The Role of Co-Interpretation of Routine Data in the Evaluation of a Virtual Research Environment

Richard Procter, Vito Laterza and Patrick Carmichael
Centre for Applied Research in Educational Technologies, University of Cambridge.

Abstract

This paper reports a novel research approach in which routine server data collected from online environments provides a 'focus' for semi-structured interviews in the course of which research participants are involved in a process of ‘sense-making’ (Weick, 1993). The TLRP Virtual Research Environment (VRE) is based on the Sakai platform which provides an access controlled online suite of interchangeable modular collaboration tools that geographically distributed research teams can use to support their work. It is highly customisable and tools can be deployed within the environment to suit the needs of individual research projects. Sakai generates very detailed 'logs', and the aggregation of the data from these server logs and database tables provided insights into how specific users and groups were engaging with the environment. These largely quantitative data are useful in determining trends in behaviour, broad patterns and sequences of events, but tell us little about context other than 'who', 'when', and 'where'. We show that it is possible to construct rich and illuminating qualitative accounts which help to 'make sense' of: single events (such as sudden increases, decreases or changes in online activity); repeated sequences and 'activity-flows' (such as the online publication of documents followed by moderated periods of peer-review, or sequences of data collection or analysis); and long-term trends in activity over the project life-cycle. Our findings suggest that the approach of using suitably reduced quantitative data supports reflective accounts which are more useful to researchers, participants and support staff than more general approaches or those which simply use routine data as part of a triangulation process.

Corresponding author

Richard Procter, Centre for Applied Research into Educational Technologies, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB. Email: richardp@caret.cam.ac.uk
Introduction

This paper reports on one of the research approaches developed and implemented in the course of the ‘VRE for Educational Research’ project. This project was part of the Joint Information System Committee (JISC) Virtual Research Environments (VRE) research programme. The project was run in collaboration with ten projects from the Economic Social Research Council’s (ESRC) Teaching and Learning Research Programme (TLRP). The partnership with the TLRP enabled a broad range of educational projects to be recruited in terms of number of institutions involved in each project, the geographical distribution of those institutions, the range of research methods used by those projects, the educational sectors of the projects, and at what stage the projects were at in their life spans. The Sakai VRE project’s overall aim was to evaluate the use of the virtual research environment for collaborative work and to understand the social processes, practices and collaborative patterns used by groups and individuals when engaged in collaborative research activities.

Sakai is an open source ‘Virtual Collaboration Environment’ community development project with the aim of building a virtual collaboration environment, initially developed in the USA by four leading universities (Michigan, Indiana, Stanford and MIT). It can be set up as either a Virtual Learning Environment (VLE) or a Virtual Research Environment (VRE). It provides a modular, access-controlled online environment with a customisable set of collaboration and communication services and tools: these include email, chat rooms, discussion forum, file store and wiki tools. Users can set up the environment as a communications hub or a digital data repository depending on the needs of their course, project or group. This paper reports how the information generated by users in the Sakai environment was used to further obtain information from project members to find out about the collaborative practices of the projects participating in the wider VRE for Educational Research project.

Routine or ‘Usage’ Data

Many studies have used user generated data to discover the activities of web page users (Chakrabarti, 2002; Pendharker, 2003). The data about a user’s interaction with the Sakai environment can be collected in a number of ways. The two major sources of information are the database tables stored inside Sakai itself for each user and from the logs written by the web server on which the instance of Sakai is running. The information from these separate

1 http://www.sakaiproject.org
sources can be aggregated together so that a more complete picture of a user’s activity can be built.

Cooley, Tan and Srivastava, (1999) describe web usage mining as having “three types of domain information available; usage, content and structure” they go on to add a further type in a later paper (Srivastava, Cooley, Dehpande and Tan, 2000) which classifies data into the following types:

- **Content**: The real data in the web pages
- **Structure**: Data which describes the organisation of the content
- **Usage**: Data that describes the pattern of usage of web pages
- **User profile**: Data that provides demographic information about users of the Web

In this paper we are interested primarily in the third classification “data that describes the pattern of usage of web pages” and to a lesser extent with demongraphics data. Whereas Srivastava et al. (2000) try “to find inter-session patterns such that the presence of a set of items is followed by another item in a time ordered set of sessions or episodes” (Srivastava, Cooley, Dehpande and Tan, 2000), we were able to relate usage patterns in user sessions directly to actual users and carry out further interviews with these users to find out about the contexts and collaborative practices that lay behind these usage patterns. It should be stressed that this level of scrutiny was undertaken with the knowledge of users, who were informed in advance that their activities online would be monitored and might form the basis of further research.

The usage information gathered about a user’s interaction with a web site can be analysed in a number of ways and for different purposes. Firstly, data can be used to characterise online activity so that the user can be offered an enhanced or personalised online experience. Sites like the Google home page can be personalised for the user in terms of their location, for weather reports, or preferences, such as stock market updates. This type of personalisation can also be very useful for users who may have specific disability requirements, thus page elements such as text sizes and colour contrast can be personalised. Secondly, usage data can be used to hypothesize user’s offline activities and then the websites online content can be adapted accordingly. Examples would include sites such as Amazon where lists of books under heading such as ‘other books you may enjoy’ appear and ‘user who brought this also brought the following’, also sites like iTunes offers ‘other music’ and Ebay offers ‘similar
items to those you bid on and lost’. Thirdly, usage data can be used in conjunction with other data sources to produce data triangulation. Leach (2002) used both log file data generated by users of the First Class System and online questionnaire data to analyse the potential of electronic conference environments for the professional development of teachers.

Our approach is a hybrid of these. In effect, usage data generated by the user and gathered from both the Sakai database tables and the web server log files was treated as (very dull) qualitative data and used to find ‘stories’ and concordances which gave us indications about how individuals and teams had interacted with the VRE. Our data enabled us to identify individual users; this gave us the ability to present our data to the research teams as the focus of a sensemaking activity (Weick, 1995) within a broader framework of co-interpretation with these identified project participants.

**Data Reduction**

When a user interacts with the Sakai environment, information is collected about that user’s session; this information is recorded in a number of different data base tables and in the web server log files. The information from these separate sources was collected into text files and then aggregated together using custom built Perl scripts and displayed on a series of web pages.

The usage information was displayed by month, both for individual users and for groups of users in the same collaborative ‘worksite’. Trends in behaviour of groups of users such could be identified, and sequences of events could be seen for example, the processes leading to collaborative writing and peer review of a project document. Although this information was very helpful in identifying patterns of engagement and collaboration, there remained two major disadvantages to this approach. First, this information, although very comprehensive and informative, it revealed little of the context of these interactions, so that even though we knew the ‘who’, ‘where’ and ‘when’ of a user interaction we did not know the ‘why’ of the interactions, or details of both their social practices and context.

The second disadvantage was the amount of data available to us was overwhelming. There were on average about 100 unique users per month in 2006, (with around 700 logins per month), with each user action (visiting a page, adding a chat message, uploading or downloading a file) generating a separate line of information in a data base table or log file (user ‘visits’ typically produced 20-50 lines). Over the course of a month this would produce
tens of thousands of lines of information. Thus although the information was comprehensive the sheer volume of information precluded any useful engagement with the data. There was thus a need to both reduce the amount of data so that it could be engaged with in a more meaningful way, for example, by presenting it as graphs of online activity over time (see Figure 1).

Figure 1: An example of a routine data graph, showing online activity over time

The availability of these data (suitably reduced) presented an opportunity for researchers to triangulate online activity with other data (for example: prior experience of ICT; information about evolving roles within projects; and interviews about the progress of the project). But they also allowed a more sophisticated ‘sensemaking’ activity as part of a broader strategy of ‘co-interpretation’, with researchers and VRE users working together to establish the broader meanings and significance of what had been observed online.

Strategy 1: Triangulation of Interviews and Routine Data

Conceptualised as a triangulation process, we can represent this process as comprising two ‘streams’ of data and interpretation (Figure 2) below shows how day-to-day reflection on their experience of using the VRE feeds into an overall review of their patterns of use and an
evaluation of its value in supporting various research activities. The researcher, involved in this review, also has access to the appropriately reduced data and may already have some hypotheses about the participant’s offline activity based on this. The information from the graphs could then be triangulated with the data from interview so that an overall picture of offline activity could be built up – but this would be largely (or entirely) a task undertaken by the researcher.

**Strategy 2: Sensemaking and Co-Interpretation**

The reduced data, however, could also be used in the context of a more reflective ‘sense-making’ activity, as part of a broader co-interpretation strategy which involved a dialogue between participants and researchers.

For Weick (1995) sensemaking is “literally, it means the making of sense” for others it is a process that is used to transform what is unknown into what is known (Huber and Daft, 1987; Waterman, 1990) it has been used in organisational studies (Gioia and Chittipeddi, 1991; Weick, 1995) and crises (Weick 1993). Weick states that sensemaking as an activity is distinct from interpretation:

> It is common to hear someone made an ‘interpretation’. But we seldom hear that someone made a ‘sensemaking’. We hear, instead, that people make sense of something, but even then, the activity rather than the outcome is in the foreground. A focus on sensemaking induces a mindset to focus on process, whereas this is less true with interpretation (Weick, 1995: p13)

We were aware that the activity that we were involved in with our participants was making sense of the graphs rather than interpretation of the graphs. “The key distinction is that sensemaking is about the ways people generate what they interpret” (Weick, 1995: p13) thus the sensemaking activity would generate the data that we were to interpret, and so may be considered a particular type of focussed and reflective activity. Sensemaking is also a social process in our case involving a dialogue between the researcher and the participant. Weick says that “the concept of sensemaking highlights the action, activity, and creating that lays down the traces that are interpreted and then reinterpreted” (1995: p13). In our case the researcher and the project participants engaged in sensemaking so that the offline context and practices behind the online activities could be uncovered and then reinterpreted.
The use of this reduced routine usage data in a graphical form (figure 3), for the sensemaking activity allowed participants to reflect on the activities that they were engaged in and to make sense the peaks and troughs of their online engagement for the researcher. The recorded conversation between researcher and participant of this activity then became our data to be interpreted. Thus for us in this case sensemaking is seen as a specific phase in a more general researcher and participant “co-interpretation” process.

Figure 3 shows this alternative model of co-interpretation of routine data through a sense-making activity. The main participant and researcher tracks are as before but the interview has been replaced by a more focussed sense-making activity which itself is capable of initiating a much more focussed and informed reflective process on the part of both researcher and participant. Weick states that:

To talk about sensemaking is to talk about reality as an ongoing accomplishment that takes form when people make retrospective sense of situations in which they find themselves and their creations. There is a strong reflexive quality to this process.

During the activity the participants would try to “make retrospective sense [of their] creations” - in this case the routine usage data. This approach involved the researcher and the participant in a joint process of sensemaking which promoted individual reflection on the entire process of learning to use the VRE and fitting it into the work patterns and social practices of their projects.

What emerged from the sense-making activity, and therefore contributed to this broader co-interpretative strategy, were insights into both the ‘day-to-day’ life of projects, refracted through the lens of VRE use; and reflections on the VRE itself. What is clear from these data, which we will now present, is that this goes beyond a triangulation process in which the research has any sense of ‘checking’ one set of data against another. Rather, both research and participants emerged from the process with new insights into patterns of activities both offline and online.
Figure 2: Triangulation of Routine Usage Data

Participant track:
- Qualitative lived experience of using the VRE
- Weakly reflective process
- Interpretation of experiences

Researcher track:
- Sever logs and Sakai database logs
- Data reduction of logs into graphical form
- Interpretation of the graphs
- Reporting processes

Interview
Figure 3: Sensemaking and Co-interpretation
Windows into Project Life

The data from the sensemaking activity provides us with an insight into the daily workings of project life, not just the individual practices or collaborative practices of the project members. The following project, which was based in a single institution but was, working with a number of primary and secondary schools, it used the VRE for a lot of the initial planning and setting up of their project between the principal investigator and a research associate.

*R: Well that time in January is also I think coincides with the fact that D went off on sick leave because hers goes way down so mine had to go up proportionately ... what you’ve got are the two main people who are taking responsibility for the project and one compensating for the other

As this quote shows the work of setting up the project does not disappear when one person is ill, but moves on to other people within the project and this was highlighted within the usage graphs.

The same project had standard letters for school access and consent forms for pupils uploaded within the resources section of their VRE, thus they was created a central online repository of a project’s working documents.

*L: But you can see I think that’s me coming in and I think that little peak would have been about October ... that would have been about time to initially just identify the schools and the paperwork around, drafting up letters ... this is doing consents, I can see those little peaks are the times I was trying to get access to schools.

In many projects the difficulty of learning how to use and engage with the new VRE environment, and the time that was needed to do that, was highlighted.

*S: ... there would have been a flurry of activity there, where in the February, where I maybe had a little window of time and I thought right, I’m going to, I’m going to embrace this technology and I’m going to find out how it works and how I can use it and so on and then getting completely overwhelmed with
field work and getting completely embedded in my own writing and my own analysis.

The above quote shows how project members would have a “flurry of activity” when they first engaged with the VRE which would then fall away as the other pressure of their work mounted up. This type of engagement was seen in many projects.

The following project based in two universities and working with four further education colleges, organized weekly chat room sessions where issues of the project’s theoretical engagement were discussed, these chat sessions involved all project members including teacher-researchers from the further education colleges:

AB: Well I tend to log on it mainly for the chat room, the chat sessions and that takes place on Monday mornings at ten o’clock, so I would imagine that when it like peaked here we tried to have a Monday morning session on the VRE every Monday.

At the end of the chat session one of the group within the session volunteered to produce a summary of the session which was then uploaded into the worksite resources folder for other members of the team to refer back to. This type of activity could then be seen in a more in depth analysis of the logs. One ‘story’ that was often seen involved participants uploading documents into the file store of the VRE and then downloading them twice just to check that the document had been uploaded correctly.

Projects also used the VRE as a secure online area for their project’s documents, these may be in the form of gathered data, access forms and protocols or research instruments. The advantage of using the VRE for these resources is that project members in any institution can have access to them, rather than being in an institutional shared drive which can only be accessed from that single institution. One of the advantages of the TLRP VRE is that it is not based in any one project member’s home university, which may cause access problems for project members who are from other universities. The following quote show how a project moved from the institutional shared drive to keep project data towards the VRE which is not access controlled from any institution but is set up by the TLRP.
MA: We had a similar procedure if you like prior to the VRE, we used something called Gspace which was set up by the University which was just a separate drive where we could upload and I had quite a lot, well a great deal of data in that particular Gspace, but that isn’t accessible from outside the University so the advantage from the point of view of the VRE is that the people who are now working on in terms of analysis and dissemination will be able to individually and independently access it so that’s why I’m going to be moving all the data into the VRE.

It's also seen from the following quotes the importance of setting up a structure for online documents that everyone in the project can understand.

CS: I think Z put most of her stuff in there, or before she went on maternity leave, I think she’d put most of her resources into there, in fact she you know, she created the, that system, you probably knew because they’ve got her name next to most of those...[folders]

The above quote shows how the one person can champion the use of the VRE or be the ‘technology steward’ (Wenger et al. 2005). A ‘technology steward’ is a member of the community with a grasp of how and what technologies can enhance the community, they act as a guide so that other members of the project team can utilise and benefit from the VRE.

This last quote highlights how the value of the use of the VRE had been perceived differently by different members in different institutions across the same project.

RI: But after Z had set up the structure, and had inputted her own files, some of the other researchers [in other institutions] said “it’s too tedious and boring”, “it’s an admin task”, “I shouldn’t be wasting my own time doing it”.

Reflecting on Using the VRE

The data from the sensemaking activity also provided us with information on how the project participants reflected on their own use of the VRE.
This first quote shows how the participants are now using the VRE for a shared store of presentations but also for the collaborative development of new project resources.

CS: “I mean another shared collaborative piece of work has been developing a teacher training resource which, which is partly a presentation to Colleges or managers or whatever ... and because we’ve ... been doing a lot of that lately, that kind of assimilation, we’ve had a, we’ve been sharing our ideas, so we’ve put different presentations in the VRE and then we’ve combined them.”

In the following quote the participant highlights the “shift in thinking” that has occurred within the project so that there is a move away standard tools such the local file system and email, to a more collaborative shared space within the VRE.

L: “I think it was about eventually realizing everything’s going to be there and shifting your thinking from just emailing documents and creating your files on your computer and the kind of normal way of doing it, but the beauty of it is you know, we did consents originally for the focus groups and then whenever I needed consent forms for LE [...] I knew exactly where to go on the VRE to get them.”

The following quote again highlights the shift in thinking but this time about how the “My Workspace” part of the site may be used as one’s own online digital repository system.

S: “So I, so I could actually upload the whole [of my PhD data] up onto [the VRE] well that is a brilliant idea. I could upload all my images to because I’ve got a load of images ...”

**Ethical Issues**

This study highlights again the fact that nothing is private on the web, and that somewhere a record of activity is kept (see Carmichael and Youdell, 2007 for a further discussion of the nature of ‘anonymity’ in virtual environments). We were able to generate very detailed accounts of user activity over two years, so detailed that we struggled with the volume of data we had. In those two years and even though the project participants knew that their use of the VRE was been studied, we were not
asked about if or how we would use this data, even when participants were presented with this information they were not shocked that it had been collected. Interestingly, during the data collection phase, participants had very little interest in the fact that routine data was collected – and had we sent them unreduced data on a regular basis this would probably not have been well-received. But when reduced data were presented in the context of the sense-making activity most participants were very interested to see what the graphs looked like (particularly in contrast to the other members of the same project, which may itself be instructive!), and reflected on the process that this might have impact on they ways in which they might use online technologies in support of project activities in the future.

**Discussion**

In this paper we have shown how a sense-making activity may continue to an overall strategy of “co-interpretation” and participant involvement. That process is not only illuminating for the researcher: it is also a potentially powerful driver of reflective processes for the participant.

We have broadly thought of this process of co-interpretation in the following way:

\[ \text{data reduction} + \text{interpretation} + \text{sensemaking} + \text{reinterpretation} = \text{co-interpretation} \]

In this case, our co-interpretation strategy involved each of these steps: data reduction and representation of that data in graphical format contributed to the focal sense-making process. This approach generated interview data that was much richer and focussed than more general discussions might have produced. More critically, it has initiated reflection as to how use of the VRE might be employed (by users)and users supported (by researchers and technical support staff), and has informed subsequent patterns of support and training provided for groups of researchers as they have begun to use the VRE; both participants and researcher appear to be better prepared for ‘next time’.
References


