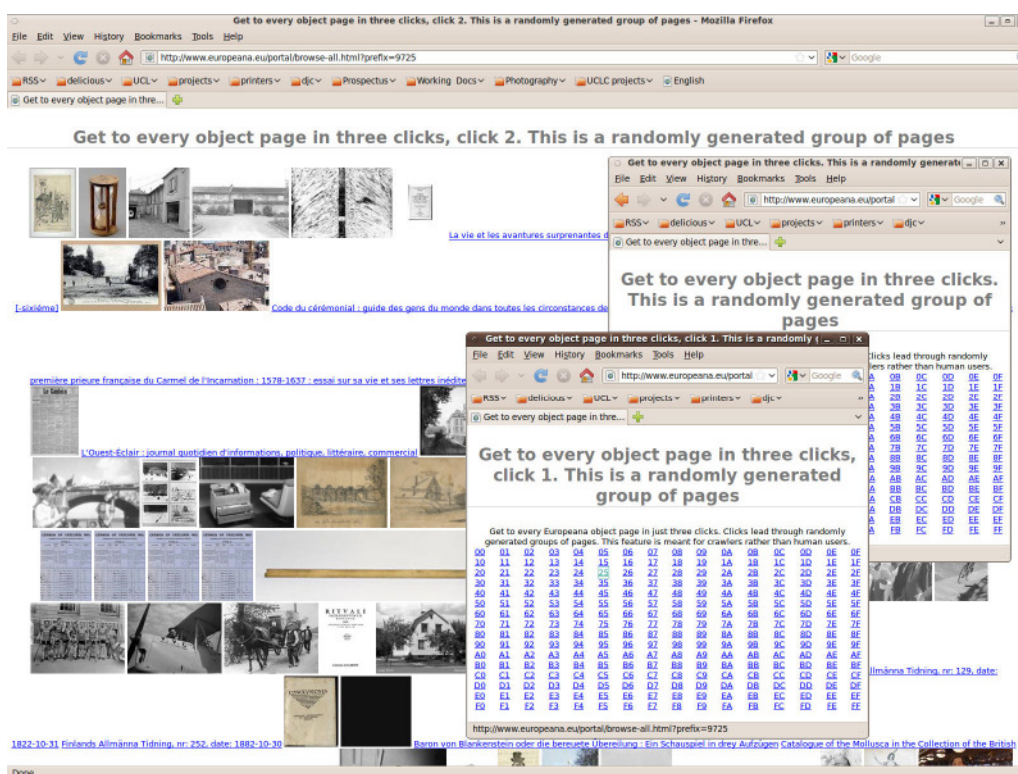
```
http://www.europeana.eu/portal/search.html?query=sole+bay
```

Robots ignore the important part, so the search results appear as one page with never the same content twice.

So to recap, the thumbnail display 'search.html' (aka brief-doc.html) leads to www.europeana.eu/portal/record/... (aka full-doc.html), which leads to the provider site (redirect.html). But to a search engine all the brief-doc and full-doc pages appeared as a single page with ever changing content. The search engine robot is caught in a maze with no clear navigation. The content [full-doc and providers site] is never properly indexed.

Browse-all

The second important change was the 'browse-all' page designed for search engines and bypassing the brief-doc page. The search engine can now use browse-all as a gateway to the full content of Europeana, each indexed as a distinct identifiable and retrievable page.



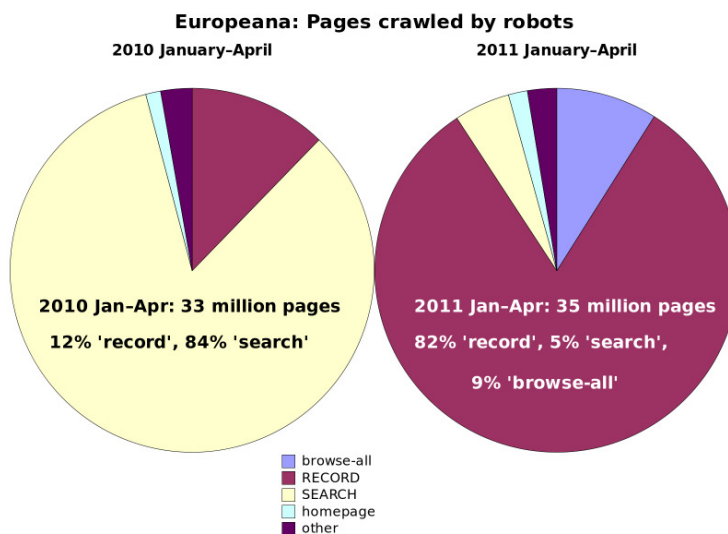
Since January 2011 there has been a huge increase in referrals from Google (The only search engine that counts: whatever their share of the search engine market, we see negligible referrals from other search engines —0.2% Yahoo, 0.1% Bing against 68.5% Google.). They go direct to the record rather than the home page. And from the record there is link to the provider site (i.e. the institutions the board represent). So we are seeing an increase in 'conversion rate' driving traffic to provider sites, making content more widely available.

Here are some key findings:

- Four key pages account for 94% of all page-views in 2010: Home-page(5%), brief-doc or thumbnails (50%), full-doc/record (36%) and redirect or click-through (3%)
- Jan-Apr 2010: Google was responsible for 23% of all Visits (150,000 out of 656,000); 69% were to homepage (104,000), 8.6% to record [full-doc] (25,000);
- Jan-Apr 2011: Google was responsible for 57% of all Visits (1,775,000 out of 3,141,000); 5.5% were to homepage (97,500) and 94% were to record (1,673,500),

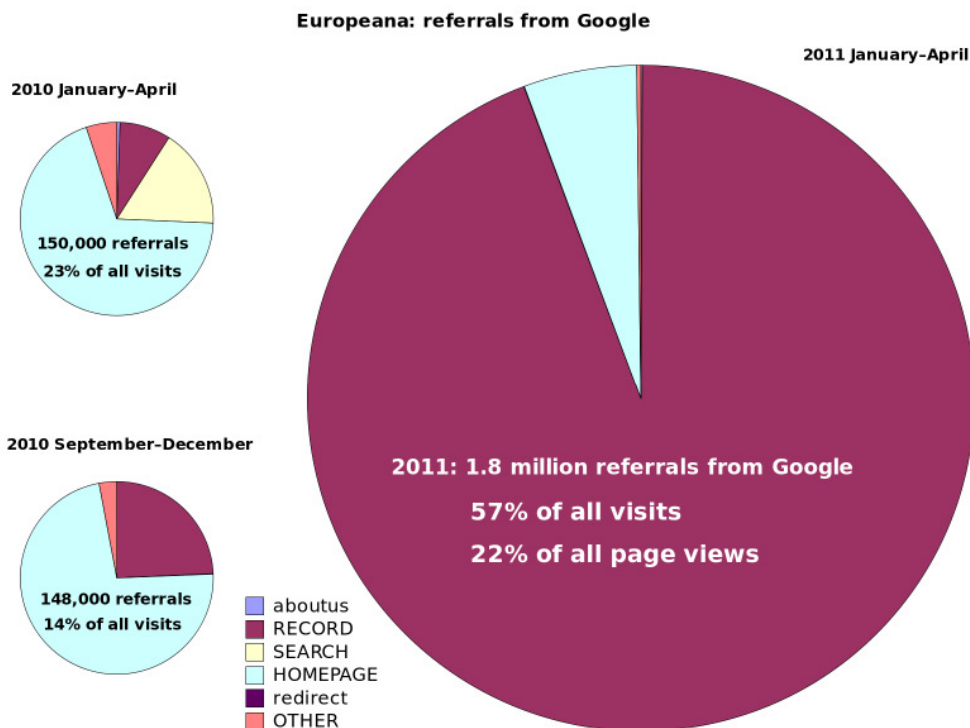
Robot use

Now that Europeana records are being indexed, a lot more robots (in fact the usual team of Googlebots) are following browse-all and going straight to the record, by-passing brief-doc, as the following graphic shows. The impact of search engine optimisation is very dramatic.



Human use

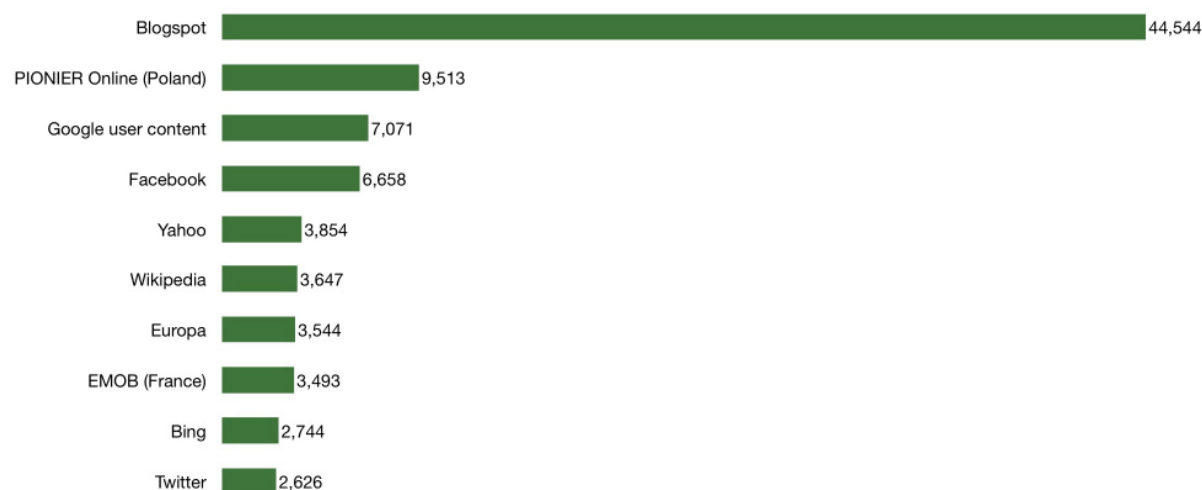
As a result of this indexing at the record level, Google referrals have increased tenfold in terms of absolute numbers of visits. There are a lot more users, a lot more bouncers, but also a lot more 'conversions'. A small slice but a much bigger pie.



Most frequent referring sites

As noted, Google is the most frequent referring site by some margin, accounting for 57% of all externally-generated traffic to Europeana.

The top ten referring sites, *excluding Google*, for the period January to April 2011 are shown below. The units are numbers of visits. Note the significant traffic from various blogs hosted by Blogspot and the traffic directed to Europeana from PIONIER online, a catalogue of the Federated Digital Libraries of Poland.



User navigation within Europeana

As noted earlier, the classic pathway that one might expect a user to take through Europeana would be homepage -> search -> [more searches] -> record -> redirect (i.e. the provider site). We have already noted, however, that opening up the site to Google indexing has changed the game, with many more bouncers going straight to a record.

Evidence that the classic mode of navigation is now less relevant than it was can be seen in this table. It shows how users actually navigate their way around Europeana in terms of transits from page (rows) to page (columns). The data are for the period January–April 2011, post-search engine optimisation, and the numbers in the table are percentages of all page-to-page transits.

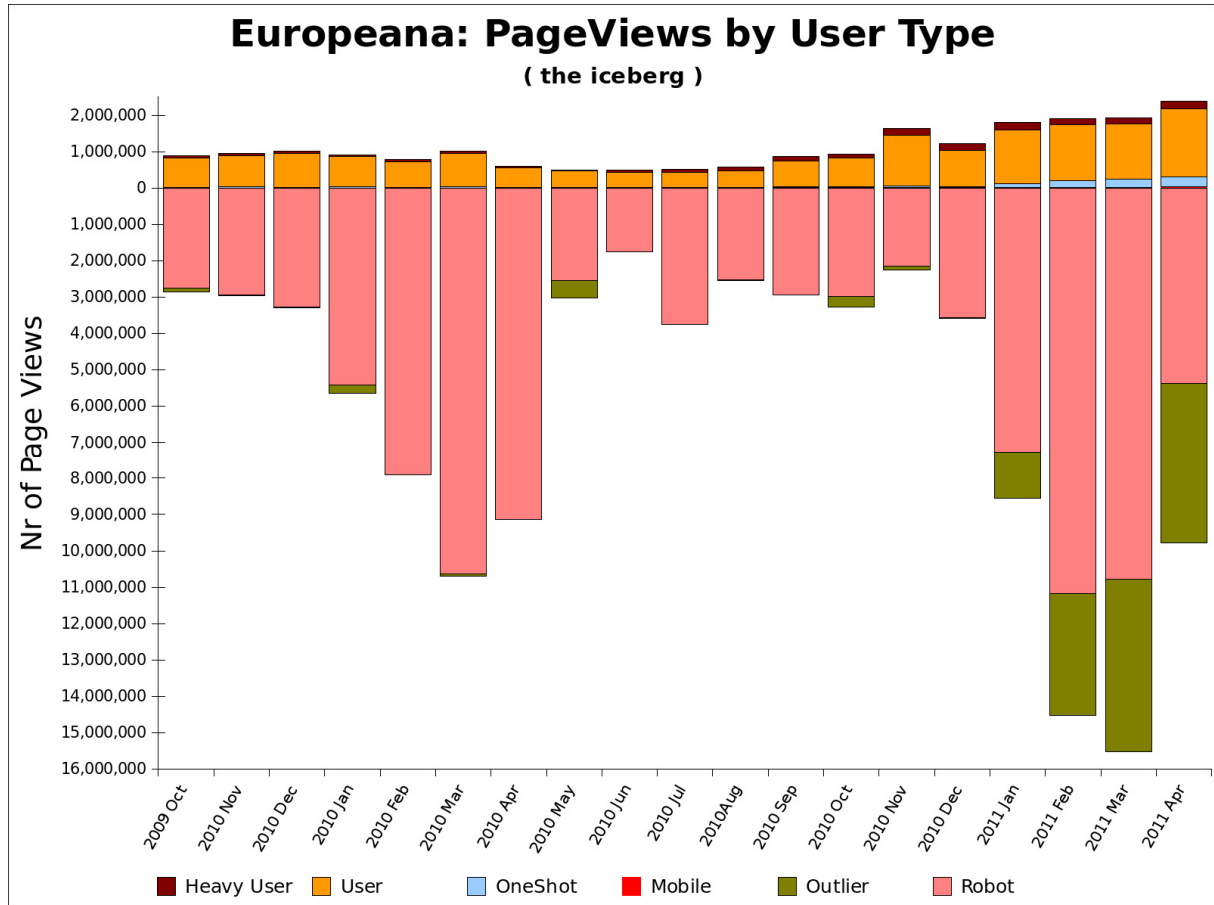
Nearly a quarter of all pageviews are now referred into the Europeana site direct to a record: predominantly by referral from Google. This is twice the number of visitors who commence their visit by starting at the home page (where a leading referrer is Blogspot). Many of these visitors will be bouncers; they do not view any other pages. Where the visitor does view multiple pages the predominant flow is much as expected: from homepage to a search, possibly multiple views of the search result thumbnails, then to a record, and from a record to a redirect to the provider site. But since such a flow represents a minimum of four page views, few visitors stay the course.

to these pages	→	record	search	homepage	redirect	aboutus	login	myeuropeana	other
↑	OFFSITE LINK	24.4	1.0	11.7	0.6	0.2	0.3	0.3	0.6
	search	7.7	22.3	0.2	<0.1	0.0	0.1	0.1	0.1
	record	6.6	1.8	0.2	11.6	0.0	0.1	0.1	0.2
	homepage	<0.1	2.6	1.1	0.1	0.6	0.1	0.1	0.7
	aboutus	0.1	0.1	0.1	-	0.8	0.1	0.1	0.2
	timeline	0.1	0.4	0.1	-	0.0	0.0	0.0	0.1
	login	-	0.1	0.1	-	0.1	0.2	0.0	0.1
	other	<0.1	0.2	0.2	<0.1	0.1	0.1	0.2	0.9

users go from these pages

[2] Temporal patterns:

In terms of page views Europeana use is dominated by Robots: like an iceberg nine-tenths of pages are hidden from view, they go unrecorded by GoogleAnalytics but they are nonetheless essential, as we have seen above, effective search engine optimisations essential to the visibility of Europeana on the world wide web.



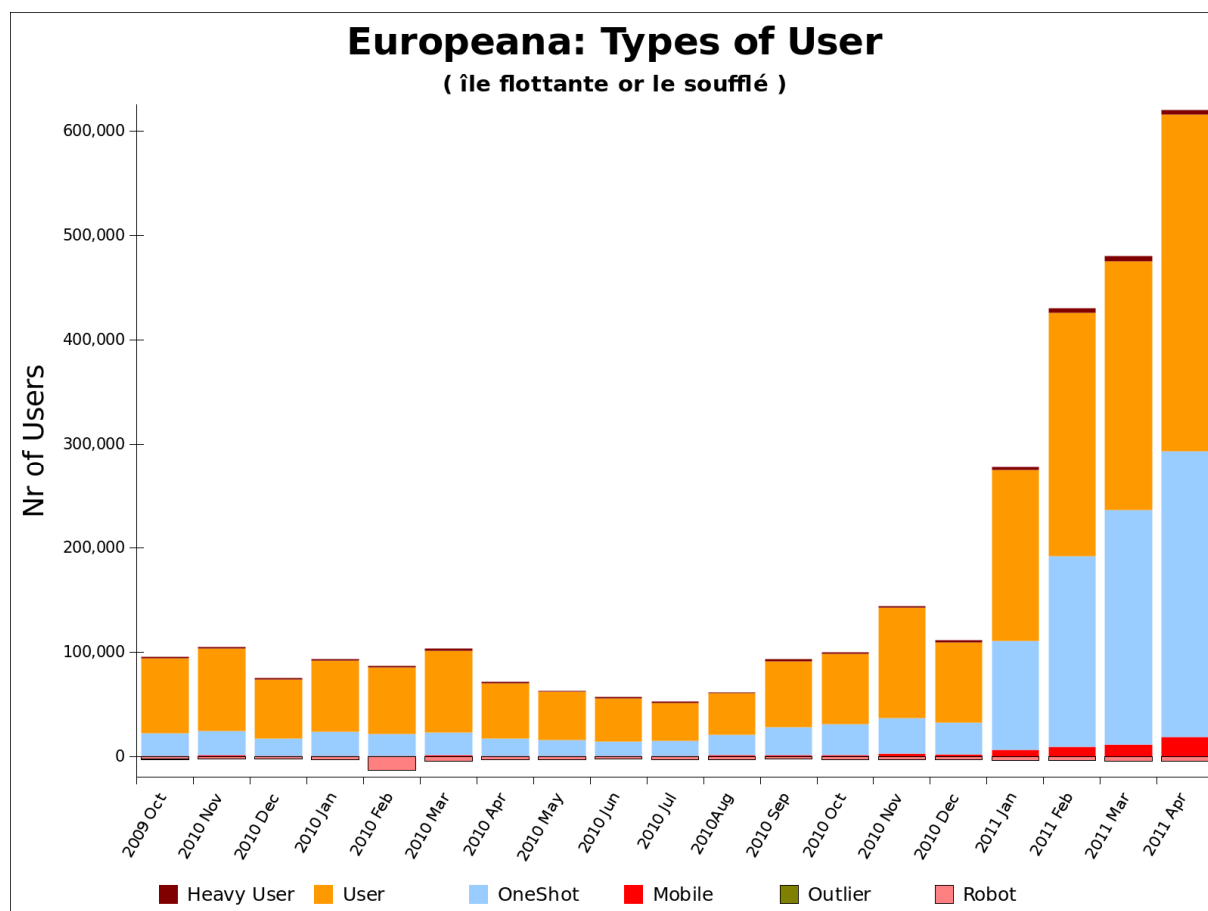
What may be less useful however are the 'outliers'. In the first three months of 2011 three IP addresses allocated to an ISP in Spain accounted for 20% of all page-views. This level of activity is not credible as coming from a lone genuine user and there is no evidence that this is a NATed or proxy connection. If, in our opinion unlikely, this is a case of real use then it would seem appropriate for Europeana to know more about such a significant customer. On the other hand if this is an automated agent, it is not clear what benefit Europeana derives from this activity which may be costing 20% of Europeana's bandwidth. In the case of genuine search engine robots (e.g. Googlebot's 40% of pageviews) the benefits are clear, but in the case of these outliers it is not. Certainly we do not see an increase in visitors from Spain that we might expect as the result of such intensive localised search engine activity.

Another pattern is seen if we consider not page-views but Visitors. While insignificant in terms of page views, One-shot users are far more visible when we count users. Most of these are the result of Google referrals and there is some evidence that French users, already well represented in Europeana content and usage, are also the most significant of the One-shot visitors.

The level of activity in May 2011 shows some falling off though it is too soon to say how much of this is a seasonal pattern or a significant but temporary aberration due to technical problems with



Google indexing in April 2011. But it seems there may be a very high seasonal use from French schools in January–April each year, much of it mediated by Google search.



There is a challenge for Europeana in this: how much of the growth in One-shot use in particular can be converted to a sustained interest in Europeana. Search-engines index pages bring in visitors, but what makes a site 'sticky'. What sort of growth have we seen these past four months: *Île flottante or le soufflé*?

By analysing the growth trends for the different user types in the graphic above, we can extrapolate and forecast the number of unique visitors, in thousands, which Europeana might reasonably expect by the end of 2011 and 2012.

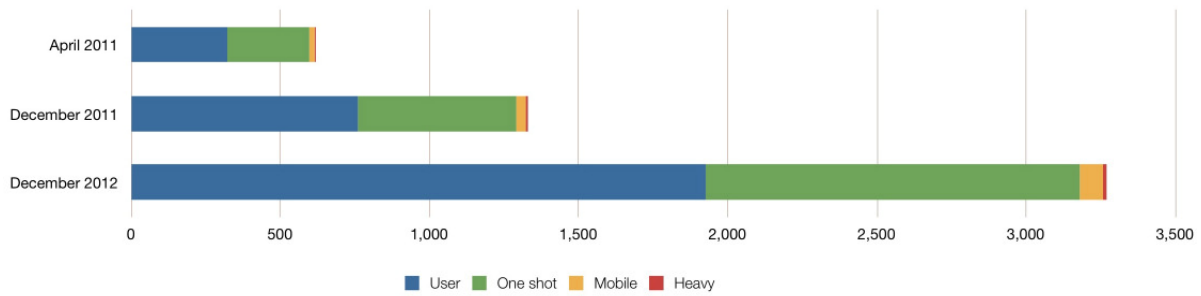
User type	April 2011 (000s)	Dec 2011 (000s)	Dec 2012 (000s)	Estimated compound annual growth rate
User	322	760	1,926	153.54%
One shot	275	531	1,254	135.93%
Mobile	18	32	78	140.15%
Heavy	4	7	12	79.37%
All users	620	1,330	3,270	145.80%

These are captured in the table above, with estimated compound annual growth rates. The mathematical (polynomial) models that underpin these forecasts fit the data very well, typically

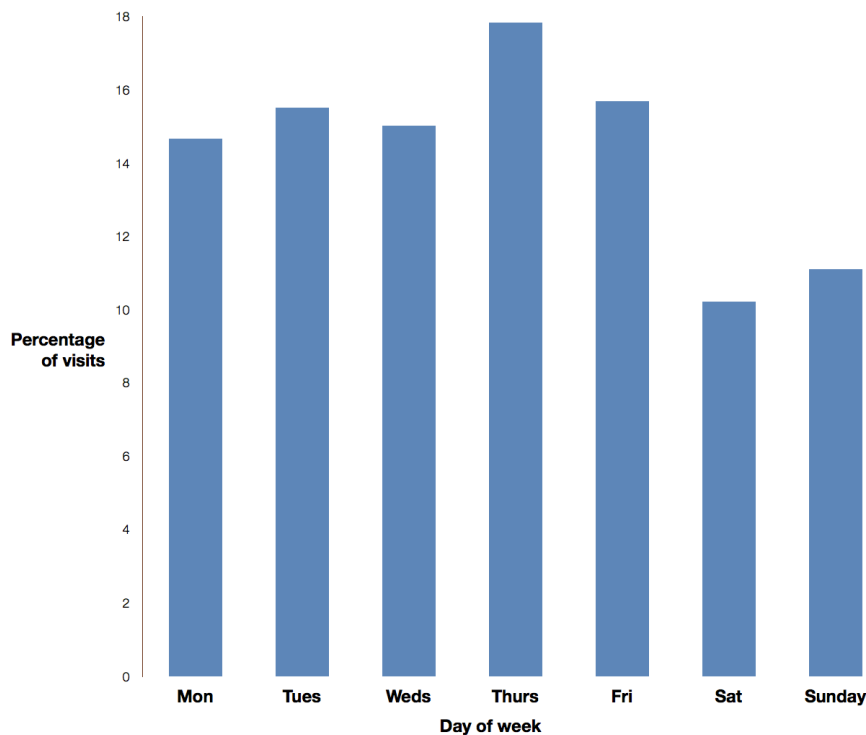


with values of R-squared in the range 0.84 to 0.96, so we can be reasonably confident that they will generate reasonably accurate forecasts in the short term. Europeana appears to have broken through and we expect the number of users to approximately quadruple over the period April 2011 to December 2012 and break through the three million mark.

The graphic below displays the same data as in the table for the numbers of unique visitors by user type.



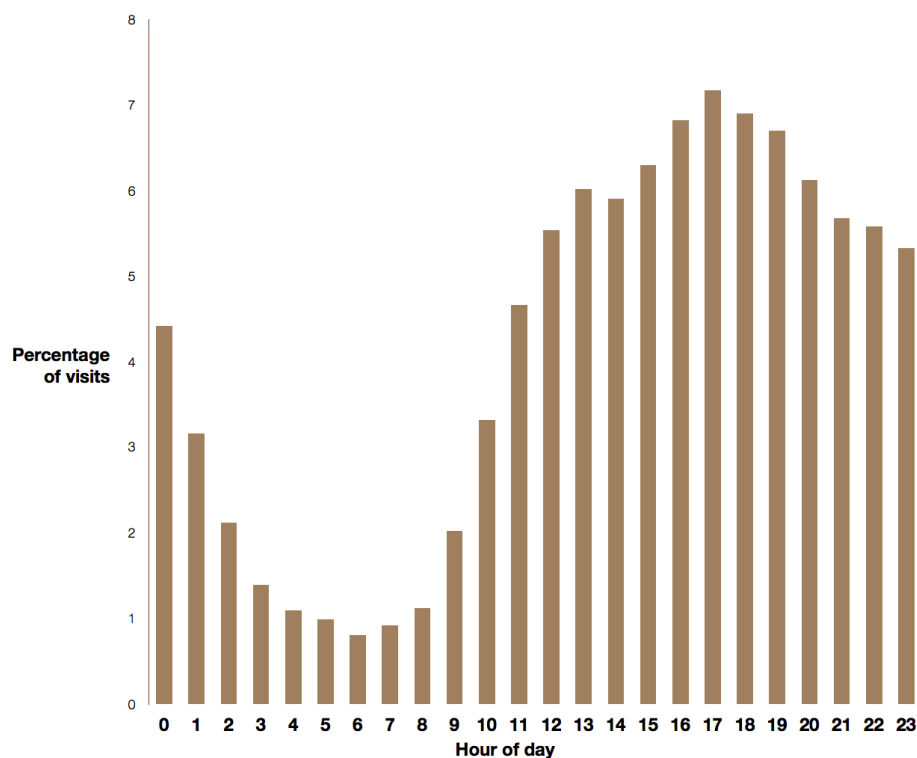
With regard to temporal patterns at the weekly level, usage shows a distinct peak on Thursday and lower but still significant levels over the weekend, at about three quarters of the level seen during the working week.



Percentage distribution of Europeana use across the week.

Mon	Tues	Weds	Thurs	Fri	Sat	Sunday
14.66	15.50	15.02	17.82	15.69	10.21	11.10

During the average day, visits to Europeana peak in the late afternoon but activity never ceases with a significant amount of `night time` traffic much of which is probably from outside the EU time zone.



Percentage distribution of Europeana visits during a typical 24-hour period.

0	1	2	3	4	5	6	7	8	9	10	11
4.42	3.15	2.11	1.39	1.09	0.99	0.80	0.92	1.12	2.02	3.31	4.66
12	13	14	15	16	17	18	19	20	21	22	23
5.54	6.02	5.90	6.30	6.82	7.17	6.90	6.70	6.12	5.68	5.58	5.33

Case study 1: Visit (session) timing

Time metrics

There are two time metrics that can be used to describe online behaviour: visit (or session) and page-view time. Traditionally time metrics, of whatever type, are used to show site 'stickiness' as a proxy for interest and satisfaction, therefore the longer the better. Even in the case of publisher platforms, like ScienceDirect or Oxford Scholarship Online, it is questionable whether these metrics do actually demonstrate interest and satisfaction. In the case of a gateway, portal or search-centric site like Europeana the opposite may be argued — the faster people move through the site the better and more efficient the indexing and navigation.

Time metrics are best used as a relative measure to compare like with like: the impact of change or seeing how things have bedded down; for differentiating the behaviour of various audiences.

Calculating time metrics

1. It is not possible to calculate times for Europeana visits of just one page and 65% of visits fall into this category.

2. There is no one-true-definition of a 'visit' or 'session'. Using alternative definitions changes the total number of visits, the maximum duration, the number of pages, and the proportion of one page (non-timed) visits.

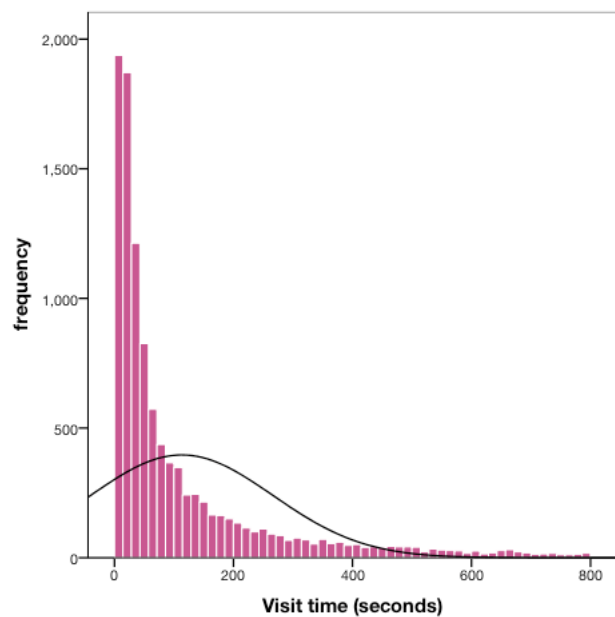
3. Google Analytics (GA) sessions are defined by session-cookies. Without these session-cookies we cannot perfectly reproduce the session data used by Google; these cookies are not currently logged so we calculate sessions from the bare logfile. Our calculations may not be strictly comparable. (*We requested the cookies be included in the log-files with our M3.1.1 Recommendation on use of logging analysis tools in Europeana v1.0 of Sept 2010*).

4. Usage time data are strongly skewed and therefore a simple mean (average) calculation as Google Analytics and many similar services provide, can be very misleading. The arithmetic average is a very poor statistic to use.

Europeana visits: the important context

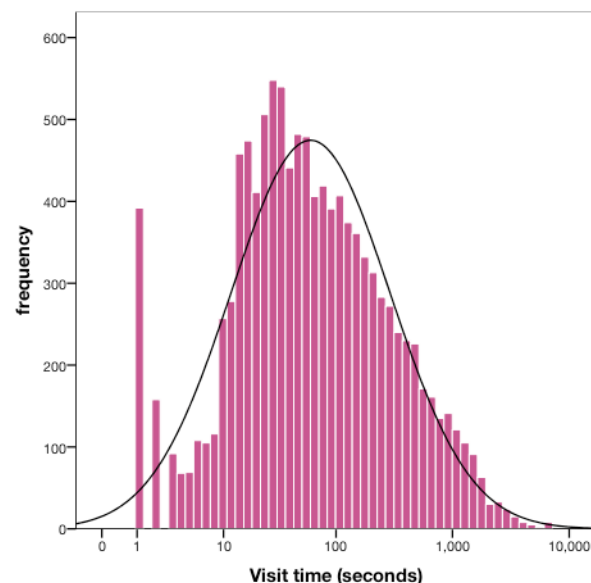
Europeana visit time typically follows a lognormal distribution, in other words there are lots of very short sessions and a few extremely long sessions.

This highly skewed pattern can be seen in the graphic below. It is very different from the normal or bell-shaped curves that we see in nature.



Many visits (65%) involve just a single page view and cannot be timed at all. Of those we can time, two-page visits of around 8 seconds represent the peak of the curve. However, a small number are very long: the longest time recorded is more than eight hours, the greatest number of pages in a single visit nearly 6,000. The presence of such extreme values causes the 'average', which is usually associated with a normal distribution, to give a misleading impression of session length.

However, if we plot the same visit time data using a logarithmic scale for the vertical axis, we get a classic bell-shaped curve (hence the description loglinear). By using this transformation, we can calculate a much more meaningful average, as we shall see.



Google Analytics session time

The GA session figures we have (for May-Jul 2010) are around the 4 minute mark (3:56, 4:12 and 4:14). Using a definition of `visit` that attempts to reproduce the effect of the GA cookies, we calculate the arithmetic average (mean) session time to be 4 minutes 20 seconds. Our guess is

that the GA average is a simple arithmetic mean with no sensitivity to the way the underlying data is distributed. It differs considerably from the median value, which we calculate to be 1 minute 26 seconds.

CIBER-research.eu visit time

The simple arithmetic mean is much longer than is reasonable as a picture of the duration of the 'typical visit'. An 'average', whether median or mean, needs to be qualified by the context of its distribution.

In the table below, we calculate the arithmetic mean by averaging the natural logs of the data (because we are dealing with a lognormal distribution). Using this method, the mean and medians converge to near identical values.

Method	Mean	Median
Typical analytics software	3 mins 21 secs	(rarely if ever quoted)
CIBER log transform	1 min 26 secs	1 min 24 secs

Since the natural logs of the data are normally distributed, we can go a stage further. If we add three standard deviations to that mean, we can identify an upper bound on what might be considered normal behaviour (capturing 99.7% of what might be considered normal human behaviour).

Finally we note that session time varies significantly by user type: 70.4 seconds (mobile users), 80.9 seconds (ordinary users) and 63.4 seconds (heavy users).

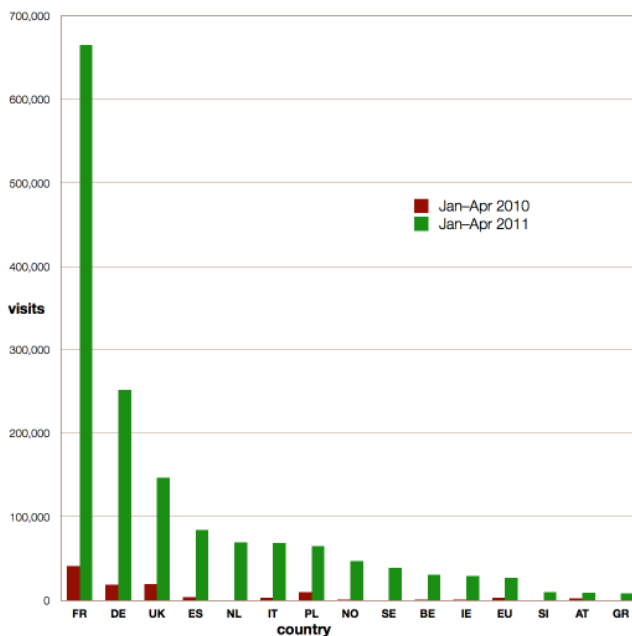
[3] Local patterns

Click-through: popular providers and collections

In our last report we compared countries of user and country of collection. We repeat that analysis here but with a modified method, counting visits rather than pages. As noted above, most visitors view very few pages, and only a very few view the `record` page. So this is an analysis of a subset of visitors: those who viewed at least one record page. But with so many users now diving straight into the record page from a Google referral that is a significant though perhaps not fully representative sample. In the few cases where more than one record was viewed in a visit, the country is the one most frequently occurring per visit.

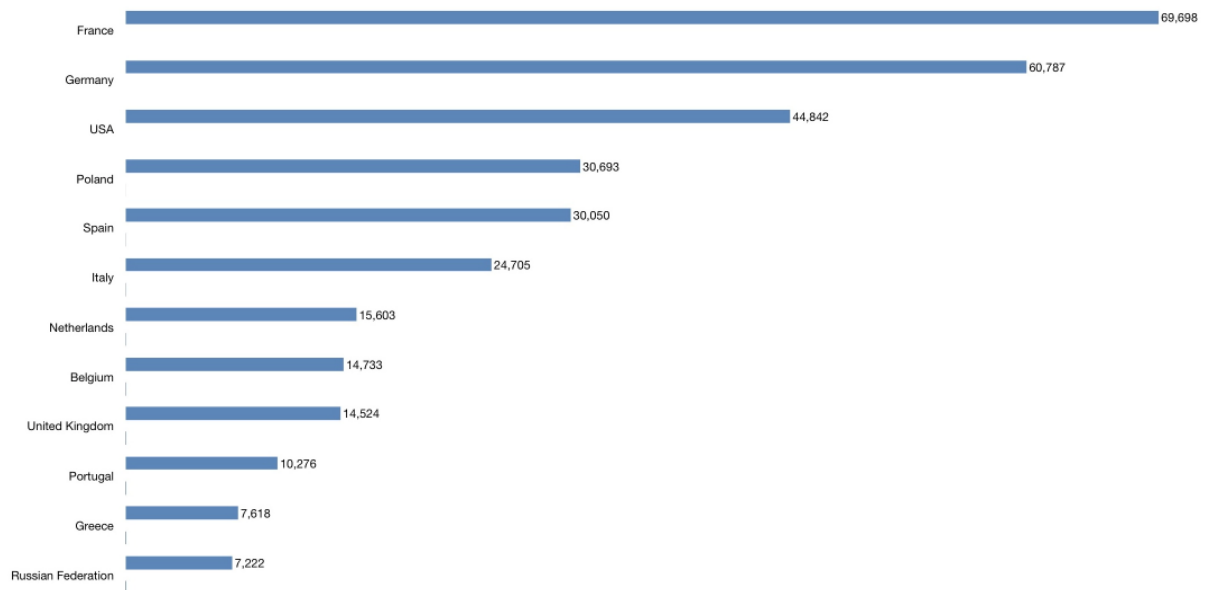
In 2010 76% of visits did not result in a record view. In 2011 this was down to 56%. This can be interpreted as a sign that visitors are going deeper into the site; drilling down from search results to the collection record; but we need also to consider that most visitors are bouncers, viewing only a single page. Since 2011 that single page has most often been a record. The big increase in record views can be attributed to the deeper indexing of the site that is sending Google users in particular straight to the record. It appears that this trend toward direct access to records has been most notable among users located in France: in 2010 they were not the most frequent bouncers (Germany, and USA being slightly ahead), in 2011 there were 360,000 single page visitors from France, thirteen times the number a year earlier. By contrast single page visits from Germany numbered 144,000 in 2011 a less than fivefold increase over 2010. Nonetheless the range of provider countries has grown: in 2010 99% of content came from 10 countries, by 2011 that 99% of content was provided by 19 countries.

Number of visits by country: January to April 2010 and 2011 compared



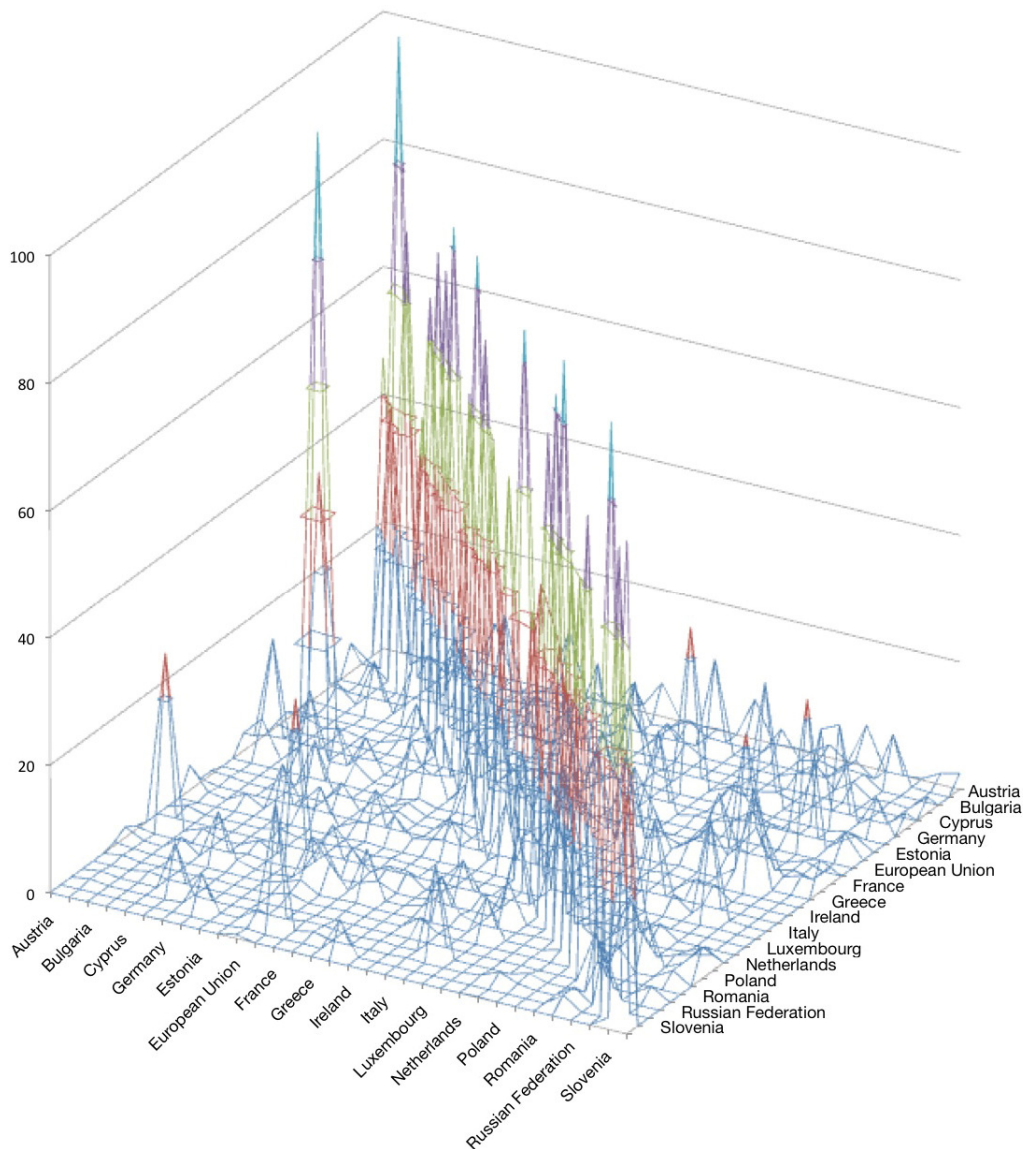
The graphic below shows the top ten main national markets for Europeana based on the number of visits. France is the largest single user of Europeana, accounting for 16% of all visits. The next highest destinations are Germany (14%), the USA (10%), Poland (7%) and Spain (7%). These five countries now account for over half of all (54.1%) of Europeana usage; the top ten countries shown in the graphic for more than three-quarters (75.8%) of visits.

Percentage of total time spent on Europeana by top user countries (January to April 2011)



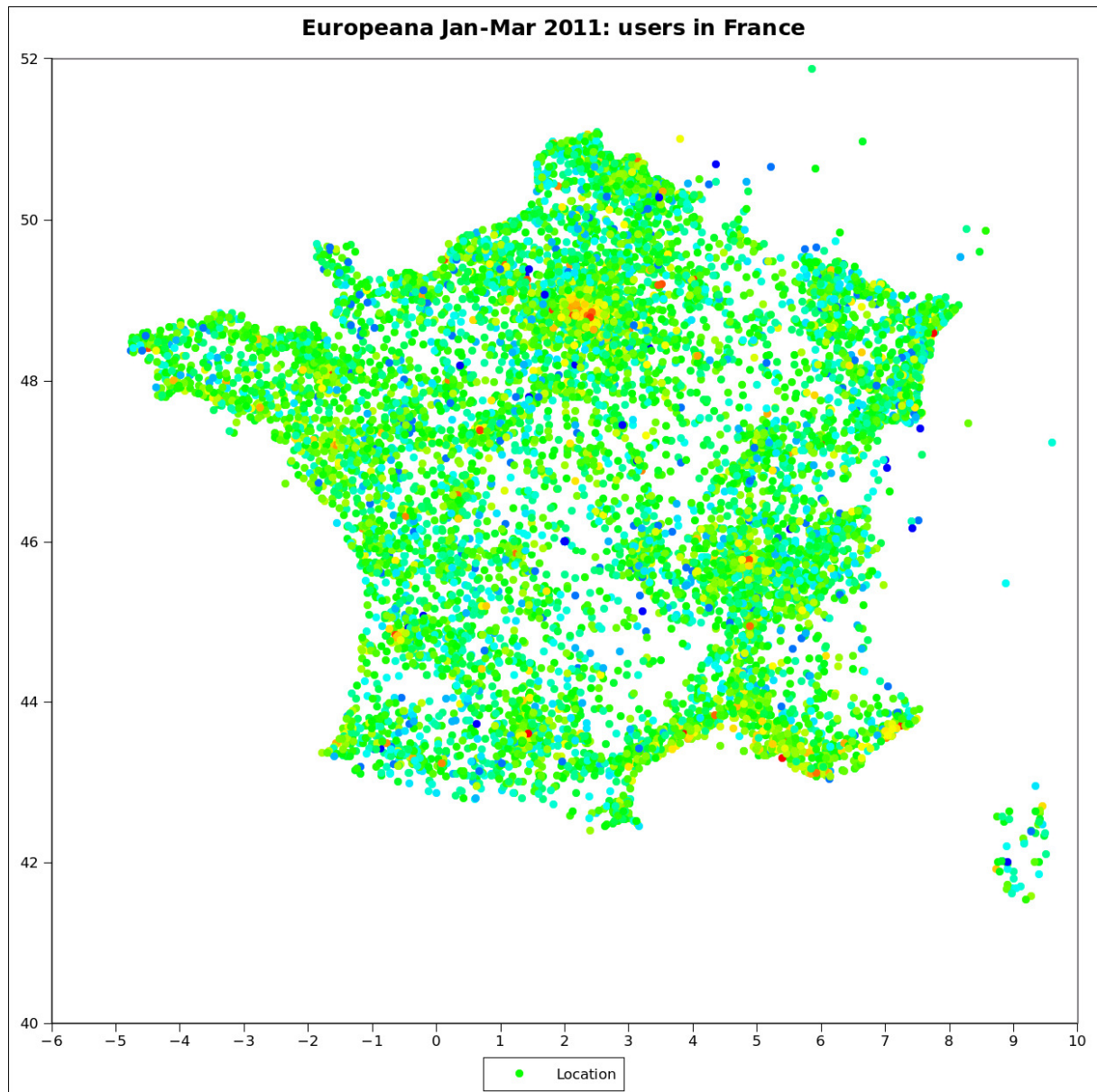


This graphic below shows how users tend to focus their interest on collections maintained within their own country. The bottom axis is the country location of the user, as indicated by their IP address. The axis on the right hand side indicates the country where the collection is housed and the vertical scale is the percentage of those collections visited within a particular user country. Nearly all the major spikes lie on the diagonal (Austrians looking at Austrian collections, Slovenians looking at Slovenian collections, etc.).



Case study 2: The French question

As we have seen, French users are, relatively speaking, very heavy users of Europeana. In the graphic below the dots represent the location of visitors based on their IP address. The colours of the dots reflect numbers of users, with red indicating high density and blue low density. Paris and other cities (and indeed heavy institutional users outside main cities) can easily be picked out in red.



It appears that French users (including DOM-TOM) are over represented relative to material in general. But as we observed in past reports we do see a strong preference for video from all users and it appears that France contributes the majority of the Video Content. It is difficult to classify content type from the log data so we cannot say at present if the majority users of video are French, and we will take a further look at this for a future report.

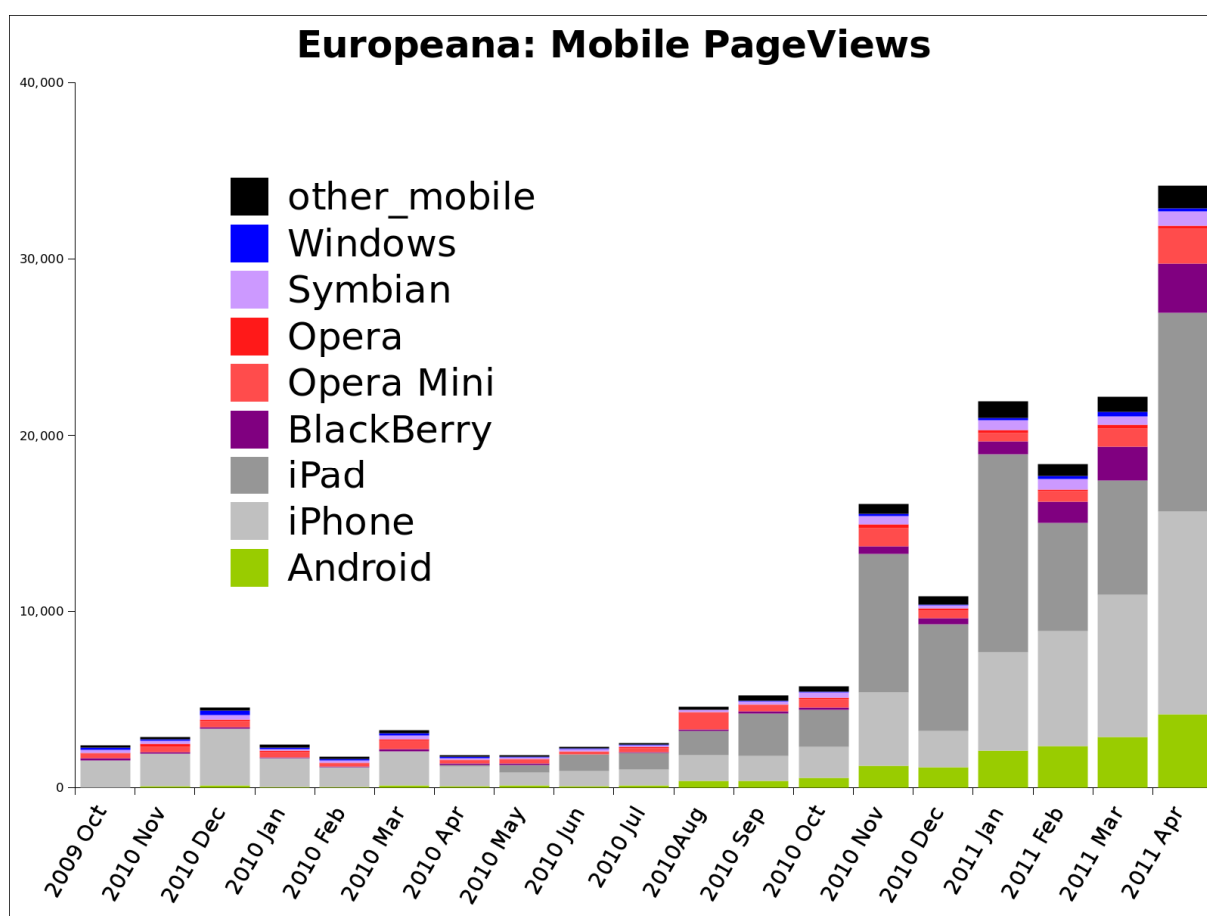


[4] Media

Case Study 3: Mobiles and other iThings

The availability of mobile platforms such as the iPad and iPhone for viewing web content has expanded dramatically and this reflects in a rapidly growing share of Europeana page views. In 2010 mobile use was 0.5% of all page views, in the first four months of 2011 1.75%

The growth of pages viewed on mobile devices is estimated to be of the order of 191% per annum (compound growth).



	April 2011 (000s)	Dec 2011 (000s)	Dec2012 (000s)	Estimated compound annual growth rate
Mobile page views	34	77	225	191.54%

As of April 2011, the fastest growing segments were the iPhone with 33.6% of all mobile page views and the iPad, with 33.0%. Apple devices have therefore captured two thirds of all Europeana mobile use.



The impact of social media

As well as search engines, social media potentially offer pointers and another route in to web content. How important is this channel and have social media yet impacted upon the use of Europeana?

An analysis of some selected and very familiar social media sites below (and these are the most important for Europeana) shows that blogging sites are beginning to make a very significant impact, accounting for around one referred visit in ten between January and April 2011, up tenfold for the equivalent period in 2010. In both absolute and relative terms, the incidence of visits referred to Europeana from other, non-blogging, social media sites remain small albeit with signs of growth.

	2010 Number of referred visits	2011 Number of referred visits	2010 % of all referred visits	2011 % of all referred visits
Blogging sites*	2,608	47,079	1.23%	10.80%
Facebook	971	6,658	0.46%	1.53%
Wikipedia	3,513	3,647	1.65%	0.84%
Wordpress	543	1,819	0.26%	0.42%
Twitter	265	1,112	0.12%	0.26%
Netvibes	1,038	915	0.49%	0.21%
Delicious	476	330	0.22%	0.08%
All referring sites	212,831	435,828	100.00%	100.00%

**Includesblogspot, WordPress, and other blogs*

The absolute numbers are possibly a better guide to understanding trends since the impact of search engine optimisation has vastly increased traffic directed via Google. At this stage of the game, it is definitely worth keeping a watching brief on social media, which look to become increasingly important in the future.



[5] Future analysis of Europeana users and information seeking

1. Determining whether use and information seeking behaviour stabilises after SEO and what it then looks like; in other words do we have a norm forming?
2. Establishing future growth limits
3. Watching and evaluating mobile use
4. Evaluating the initiative with Wikipedia and watching impact of social media more generally
5. Get closer to the Europeana users; develop user stereotypes and digital signatures using cluster analysis



Appendices

Appendix 1: Numbers of page-page transits – FROM row TO column (January-April 2011)

page views	Total	record	search	homepage	redirect	aboutus	login	myeuropeana	Other
Total	8018930	3118984	2282566	1084421	982394	151230	88985	71448	238902
OFFSITE	3139690	1953843	82527	936019	50573	17020	27393	25579	46736
search	2443442	617608	1790905	12276	5	3160	6094	4854	8540
record	1656664	528789	147425	14621	931560	3753	6166	4701	19649
homepage	416306	1380	204522	88818	80	44574	11382	11691	53859
aboutus	120494	6499	7151	6487	0	61887	9209	9550	19711
timeline	54324	7338	34232	1181	0	481	1079	812	9201
login	43294	0	31	8578	0	11375	16129	1604	5577
other	144716	3527	15773	16441	176	8980	11533	12657	75629

Appendix 2: Number of visits by referrer site and Europeana page type (January-April 2011)

Visits	#_Total	record	homepag	searc	redirec	aboutu	login	myeuropean	other
GOOGLE	138971	132190	65256	270	27	714	13	31	1498
UNKNOWN	435828	85918	310080	18763	3167	5457	1464	1089	9890
_KNOWN_OTHER_	95922	30607	56470	4652	1525	1507	33	4	1124
EUROPEANA	80264	18418	8935	18826	25006	3329	749	848	4153
BLOGSPOT	44544	66	44019	126	269	36	1	1	26
fbp.pionier.net.pl	9513	7450	455	1608	0	0	0	0	0
GOOGLEUSERCONTENT	7071	1799	480	769	3404	229	38	37	315
FACEBOOK	6658	3344	2343	557	18	127	15	1	253
YAHOO	3854	1641	1416	176	20	66	2	1	532
WIKIPEDIA	3647	288	3063	235	46	5	0	0	10
EUROPA	3544	1	3542	0	0	1	0	0	0
www.emob.fr	3493	0	3493	0	0	0	0	0	0
BING	2744	1087	1616	1	0	16	6	1	17
www.bnf.fr	2616	0	2616	0	0	0	0	0	0
scd-sfx.u-strasbg.fr	2577	0	0	2577	0	0	0	0	0
LIVE	1981	673	733	110	30	46	0	0	389
WORDPRESS	1819	175	757	14	850	21	0	0	2
www.nytimes.com	1521	1	1520	0	0	0	0	0	0
www.kb.nl	1294	0	154	0	0	1140	0	0	0
TWITTER	1112	339	440	129	0	111	2	1	90
www.heise.de	979	0	979	0	0	0	0	0	0
www.netvibes.com	915	47	834	11	0	15	0	0	8
www.bn.org.pl	808	0	808	0	0	0	0	0	0
www.digmap.eu	661	0	661	0	0	0	0	0	0
BLOG	577	35	444	91	0	6	0	0	1
roai.mcu.es	547	0	547	0	0	0	0	0	0
www.elgrancapitan.org	483	0	0	0	483	0	0	0	0

Visits	#_Total	record	homepag	searc	redirec	aboutu	login	myeuropean	other
www.service-public.fr	479	0	479	0	0	0	0	0	0
www2u.biglobe.ne.jp	416	416	0	0	0	0	0	0	0
www.lecdi.net	387	0	46	341	0	0	0	0	0
www.e-book.com.au	385	0	385	0	0	0	0	0	0
hispana.mcu.es	288	0	288	0	0	0	0	0	0
www.deutsche-digitale-	284	0	284	0	0	0	0	0	0
www.culture.gouv.fr	272	2	247	0	23	0	0	0	0
doucetpiquante2.canalblog.co	189	0	4	0	185	0	0	0	0
app.e2ma.net	177	24	134	13	0	0	0	0	6
BLOGS	139	3	125	3	0	1	1	2	4
www.photo.rmn.fr	41	4	0	0	37	0	0	0	0
tek.sapo.pt	26	0	26	0	0	0	0	0	0



Appendix 3: Number of visits by day of week and hour (January to April 2011)

Hour	Total	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	281425	35294	44203	42233	41387	49150	39071	30087
1	200307	26573	31873	30311	27693	34525	25964	23368
2	134015	17443	21915	19836	19069	22378	15985	17389
3	88700	11120	14252	12541	12777	15661	11590	10759
4	69324	8924	10854	9828	8779	12512	9626	8801
5	62883	8897	9102	8321	10226	9696	7869	8772
6	50980	7004	8672	7635	7715	8422	5380	6152
7	58684	7354	9433	8875	9964	9821	6121	7116
8	71065	9784	12450	11063	10652	13079	7129	6908
9	128324	16190	22431	21442	18929	24410	12526	12396
10	210548	29478	36187	35636	33046	40054	17344	18803
11	296377	43121	47145	46463	48144	55791	27371	28342
12	352409	51521	57565	55800	53913	61686	34476	37448
13	383322	50831	58912	58639	77132	62323	35733	39752
14	375831	54696	59692	59826	68863	60427	35182	37145
15	401033	59842	61721	63649	75055	67108	35526	38132
16	434432	63531	66941	65932	85692	71301	39753	41282
17	454163	70695	66823	66337	95019	69540	40279	45470
18	439534	68856	62581	64906	86138	62322	43667	51064
19	426393	66623	64578	61438	79593	57684	43500	52977
20	389558	60598	59790	55549	75972	46181	40353	51115
21	361463	58086	53600	51826	63837	47459	38582	48073
22	355501	54541	53107	50828	64560	48910	38604	44951
23	339556	52054	53033	47147	60284	48403	38254	40381
Total	6365827	933056	986860	956061	1134439	998843	649885	706683

Appendix 4: Number of visits by collection country (periods as defined)

Collection Country	Jan–Apr 2010	Jan–Apr 2011	Jan 2011	Feb 2011	Mar 2011	Apr-2011	Apr 2011
Did not View Record	346864	526574	130327	123599		131035	141613
FR	40563	665315	145021	222727		138216	159351
DE	18880	251261	20609	47179		73699	109774
UK	18978	146739	4526	17664		50363	74186
ES	3503	83925	11197	23423		19132	30173
NL	0	69362	3001	5794		29856	30711
IT	2738	68286	4495	10697		23483	29611
PL	9635	65040	4636	11019		19049	30336
NO	732	46635	1907	9080		12226	23422
SE	71	38754	2115	9082		10698	16859
BE	1115	30803	2746	5782		8112	14163
IE	457	29197	1121	7324		7130	13622
EU	2823	26496	3481	3925		7883	11207
SI	170	9928	1252	2087		2730	3859
AT	2596	8566	407	812		2426	4921
GR	346	8232	1109	1673		2535	2915
SK	0	7647	1154	2315		1792	2386
CH	2302	6491	951	1611		1668	2261
RO	1509	5260	737	1330		1322	1871
PT	0	4550	232	577		1512	2229
RS	0	4265	15	449		1100	2701
IL	0	2453	176	747		481	1049
DK	0	2064	119	128		646	1171
FI	136	1794	135	318		491	850
HU	0	1408	416	268		379	345
CZ	0	1355	96	101		436	722
LV	36	1239	46	138		427	628



Collection Country	Jan–Apr 2010	Jan–Apr 2011	Jan 2011	Feb 2011	Mar 2011	Apr-2011	Apr 2011
BG	0	1236	202	529		304	201
EE	0	1115	35	23		426	631
LU	1535	1039	539	182		137	181
LT	0	760	130	229		186	215
RU	0	633	59	78		87	409
CY	11	78	9	21		14	34
#_Total	455000		343001	510911		549981	714607

Appendix 5: Page views by user type (January-April 2011)

	Total	Heavy-user	Mobile	OneShot	User
record	3119768	175359	54654	664005	2225750
search	2283147	343932	16328	6039	1916848
homepage	1084422	103814	15597	98973	866038
redirect	982394	69426	7412	12339	893217
aboutus	92945	4606	1006	1250	86083
login	88985	25778	8	246	62953
myeuropeana	71448	25550	45	48	45805
timeline	30912	1789	9	43	29071
communities	25077	1405	5	177	23490
partners	23676	1758	46	378	21494
kindle_competition	18286	602	471	1692	15521
thoughtlab	15402	1203	18	51	14130
newcontent	13168	911	6	43	12208
aboutus_background	12120	888	5	69	11158
aboutus_demos	11647	589	6	32	11020
register	11371	318	0	3	11050
rr-f	11167	331	0	124	10712
aboutus_jobs	11060	694	30	416	9920
usingeuropeana	10323	844	109	92	9278
aboutus_faqs	7382	526	0	24	6832
contact	5573	420	347	111	4695
aboutus_enevs	5262	459	1	10	4792
termsofservice	4512	394	351	20	3747
europeana-contributors	4474	782	0	23	3669
communities_maps	4322	289	3	18	4012
reading_europe	4225	355	38	77	3755
aboutus_press	4196	514	1	24	3657
sitemap	4124	400	14	34	3676



	Total	Heavy-user	Mobile	OneShot	User
to_exhibitions:	4096	361	0	5	3730
communities_arnouveau	4056	313	2	18	3723
thoughtlab_digitallibrarycatalogue	3964	284	1	5	3674
communities_medieval	3654	265	6	13	3370
accessibility	3602	361	15	18	3208
rr-p	2986	121	2	59	2804
aboutus_jobs_developmentteammanager	2949	147	55	258	2489
browse-all	2645	675	0	0	1970
aboutus_jobs_juniordataingestionspecialist	2641	102	27	109	2403
communities_biodiversity	2627	271	0	9	2347
thoughtlab_semanticsearching	2388	326	2	14	2046
languagepolicy	2333	332	3	13	1985
advancedsearch	1994	431	0	3	1560
thoughtlab_digitisation	1872	248	0	24	1600
privacy	1835	272	6	10	1547
thoughtlab_usergeneratedcontent	1816	274	0	15	1527
storyline	1732	54	4	9	1665
thoughtlab_improvingmetadata	1700	281	1	3	1415
thoughtlab_enrichingmetadata	1633	287	1	20	1325
logout	1502	181	0	1	1320
thoughtlab_serviceinfrastructure	1213	226	1	3	983
using-europeana	1101	64	36	119	882
thoughtlab_linkedopendata	823	118	3	18	684
thought-lab	667	88	11	63	505
change-password	481	42	0	0	439
aboutus_jobs_socialmediamarketeer	477	18	3	15	441
disclaimer	468	194	0	1	273
aboutus_jobs_businessprojectsassistant	323	18	1	1	303
year-grid	304	16	1	26	261



	Total	Heavy-user	Mobile	OneShot	User
aboutus_press_signup	227	34	0	2	191
pd-usage-guide	173	30	3	10	130
rr-r	157	16	0	5	136
new-content	91	1	0	6	84
unknown	59	19	1	2	37
thinkvideo	37	2	0	5	30
bob	29	0	0	7	22
to_exhibitions: themes	27	4	0	0	23
Total	8020070	770412	96695	787250	6365713
