Working with Text Logs: Some Early Experiences of Record and Reuse

Andy Crabtree[1] and Mark Rouncefield[2]

[1]School of Computer Science & IT, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, UK.

[2]Computing Department, Lancaster University, Infolab 21, South Drive, Lancaster LA1 4WA, UK.

Email address of corresponding author: andy.crabtree@nottingham.ac.uk

Abstract. In this paper we consider the notion of record and reuse as potential area of eSocial Science research, where the digital resources exploited by people in particular domains of action become resources for the social science analyst as well. We articulate the notion through a case study of the use of text logs in an ethnographic study of a mixed reality game, focusing on the work involved in using digital materials to develop broader requirements for the development of record and reuse applications and services. Accordingly a generic process of work is articulated through this case study and a suite of tools supporting record and reuse is identified to support social science research in the digital environment of the future.

Introduction

The vision of the Grid as a distributed network of 'virtual organizations' that enable computationally intensive tasks to be processed by a great many machines (Foster et al. 2001), is one well suited to the data intensive demands of the natural sciences (see, for example, TeraGrid).[1] The purchase of the vision is less clear with regard to the social sciences, however. This is not due to the scientific status of the social sciences, nor to the theoretical character of a great deal of research carried out by social scientists. Rather, the issue revolves around the prosaic nature of social science data. Even 'large scale' research in the social sciences – such as survey based research – exploits only very small datasets. For example, at the recent *eSocial Science Training and Awareness Day* in Manchester (November 19th 2004), a social science researcher exploiting quantitative techniques to address national trends in the UK was doing so on the basis of a dataset of a mere 110 megabytes. It is not unusual for qualitative datasets - which are often construed of as small scale or 'micro' in nature - to be larger. Some of the qualitative datasets in our own research projects are over 5 gigabytes, for example, and that of course is still computationally trivial.

The point of these mundane observations is to recognize that the demands of social science research are different to those of the natural sciences and that these will affect the ways in which the Grid vision is played out and realized in the social sciences. eSocial Science cannot be exclusively predicated on intensive multi-computer processing as 'large' and 'small', macro and micro, are not so much matters of data size but of *analytic reasoning* in the social sciences (Benson and Hughes 1991, Hughes and Sharrock 1997). Rather than focus on computationally

1 www.teragrid.org

demanding tasks in the first instance then, we think it more useful to consider the future nature of social sciences records instead. We think it particularly relevant to consider the ways in which computation might make new resources available to social scientists through the recording of new forms of data and to consider the ways in which novel datasets may be 'reused' or transformed from raw computational logs into materials that support social science analysis. It is here that computational complexity might enter into practice through the design and development of new applications and services. The issues of record and reuse stand at the core of eSocial Science Research Node at the University of Nottingham.[2] Evolution rather than revolution is a central tenet of the Node's research programme, which seeks to develop digital resources to complement established practices of social science research. Our purpose here is to provide an initial exploration of some of the issues bound up with record and reuse through case study.

The case in question is drawn from our own research which is, broadly speaking, ethnographic in character (Prus 1996). Since its migration from anthropology to domestic use (i.e., using the approach to study our own societies), ethnography has drawn on a wide range of resources to understand social phenomenon. For example, Robert Ezra Park, one of the most influential members of the Chicago School of Sociology (which pioneered the domestic use of ethnography), exploited 'concentric zone maps' along with ethnography to chart the various social divisions and boundaries shaping the urban environment (Park et al. 1925). The ethnographic tradition draws on a multiple resources to get the job of observing and analyzing social life from 'within' done, routinely exploiting biographical resources (e.g., Wieder and Zimmerman 1977), visual resources (e.g., Prosser 1998), technological resources (e.g., Heath and Hindmarsh 2002), and a great many other resources besides. The point here is not to assemble an exhaustive list of the resources used by ethnographers but to recognize that ethnography is done not only through the immersion of a researcher in a setting but through *the use of material resources* as well. These resources may be external to the fieldwork setting, such as above where diaries, or photography, or video cameras are introduced into the field, or they may be internal and consist of such things as the objects of work, working documents, electronic media, etc.

External and internal resources co-exist in the course of ethnographic research. Thus, and for example, a person working on a computer may find her work the object of video recording. Her work, both on the computer and on the phone, is also likely to be logged, monitored and manipulated in various organizationally significant ways with those logs being used as a resource for other organizationally relevant activities (Hughes et al 2002). Our concern with resources here is novel. We are particularly concerned to explore the possibility of making resources that are internal to a setting of action under study available as resources that permit closer inspection and analysis of that setting of action. Thus, the *resource for getting the work done becomes a resource for conducting ethnographic study of the work* (and action more generally). We believe that eSocial Science opens up this possibility and in this paper we examine in it detail through the use of text logs generated in the playing of a mixed reality game (Crabtree et al. 2004). Elsewhere (Cheverst et al 2004; Graham et al 2005) we have examined the use of text logs in facilitating the understanding aspects of everyday work. It is worth pointing out, however, that we do not offer text logs as a solution to problems of ethnographic inquiry or social science research more generally. Nor is the use of text logs an attempt to extend the 'virtual ethnography' programme (Hine 2000). The text logs simply offer an acknowledgement of the growing importance and pervasiveness of various forms of

2 www.ncess.ac.uk/research/nodes/index. shtml#rodden

digital data as well as a demonstration of the possibility of eSocial Science to transcend traditional resource boundaries and make tools supporting action in particular domains of human activity into tools supporting research as well.

It is also important to stress that the text logs do not stand alone but are married to other resources that are internal and external to the digital environment. The text logs gain their purchase as part of broader ethnographic inquiry that exploits traditional techniques of immersion and direct observation, and the gathering of data through fieldnotes and video recording. When considering the use of text logs we focus on the interplay between resources, particularly on the marriage of text logs and audio files generated within the digital environment and their relationship to video recordings external to the digital environment. In many respects the text logs themselves are not of primary interest. Rather it is the *work* involved in marry these resources together that is of concern. The work of using multiple data sources generates requirements for record and reuse applications that go beyond text logs to incorporate a rich mix of digital resources employed by users in their activities. The vision here is that those self same resources may in the future be exploited by ethnographers studying those activities, and it is towards unpacking requirements for the development of eSocial Science applications that support record and reuse that we turn to in conclusion.

Working with Text Logs

Exploiting the resources that users employ in the course of their interactions as a resource for conducting ethnographic studies can clearly add value to analysis (Chevherst et al. 2003, Benford et al. 2005). Exploiting such resources takes work, however, and in this section we wish to explicate the work involved in making text logs work. The text logs in this case are taken from a mixed reality game called *Uncle Roy All Around You*.[3] To play the game, street players must navigate the streets of a city and locate Uncle Roy's office. To do this they use a handheld computer or PDA that provides a map view of the city streets. The players receive automatically generated clues providing them with directions to particular places in the city and further clues are provided when they declare their current position. Street players also receive text messages from online players, who can track a street player's movements through a virtual model of the actual city streets. Street players can respond to online player messages by recording a short (7 second) audio message. Working via text and audio messages online players and street players collaborate to find a postcard. Once found, online players are provided with information that they use to guide street players to Uncle Roy's office. The technology employed in the game captured the clues sent to street players, the text messages sent by online players, and the audio messages sent by street players, and combined them together in a single text log with associated audio files. An ethnographer also accompanied several street players as they made their way around the city (Manchester in this case) and recorded their interactions on video. In order to develop an understanding of the social organization of the game it was necessary to combine these internal and external resources to furnish coherent instances of interaction for analysis. Below we consider the work involved in using the text logs to assemble coherent representations that permit analysis of the technology-in-use (Button 1992).

Although the text logs capture the features of interaction described above, those features are not readily accessible, as can be seen in Figure 1. The log consists of wide range of information, ranging from temporal and spatial to interactional. In order to extract usable data

---

3 www.blasttheory.co.uk/bt/work_uncleroy.html

from the log – i.e., data that will support and afford particular kinds of analysis - it is necessary to clean the log up so that *salient features can be identified*. In the case of understanding interaction between online and street players, identifying salient features requires that spatial, temporal and numerical information be stripped out of the log and salient features be extracted, as in Figure 2.

5478	1.08387E+12	82537	venom_89	what up patrick			
5479	1.08387E+12	18928	ant_49	can you speak now			
5480	1.08387E+12	82537	555	396	0	0	
5481	1.08387E+12	18928	462	405	0	0	
5482	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Cantonese warning. (21 minutes remaining)</m><h>0</h></c>				
5483	1.08387E+12	18928	462	405	0	0	
5484	1.08387E+12	82537	Nicole_20	hi patrick			
5485	1.08387E+12	18928	452	401	0	0	
5486	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Strangeness can be comforting when you are in the right mood. (21 minutes remaining)				
5487	1.08387E+12	18928	452	401	0	0	
5488	1.08387E+12	82537	Nicole_20	that jacket looks cool			
5489	1.08387E+12	82537	audio_id-82537_time-1083871002203.mp3				
5490	1.08387E+12	82537	509	438	0	0	
5491	1.08387E+12	18928	453	400	0	0	
5492	1.08387E+12	18928	dave_31	record a msg emma I m trying to help you...			
5493	1.08387E+12	82537	audio_id-82537_time-1083871041921.mp3				
5494	1.08387E+12	18928	audio_id-18928_time-1083871046968.mp3				
5495	1.08387E+12	82537	dave_31	go into the graffitied phonebox by the railings			
5496	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Head for Princess Street. (20 minutes remaining)</m><h>0</h></c>				
5497	1.08387E+12	18928	453	400	0	0	
5498	1.08387E+12	18928	430	429	0	0	
5499	1.08387E+12	82537	dave_31	have you found the postcard ontop of the phone			
5500	1.08387E+12	18928	414	444	0	0	
5501	1.08387E+12	18928	398	466	0	0	
5502	1.08387E+12	18928	Nicole_20	hi emma. this phone box is where my postcard is			
5503	1.08387E+12	82537	audio_id-82537_time-1083871093328.mp3				
5504	1.08387E+12	82537	Nicole_20	hi patrick my postcard is at this phone box			
5505	1.08387E+12	18928	<c><n>local</n><t>2</t><m>I never use a disguise: no need. (19 minutes remaining)</m><h>0</h></c>				
5506	1.08387E+12	18928	398	466	0	0	
5507	1.08387E+12	18928	<c><n>local</n><t>3</t><m>Go immediately to 20 Dickinson St.</m><h>26</h></c>				
5508	1.08387E+12	82537	Nicole_20	there is something on top of the phone box			
5509	1.08387E+12	82537	Nicole_20	ur cool thank u			
5510	1.08387E+12	18928	audio_id-18928_time-1083871136531.mp3				
5511	1.08387E+12	18928	398	466	0	0	
5512	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Head for the tram lines. (19 minutes remaining)</m><h>0</h></c>				
5513	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Head for the tram lines. (19 minutes remaining)</m><h>0</h></c>				
5514	1.08387E+12	18928	398	466	0	0	
5515	1.08387E+12	82537	dave_31	u need the phonebox on portland st by the tower			

Figure 1. The raw text log

Venom: what's up Patrick
Ant: can you speak now
Clue: Cantonese warning (21 minutes remaining)
Nicole: hi patrick
Clue: Strangeness can be comforting when you're in the right mood (21 minutes remaining)
Nicole: that jacket looks cool
Audio ID 82537
Dave: record a msg emma … I'm trying to help you
Audio ID 82537
Audio ID 18928
Dave: go into the graffitied phonebox by the railings
Clue: Head for Princess Street (20 minutes remaining)
Dave: have you found the postcard on top of the phone
Nicole: hi emma. This phonebox is were my postcard is
Audio ID: 82537
Nicole: hi patrick. my postcard is at this phonebox
Clue: I never use a disguise: no need (19 minutes remaining)
Clue: Go immediately to 20 Dickinson St
Nicole: there is something on top of the phone box
Nicole: u r cool thankyou
Audio ID: 18928
Clue: Head for the tram lines (19 minutes remaining)
Clue: Head for the tram lines (19 minutes remaining)
Dave: u need the phonebox on portland street by the tower

Figure 2. Cleaning up the text log

While accurately representing the work that is involved in cleaning up the log, this is something of a fraudulent representation, as the work of cleaning up is directed. Specifically, and in terms of ethnographic study here, it is directed towards extracting *conversational threads* that are relevant to particular aspects of the study. In this case, what is of interest is the talk that pertains to Patrick (a player whose interactions were recorded on video), so cleaning up the log involves extracting conversational threads between particular participants from out the flow of overlapping threads between the multiple participants represented in the log. The cleaned up log can be seen in Figure 3.

Venom: what's up Patrick
Nicole: hi patrick
Nicole: that jacket looks cool
Audio ID 82537 (Patrick)
Audio ID 82537 (Patrick)
Dave: go into the graffitied phonebox by the railings
Dave: have you found the postcard on top of the phone
Audio ID: 82537 (Patrick)
Nicole: hi patrick. my postcard is at this phonebox
Nicole: there is something on top of the phone box
Nicole: u r cool thankyou
Dave: u need the phonebox on portland street by the tower
Audio ID 82537 (Patrick)

Figure 3. The cleaned log

The point of the fraudulent example is to show that cleaning up the log relies on *making determinations of relevance* – at its simplest level, determining what's 'noise' and what's not (a judgement that is relative to the kind of analysis to be performed). For the ethnographer such determinations rely on the use of other materials: in this case, on Patrick's audio files (which have a unique ID) and video recording of Patrick's interactions on the street. It is with reference to this other material that salient features of the log become identifiable and *start* to make sense. Emphasis is placed on *start* to make sense as there is a job of work to be done here to make the log extracts actually make sense. Thus, the audio files and video recording must first be transcribed. Four audio files are implicated in the logged sequence of interaction and transcription of their content can be seen in Figure 4.

Venom, how you doing? Listen, I'm heading to China Town but if you find your phone box give us
Nicole, it's very comfortable and it's keeping me very warm tonight, so thanks for that
Can you direct me to it. I'm outside the red phonebox outside er – oh bollocks – outside Reyner Street
Tell me about someone from your past who never leaves you – that's what it says on your postcard

Figure 4. The transcribed contents of the audio files

Following transcription, the audio files must be synchronized with the log extracts and it is in doing this job of work that the emphasis placed on *start* to make sense gains particular purchase, as can be seen in Figure 5.

Venom: what's up Patrick
Nicole: hi patrick
Nicole: that jacket looks cool
Patrick: 'Venom, how you doing? Listen, I'm heading to China Town but if you find your phone box give us'
Patrick: 'Nicole, it's very comfortable and it's keeping me very warm tonight, so thanks for that.'
Dave: go into the graffitied phonebox by the railings
Dave: have you found the postcard on top of the phone
Patrick: 'Can you direct me to it. I'm outside the red phonebox outside er – oh bollocks – outside Reyner Street.'
Nicole: hi patrick. my postcard is at this phonebox
Nicole: there is something on top of the phone box
Nicole: u r cool thankyou
Dave: u need the phonebox on portland street by the tower
Patrick: 'Tell me about someone from your past who never leaves you – that's what it says on your postcard'

Figure 5. Synchronizing text messages and audio files

As we can see, the synchronized log does not make sense. The log provides an *inaccurate representation of the orderliness of interaction* – it may look orderly but it is not as can be determined by the nature of the participants' utterances (it simply does not make sense for Nicole to thank Patrick for finding her postcard before he has told her that he has found it, for example). The job of synchronizing the text messages and audio files relies, then, on splicing the audio file transcripts and log extracts together in such a way that make *conversationally recognizable sense*, as in Figure 7.

Venom: what's up Patrick
Patrick: 'Venom, how you doing? Listen, I'm heading to China Town but if you find your phone box give us'
Nicole: hi patrick
Nicole: that jacket looks cool
Patrick: 'Nicole, it's very comfortable and it's keeping me very warm tonight, so thanks for that.'
Dave: go into the graffitied phonebox by the railings
Patrick: 'Can you direct me to it. I'm outside the red phonebox outside er – oh bollocks – outside Reyner Street.'
Dave: have you found the postcard on top of the phone
Nicole: hi patrick. my postcard is at this phonebox
Nicole: there is something on top of the phone box
Patrick: 'Tell me about someone from your past who never leaves you – that's what it says on your postcard'
Nicole: u r cool thankyou
Dave: u need the phonebox on portland street by the tower

Figure 7. Synchronized text messages and audio files

The job of synchronization continues, using the same principle to marry the text log extracts, audio file transcripts and video transcripts together and assemble a coherent representation of interaction between street players and online players, as in Figure 8. Regular text indicates Patrick's talk on the streets that was recorded on video, underlined text indicates messages received from online players, and bold text indicates audio messages sent by Patrick to the online players (it is notable that what Patrick says and what the online players hear is different due to the limited audio recording facility).

Patrick: Oh, someone's sent me a message here, Venom.
Message from Venom: What's up Patrick
Patrick (selects record audio): 'All right **Venom, how you doing? Listen, I'm heading to China Town but if you find your phone box give us** a shout, yeah'.
Patrick: OK, right, China Town.

Patrick is making his way to China Town.
Patrick (looking at PDA and laughs): Someone's telling me I've got a nice jacket on.
Message from Nicole: Hi Patrick - that jacket looks cool.
Patrick (selects 'audio record'): 'Thanks **Nicole, it's very comfortable and it's keeping me very warm tonight, so thanks for that**.'
Patrick: It's bizarre in it.

Patrick is making his way towards China Town and approaches a couple walking down the street towards him.
Patrick: Excuse me? Do you know where China Town is?
One of the couple turns to his right and points across the street saying 'That way'.
Patrick: Right, down there, wicked, cheers.
Patrick: It's where there's the big gate.

Patrick is making his way towards China Town.
Patrick: Right, Dave's just sent me a message – 'Go into the graffiti phonebox by the railings.' So I'm going send him a little message now.
Patrick (selects 'audio record'): 'Dave, **can you direct me to it. I'm outside the red phonebox outside er – oh bollocks** (looks around for street sign) **– outside Reyner Street**. So yeah, if you want to direct me there that would be wicked. Cheers mate.'
Patrick crosses the road, heading towards two red phone boxes.
Message from Dave: Have you found the postcard on top of the phone?
Message from Nicole: Hi Patrick - my postcard is at this phone box.
Patrick: Is he trying to goose me here. He's telling me that I've got to have a look on top of the phonebox. And Nicole, who likes my jacket, is saying that her postcard's here as well. So …

Patrick opens the phonebox and looks inside.
Patrick: There's nothing in there.
Patrick sees a postcard on the floor outside the phonebox and picks it up.
Patrick: All right (it says) 'Tell me about someone from your past who never leaves you.' Right, OK.
Patrick (selects 'audio record'): Right. 'Nicole, **tell me about someone from your past who never leaves you – that's what it says on your postcard** and it's on (looks around him for street sign) Charlotte Street.'
Message from Nicole: U R cool thankyou.
Message from Dave: U need the phonebox on Portland St - by the tower.
Patrick (looking at PDA): Right, and apparently there's another one. There's a phonebox – Dave's telling me there's a phonebox over by Portland Tower.

Figure 8. Assembling a coherent representation of interaction

Once assembled, internal and external resources - audio files and video sequences and stills in this case - may be linked to or embedded within each representation and a corpus of exhibits may then be assembled (e.g., Crabtree 2004) to support cooperative analysis between social scientists and computer scientists of the social organization of the technology-in-use (e.g., Crabtree et al. 2005). Whatever the added extras it is hopefully clear that text logs do not, in and as of themselves, contain data. Rather, the data must be produced through the analytically oriented working of resources internal to situated action and their combination with resources external to the setting of action. Data is constructed then and produced through the work-practices of the analyst. In this case those practices involve cleaning up the text log to identify salient features of interaction such as conversational threads between particular participants; a job of work that relies on exploiting and combining audio files and video recordings to make

determinations of relevance. The identification of salient features enables the analyst to start to make sense of the text logs. Making sense of the logs consists of transcribing audio files and the interaction recorded on and then splicing these texts together with text logs using the principle of conversational recognizability to assemble coherent representations of interaction that permit analysis.

Requirements for Record and Reuse

The themes of record and reuse open up the potential to develop eSocial Science applications and services that make the resources employed in particular settings of action into resources that simultaneously support social science research. Through case study we have sought to explicate something of the work that is implicated in using digital resources. The work involved in 'making the logs work' articulates broader challenges for the ongoing development of record and reuse applications that may, in the longer term, go beyond social science studies of technology. It is towards unpacking those challenges that we now turn.

In his classic account of sociological research methodology "Documents of Life" Plummer (2001) carefully sets out how a range of comparatively commonplace documents such as diaries and letters can be interrogated an analysed in order to afford various insights into the character of social life. Without buying into the 'narrative turn' that often accompanies such insights we want to point to the growing prevalence of 'digital documents of life' – of which text logs are but one example – as unavoidable and interesting features of everyday life and work. The use of computer-based applications and services has become a feature of everyday life in a great many settings and the emergence of Ubiquitous Computing makes it possible to consider instrumenting the digital environment of the future with record and replay components (e.g., ECT 2004). Future digital environments will extend beyond the desktop PC to embed computing in the wider environment, marrying computing applications to an array of sensors and mobile devices. Mobile phones are already commonplace and understanding the impact of new technologies such as these is already a matter of diverse social science concern (Rheingold 2002). When considering the potential of record and reuse we wish to move beyond substantive research interests, however, to consider the digital environment of the future as a resource for social science research more generally, for whether we like it or not human activities, interactions and relationships are, and will continue to be, increasingly mediated by technology. 'Getting on the inside' of technologically mediated action in circumstances where interaction is increasingly distributed, mobile and mediated through small (even invisible) devices would seem to be an important, if not an essential, part of social science research.

'Getting on the inside' has certainly been of value to our own research endeavours (Chevherst et al. 2003, Benford et al. 2005), but record and reuse is not limited to ethnographic concerns. Other researchers bring a broader set of concerns to the resources we have been working with. For example, computer scientists have drawn on the same sets of logs exploited by ethnographers in order to understand the frequency, resolution and accuracy of location-based technologies (e.g., Benford et al. 2004). This kind of analysis involves exploiting the spatial and temporal information the logs contain to generate visualizations of such things as GPS and WiFi coverage. Tied to ethnographic studies, these kinds of analysis provide resources to analyse the effects of technology on interaction and the ways in which technological effects are managed in interaction (e.g., Crabtree et al. 2004).

Whether designed to support studies of technology-in-use or for more general deployment in the digital environment of the future, the development of record and reuse applications will need to support a distinct process of use that is elaborated by the work involved in the ethnographic use of text logs and by alternate uses as indicated above. First it is necessary to support the capture of salient information, whether that is text messages and audio files or spatial and temporal information, etc. Obviously there is a need here to instrument as broad a range as possible of the digital components that are implicated in users interactions and to make them available as resources for all manner of analysts. There is also a need to develop editors that enable analysts to select a range of components that support the kinds of study they have a particular interest so that they might configure and tailor recording and capture in real time to meet their needs. It is also necessary to support extraction from the logs in order to enable the analyst to identify information that is salient to his or her studies. At the current moment in time extraction is largely a manual task, a matter of teasing out such things as conversational threads from the logs, for example. Clearly there is a need to support extraction, to make it more effective and efficient but we should not assume that it is a straightforward matter of automation.[4] Extraction makes it is necessary to support the manipulation of multiple media from within the digital environment and to support the combination of internal media with external media. The use of multiple media also makes it necessary to support synchronization so that extracted media from multiple sources (e.g., text messages, audio files, and video sequences) may be married together. Synchronization is also, currently, a manual task, a matter of extracting salient features from multiple sources by hand as it were through the use of a variety disparate applications (video cameras, movie capture and editing tools, audio players, spreadsheets, word processing applications, etc.) and interleaving them through the construction of unifying representations (such as transcripts or visualizations). The development of computer support in this area should make it possible to group extracted media together and play them besides one another in a single environment to support the more efficient production of data. Interleaving multiple media will require a further level of development, however, which is concerned to support the generation of a variety of unifying representations. Not only might the generation of transcripts or visualizations, or statistical charts, etc., be supported but those representations might also be married to the resources from which they were generated. This would support the probity of findings, enabling re-examination and analysis by others.

The process of record and reuse as articulated in Figure 9 very much reflects how things are done now, though the 'doing' is largely manual. While capture is automated (though editors are not yet available), extraction, the manipulation of multiple media within the digital environment, combination of internal with external media, synchronization of media, and production of representations is at the current moment a time consuming and laborious task. Consequently, and while recognizing the attendant challenges of extraction, there is a need to develop a suite of tools to support record and reuse. These should include:

- **Capture tools.** Tools that enable researchers to record events in digital environments and editors to tailor recording. Developments in Ubiquitous Computing have seen the emergence of the Equip Component Toolkit (ECT) at the University of Nottingham.

---

4 It is not at all clear, for example, just how conversational threads might be teased out by the machine as it often takes work on the analyst's part to recognize which utterances are related to which. As we have seen, the logs themselves do not represent the orderliness of conversation so any attempts at automating abstraction will have to reckon with and respond to issues of analytic complexity. How, for example, does one encode the principle of conversational recognizability?

While the toolkit has been designed to support the rapid configuration of future and emerging digital environments (Greenhalgh et al. 2004), efforts are underway to instrument components to support recording of events and states.

- **Extraction tools.** Tools that enable researchers to extract multiple sources of information from recorded logs, and which allow them to edit extracted information and combine it with media from external sources to produce unique datasets.

- **Synchronization tools.** A replay system that exploits time stamps to coordinate the use of the multiple media in a dataset, which enables cross referencing and indexing to support the splicing together of multiple media, and which enables multiple media to played side-by-side.

- **Representation tools.** Tools that support the production of representations from datasets and which preserve the relationship between representations and the media from which they are derived, and which enable source media to be recovered and viewed.

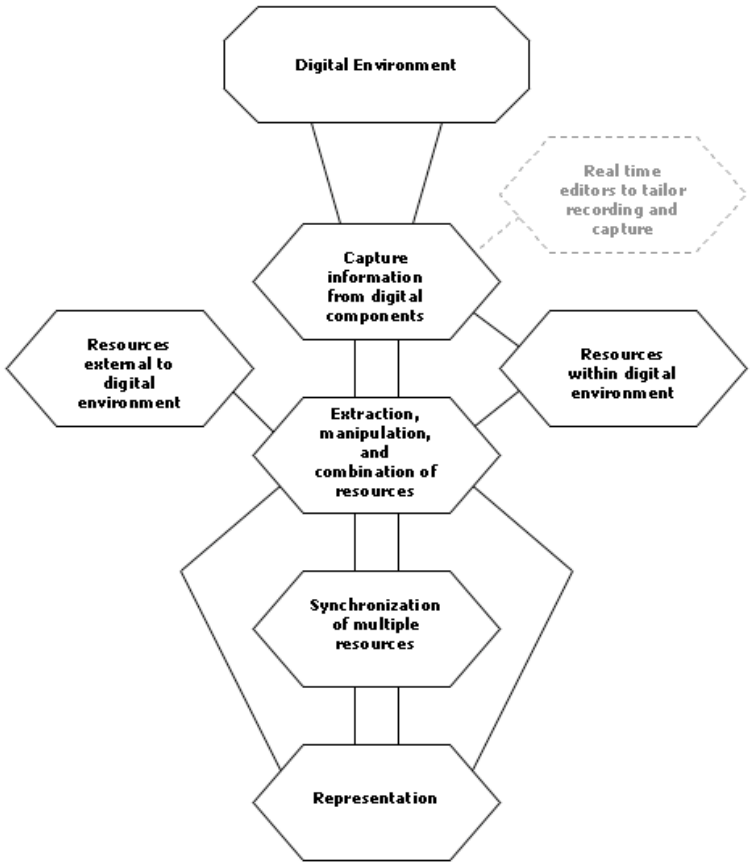

Figure 10. Process of record and reuse

These are, of course, high-level requirements. Working out just what they mean, and just what is to be built, will be one of the central tasks of the eSocial Science Research Node at the University of Nottingham and be the focus of future reports. Broadly speaking, the aim is to start with a modest and manageable challenge – working with text logs – and see what can be done to support the process of record and reuse through design before extending the approach to more diverse experiences (which are currently under development at the Mixed Reality

Laboratory at Nottingham). Through a process of incremental design the intention is to develop a suite of components and tools that may be exploited not only in the study of technology-in-use but more generally in social science studies of the digital environment of the future. An important aspect of technological development in this area revolves around the ethics of record and reuse. While the Research Node will seek to inform the development of an eSocial Science code of practice, ethical concerns extend beyond issues of managing digitally recorded materials that directly derive from activities 'in the wild'. In this respect the Research Node will also explore the development of tools that support the management of resources derived from public domains, controlling access to and restricting the use of material which has not been formally approved. However, the ethical challenges of record and reuse go much further than that. Social science researchers are today well aware of issues of informed consent but the digital environment of the future - where source materials may be exploited by multiple researchers for multiple purposes to produce a potentially wide variety of datasets, representations, analyses, and findings - introduces a new level of complexity to the debate. Just what informed consent means in such circumstances, just what someone is consenting to here, is not at all clear and a thorny problem to be reckoned with as eSocial Science develops into a recognizable field of research.

In conclusion, we have suggested that area of major potential for eSocial Science lies in the development of record and reuse applications and services. We have sought to articulate the nature of record and reuse through a case study of the work involved in using text logs generated in different digital environments. Reflection on the nature of that work has in turn been used to outline the process of record and reuse and to identify requirements for the continued development of record and reuse. In particular we have identified a process of work – capture, extraction (including manipulation and combination of internal and external resources), synchronization of resources, and the generation of representations to support analysis – which articulates requirements for the development of a suite of supporting tools. The development of these tools is being explored by the eSocial Science Research Node at the University of Nottingham. While the social science studies presented here have been ethnographic in character and concerned to understand the social organization of technology-in-use, the ultimate ambition of this programme of research is develop a broad range of components and tools that provide new data resources of more general purchase.

Acknowledgments

The research on which this article is based was funded by the eSocial Science Research Node at the University of Nottingham, the Equator Interdisciplinary Research Collaboration (www.equator.ac.uk), and the Interdisciplinary Research Collaboration in Dependability (www.dirc.org.uk). Thanks also go to Stuart Reeves for his work on documenting the record and reuse process.

References

Benford, S., Seagar, W., Flintham, M., Anastasi, R., Rowland, D., Humble, J., Stanton, D., Bowers, J., Tandavanitj, N., Adams, M., Row Farr, J., Oldroyd, A. and Sutton, J. (2004) "The error of our ways: the experience of self-reported position in a location-based game", *Proceedings of 6th International Conference on Ubiquitous Computing*, pp. 70-87, Nottingham, UK: Springer.

Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M. and Row Farr, J. (2005) "Can You See Me Now?", to appear in *ACM Transactions on Computer-Human Interaction*, ACM Press.

Benson, D. and Hughes, J.A (1991) "Method: evidence and inference", *Ethnomethodology and the Human Sciences* (ed. Button, G.), pp. 109-136, Cambridge: Cambridge University Press.

Button, G. (1992) "The curious case of the vanishing technology", *Technology in Working Order: Studies of Work, Interaction, and Technology* (ed. Button, G.), pp. 10-28, London: Routledge.

Cheverst, K., Clarke, K., Fitton, D., Rouncefield, M., Crabtree, A. and Hemmings, T. (2003) "SPAM on the menu: the practical use of remote messaging in community care", *Proceedings of the 2003 ACM Conference on Universal Usability*, pp. 23-29, Vancouver, Canada: ACM Press.

Cheverst, K., Fitton, D., Rouncefield, M., and Graham, C. (2004) 'Smart Mobs' and Technology Probes: Evaluating Texting at Work In Proceedings of ECITE 2004.

Crabtree, A. (2004) "The social life of Uncle Roy: field report", *Equator Technical Report*, www.mrl.nott.ac.uk/~axc/documents/The_Social_Life_of_Uncle_Roy_Field_Report.pdf

Crabtree, A., Benford, S., Rodden, T., Greenhalgh, C., Flintham, M., Anastasi, R., Drozd, A., Adams, M., Row-Farr, J., Tandavanitj, N. and Steed, A. (2004) "Orchestrating a mixed reality game 'on the ground'", *Proceedings of the 2004 CHI Conference on Human Factors in Computing Systems*, pp. 391-398, Vienna: ACM Press.

Crabtree, A., Rodden, T. and Benford, S. (2005) "The social life of Uncle Roy", under review for *Proceedings of the 9th European Conference on Computer Supported Cooperative Work*, September 18th-22nd December, Paris: Kluwer Academic Publishers.

ECT (2004) *The Equip Component Toolkit*, www.crg.cs.nott.ac.uk/~jym/ect/ect.php?page= Component_ Documentation

Foster, I., Kesselman, C. and Tuecke, S. (2001). "The anatomy of the Grid: enabling scalable virtual organizations", *International Journal of High Performance Computing Applications*, vol. 15 (3), pp. 200-222.

Graham, C., Cheverst, K., Fitton, D., and Rouncefield, M.,(2005) "How Do You Turn A Duck Into A Soul Singer? Put It In The Microwave Until Its Bill Withers": Some social features of a simple technology. In Proceedings of "Less Is More' Forum, Microsoft Cambridge UK.

Greenhalgh, C., Izadi, S., Mathrick, J., Humble, J. and Taylor, I. (2004) "ECT: a toolkit to support rapid construction of ubicomp environments", *Proceedings of 6th International Conference on Ubiquitous Computing* (Demonstration), September 7th-10th, Nottingham: Springer. http://ubicomp.org/ubicomp2004/adjunct/demos/greenhalgh.pdf

Heath, C. and J. Hindmarsh (2002) "Analyzing interaction: video, ethnography and situated conduct", *Qualitative Research in Action* (ed. May, T.), pp. 99-121, London: Sage.

Hine, C. (1998) *Virtual Ethnography*, London: Sage.

Hughes, J. and Sharrock, W. (1997) *The Philosophy of Social Research* (3rd edition), London: Longman.

Hughes, J., Rouncefield, M., abd Tolmie, P. (2002) Representing Knowledge: Instances of Information Management. *The British Journal of Sociology* Vol 53, No. 2 pp 221-238

Lynch, M. (1993) Scientific Practice and Ordinary Action: Ethnomethodological and Social Studies of Science, Cambridge: Cambridge University Press.

Park, R.E., Burgess, E.W. and McKenzie, R.D. (1925) *The City*, Chicago: University of Chicago Press.

Plummer, K. (2001) *Documents of Life2: An invitation to critical humanism*. London. Sage.

Prosser, J. (ed.) (1998) Image Based Research: A Sourcebook for Qualitative Researchers, London: Falmer Press.

Prus, R. (1996) "The ethnographic research tradition", *Symbolic Interaction and Ethnographic Research*, pp. 103-140, New York: State University of New York Press

Wieder, D. and Zimmerman, D. (1977) "The diary: diary interview methods", *Urban Life*, vol. 5 (4), pp. 479-498.