Research 2.0: how do we know about the users that do not tell us anything?

Patrick McAndrew, Steve Godwin, Andreia Santos

The Open University

Institute of Educational Technology
Walton Hall
Milton Keynes
MK7 6AA

{p.mcandrew, s.j.godwin, a.i.santos}@open.ac.uk,

The OpenLearn initiative at the Open University (http://www.open.ac.uk/openlearn) offers free and open access to online material across a wide range of subjects. This material has been placed in on online environment based on the Moodle learning environment together with additional tools for communicating with other users and creating knowledge maps. One of the design aims of the initiative was to be low barrier to access so all content is available without registration, though some tools and features will only work once registered. The result is that we are seeking to research a site that is publicly accessible and has a majority of users that do not identify themselves, many of whom spend a short time on the site. As a further challenge the content itself is openly licensed using Creative Commons (http://creativecommons.org/) and can be taken and relocated on mirror servers, or accessed remotely through content feeds. The initiative has had to face this challenge and implemented a mixture of tracking, simplified surveys and the gathering of interesting stories. This approach has enabled us to spot interesting trends while remaining unsure about many of our users and their aims. The methods that we find we are using indicate a new style of research that can be related back to Web 2.0 as Research 2.0.

Research challenge

Working as a researcher in a mobile or open environment can mean accepting a loose and remote connection with end users. In an initiative such as OpenLearn, where the aim is to provide free and open access, the definition of who is the user is problematic; our users can be anyone. From a research point of view, the challenge was that, while it would be good to know as much as we could about those users, it was clear that in such an open system we could not expect everyone to provide us with all the data [ref something on user sampling]. With more than 2million unique users over two years of operation we also needed to accept that we would not be able to handle full data from all users in any case. Our solution to this problem involved breaking down the users into categories, and adopting different strategies for those different categories; building up broad pictures to help us understand results; and, accepting that we were in a dynamic changing research world with realistic methods and aims.

User categories for open research

As a starting point for our research approach we planned an approach where we did not see all users as equal. We developed a three-level approach to studying our users seeing them as enthusiasts, registered users and visitors. These three categories can be considered further as:

The **enthusiasts** are those who are prepared to tell us what they do. For some of these enthusiasts we have found that simply giving the avenue to report back data to us has enabled us to capture stories and investigate new ways to use OpenLearn. In other cases enthusiast activity has come to our attention through contacts, the impact of their work or through self-descriptions in blogs that mention our site. To help us to do this we have established watches against mentions of our own site using Google's Alert feature (http://www.google.com/alerts), and carried out scans of activity through the technorati blog analysis site (http://technorati.com/).

For **registered** users we can identify both their activity on the site, through logs in the Moodle system, and we request during registration that they indicate to us if they can be approached for research purposes. In practice about half of those who register on the site give this permission and can then be approached with requests to supply additional information through surveys, email interview, or direct contact. While after over a year of operation there are more than 60,000 registered users of OpenLearn, these represent less than 3% of the overall users as measured by software tracing machine access.

Visitors are the general users of the site, including the 97% of our users for whom we have no direct measure of their activity and are only left with the tracks left from IP addresses and search engine hits. These are crude tools but should not be ignored in analysing use (Harley & Henke, 2007). In the case of OpenLearn, both web analytics software and custom software created to work with the Moodle learning environment logs helped us monitor overall behaviour of visitors.

Spotting patterns in user behaviour

The three-level approach gives us scope to address a variety of research questions and then to shape how our site operates. For example one possible question is "How are people using the OpenLearn site?". OpenLearn does not apply a value-judgement in the access to the content, however content was designed with expectations about user engagement and it is easy to take the view that those who have spent more time on the site, registered and carried out tasks that leave evidence have gained more than those who visit once and do not return.

The evidence that we can use to address this comes first from considering visitor data. For example we can see from our analytics data that over a six month period (February-July 2008) we had approximately 60% of visitors who viewed only one page on the site. The proportion of such "bounce visitors" was highest if users arrived from a search result (66%) and lowest from direct access to the site (50%). So those who arrive from a search and may have no expectations of the site could well have found the simple answer to some questions, but we also suspect that the site needs to do more to appeal to this sort of visitor.

The next stage is to investigate through registered users actions. Using software developed in the initiative log data stored by the Moodle learning environment is converted into traced visits depending on machine address. The software then enables overall trends to be calculated and also visits to be examined. For all visitors machine addresses indicated by IP (Internet Protocol) numbers provide an approximate way to track continuous usage, for registered users their coded identity allows more reliable tracking while they remain logged in. Log data only enables visit times to be estimated based on the time between page impressions from the Moodle server. A conservative measure is used so that if the user only visits one page no time on site is recorded, though the user could have spent much longer reading that page

Examining the data for registered users showed that there was a distinction between those who spend a significant amount of time on the site and those who visit quickly. Figure 1 below illustrates this for a particular sample over a 6-week period and is indicative of the overall shape of use. The majority of users (more than half of all the registered users in this sample period) spend less than 30 minutes in total using the site, however a significant minority (nearly 10% of registered users) have spent more than four hours on the site. We can expect both the experience of these users and their willingness to report details to be different.
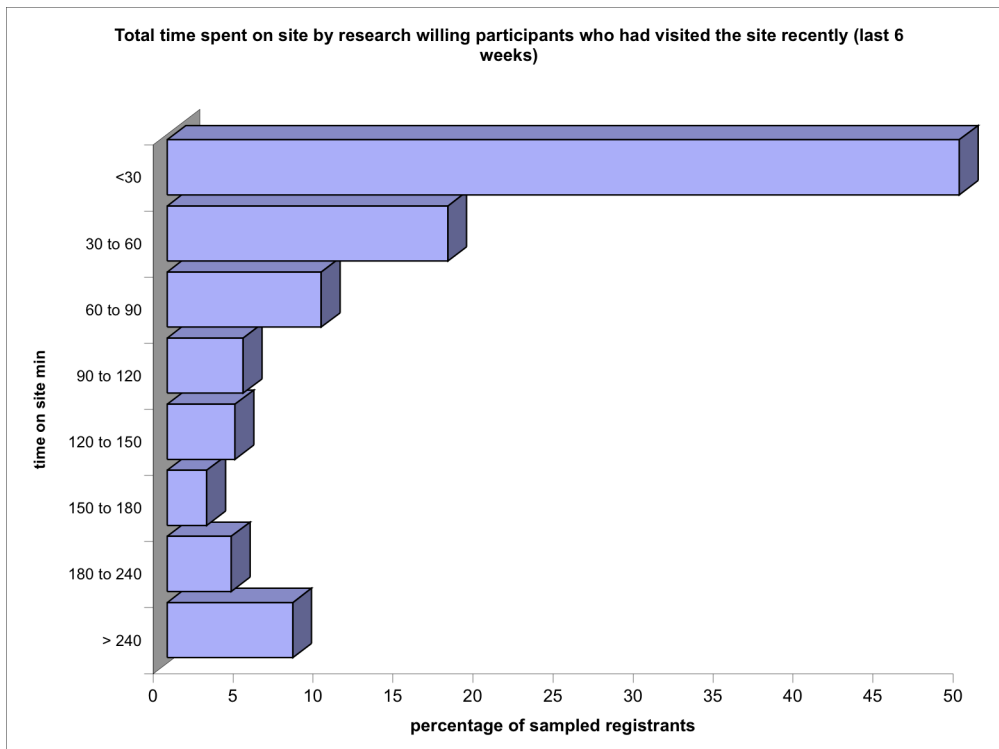

Figure 1. Pattern of use of OpenLearn – based on 6-week sample of registered users.

The logged data enabled a division in the sample to target the lower end users from amongst those who are registered with a survey and those with higher time on site users with a more extensive questionnaire and follow up questions. For some of the registered users we also have evidence of their engagement and use of the site through the artefacts that remain after they have made forum entries, enrolled in units and posted to their own learning journal. Clustering the data returned from the questionnaires has enabled us to propose further types of users that are using our sites, for example classifying some as "volunteer students" and others as "social learners" (Godwin & McAndrew, 2008).

We are now reviewing what it means to be an OpenLearner, and so offer greater reason to register with the site and prepare to more fully use the opportunity to learn. Features associated with social activity are being explored that will separate out the dependence on subject-based content. In particular all users can record interests, collect things that interest them, spot what others like to do. To support this we have added in a personal view of the site, tagging of content and personal interest, and a record of the user's own actions. These changes are primarily designed to further lower the barriers to use and increase the value of the site to learners. However they have the secondary benefit of making user actions apparent to the researcher so that we can understand where interests lie and the paths that users take through content.

Learning from the enthusiasts as lead users

OpenLearn offers a "LearningSpace" where we expect users mainly to access the content as learners. At the same time we are giving permission to users to work with the content in any way they wish. This was made explicit in the provision of a separate "LabSpace" with extra facilities and the invitation to users to make changes to the content. What we did not expect was how innovations in use would take place away from our own site. Two examples from outside the OpenLearn team are the extraction of our content for reuse in distributed CDRoms/DVDs to provide local personalised learning environments in remote parts of the world (Esslemont, 2007) and the transfer of OpenLearn content through RSS feeds into other environments (Hirst, 2007).

These enthusiast users provide innovations that we did not plan for or had envisaged having a different purpose. The model of users as innovators is considered by von Hippel (2005) as an extension of his view of "lead users" that are going beyond the mass of users. While the number of such users we can identify is small, we have clear examples of such lead users and been able to draw on their experience and change our own work to benefit others. What we do not know is whether we have a greater mass of lead users amongst those who have not made contact with us. Attempts to monitor this have included automated notification of blog entries that refer to "openlearn", encouraging contact and being aware of potential connections, however it remains difficult to make an assessment of the level of participation and identify interesting activities. Direct appeals to draw innovators to the site announced as opportunities and competition have had some success at a small scale to encourage educators to edit materials on the site. The result has been new material of benefit to all of our users, for example a translation into Catalan of an existing unit on genetics, and a connection with users who are trying out new ideas. This suggests a model where we can further our research by offering authentic actions on site that can also provide us with data.

Research 2.0

The discussion so far has looked at the particular example of OpenLearn, however we believe there are lessons that are transferable to other similar projects and researching informal learning in the mobile environment. "Web2.0" (O'Reilly, 2005) had been used as a label to characterise the change in expectations on the support for interactions between web sites, increased flexibility and greater value of placed on user generated actions. Arguably, in the research area we need to change the way we research and adopt a "Research 2.0" approach. O'Reilly identified 8 distinct patterns for web developers to embrace as they worked towards Web2.0 behaviour. In Table 1 we offer 8 variations on those patterns to act as advice for researchers trying to work with informal learning, the table includes shortened versions of O'Reilly's list as a comparison.

	"Research 2.0"	"Web 2.0"
1.	Study the interesting things that happen	The Long Tail

	In the long tail (Anderson, 2006) those who work with Web 2.0 can expect to find just as much value as with the high volume aspects at the head and it becomes counter productive to plan too much rather switching from filtering on the way in to filtering on the way out (Weinberger, 2007). For researchers we need to spot the interesting and unpredictable that can be found as much in the unexpected actions of individuals as the mass actions that we planned for.	
2.	Look for patterns that can apply more widely	Data is the Next Intel Inside
	The value of data about interactions over the designed hardware or software that runs it for researchers translates to the patterns that can be extracted and lessons drawn from use at scale. Without trying to draw out patterns though we can miss on ways to feed this information back into our research.	
3.	Encourage all to be part of the experiment	Users Add Value
	That users add value is a cornerstone of taking a user-centred approach to research, passive observation can give some of that value but extending an invitation to all involved, end-users and producers, will help maximise this value and also enable routes to get extra information.	
4.	Build valuable activities that give data	Network Effects by Default
	If we can find a way to draw on activity by default then we will have less need to ask users to provide us with data just for our research. For example, activities that give that help support the user, such as reflective logs, also give the researcher access to those reflections.	
5.	Recognise openness has a lot of benefits	Some Rights Reserved
	The decision to release rights as far as possible has brought great benefits to OpenLearn that should extend to the research. In particular allowing transfer between systems, using early discussion, and open publication.	
6.	Draw conclusions though you wish you had more data	The Perpetual Beta
	No computer system is ever completely finished or perfect, rather it is always a beta release that can be refined. Similarly no research project ever has all the data and can be sure of the results, but to be of value indicative results need to be available in early forms. This challenges the peer review process and encourages the use of forms of self-publication such as internal technical reports and blogs.	
7.	Be prepared for the user that arrives anywhere	Cooperate, Don't Control
	The design pattern to cooperate rather than control is a reminder to avoid imposing rigid paths, instead encouraging and cooperating with users. For research purposes this means we need to allow for users that start at any point. Data gathering that depends on users reading advice or passing through other points (e.g. logins) will not work in all circumstances.	
8.	Realise there is no way to control all access	Software Above the Level of a Single Device
	The content that enables us to build our site can also be used to build alternative sites, and indeed alternative ways to access the content from other devices. In some cases it will be	

> feasible to track some remote usage, but at some point there will be activity that we know nothing about. As researchers we must therefore operate with incomplete data, and be happy if we find out even a little of what is happening in these remote sites.

Table 1: Design patterns for Research 2.0 (Web 2.0 column based on O'Reilly (2005))

This advice is in itself tentative but has provided a structuring framework and evidence can be found for the operation of each of these patterns within in OpenLearn. We believe such a set of patterns can help to shape the interests of those involved in the production, use and reuse of open content and encourage informal learning. Such research also challenges standard approaches to ethics; we cannot expect someone who briefly visits a web site to then read through conditions and confirm statements. For OpenLearn a strong ethical framework was established and followed but in practice occasional interesting anomalies arose. If someone has registered on the site and indicated that they do not wish to be approached for research purposes (the default option during registration) then we also do not use their postings as illustrations or in analysis for research. This at first seems reasonable as the best way to comply with their wishes. However, their postings are public and it was the clear majority view during the workshop when this paper was first presented that others looking at OpenLearn as a research resource would feel able to quote and use the same posts as they would be unaware of the user's position. In practice we feel a weaker approach to ethical treatment of user activity is needed and Internet users in general are likely to accept that public activity can be both reused, and be the subject of research without their direct agreement. In all cases users would retain their own rights, which would for example prohibit malicious use.

Reflections for informal and mobile learning

The definition of informal learning established by Livingstone (2001) as "any activity involving the pursuit of understanding, knowledge of skill which occurs without the presence of externally imposed curricular criteria" seems to encompass OpenLearn. Among our users there appears to be a continuum from chance arrivals who may pick up some knowledge, to those who are preparing to register for a paid-for course. Open content that has no licencing restrictions and can be transformed is also a suitable base for use on mobile devices, with sample content transferred into mobile content sites and the underlying XML format suitable for automated transformation. However it is the open availability of the content that allows general mobility through no need to belong to a group associated with an institution (e.g. be a registered student) rather than mobility in the device it is offered on. We expect advances in open learning approaches to focus on the provision of a ubiquitous social environment rather than supporting particular devices. In this view we align with Taylor, Sharples and Vavoula (2005) in seeing "the learner that is mobile, rather than the technology." We also feel that we have faced common research challenges with those who are studying the use of mobile devices, the suggestions of design patterns for Research 2.0 may at first appear to imply an acceptance of reduced rigour, however instead they are an indication of the need for research itself to evolve and become more agile, in line with the technologies it studies.

References

Anderson, C (2006) *The Long Tail: How Endless Choice Is Creating Unlimited Demand*. Random House.

Esslemont, C. (2007) 'Bridging the abyss: open content to meaningful learning' in *OpenLearn: Researching open content in education, Proceedings of the OpenLearn2007 Conference*, 30-31 October 2007. pp44-46. http://kn.open.ac.uk/workspace.cfm?wpid=7979

Godwin, S.J. & McAndrew, P. (2008) 'Exploring User Types and What Users Seek in an Open Content Based Educational Resource' to appear in *Proceeding of Ed-Media 2008*.

Harley, D. and Henke, J. (2007) Toward an Effective Understanding of Website Users: Advantages and Pitfalls of Linking Transaction Log Analyses and Online Surveys. *D-Lib Magazine*.13(3/4), http://www.dlib.org/dlib/march07/harley/03harley.html

Hirst, T. (2007) 'Feeding from open courseware: exploring the potential of open educational content delivery using RSS feeds' in *OpenLearn: Researching open content in education, Proceedings of the OpenLearn2007 Conference*, 30-31 October 2007. pp15-16. http://kn.open.ac.uk/workspace.cfm?wpid=7979

Livingstone, D.W. (2001) Adults' Informal Learning: Definitions, Findings, Gaps and Future Research. NALL Working Paper No.21, CSEW, OISEUT, University of Toronto, Canada.

O'Reilly, T. (2005) 'What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software' http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Sharples, M., Taylor, J., and Vavoula, G. (2005) Towards a Theory of Mobile Learning. Proceedings of mLearn 2005 Conference, Cape Town. http://www.mlearn.org.za/CD/papers/Sharples-%20Theory%20of%20Mobile.pdf

von Hippel, E (2005) *Democratizing innovation*, MIT Press: Cambridge, MA. http://web.mit.edu/evhippel/www/democ1.htm

Weinberger, D. (2007) *Everything Is Miscellaneous: The Power of the New Digital Disorder*. Times Books.