社会媒体对多媒体网络平台的推荐研究:基于 Europeana 的案例分析

Social Media Referrals on a Multi-media Platform:Case Study Europeana

David Nicholas and David Clark
(CIBER Research Ltd. Newbury, UK)

翻译整理:姜婷婷　迟　宇　史敏珊　(武汉大学信息管理学院)

Europeana.eu 是一个多语言网站,提供了数以百万计的来自于欧洲博物馆、图书馆、档案馆以及多媒体资源收藏的数字化条目。本文研究了该网站的使用情况,尤其是社会媒体对该网站的推荐。我们采用了三种方法搜集相关数据,包括 Google Analytics、ClickStream 日志及 http 服务器日志。本文特别展示了利用 Google Analytics 分析得到的结果。

研究背景

Europeana 于 2008 年推出网站原型,自 2010 年开始全面运行。作为一个欧洲数字资源的门户网站和搜索引擎,它为各语言用户提供了来自于 34 个国家、2200 家机构的 2 千 600 万件欧洲文化收藏品,其中包括图书和手稿、照片和画作、电视和电影、雕塑和手工艺品、日记和地图、乐谱和音像等等,应有尽有(图 1)。目前 Europeana 已经获得了人文学者、文物专家甚至是游客的频繁使用,访客总量已经达到了约 500 万。

CIBER 自 2009 年开始分析 Europeana 的使用情况,现在已经积累了长达 3 年的数据,用以评价该网站的发展、变化与革新,这些日志数据是我们理解并预测网站数百万用户的目标和动机的基础。本文报告了我们的最新研究成果,主要关注的是社会媒体对 Europeana 的推荐情况,即 Facebook、Twitter 等网站所带来的访问流量的大小和特征,这类网站有可能为 Europeana 带来可观的流量并促使人们反复使用 Europeana 的内容。这项研究的首要目的是要看看 Google Analytics 所提供的网站使用数据在稳健性和精确性方面的表现如何,同时将其与传统的数据来源(即 http 服务器日志和 ClickStream 日志)进行比较。

研究方法

为了对 Europeana 的使用情况进行最好最全面的分析,我们采用了多个数据源,它们各有一些优缺点。基于网络服务器的日志方法并不记录用户端所发生的事件,这可能造成 10～15%的流量无法追溯。我们发现 http 访问日志和 ClickStream 日志的符合程度是可以接受的,2012 年 6 月至 12 月间前者的页面浏览总数只高出了 1%,但这两个数据源与 Google Analytics 的差别比较大。Google Analytics 要求用户浏览器支持 JavaScript 并且用户是接受 cookies 的。2012 年 6 月至 12 月间 Google Analytics 和 Europeana 日志的页面浏览数据相差 26%,2011 年 1 月至 5 月间的差别更大。因而我们引入了异常值的概念:来自于单个 IP 地址的一系列页面请求,对于这样的情况,Google Analytics 一般都不予记录,因为这可能是自动的代理程序造成的,它们检索网络内容但不运行 JavaScript。我们在 2011 年的日志数据中识别出若干这样的代理程序,从而使其接近 Google Analytics 数据;我们也对 2012 年的 ClickStream 数据进行了类似的处理,通过对异常值的修正将其与 Google Analytics 之间的差别控制在可以接受的误差范围之内。值得注意的是,

[基金项目]
[作者简介]

Google Analytics 对 cookies 和 scripting 的依赖也存在问题，假如用户阻止了这两项功能，那么他们的行为将无法记录。

研究结果

Europeana 在 Google Analytics 中自行定义了社会媒体“高级群组”，包括 20 个推荐网站域名，其中 Facebook 和 Wikipedia 是社会性流量的主要来源，其次是 WordPress、Blogspot、Twitter，而 2013 年 Europeana 连续数周在主页上推广的 Pinterest 以较大的差距排在它们后面，当然还有的社会媒体网站在过去六个月时间内（2012 年 10 月至 2013 年 3 月）所带来的流量微不足道甚至为零，见表 2。Google Analytics 在“流量来源”下也提供了“社会性”分析，从这份报告的“网络推荐”部分来看，Google Analytics 对“社会性”的定义明显比 Europeana 的社会媒体“高级群组”定义或 CIBER 的修正扩展定义都要广泛得多。我们在分析时所使用的是 CIBER 对 Europeana 的社会群组进行修正后的扩展版本。

流量规模与增长

在过去的一年里（2012 年），Europeana 的 450 万访客中有 70% 来自于搜索推荐，而他们又几乎全部来自于 Google（97%）。Google Analytics 在 2011 年 10 月以前还没开始报告社会推荐情况，因此我们能够获取的数据时间段是有限的，但这在某种程度上可以通过日志数据来进行扩展。我们手头上的有限数据表明，社会推荐在 2011 年 10 月网站新门户发布时出现了一个小小的高峰，而那之后就保持在每周 1,000 次左右，并从 2012 年 8 月开始出现了不规则的增长，现在已接近每周 1500 次。在 2013 年 4 月的时候，社会推荐只占了网站所有访问量的百分之一，与一年前相比仅仅高了 0.02%。这可能是因为与 Europeana 相关的社会性活动只发生在社会媒体站点内部，完全绕开了 Europeana.eu。在这种情况下我们无法评价“社会媒体”的效力。

个人社会媒体

自 2011 年 10 月 Europeana 网站新门户发布以来，Facebook 以接近 30,000 次的推荐成为了最具影响力的社会媒体网络。根据 Google Analytics 提供的数据，来自 Facebook 的访客的“平均访问时间”略超过 3 分钟，比所有访客的平均访问时间 2.5 分钟要长。排在 Facebook 之后的依次是 WordPress（9,000 次推荐）、Blogger（4,200 多次）、Twitter（接近 3,300 次）以及 Netvibes（略超过 2,000 次）。当我们只考虑并比较相对稳定的秋季月份（2011 年及 2012 年的 9 月至 12 月）时，网站整体流量的翻倍与社会推荐的年同比增长无法匹配，因为来自 Facebook（2012 年接近 10,000 次推荐）和 Twitter（1,650 次）的流量仅增长了 12%。只有 WordPress 保持了与网站整体一致的增长速度，尽管它带来的流量只有 Facebook 的三分之一（2012 年 3,037 次推荐；年同比增长 162%）。比较有意思的是 Twitter，尽管推荐次数几乎没有什么增长，但是停留时间翻倍了，2011 年秋季的平均时间为 2.5 分钟，而 2012 年达到了 5 分钟。令人惊讶的是 Europeana 主页上特别推广的 Pinterest，在所有社会媒体中排名第 6，2012 年 9 月至 12 月的流量仅为 681，而较高的每次访问浏览网页数量（平均 12 个）以及超长的停留时间（12 分钟）可能是由网站内部开发或测试活动引起的，并不代表“真实用户”的行为。

对于整个 Europeana 网站来说，大部分入站流量（大约一半的非搜索引擎推荐）都直接指向了收藏品条目本身，另有 12% 指向“搜索”。有趣的是，如果只考虑社会推荐，有一半的入站流量指向了主页，而另有大约 17% 指向“搜索”。Google Analytics 提供的非正式“回溯”分析表明，大部分的社会性流量可能来自于从事数字人文及相关领域研究的人员，而这些业内人士并不具有代表性。我们将“My Europeana”这样一项个人定制功能也划归入“社会性”范畴。2012 年 6 月至 12 月间，记录显示共有 1,400 位用户进行了 ID 登录，即使是最繁忙的 11 月也只有不到 300 位。

种种迹象表明，社会媒体并没有推动 Europeana 的增长，也没有表现出这样的趋势。当然，所有这些社会媒体的行动与“黄油配面包”的搜索引擎推荐相比都是微不足道的。这不仅是相对 Google 而言的，比如波兰的聚合站点 pionier.net.pl 就带来了 37,000 次（5%）推荐，高于 Facebook 的 29,000 次（＜4%）。Eu-

ropeana 的一项研究显示,这个波兰合作伙伴通过 API 成功地向 Europeana 贡献了流量。

国家分析

令我们有点惊讶的是,向 Europeana 贡献流量最活跃的国家是西班牙。在最近的6个月中,根据 Europeana 自己定义的社会媒体高级群组,8.8%的社会性流量来自于西班牙,美国只排在第二位(7%)。考虑到美国的人口规模以及人们对社会媒体的接受程度,这是我们始料未及的。

社会行动与社会媒体

为了研究来自于社会媒体的用户是否更愿意分享,我们首先必须定义"愿意分享"。在 ClickStream 日志中,'SAVE_SOCIAL_TAG'动作在 2012 年 6 月至 12 月间(我们只有这段时间的 ClickStream 日志)只发生了 189 次,相对于 960 万次的对象页面访问、480 万次的搜索结果显示以及 160 万次的主页浏览,这几乎可以忽略不计了。如果采用 Google Analytics 中"社会性插件"分析,这一数量依然比较低,但会相对较好一些,2012 年 9 月至 2013 年 3 月间共发生了 3,945 次"特定社会性动作"。如果用 Europeana 自己的社会媒体高级群组加以限定,"特定社会性动作"减少为 291 次。这样一来,来自于社会媒体的访客更倾向分享,分享率为 0.68%。当然,社会媒体用户会进行分享是因为他们知道这就是社会媒体的主旨。

研究结论

社会媒体的使用是一个复杂的研究领域,受困于人们对社会媒体的识别、定义、好奇心及阐释等问题。鉴于 Europeana 规划人员对社会媒体的重要性的认识以及人们对该领域的研究热情,我们还需要进一步研究以下问题:是什么导致了社会媒体只给 Europeana 带来了相对较低的流量(约为总流量的 1%)? 为什么社会媒体的使用相对来讲没有发生增长? 社会媒体是否从更乐于分享的用户那里创造了更有"质量"的流量? 我们对 Europeana 数据的"离岸"研究具有什么意义,比如放到 Facebook 这样的网站中来研究? 尽管社会媒体的本质是分享,但分享活动本身却并不常见。

Social Media Referrals on a Multi-media Platform: Case Study Europeana

David Nicholas and David Clark

(Professor David Nicholas, Director, CIBER Research Ltd, I Westwood Farmhouse, Newbury RG14 7RU;
College of Communication and Information University of Tennessee, 1340 Circle Park Drive, 423 Communications Bldg Knoxville, TN 37996-0341, USA)

Abstract

The article examines the usage of the Europeana. eu website, a multi-lingual online collection of millions of digitized items from European museums, libraries, archives and multi-media collections. It concentrates on the social media referrals to the site. Three methods for obtaining the data are examined: Google analytics, ClickStream logs and http server logs. The analyses produced by Google Analytics are highlighted in the article.

Background

Europeana, launched in 2008 as a prototype and operating as a full service since 2010, is a gateway, portal or search engine to the digital resources of Europe's museums, art galleries, libraries, ar-

chives and audio-visual collections (Figure 1). Europeana is regarded as trusted (curated) source connecting users directly to authentic and curated material. It provides multilingual access to 26 million European cultural objects in 2200 institutions from 34 countries. Books and manuscripts, photos and paintings, television and film, sculpture and crafts, diaries and maps, sheet music and recordings, they're all there. Europeana claim that there is no longer the need to travel the continent, either physically or virtually. If you find what you like you can download it, print it, use it, save it, or share it (http://www.europeana.eu/).

While Europeana is essentially a portal it also has aspirations well beyond that; it believes it can help stimulate the European digital economy; it also mounts online exhibitions and takes part in crowd sourcing experiments (World War 1 is currently the subject of such an experiment). Europeana is also working with other digital channels to distribute their content, most notably Google, Wikipedia and Facebook.

It is a site that currently attracts around five million visitors andis used heavily by humanities scholars, heritage professionals and even tourists. CIBER have been analysing usage of Europeana since 2009 and have now amassed a three-year long-series of data to evaluate Europeana's growth, changes and innovations. As a consequence we have assembled a large evidence base showing how a whole range of people use cultural collections and artefacts, in a virtual environment. Thus we use logging as the basis of insight and prediction about the purpose and motive of the millions who use Europeana.

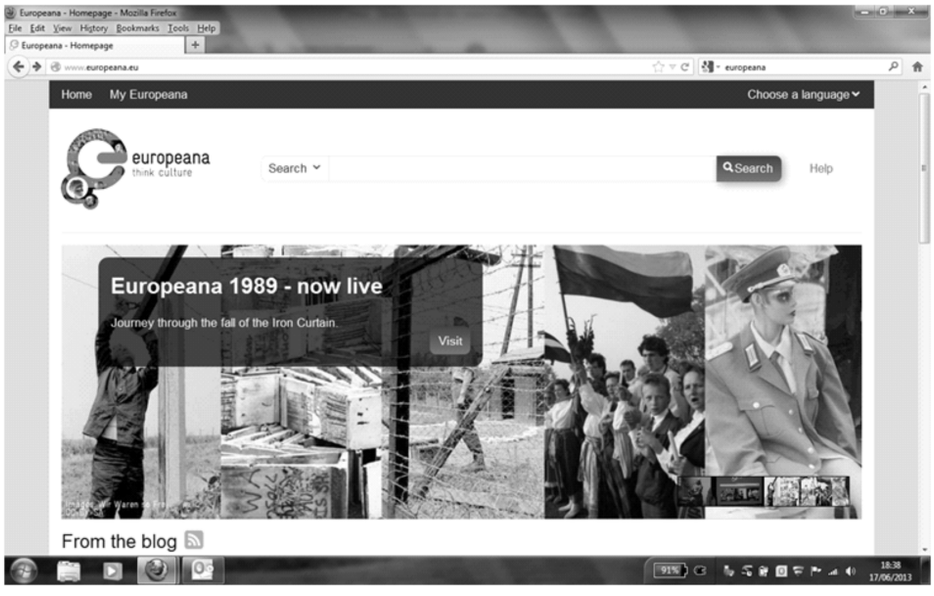

Figure 1 Europeana home page

Aims and objectives

The study reported here features our latest research which focuses on the social media referral: volume and characteristics of the traffic coming from Facebook, twitter and the like, which could potentially drive a lot

of traffic to Europeana and encourage much re-use of Europeana content. It is a type of digital behaviour prevailing in Europeana which we regard to be particularly significant and strategic forms of digital behaviour not only for Europeana, but for all information providers on the Web.

A prime objective of the study is to see what Google Analytics could provide in regard to robust and precise usage data and how it compared with our traditional usage sources – http server logs and ClickStream logs.

While textual websites, like those of scholarly publishers and libraries, have been well researched from a digital usage point of view, very few studies of multi-media platforms have been undertaken, and so in the way the paper is quite unique.

Methodology

For CIBER's earlier Europeana work we relied upon server http request logs using CIBER's own 'deep log' methods[注1]. However, for the study reported here we wanted also see to what extent, the now ubiquitous, Google Analytics (GA) data could undertake key information seeking analyses more cheaply and effectively. This is important given the fact that Europeana, like many organisations, are relying increasingly heavily on GA for all their usage and marketing needs. While we have utilised GA heavily in this paper as we will learn GA cannot always supply the data required in a convenient form and have thus supplemented it with our own tried and trusted deep log methods. There is great potential to make better use of GA but it requires considerable investment and effort, not only to interpret the output but in experimental design, preparation and configuration of event tracking code, and this is generally not undertaken by institutions and analysts.

In addition, to the http logs and GA data, we also had access to a series of ClickStreamer logs which had become recently available. However, we only had access to the ClickStreamer series of Portal logs from June to December 2012 (the minimal time-scale necessary for a robust analysis given the seasonal/diverse nature of usage data). As a result we have sometimes used the old series of raw http-request logs for a broader overview and perspective.

Thus, to provide the best and most comprehensive analysis of Europeana usage we have used a variety of data sources. And it is worth pointing out their various strengths and weaknesses. There are, in essence, three points at which we can take the pulse of a website. On receipt of a request by the server; by tapping into the internal process of the site's content management system (CMS); by causing the browser to send an acknowledgement when content is received. The first of these, monitoring incoming traffic, has been used since the web's inception. It relies on http server request log files originally intended for server management and software maintenance. Not being intended for market research purposes, means the record is not always in the most convenient form. On the other hand it may hold information that would not otherwise be collected because it did not seem relevant at the relevant at the time.

Figure 2 outlines the web-server process and the points at which usage can be measured. For a very simple website with no CMS the URL requested (e. g. a link in the clickstream) maps more or less directly to a web-page file, which is despatched by the server back to the client (browser). In this case the traditional server log is in effect also the CMS log. But today, CMS is the norm and the request no longer maps direct to a file but is interpreted by the CMS. As a result records are retrieved from a database and a web page constructed on demand. The cost of this flexibility and complexity is that the incoming request is no longer a straightforward and reliable indication of what was served in response. Interpretation of request logs becomes a matter of 'reverse engineering' the programming of the CMS. In such cases logging from within the CMS becomes attractive. For some purposes this is obvious and inherent to the application area: an online shop for example will almost certainly be linked to stock control, and accounting records. These can be considered specialised varieties of 'log file'; they can be used for analysis in similar ways to the server log. Or, a specific form of log may be kept for market research and data mining. For any special logging the problem is to specify in advance what needs recording.

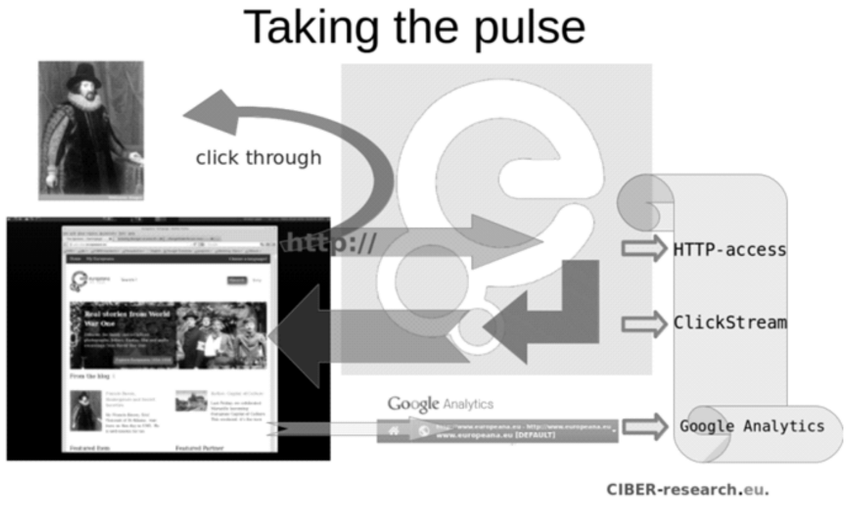

Figure 2 Taking the pulse of a website using logs

The difficulty of web-server based logging, wherever the monitoring point, is that it does not record what happens at the user end. A web-page is served but there is no record of its receipt. The solution is to insert scripting into the web-page so that on receipt a secondary request is despatched to report back to a logging system. This is the method employed by Google Analytics and others similar solutions such as the open source Piwik. This can, like CMS logging, resolve the 'reverse engineering' problem, but the task of deciding what to track and of deploying the necessary web tracker 'events' to best effect remains. It also needs to be noted that this approach depends on the end-user accepting and not deleting the tracking cookies and scripts. Our research suggests that for this reason significant traf-

fic, perhaps 10～15%, may be untracked by such browser based methods. This could make a big impact on some analyses, especially those regarding relatively lowly used activities and behaviours.

Taking measurements at various stages of what should, in principle, be a single transaction, raises the problem of reconciling the various accounts. Even if the numbers do not agree we should be able to account for differences. The agreement between http-access log and ClickStream is acceptable: over the period June to November 2012 the http-access log shows a page view count higher by 1%. However, as we shall learn in greater detail, agreement between either of these sources and Google Analytics is much harder to establish.

Google Analytics depends on JavaScript being active on the client browser and the acceptance of the Google cookies. Without JavaScript the logging data will not be recorded. Without the cookies it is not possible to identify returning visitors, nor gather reliable information about the sequence and timing of page views. Based on the six-month ClickStream series between 15～30% of visits have a Google cookie set when requesting the landing page, this implies a previous visit to Europeana and retained cookies. For visits comprising more than a single page, the GA cookies are present in 85～90% of page views, thus we think it is highly probable that the remaining 10～15% (possibly one in six visits) have blocked cookies and possibly JavaScript and would not therefore be tracked by Google Analytics.

Unfortunately this estimate of 10～15% untracked visits by Google Analytics does not account for the massive gap between the page views reported by GA and those from the Europeana logs which is 26% in the period June-December 2012. In only one month (September) is the difference (16%) low enough to be plausibly attributable to user blocking of GA. For June the figure is 54%: some further explanation is required.

In the period January-May 2011 a much greater mismatch of page view counts between Google Analytics and the http-access log was observed: the uncorrected figure exceeding 250%. In that case we introduced the concept of an outlier: a series of page requests from a single IP address, often over many days, far too numerous to be the efforts of a single user. Thus the 'visitor' displays all the characteristics of an automated agent or robot bar the user-agent identifier. It could be a cloaked robot. Significantly, such cases tend to go unrecorded by Google Analytics as automated agents retrieve web content but do not run JavaScript. In the early months of 2011 identifying less than a dozen such agents was sufficient to bring the logs and GA into near-enough agreement. A similar process can be applied to the 2012 ClickStream series. For example in August 2012 8.2% of all page-views originated from a single IP address located in Beijing. China has a large population, they may have a considerable interest in European culture, the single IP address could be a proxy for many individual users; on the other hand such heavy and sustained use does not display the irregular pattern of use expected of normal users. If an outlier correction is applied then the difference between GA and the ClickStream data can be coerced into an acceptable error band.

In sum: Google Analytics' reliance on cookies and scripting is effective in suppressing the effect of

cloaked robots and other automated agents that would distort the profile of a normal sentient user; but the same feature will also miss genuine users who have blocked cookies and JavaScript.

Results

With so muchEuropeana (and scholarly publisher) planning (and hopes) resting on social media use for growth and re-use it is worth first pointing out that there are substantial problems in defining 'Social Media', which need to be clarified in order to make a fair and accurate evaluations and comparison of growth rates and contribution to overall traffic.

The Google Analytics 'advanced segment' for social media, aspersonally defined and used by Europeana, contains 20 sources (referrer domains), some of which have registered insignificant or even no traffic at all during the last six months (October 2012 — March 2013) See Table 2. The major sources of social traffic are Facebook, and Wikipedia; there is also significant traffic from WordPress, Blogspot, twitter and, a considerable way behind, Pinterest, the latter being publicised on the Europeana homepage for many weeks during 2013. We shall return to individual performance later in this section, here we shall confine ourselves to the problems of definition.

Table 2 Social segment definition(GA)

Social	Page views
facebook. com	16928
tweet	0
LinkedIn	361
YouTube	42
reddit	167
digg	14
delicious	83
stumbleupon	0
Flickr	222
MySpace	0
hootsuite	28
retronaut	167
Wikipedia	10882
bit. ly	0
tinyurl	0
t. co	9993
wp. me	0
blogspot	4044
wordpress	4253
Pinterest	1265
sum	48449
intersect	44085
Oct-Mar	3473308

An interesting definitional case isTwitter. The Twitter traffic is identified by "include Source containing 't. co'". Patently, this is too loose a definition as it will not only pick up 't. co', but any domain containing that sequence of characters e. g. search. bt. com. The result is that that the number of visits captured by this method is at 9,993 (for the most recent six months) four times greater than the actual number of visits from t. co (Twitter). The true total of social sources (40,791) is inflated by 19% (48,449). The overall effect on the visit count for the social segment is to some extent mitigated by the fortunate chance that the loose 't. co' rule will pick up blogspot. com which is already included by its own rule. The problem can be fixed by replacing the rule "include Source containing 't. co'" with "include Source Exactly matching 't. co'" or with "include Source Matching RegEx ^t\. co".

Table 3 Social segment blogs (selection only). Source:GA

	Oct 1,2011—Mar 31,2012	Oct 1,2012—Mar 31,2013
agioritikesmnimes. pblogs. gr	64	0
bazoga. over-blog. com	43	0
bgpw. blog. pl	21	21
blog. bnf. fr	21	21
blog. crdp-versailles. fr	43	0
blog. daum. net	21	43
blog. euscreen. eu	0	64
blogs. ec. europa. eu	21	21
blogs. helsinki. fi	21	0
blog. sina. com. cn	0	64
blogs. law. harvard. edu	64	0
blog. slub-dresden. de	21	0
blogs. sch. gr	43	43
boqo. over-blog. com	21	0
cblog. culture. fr	43	0
christypato. blog. br	21	0
deal. blog. kazeo. com	0	128
deal. blog. mongenie. com	213	0
digiblog. hu	21	0
enfinlivre. blog. lemonde. fr	43	0
estudamais8. blogs. sapo. pt	21	0
eueublog. wordpress. com	0	64
fablog. iransalamat. com	43	0
formacion. universiablogs. net	43	0
googleblog. blogspot. com	21	0
kluwercopyrightblog. com	21	0
konzervativci. blog. com. mk	21	0
leblog-ffg. over-blog. org	43	21
libblog. ucy. ac. cy	21	0
pisani. blog. lemonde. fr	43	43
sog. blog. so-net. ne. jp	0	21
somewhereinblog. net	43	43

Blogspose definitional problems too (Table 3). The social segment includes blogs but only those from WordPress and Blogger. There are many other blogs hosted elsewhere that are not included. On the other hand treating all referrals originating from a WordPress or Blogger domain may be too broad a definition of a blog. WordPress in particular is a popular hosting platform for photographers' and artists' galleries. No method of classification will be entirely satisfactory but on balance we think the 'social' classification should be broadened to include any domain containing the subdomain 'blog'. or 'blogs.', but excluding blog. europeana. eu. The result is that another 1,085 visits can be added to the social segment.

Google Analytics provides under "Traffic Sources" a "Social" analysis. Looking at the "Network Referrals" section of this report it is clear that the GA definition of 'social' is again far broader than either Europeana's own 'Advanced Segment' definition or the corrected and extended version used by CIBER. How many networks are included depends on the period of the report: for March-April 2013 it includes 48, Jan—May 2012/13 includes 78 etc. The definition is as long as a piece of string and makes social network behaviour very difficult to delineate.

To conclude there are three 'social' definitions at work here: Google's, Europeana's social segment and CIBER's own expanded version based on a corrected version of the Europeana social segment and used for the following analyses.

Size and growthin traffic

To place social media referrals in context it is worth first looking at all referrals. Seventy per cent of the 4.5M visits to Europeana in the past year(2012) were search referrals, nearly all (97%) from Google. By contrast, runner-up Bing accounts for just 0.5%. Eighteen per cent of visits originate as links from other sites, 11% are direct —typed-in or bookmarked— and campaigns (newsletters etc.) contribute a little over one per cent.

Google Analytics was not reporting social referral before Oct 2011, sothere is a limited time series, which we can to some extent enhance with log data. The limited data we have show that there was a slight peak in social referrals around the time of a new portal launch in October 2011 (thanks to associated publicity one presumes), but after that it settles down to around 1,000 per week; since August 2012 there has been some irregular growth and the base-rate is now nearing 1500 per week. Between Oct-March 2011/12 and 2012/13 the overall year-on-year general visitor growth is 90%. However if we look at the 'social segment' the visitor growth is 34%. Exclude blogs and visitor growth falls to 25%. Looking at blogs alone [the visitor growth rate is 58%. The social element is a little more significant on the exhibitions site and predictably significant for blog. europeana.

In April 2013 Social Referrals only accounted for one per cent of all visits to the site, a bare 0.02% higher than a year previous. It could be that Europeana's social media activity takes place solely within the context of these sites and entirely by-passes Europeana. eu. In such acontext we cannot re-

fute claims for the efficacy of 'social media', nor can we support them. In the context of the Europeana. eu web-site however social referral is not at present significant and is not growing above the trend for the site as a whole. So the action has to be happening elsewhere, on the social media sites themselves.

Individual social media

The dominantsocial media network is Facebook with nearly 30,000 referrals in the year since the new portal launch (Oct 2011). The 'average visit duration' of these Facebook sourced visitors is, according to Google Analytics just over 3 minutes. Although 'average' is a poor single metric to use in this context —the distribution being log-normal— the duration is slightly higher than the 2.5 minute average for all visitors. So more dwell time for social media users, but not really sufficient to build a strong case for more committed users, and anyway see our earlier comments about the problems of using dwell time in isolation as a metric.

Facebook was followed by WordPress in popularity, nearly 9000 referrals, Blogger (over 4200), Twitter (nearly 3300) and Netvibes (just over 2000).

When we consider and compare only the relatively stable Autumn months (Sept - Dec, 2011 and 2012) the overall doubling of traffic on the site is not matched by a corresponding growth in social referrals year on year: Facebook (nearly 10000 referrals, 2012) and Twitter (1650 referrals) traffic in particular shows only a 12% increase in visits. Only WordPress, with only a third of the Facebook traffic (3037 referrals in 2012; 162% year-on-year growth) has kept pace with the overall pace of the site. However, Twitter is an interesting case because while there is little growth in referrals, dwell time has in fact doubled. The average for Twitter was 2.5 minutes in autumn 2011, 5 minutes in 2012. Pinterest, Europeana's latest social media venture, a content sharing service that allows members to "pin" images videos and other objects to their pin board, currently featured on the Europeana homepage (and so attracting considerable publicity), surprisingly perhaps comes in at 6th in the social media ranking, with a light traffic flow (681 visits Sep - Dec 2012). The high number of page views per visit from Pinterest (average 12) and very long dwell time (12 minutes) suggest 'unreal user' activity, something odd is happening here. We suspect, as this feature on the home page is quite recent, that this may be internal development or testing activity. This should be checked, otherwise a false impression might be provided.

We can contrast the traffic flow for the siteas a whole with the flow of social media visits using Google Analytics. For the site as a whole most inbound traffic goes direct to a record (about half of all non-search engine referrals) and twelve per cent to 'search'. Interestingly, for social referrals half the inbound traffic goes to the homepage and around seventeen per cent to 'search'. An informal analysis of 'trackbacks' provided by Google Analytics suggests that much of the social traffic may be by people involved in development or research in digital humanities and related fields, not a very representative

group: insiders. During the period, 30 Dec 2012 - 29 Jan 2013, when there were 6,628 social referrals, blog.europeana.eu had 8,000 visitors. It is probable the blog users are already familiar with Europeana in which case it is probably not bringing in many new users.

'My Europeana', a personal/customised facility, may also be considered to belong in the 'social' category. Between June and December 2012 a total of 1400 users were recorded as having logged in with a userid, less than 300 even in the busiest month of November. Though a few users appear to log-in and view many hundreds of pages in a month there is little evidence of regular and sustained use of 'My Europeana', less than 50, a tiny amount, have used the feature for three or more months in seven. The majority appear in the record one month only during which they view less than forty pages. Overall logged-in users account for one-half per cent of all page views.

Social media, then, is not driving Europeana growth, and unlikely to do from the evidence we have to hand. The example of Pinterest is illustrative. Consider the featuring of a link to Pinterest on the Europeana homepage: this would appear to be of net benefit to Pinterest. Europeana has over 2,000 'followers' on Pinterest and over 600 'pins', but referrals back to Europeana during the last four months of 2012 amount to 680. The big question is what, in the context of Europeana, is social media for? Should we expect it to drive traffic to Europeana, or is Europeana the glue layer that enables Pinterest to be a showcase for Europeana's provider institutions? There is scope for a more comprehensive research programme in this area, linking together the traffic analysis of the Europeana web-presence (including blog, exhibitions, API) with similar data drawn from Europeana's providers.

Of course, all these social media initiatives are insignificant compared to 'the bread and butter' search engine referral; and not just via Google: pionier.net.pl the Polish aggregator site brought in 37,000 (5%) referrals, compared to facebook.com 29,000 (<4%), with all its millions of members. A study by Europeana (http://pro.europeana.eu/pro-blog/-/blogs/1660413) shows that API use by the Polish partners is proving very successful in sending traffic to Europeana.

Most of the social media (narrowly defined) traffic appears to flow into the home page rather than to specific items. This is in marked contrast to referrals from blogs which are more often to a specific page.

Country analysis

First a note of caution: determining the user's location is only approximate and, particularly when looking at the standard Google Analytics report, language choice and country are not the same: 'Language: en-us' is not the same as 'Country/Territory: United States'. The language indication is merely the default setting of the browser and cannot be relied upon. Location, which is based on IP address allocation, can also cross borders.

Taking this into account it is still somewhat surprising to find that the most active country for social media traffic to Europeana.eu is Spain. In the most recent six-months Spain accounts for 8.8% of

social media traffic as defined by Europeana's own Social Media advanced segment, the USA is second (7.1%). Taking into account the much larger population of the USA and the mature state of social media uptake there this is unexpected. However, as we have already observed Social Media accounts only for 1 per cent of visits and visitors, so the statistics are likely to be unstable and be perturbed by factors which can be difficult to identify.

Social actions and social media

In order to find out whether users coming from social media are more likely to share(thought to be a positive by information providers) you first have to define 'likely to share'. The clickstream logs show negligible use of the 'SAVE_SOCIAL_TAG' action. For the period June - December 2012 (the only period for which we have clickstream logs) the action occurred 189 times. Set against 9.6million accesses to object pages (FULL_RESULT_HTML) and 4.8million presentations of search results (BRIEF_RESULT:search), and 1.6 views of the homepage (INDEXPAGE) it is clear that not much sharing goes on; so insignificant that we need to look for another definition of 'social media sharing'.

If we turn to the Google Analytics equivalent, 'Social Plugins', the numbers are still low, but better: September 2012 - March 2013, 3,945 'Unique Social Actions'. Set against the 3.4million visits in that period a social sharing action occurs at a rate of one per 866 visits (0.12%). When that report is restricted using Europeana's own 'Social Media advanced segment' the number is reduced to 291 'Unique Social Actions'. There is indeed a greater propensity to share by visitors coming from social media: a rate of one per 146 visits (0.68%). But the actual numbers are very small, in fact of the 142 Social sharing sources used by all visitors only three —Facebook, Google+, and Twitter— appear when the report is restricted to 'social segment' referrals. One reason for this may be that the 'advanced segment' has been defined too narrowly— inputs should match outputs; all the social sites recorded by Google Analytics as 'social sources' should be included in the segment. The alternative is to restrict the Social Plugins report to match the advanced segment. In that case the 'all users' figure declines to one in 1,104 (0.09%) [the Social segment is, of course, unchanged at 0.68%]. So, users coming from social media are more likely to share. However there might be a strong element of autocorrelation here, a tautology: social media users share because that is what social media is about.

Conclusions

For the very first time CIBER has been able to evaluate Europeana usage by all the available quantitative methodologies: deep log analysis, ClickStreamer logs and Google Analytics. In fact we believe this is the first time the three methodologies have been employed in regard to usage of a single website. We were especially interested to find out whether Google Analytics' popularity is matched by its capabilities and this article produces many useful GA derived analyses. GA proved to be a very useful usage tool, albeit one which sometimes underestimates usage, and also one which needs careful calibration

and interpretation to obtain full benefits.

Of course, using multiple sources of data has a downside as it highlights differences and divergences which need to be resolved. Considerable effort has gone into ironing out the resulting confusion caused. If you only have one clock you either trust the time it tells, compensating for known errors, or do without. If you have two clocks that tell different times, you cannot trust either: you know less not more.

In respect to the resultson social media, taking Europeana's definition ('social segment') the overall year-on-year visitor growth is 34%, compared to an overall visitor growth for European of 90%. Exclude blogs and visitor growth falls to 25%. Looking at blogs alone the visitor growth rate is 58%. Social media use is a complex area which is bedevilled by problems of identification, definition, novelty and interpretation. Given the importance accredited to it in Europeana planning circles, and the passions typically associated with it, there is a need for a detailed investigation to discover why it has driven relatively low volumes of traffic towards Europeana (around 1% of all traffic), why usage is not growing relatively speaking, whether it is generating more 'quality' traffic from users with a greater propensity to share and what significance can be read into use of Europeana data 'offshore', on sites like Facebook. There is a greater propensity for social media to share, but the activity itself is very uncommon.

注释

1 Nicholas D, Clark D and Rowlands I. Information on the go: case study of Europeana mobile users. Journal of the American Society of Information Science, 64(7), 2013:1311-1322

(收稿日期:2013- -)