Exploiting Digital Records: New Resources and Tools for Qualitative Research in Contemporary Social Science *

Andy Crabtree, Chris Greenhalgh, Andy French, Tom Rodden and Steve Benford

Mixed Reality Laboratory, School of Computer Science & IT, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, United Kingdom.
{axc, apf, cmg, sdb, tar}@cs.nott.ac.uk

Abstract. A wide range of computational tools are currently available for use within qualitative research, yet very few of these have actually been designed to support the diverse needs of the social sciences. The emergence of Grid computing and e-Social Science raises new possibilities for social science research, however. e-Social Science provides the opportunity to move beyond existing technologies and the promises of hypermedia to consider the production of *digital records*. Digital records consist of two distinct components: 1) resources internal to computational environments (such as text messages, voicemail, email, etc.), and 2) external resources gathered by a field worker (video and audio recordings, photographs, etc.). We present *Replay Tool*, which enables qualitative researchers to combine internal and external resources and generate faithful representations of social order in the media rich computational environments that populate everyday life.

1. Introduction

A wide range of computational tools are currently available for use within qualitative research,[1] yet very few of these have actually been designed to support the diverse needs of the social sciences. Tools offered for qualitative research have been developed across education, nursing, disease control, animal behaviour, time and motion study, and (by far the largest category) generic document and text analysis, for example. The consequence is that many of the qualitative software tools offered by developers are of extremely limited value to social science researchers. A much smaller number of software packages than initially meets the eye actually respond to the needs of the qualitative social scientist. Tools such as Anvil, Atlas.ti and The Observer enable the social science researcher to exploit text, photography, audio, and video and support widespread practices of annotation and coding.

The emergence of Grid computing (Foster et al. 2001) raises new possibilities for social science research. The Grid, with its emphasis on high performance computing,

* © Andy Crabtree, 2006. Paper presented at Words Instead of Numbers: The Status of Software in the Qualitative Research World, September 29, Imperial College London: Association for Survey Computing.

[1] See, for example, www.eval.org/Resources/QDA.htm

naturally lends itself to quantitative research. The vision, however, is beguiling as even large-scale surveys produce computationally trivial amounts of data (Crabtree and Rouncefield 2005); indeed, qualitative datasets rich in multimedia are routinely larger than their quantitative counterparts. The problem of scale in the social sciences revolves not around data size then, but around *analytic* matters to do with the 'macro' and the 'micro' and the perennial arguments that inhabit scientific debate between quantitative and qualitative researchers (Hughes and Sharrock 1997). Consequently, the emergence of the Grid opens up new possibilities not only for quantitative research but for qualitative research as well, and for the development of new resources and tools in particular.

In this paper we wish to consider the development of new resources and tools that support qualitative research in the digital age. In contemporary society computing permeates everyday life, cutting across work, education, leisure and play. Understanding the impact of computing on everyday life has been a longstanding concern in the social sciences, yet remarkably few resources and tools exist to support qualitative research in this area. Throughout its historical development, qualitative research has incorporated the technology of the times into its investigations, ranging from the humble pen and paper to photography, audio recording and video. While computational media are a primary focus of contemporary qualitative research (e.g., Hine 1998), the incorporation of computational media into qualitative research is still in its infancy. Qualitative researchers may well exploit digital audio and video devices and the Internet to gather data (such as emails, web pages, blogs, etc.), but the power of computing offers a great deal more than that (Dicks et al. 2005).

The emergence of e-Social Science as a feature of Grid development provides the opportunity to move beyond existing technologies and the promises of 'hypermedia' (ibid.) to consider the production of *digital records*.[2] Digital records may be seen as a natural evolution of the 'ethnographic record' (Wolcott 1999) or that collection of materials that the qualitative researcher gathers in the course of his or her inquiries, which become the subsequent focus of analysis. Digital records consist of two distinct components: 1) *external* records such as video, field notes, photographs, working documents, and the other material gathered by the qualitative researcher from a setting; and 2) *internal* records which include the set of digital media (text messages, voice mails, emails, etc.) exploited by users in the course of their collaborations within media rich computational environments.

Digital records make internal resources available to qualitative research by *logging* or recording the digital media that people use in the course of their collaborative activities, providing an unprecedented level of access to everyday interaction in media rich computational environments (Crabtree et al. 2006a). Added to this, software developments in e-Social Science (French et al. 2006) make it possible to *replay* the contents of digital records much as one might replay a video recording to support (re)inspection and analysis. Computer support also extends to external resources and enables the qualitative researcher to replay them *alongside* internal resources to provide comprehensive real-time views on digitally mediated interaction and collaboration.

[2] www.ncess.ac.uk/nodes/digitalrecord

Our interest in the development of new computational resources and tools for qualitative research is *practical* rather than conceptual. Predicated on co-design between qualitative researchers and computer scientists, we examine qualitative research practices which turn internal resources into *usable* resources and present computational tools that enable qualitative researchers to marry internal and external resources together. Specifically our bespoke software, *Replay Tool*, enables qualitative researchers to:

- Visualize the content of internal (system) recordings of interaction;
- Extract sequences of interaction and collaboration of relevance to analysis;
- Remove non-relevant features from internal records;
- Synchronize internal records with external records;
- Add annotations;
- Re-order synchronized internal and external records for purposes of analysis.

The need to re-order digital records is brought about by the nature of recording in digital environments. Computer systems log or record events – the sending of a text message, say – according to system time or the time when the event is registered by the system. System time is not the same as interaction time, however. Thus, when an event is registered by a system is not the same as when that event is *acted upon* by the recipient. There is, then, a fundamental difference between the logged order of events, or the order of interaction recorded by a computer system, and the interactional order of events itself. Digital records offer a seductive representation of social order and there is a real need for qualitative researchers to be aware of, and for designers to respond to, the immanent risk of reification (Crabtree et al. 2006b).

Replay Tool provides tools which enable qualitative researchers to construct faithful representations of social order in media rich computational environments. The development of *Replay Tool* also highlights a number of future challenges, including the development of novel mechanisms and visualizations for recording and indexing digital records; the development of mechanisms to better handle temporal issues such as representing interactional order and managing temporal slippage in recordings; and the development of grid-based mechanisms that enable qualitative researchers to perform structured forms of analysis. While still in the early stages of development, existing *Replay Tool* functionality demonstrates the potential to support qualitative research across a burgeoning array of digital environments that are emerging to populate everyday life.

2. Working with Digital Records

Our own experiences of working with digital records emerges from qualitative studies of emerging technological environments, and what are referred to by the technological community as 'hybrid ubiquitous computing environments' in particular

(Chalmers and Areti 2004). These environments support collaboration through the use of heterogeneous interaction mechanisms. They distribute interaction between physical and digital settings, so that (for example) persons online exploit virtual environments and text messaging to interact with persons in the physical world, who exploit mobile location-based technologies and audio messaging to interact with their online counterparts. Hybrid ubiquitous computing environments exploit broadband and wireless infrastructure alongside invisible sensing systems (such as GPS and WiFi) to provide new forms of computing that move users away from the workplace and the desktop and situate computing in the physical fabric of everyday life (Weiser 1991). Their development is still in its infancy, although the underlying technical infrastructure is already a commonplace feature of everyday life: GPS systems are widely used by drivers today, broadband is a common feature of computing in the home, and WiFi 'hotspots' are increasingly found throughout towns and cities around the world, for example. Hybrid ubiquitous computing environments seek to further develop this infrastructure through the construction of a novel range of 'experience' projects (Equator 2006) that deploy heterogeneous technical arrangements in the wild where they are available to members of the public.

Our role as qualitative researchers in this domain is to study these public deployments and unpack their social features in order to inform their continued development and ensure that emerging systems resonate with the practical circumstances of their use. To do this, we exploit internal and external recordings of interaction – i.e., recordings of interaction that are system-based and detail interaction within the digital environment, and audio-visual recordings of interaction that are produced by a field worker tracking members of the public in their interactions. In turn, this provides us with a rich picture of interaction across physical and digital domains. What we want to do here is articulate the work that is involved in combining system-based recordings, or 'system logs' as they are often called, with external recordings, particularly video. As a result of understanding the manual work that is involved in exploiting system recordings, we have developed a computer-supported approach which we call 'record and replay' (Crabtree et al. 2006a). Below we consider the manual work that is involved in exploiting system-recordings alongside external recordings before presenting our record and replay system: *Replay Tool*.

It is worth noting, before we proceed, that qualitative research is a heterogeneous enterprise that encapsulates a diverse range of approaches. For our own part, we work in a 'naturalistic' tradition and so seek develop representational accounts of collaboration and interaction that display naturally accountable features of interaction (Garfinkel 1967); that is, which display the observable, reportable, remarked upon, and responded to features of collaboration as visibly manifest in participants' interactions. We suspend a concern with the theoretical features of interaction then, and while this is perhaps an uncommon approach to qualitative research (Heath and Button 2002), there is nevertheless a strong element of generality to the work involved in working with system recordings whatever the analytic persuasion: system logs must be parsed, for example, relevant data must be located, multiple recordings must be synchronized, annotations (whether theoretical or naturalistic in nature) will invariably be made, etc.

2.1 Working with System Recordings

In order to elaborate the work involved in working with multiple recordings we focus on the use of recordings from a recent hybrid ubicomp game. Studying games might seem like a trivial affair for social scientists to be involved with. However, games offer a safe and engaging environment to trial future and emerging technologies and to understand many of their social characteristics (Benford et al. 2006). There is a methodological value to constructing and deploying games in the wild then (Crabtree 2004). This particular game has been deployed at a number of public venues across the UK for 2 weeks at a time, attracting over 1000 members of the public each time. The game was co-designed by the performing arts group Blast Theory (www.blasttheory.co.uk) and members of the Mixed Reality Laboratory at the University of Nottingham. The game is both an artistic exploration of the theme of trust and a technical exploration of mobile location-based technology. Technically, the game exploits commercially available GPRS data services to extend prior work on location-based experiences by exploring the use of self-reported positioning, in contrast to automated GPS positioning. For a full description of the game and its technical arrangement see Benford et al. (2004).

The game is played on the streets and online. Street players pay £3 to take part in the experience. They are equipped with a GPRS-enabled PDA and roam 1 sq. km of city streets exploiting a map on the PDA to find their way to various locations 'pointed to' by clues issued by the game server.

Figure 1. Street players mobile interface.

Online players play for free, inhabit a virtual facsimile of the city streets, and roam around through the use of avatars. When street players declare their location on the

PDA map a street player avatar appears in the virtual world alongside a player 'card', which provides a photograph and physical description of the street player (Figure 2). Online players can now see the street player and collaborate with them by sending text messages. Street players may respond to online players by sending short (7 second) audio messages from their PDAs.

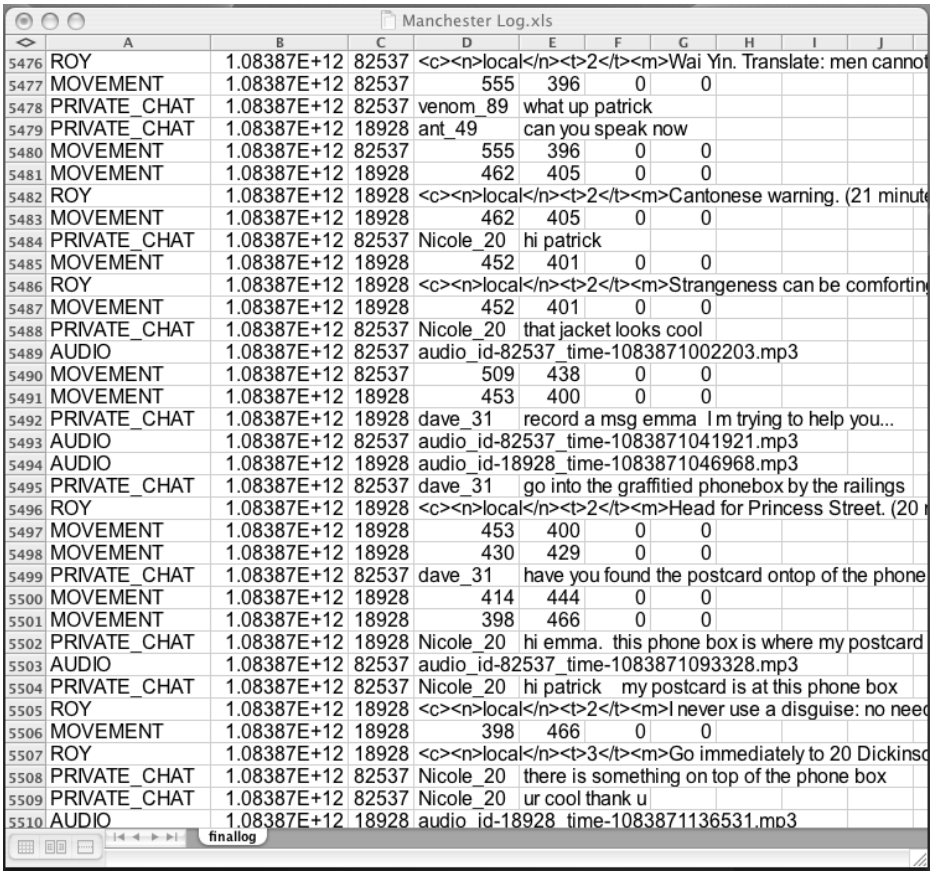
Figure 2. Online players interface showing street player declaring position.

Street players and online players are obliged to collaborate with each other if the game is to be completed. A small number of physical postcards are assigned to each online player. These are distributed throughout the physical city and the so the online player must enlist the help of a street player to locate one of the cards. In return, the online player is provided with instructions from the game server that he or she must use to guide the street player to a certain location in the city where the two take part in an orchestrated event (see www.blasttheory.co.uk/bt/work_uncleroy.html for details).

In order to support qualitative study of interaction between street players and online players, the system recorded location reports (i.e., places where street players declared their positions), clues and instructions from the game server, text messages sent by online players, and audio messages sent by street players. This data was complemented by video recordings of particular street players journey through the physical city, gathered by a field worker. The dataset provided the focus for subsequent analysis of interaction in the game. We focus here on a small subset of the dataset – a short sequence of interaction that spans internal and external recordings – for obvious reasons of space. The work elaborated by the subset is generic and applies across the entire dataset and others besides (e.g., Flintham et al. 2003, Crabtree et al. 2004, Benford et al. 2006).

The raw data recorded by the system was exported to a spreadsheet (Figure 3), with events being ordered according to the time they were registered by the game server. Audio messages recorded by street players were referenced in the spreadsheet and provided separately in mp3 files, each labeled as it is in the spreadsheet (e.g., audio_id_82537_time_1083871041921). Although the system log captures internal

features of interaction or those features of interaction that occur within the hybrid computing environment, it is evident in Figure 1 that they are not readily accessible or amenable to analysis. The log is messy, largely unintelligible, and peppered with computational noise, such as obscure codes, symbols, and machine-readable time references. In order to turn it into a usable resource it is necessary to clean the log so that the researcher can identify salient features. Naturally, salient features may vary depending on the kind of analysis undertaken. What is much less variable is the need to 'clean the log', to make it accessible to qualitative researchers.



	A	B	C	D	E	F	G	H	I	J
5476	ROY	1.08387E+12	82537	<c><n>local</n><t>2</t><m>Wai Yin. Translate: men cannot						
5477	MOVEMENT	1.08387E+12	82537	555	396	0	0			
5478	PRIVATE_CHAT	1.08387E+12	82537	venom_89	what up patrick					
5479	PRIVATE_CHAT	1.08387E+12	18928	ant_49	can you speak now					
5480	MOVEMENT	1.08387E+12	82537	555	396	0	0			
5481	MOVEMENT	1.08387E+12	18928	462	405	0	0			
5482	ROY	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Cantonese warning. (21 minute						
5483	MOVEMENT	1.08387E+12	18928	462	405	0	0			
5484	PRIVATE_CHAT	1.08387E+12	82537	Nicole_20	hi patrick					
5485	MOVEMENT	1.08387E+12	18928	452	401	0	0			
5486	ROY	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Strangeness can be comfortin						
5487	MOVEMENT	1.08387E+12	18928	452	401	0	0			
5488	PRIVATE_CHAT	1.08387E+12	82537	Nicole_20	that jacket looks cool					
5489	AUDIO	1.08387E+12	82537	audio_id-82537_time-1083871002203.mp3						
5490	MOVEMENT	1.08387E+12	82537	509	438	0	0			
5491	MOVEMENT	1.08387E+12	18928	453	400	0	0			
5492	PRIVATE_CHAT	1.08387E+12	18928	dave_31	record a msg emma I m trying to help you...					
5493	AUDIO	1.08387E+12	82537	audio_id-82537_time-1083871041921.mp3						
5494	AUDIO	1.08387E+12	18928	audio_id-18928_time-1083871046968.mp3						
5495	PRIVATE_CHAT	1.08387E+12	82537	dave_31	go into the graffitied phonebox by the railings					
5496	ROY	1.08387E+12	18928	<c><n>local</n><t>2</t><m>Head for Princess Street. (20 r						
5497	MOVEMENT	1.08387E+12	18928	453	400	0	0			
5498	MOVEMENT	1.08387E+12	18928	430	429	0	0			
5499	PRIVATE_CHAT	1.08387E+12	82537	dave_31	have you found the postcard ontop of the phone					
5500	MOVEMENT	1.08387E+12	18928	414	444	0	0			
5501	MOVEMENT	1.08387E+12	18928	398	466	0	0			
5502	PRIVATE_CHAT	1.08387E+12	18928	Nicole_20	hi emma. this phone box is where my postcard					
5503	AUDIO	1.08387E+12	82537	audio_id-82537_time-1083871093328.mp3						
5504	PRIVATE_CHAT	1.08387E+12	82537	Nicole_20	hi patrick my postcard is at this phone box					
5505	ROY	1.08387E+12	18928	<c><n>local</n><t>2</t><m>I never use a disguise: no neec						
5506	MOVEMENT	1.08387E+12	18928	398	466	0	0			
5507	ROY	1.08387E+12	18928	<c><n>local</n><t>3</t><m>Go immediately to 20 Dickinsc						
5508	PRIVATE_CHAT	1.08387E+12	82537	Nicole_20	there is something on top of the phone box					
5509	PRIVATE_CHAT	1.08387E+12	82537	Nicole_20	ur cool thank u					
5510	AUDIO	1.08387E+12	18928	audio_id-18928_time-1083871136531.mp3						

Figure 3. Extract from the system recording of interaction between players.

Salient features consist of those features of the log that enable us to understand interaction between online and street players as it naturally unfolds. Clues and messages are salient then. However, for our purposes, salience is not determined by the log alone but with reference to external resources as well, in this case with reference to video of interaction that has been captured by accompanying players on the streets. Cleaning up the record consists then of identifying parts of the log that 'sit alongside' external resources. What we are after is the parts of the log that accompany the external resources we have gathered.

In our initial cut through the data, the identification of salient features was done by searching, finding and extracting from amongst the 5000+ entries that populate the log, the clue and message sequences that matched the interaction captured on video. The extracts were then cleaned up to make them more intelligible. Obscure arrangements of numbers and symbols, which are analytically meaningless features of the log for the qualitative researcher, were deleted and players were given their 'proper' names – i.e., the names they gave themselves in the course of playing the game, which are not necessarily their real names. Cleaning up the log also required that non-relevant features of interaction were stripped out. For purposes of our analysis 'non-relevant' features consist of text and audio messages that are not related to 'what is going on' in the video we have captured. Basically, the system log can be seen to contain and represent multiple interactional threads or multiple sequences of conversation occurring between multiple parties. Only some of these threads are relevant to the discrete sequences of interaction recorded on video and so the others must be removed, as they have no work to do in the analysis of interaction here. The work of cleaning the log continues then with an eye towards identifying and extracting only those features that relate to the interactional sequences recorded on video. In this particular example, what is of interest are the interactional threads involving a street player called Patrick (audio ID 82537), online players called Venom, Nicole and Dave, and clues or instructions from the game server (Figure 4).

Manchester.xls

	A	B
1	Game Server	Wai Yin. Translate: men cannot enter. Make appropriate move. (23 minutes remaining)
2	Venom	what up patrick
3	Nicole	hi patrick
4	Nicole	that jacket looks cool
5	Audio iD 82537	Time: 1083871002203.mp4
6	Audio ID 82537	Time: 1083871041921.mp4
7	Dave	go into the graffitied phonebox by the railings
8	Dave	have you found the postcard ontop of the phone
9	Audio ID 82537	Time: 1083871093328.mp4
10	Nicole	hi patrick my postcard is at this phone box
11	Nicole	there is something on top of the phone box
12	Nicole	ur cool thank u
13	Dave	u need the phonebox on portland st by the tower
14	Audio ID 82537	Time: 1083871167375.mp4
15	Nicole	huh...um someone from my past...
16	Dave	ok mate well done I'm waiting for directions
17	Audio ID 82537	Time: 1083871209640.mp4
18	Nicole	is it a riddle
19	Game Server	Where are you? Press ACTIONS then choose I AM HERE to let me know.
20	Nicole	ok my best friend royce. he is so great
21	Audio ID 82537	Time: 1083871245828.mp4
22	Nicole	and has always stuck by me
23	Game Server	An ambulance stopped here. It looks like someone I know. I turned away towards the bank

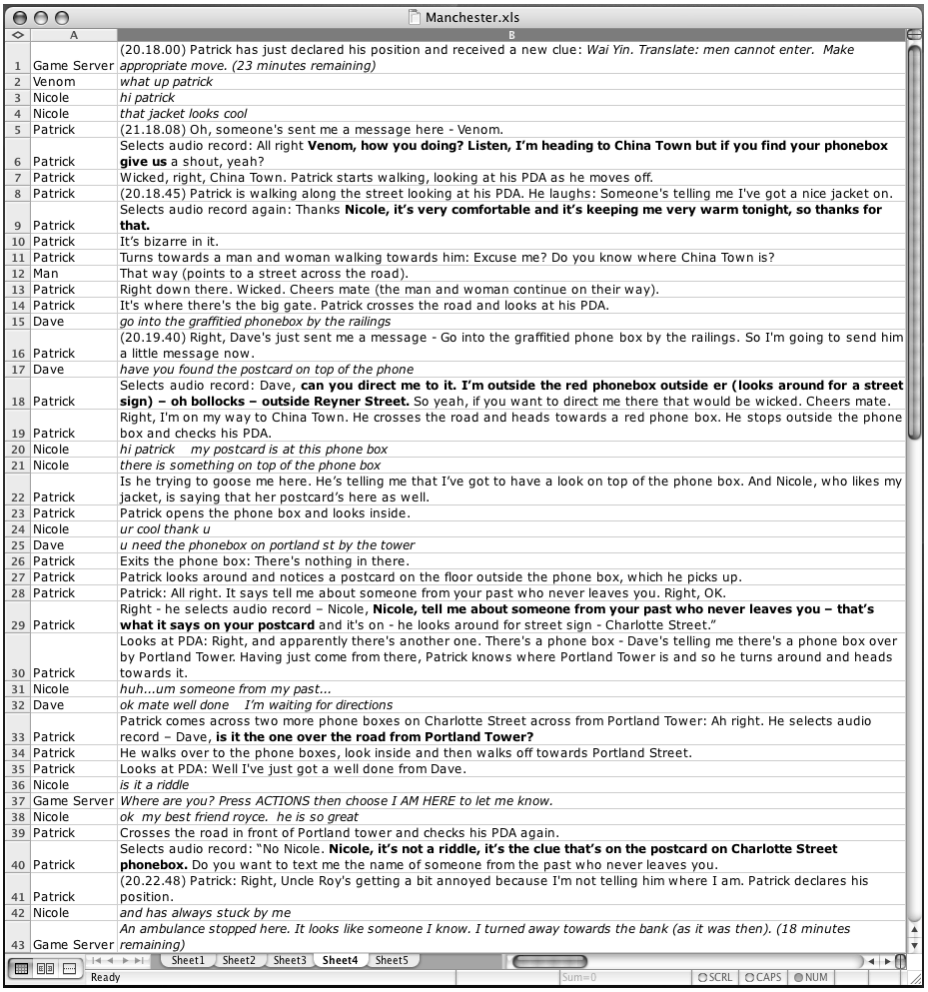
Sheet1 Sheet2 Sheet3 Sheet4

Figure 4. The cleaned up log extract.

2.2 Combining Multiple Recordings

What the cleaning process leaves us with, as can be clearly seen in Figure 4, is a rather sparse representation of interaction. There is a need now to 'thicken it up' (Crabtree et al. 2006a) or create a much richer representation of interaction by adding the

contents from other recordings. Thus, and for example, the contents of audio files might be added, which first requires that the audio files be transcribed. There are many ways in which the work of transcription may be accomplished – from simple textual renderings to highly structured renderings of the sort provided by Conversation and Interaction Analysis (Sacks et al 1974, Jordan and Henderson 1995). All might benefit by converting system time (Column B, Figure 3) into ordinary clock time, though this is essential to the more exacting approaches to transcription. Whatever approach is used, transcriptions then need to be synchronized with the contents of the system log along with the contents of video and any other resources that the qualitative researcher might gather, such as photographs and field notes. Video content is synchronized by adding further transcriptions of talk and descriptions of practical action carried out on the streets. The combination of internal and external recordings enables the qualitative researcher to reconcile the various fragments of distributed interaction that are contained in various records (e.g., in system logs, audio files and on video). In turn this provides rich descriptions of interaction between participants and enables inspection of the interaction between participants (Figure 5).



Manchester.xls

	A	B
1	Game Server	(20.18.00) Patrick has just declared his position and received a new clue: *Wai Yin. Translate: men cannot enter. Make appropriate move. (23 minutes remaining)*
2	Venom	*what up patrick*
3	Nicole	*hi patrick*
4	Nicole	*that jacket looks cool*
5	Patrick	(21.18.08) Oh, someone's sent me a message here - Venom.
6	Patrick	Selects audio record: All right **Venom, how you doing? Listen, I'm heading to China Town but if you find your phonebox give us** a shout, yeah?
7	Patrick	Wicked, right, China Town. Patrick starts walking, looking at his PDA as he moves off.
8	Patrick	(20.18.45) Patrick is walking along the street looking at his PDA. He laughs: Someone's telling me I've got a nice jacket on.
9	Patrick	Selects audio record again: Thanks **Nicole, it's very comfortable and it's keeping me very warm tonight, so thanks for that.**
10	Patrick	It's bizarre in it.
11	Patrick	Turns towards a man and woman walking towards him: Excuse me? Do you know where China Town is?
12	Man	That way (points to a street across the road).
13	Patrick	Right down there. Wicked. Cheers mate (the man and woman continue on their way).
14	Patrick	It's where there's the big gate. Patrick crosses the road and looks at his PDA.
15	Dave	*go into the graffitied phonebox by the railings*
16	Patrick	(20.19.40) Right, Dave's just sent me a message - Go into the graffitied phone box by the railings. So I'm going to send him a little message now.
17	Dave	*have you found the postcard on top of the phone*
18	Patrick	Selects audio record: Dave, **can you direct me to it. I'm outside the red phonebox outside er (looks around for a street sign) – oh bollocks – outside Reyner Street.** So yeah, if you want to direct me there that would be wicked. Cheers mate.
19	Patrick	Right, I'm on my way to China Town. He crosses the road and heads towards a red phone box. He stops outside the phone box and checks his PDA.
20	Nicole	*hi patrick my postcard is at this phone box*
21	Nicole	*there is something on top of the phone box*
22	Patrick	Is he trying to goose me here. He's telling me that I've got to have a look on top of the phone box. And Nicole, who likes my jacket, is saying that her postcard's here as well.
23	Patrick	Patrick opens the phone box and looks inside.
24	Nicole	*ur cool thank u*
25	Dave	*u need the phonebox on portland st by the tower*
26	Patrick	Exits the phone box: There's nothing in there.
27	Patrick	Patrick looks around and notices a postcard on the floor outside the phone box, which he picks up.
28	Patrick	Patrick: All right. It says tell me about someone from your past who never leaves you. Right, OK.
29	Patrick	Right - he selects audio record – Nicole, **Nicole, tell me about someone from your past who never leaves you – that's what it says on your postcard** and it's on - he looks around for street sign - Charlotte Street."
30	Patrick	Looks at PDA: Right, and apparently there's another one. There's a phone box - Dave's telling me there's a phone box over by Portland Tower. Having just come from there, Patrick knows where Portland Tower is and so he turns around and heads towards it.
31	Nicole	*huh...um someone from my past...*
32	Dave	*ok mate well done I'm waiting for directions*
33	Patrick	Patrick comes across two more phone boxes on Charlotte Street across from Portland Tower: Ah right. He selects audio record – Dave, **is it the one over the road from Portland Tower?**
34	Patrick	He walks over to the phone boxes, look inside and then walks off towards Portland Street.
35	Patrick	Looks at PDA: Well I've just got a well done from Dave.
36	Nicole	*is it a riddle*
37	Game Server	*Where are you? Press ACTIONS then choose I AM HERE to let me know.*
38	Nicole	*ok my best friend royce. he is so great*
39	Patrick	Crosses the road in front of Portland tower and checks his PDA again.
40	Patrick	Selects audio record: "No Nicole. **Nicole, it's not a riddle, it's the clue that's on the postcard on Charlotte Street phonebox.** Do you want to text me the name of someone from the past who never leaves you.
41	Patrick	(20.22.48) Patrick: Right, Uncle Roy's getting a bit annoyed because I'm not telling him where I am. Patrick declares his position.
42	Nicole	*and has always stuck by me*
43	Game Server	*An ambulance stopped here. It looks like someone I know. I turned away towards the bank (as it was then). (18 minutes remaining)*

Sheet1 Sheet2 Sheet3 Sheet4 Sheet5

Ready Sum=0 SCRL CAPS NUM

Figure 5. The synchronized log, combining audio and video with system data.

Inspection of the synchronized log is not a matter of running the data through software packages to generate statistical representations and visualizations, for example, or of exploiting text-mining techniques to locate phenomenon of interest. We do not exclude these issues from consideration in analysis or design however; we only put them on hold for purposes of our analysis. Naturalistic analysis is a matter of human skill and judgement to some large extent and consists of manually inspecting and examining thick descriptions of interaction in order to identify its socially organized features.[3] It is in respect of the socially organized features of interaction that the synchronized log *displays* that it becomes both analytically interesting and troublesome.

[3] That is thick description *pace* the philosopher Gilbert Ryle, not the anthropologist and cultural theorist Clifford Geertz. See Crabtree (2003) to unpack the distinction.

Firstly, the intertwining of text content, audio content and video content brings the log to life in that the contents of the log start to assume some kind of *recognizable sense*. Unlike the representation of interaction in Figures 3 or 4, Figure 5 allows us to see that such things as greetings and introductions are made, that collaboration sometimes goes no further, that the local knowledge of passing members of the public is drawn upon by street players to navigate the city streets, that instructions are issued by online players which orient street players to specific features of the streets (phone boxes, railings, buildings, etc.), that collaborations ensue and are directed towards finding postcards for the online players, that such collaborations lead street players off track and require intervention by the game server, etc.

We can also see that what online players hear from street players, as conveyed to them by the contents of audio files, does not represent all of what is said by street players. The bold typing in Figure 5 represents the contents of audio files (italic text represents text messages sent by online players and the game server). From this we can see that such things as requests for directions or for particular actions to be done are not always heard. This may impact upon interaction and in part account for the occasional breakdowns in communication between players. The synchronized log also makes it visible that not all 'utterances' are treated as relevant by street players to situated action on the streets. For example, Patrick's actions do not turn upon Nicole's utterances 24, 31, 38, and 42. This is not to say that Patrick does not see them but that they are not relevant to or responded to in Patrick's ongoing interactions 'here and now' on the streets. Neither is it to say that these utterances have no part in playing the game – while Patrick does not respond to all messages, some of them they are important to gameplay, with utterances 38 and 42 triggering a response from the game server which will furnish Nicole with the information needed to guide Patrick to his ultimate destination.

2.3 Handling the Reification of Interactional Order

The synchronized log reveals that interaction on the streets is not driven by each and every utterance made by players online but is instead shaped by the exigencies of the situation on the ground, and it is in this respect that the synchronized log as it stands in Figure 5 is analytically troublesome. Organizing content in terms of system time, the system log offers a seductive representation of the unfolding order of interaction. However, it is critical to appreciate that when distributed fragments of interaction enter the interactional situation is *not the same* as when they are logged by the system. System order and interactional order are fundamentally different. System order is based on the measurable linear progression of some standard unit of time (cesium resonance, for example). Interactional order, on the other hand, is based on the exigencies of practical action. This has serious ramifications for the way in which interaction is understood in existing and emerging computational environments. Consider, for example, an attempt to understand the use of text messaging with mobile phones: when a message is sent and logged by the system is not necessarily the same time as when it is read and *acted upon* by the recipient. The recipient of the text message may not reply and otherwise act upon the message until some time after they have received it. The logged order reifies interactional order then, representing interaction in terms

of system time but not *action time*, and so stands in our way of developing a real world, real time understanding of interaction.

Manchester.xls

	A	B	C
1	1	Game Server	(20.18.00) Patrick has just declared his position and received a new clue: *Wai Yin. Translate: men cannot enter. Make appropriate move. (23 minutes remaining)*
2	2	Venom	what up patrick
3	5	Patrick	(21.18.08) Oh, someone's sent me a message here - Venom.
4	6	Patrick	Selects audio record: All right **Venom, how you doing? Listen, I'm heading to China Town but if you find your phonebox give us** a shout, yeah?
5	7	Patrick	Wicked, right, China Town. Patrick starts walking, looking at his PDA as he moves off.
6	3	Nicole	*hi patrick*
7	4	Nicole	*that jacket looks cool*
8	8	Patrick	(20.18.45) Patrick is walking along the street looking at his PDA. He laughs: Someone's telling me I've got a nice jacket on.
9	9	Patrick	Selects audio record again: Thanks **Nicole, it's very comfortable and it's keeping me very warm tonight, so thanks for that.**
10	10	Patrick	It's bizarre in it.
11	11	Patrick	Turns towards a man and woman walking towards him: Excuse me? Do you know where China Town is?
12	12	Man	That way (points to a street across the road).
13	13	Patrick	Right down there. Wicked. Cheers mate (the man and woman continue on their way).
14	14	Patrick	It's where there's the big gate. Patrick crosses the road and looks at his PDA.
15	15	Dave	*go into the graffitied phonebox by the railings*
16	16	Patrick	(20.19.40) Right, Dave's just sent me a message - Go into the graffitied phone box by the railings. So I'm going to send him a little message now.
17	18	Patrick	Selects audio record: Dave, **can you direct me to it. I'm outside the red phonebox outside er (looks around for a street sign) – oh bollocks – outside Reyner Street.** So yeah, if you want to direct me there that would be wicked. Cheers mate.
18	19	Patrick	Right, I'm on my way to China Town. He crosses the road and heads towards a red phone box. He stops outside the phone box and checks his PDA.
19	17	Dave	*have you found the postcard on top of the phone*
20	20	Nicole	*hi patrick my postcard is at this phone box*
21	21	Nicole	*there is something on top of the phone box*
22	22	Patrick	Is he trying to goose me here. He's telling me that I've got to have a look on top of the phone box. And Nicole, who likes my jacket, is saying that her postcard's here as well.
23	23	Patrick	Patrick opens the phone box and looks inside.
24	24	Nicole	*ur cool thank u*
25	26	Patrick	Exits the phone box: There's nothing in there.
26	27	Patrick	Patrick looks around and notices a postcard on the floor outside the phone box, which he picks up.
27	28	Patrick	Patrick: All right. It says tell me about someone from your past who never leaves you. Right, OK.
28	29	Patrick	Right - he selects audio record – Nicole, Nicole, **tell me about someone from your past who never leaves you – that's what it says on your postcard** and it's on - he looks around for street sign - Charlotte Street."
29	25	Dave	*u need the phonebox on portland st by the tower*
30	30	Patrick	Looks at PDA: Right, and apparently there's another one. There's a phone box - Dave's telling me there's a phone box over by Portland Tower. Having just come from there, Patrick knows where Portland Tower is and so he turns around and heads towards it.
31	33	Patrick	Patrick comes across two more phone boxes on Charlotte Street across from Portland Tower: Ah right. He selects audio record – Dave, **is it the one over the road from Portland Tower?**
32	34	Patrick	He walks over to the phone boxes, look inside and then walks off towards Portland Street.
33	32	Dave	*ok mate well done I'm waiting for directions*
34	35	Patrick	Looks at PDA: Well I've just got a well done from Dave.
35	31	Nicole	*huh...um someone from my past...*
36	36	Nicole	*is it a riddle*
37	40	Patrick	Selects audio record: "No Nicole. **Nicole, it's not a riddle, it's the clue that's on the postcard on Charlotte Street phonebox.** Do you want to text me the name of someone from the past who never leaves you.
38	38	Nicole	*ok my best friend royce. he is so great*
39	42	Nicole	*and has always stuck by me*
40	39	Patrick	Crosses the road in front of Portland tower and checks his PDA again.
41	37	Game Server	*Where are you? Press ACTIONS then choose I AM HERE to let me know.*
42	41	Patrick	(20.22.48) Patrick: Right, Uncle Roy's getting a bit annoyed because I'm not telling him where I am. Patrick declares his position.
43	43	Game Server	*An ambulance stopped here. It looks like someone I know. I turned away towards the bank (as it was then). (18 minutes remaining)*

Sheet2 Sheet3 Sheet4 Sheet5 **Sheet6**

Ready Sum=946 SCRL CAPS NUM

Figure 6. The interactional order of events.

Figure 6 shows the relative positions of events when organized in terms of their observable and reportable entry into the interactional situation as articulated by the street player in his talk and practical actions on the street, as observed by the field worker, and as recorded on video. The re-ordering of recorded events to reflect interactional order has a profound effect on the representation and subsequent analysis of interaction. Consequently, players' utterances find a new *place, sense, and practical purchase* where they initiate and respond to specific actions and reflexively articulate the interactional situation. In the short (170 second) sequence we have been using as an example here, 25 out of 43 log entries have been re-positioned to reflect the interactional order of events (see Number Column and Column A, Figure 6, for positional differences). The entire system log consists of over 5000 entries over a 9-day period. Where video is combined with log contents, which only covers a subset of the entire

system log (several hours in distinction to days), the synchronized log nearly doubles in number of entries (compare Figures 4 and 5, for example). Re-ordering system logs to reflect interactional order has a profound effect on the shape of the log, on what we see and can see of interaction and, as a consequence of that, on our analysis of interaction.

Accordingly, the re-ordered log reveals that finding a postcard intersects and overlaps with the work of following clues and that both are done by exploiting local knowledge, either that of passers by or that gathered in the process of navigating the physical game space, to interpret and make sense of messages and clues. Furthermore, the re-ordered log makes it visible that online and street players coordinate the search for postcards by issuing instructions and clarifications, and that this interactional work is essentially topographical in character. Thus, in all their variations instructions and clarifications revolve around formulating adequate directions to places and objects, and coordinates locating places and objects. Thus, and for example, Dave instructs Patrick to "go to the graffitied phone box by the railings". However, it is not clear from Patrick's position on the streets just where the graffitied phone box is and he formulates a clarification framed in terms of his immediate topological relevancies – i.e., in terms of his current physical location and what to look out for from here: "Can you direct me to it? I'm outside the red phone box outside Reyner Street", which Dave responds to by saying, "You need the phone box on Portland Street by the tower." While online players track street players through a virtual facsimile of the street player's physical environment the two parties do not share the same orientation to places and objects; there is no reciprocity of perspectives and so it must be produced, constantly. Successful collaboration thus relies upon the players' ability to establish a *mutually intelligible* orientation to places and objects. Instructions are not sufficient in themselves to establish this, however, as they lack topological validity and cannot do otherwise given the asymmetry between the physical and digital settings of action. The topological validity of instructions is practically resolved, and mutually intelligible orientations produced then, through clarifications framed in terms of topological relevancies, the identification of candidate places and objects, and the subsequent reformulation of instructions furnishing coordinates that align street players and online players orientations and permit effective interaction: finding a postcard, for example.

3. Developing Computer Support

Our brief consideration of the work involved in combining multiple recordings for purposes of naturalistic analysis raises requirements for technical support that may be of more general purchase to qualitative researchers. In particular, the work highlights generic themes shaping the computational assembly of digital records:

- Visualizing the content of internal (system) recordings of interaction;
- Extracting sequences of interaction and collaboration of relevance to analysis from internal records;

- Removing non-relevant features from internal records;
- Synchronizing internal records with external records (e.g., video);
- Adding annotations (transcriptions, descriptions, etc.);
- Re-ordering synchronized internal and external records for purposes of analysis.

3.1 Replay Tool

Replay Tool provides an evolving suite of tools supporting the computational assembly of digital records. The first stage of the process is to retrieve the log files generated in computational environments. As noted above, these files often contain data that is not necessarily designed for human consumption – special code numbers may represent names and dates, etc., and the log may be populated with meta-tags and symbols designed to be machine-readable. The first component of *Replay Tool* 'parses' raw log files, replacing machine-readable codes, symbols and meta-tags with meaningful names and formats wherever possible. This process is semi-automated; it requires a description of the log format so that the parser can understand it, but once created, this allows for automatic data extraction for all logs of the same type. At the same time, links to associated data, such as links to the audio files recorded by street players in the above example, can be upgraded from text entries in the log files to hyperlinks that can be played when the log is loaded into the replay component.

In addition to parsing system logs, *Replay Tool* provides a data management and coordination tool that provides a visual overview of all internal and external records (system logs, videos, photographs, field notes, etc.) and automatically extracts information from the parsed logs to provide useful categories of information: the participants involved in interaction; text messages sent by participants; media files generated by the log and gathered by the researcher(s); spatial views of location-based data; and temporal views of the data. Temporal views are particular important as they enable the user to generate an overview of the data available at particular times and select parts of a system log that 'sit alongside' video and other external records (Figure 7; compare with Figure 3). Temporally selected log extracts are synchronized with external recordings by specifying estimated start times for each file (synchronization can be fine tuned during playback). This temporally synchronized dataset is then loaded into the replay component, where its parts can be replayed side-by-side (Figure 8).

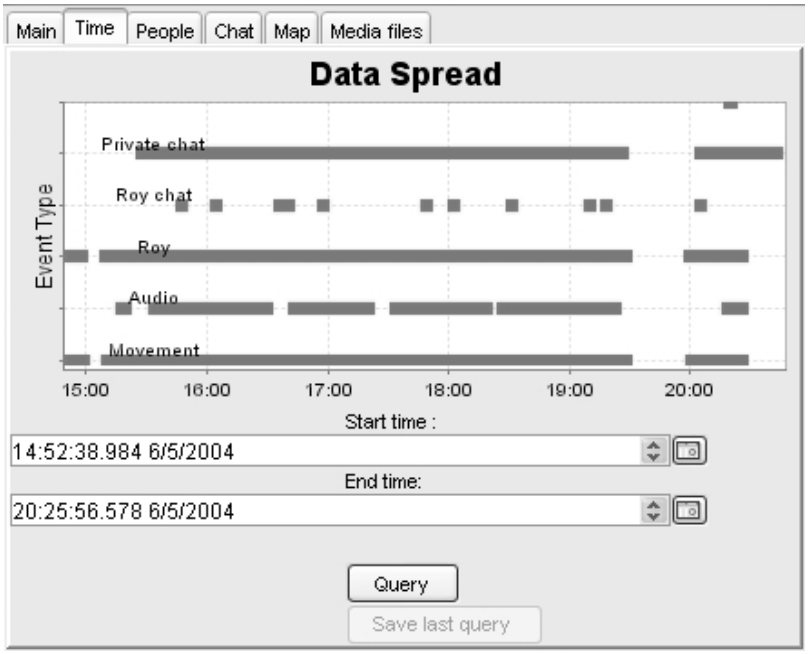


Figure 7. Finding salient data in system records.

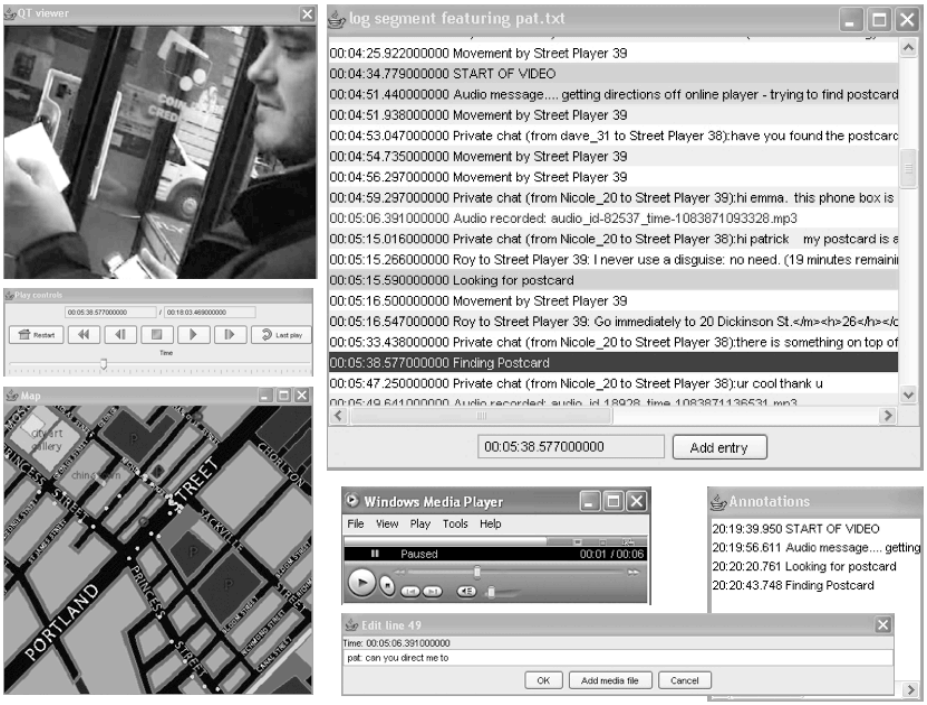


Figure 8. Loading salient data in the replay component.

Moving clockwise from the top right-hand corner of Figure 8, the replay component imports a) temporally synchronized video (version 2 of *Replay Tool* enables multiple videos to be imported); b) a temporally synchronized system log extract (the dark entry towards the bottom of the window displays the current relationship between the synchronized video and log extract so the analyst can see at a glance what system events were happening as he or she plays through the video); c) an annotations index, which allows the analyst to see and 'jump to' any transcriptions or descriptions added to the log; d) a text entry box allows the analyst to add annotations to the log; e) above that sits a media player that allows the analyst to listen to and replay audio files (as noted above, these are available via hyperlinks in the log); and f) a map view makes previously meaningless location-based codes and symbols available as meaningful maps that display a person's or persons route through a physical setting as given (in this case) by their GPRS-based reports of location. If recordings of interaction in the online players virtual environment where available, they could be added and synchronized too. Simple editing functionality allows the log to be re-ordered to reflect interactional order. Furthermore, all annotations and repositioned entries are stored in a relational database and associated with a point in time. This means that the original system log remains unchanged and is available to subsequent analysis by other researchers.

The core functionality of *Replay Tool* enables the qualitative researcher to produce representational accounts of interaction. It respects the contingent and time-based nature of analytic work, allowing researchers to work through the data and to work up an account that develops through careful inspection of multiple recordings. Furthermore, it enables multiple recordings to be inspected in novel ways; videos may be played alongside log extracts to inspect what is 'going on' in physical settings alongside what is 'going on' in digital settings; hyperlinks situate audio in its proper context; and map views enable new and previously unavailable views of location-based sensing systems, for example. Ultimately, *Replay Tool* enables the qualitative researcher to dispense with much of the time-consuming and laborious manual work involved in combining multiple recordings and provides new tools and resources to unpack interaction in existing and emerging computing environments.[4]

3.2 Developing Qualitative Studies of Technology in Everyday Life

We have already noted that a small range of tools support qualitative social science research – such as Atlas.ti, Anvil and The Observer. These allow the combination of transcripts, video, photographs, etc., and they support more structured forms of analysis as well. What they do *not* do is support the combination and replay of system records alongside external records, however. It might be argued, with some reason then, that they do not adequately support the study of interaction in existing and emerging computational environments. The ability to capture, combine, replay and reorder system records is necessary to adequate understanding of interaction in today's

[4] See Crabtree et al. (2006c) and French et al. (2006) for a more comprehensive account of *Replay Tool*.

distributed, media rich computational environments. While existing tools may be of some value they are inevitably limited by their inability to handle and exploit systems logs alongside other recordings then.

Clearly there is a need to further develop Replay Tool, however. Of particular importance is the need to implement support for more structured forms of analysis to promote the broad uptake of record and replay technologies. Accordingly, version 2 of *Replay Tool*, which is currently under development, focuses on exploiting 'semantic grid ontologies' (OWL 2006) to support coding. This enables analysts to apply the various conceptual components of theoretical or analytic schemes to the digital record, or to develop their own conceptual frameworks based on their analysis of the data. The second iteration of *Replay Tool* enables the analyst to play various recordings and annotations as separate but synchronized tracks, to create a coding scheme as he or she works through the data, and to assign codes to tracks as he or she works up a representational account of interaction (Figure 9). Thus, and for example, the analyst can code audio transcripts, video, log extracts, and annotations, developing a coherent conceptual framework that incorporates and spans across all the available data.

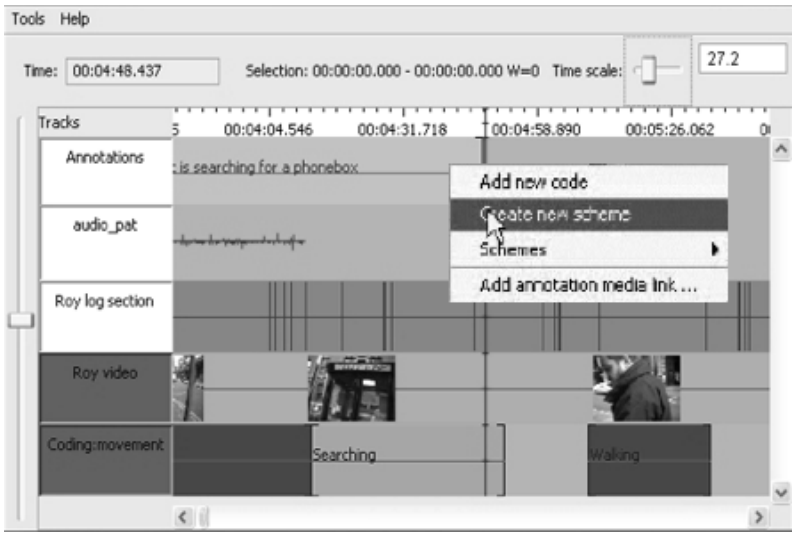


Figure 7. Coding video of street player interaction.

Tied to the development of support for coding is the development of visualizations that make visible what our colleagues describe as "the logic of coding". The aim here is to visualize the entire coding scheme and, importantly, to effect changes to the scheme which are automatically propagated across the dataset. Thus, and for example, the analyst might decide that two codes should be merged into one. Working out the complexities and ramifications of automatic propagation is a core feature of our ongoing work.

The possibility also exists, and is currently being explored, to develop a range of visualizations supporting different analytic activities. We have already mentioned the visualization of location-based events. One of the core challenges here is to make such representations much more dynamic, to reflect such things as the temporal quali-

ty of GPS or WiFi coverage. Thus, and for example, the analyst would be able to see the relationship between invisible sensing systems and interaction on the streets in the actual course of replaying and analyzing the dataset. We are also working on the development of visualizations that support the mining of systems logs. The aim here is to enable broader usage of the logs. Thus, and for example, visualizations are under development that enable the analyst to see patterns of collaboration between players, to trace their movements across physical and digital environments, and to inspect the conversational interactions that take place between participants. It is also possible to move beyond qualitative research and support the production of quantitative representations. In coding the data, the analyst essentially transforms the data into a set of countable objects. Coding becomes a basis then for generating quantitative visualizations of the data. One of the requirements here is to be able to "drill down" into the data, moving from quantitative representation to coded data, to raw data, for example. Again, working out the complexities of the matter is a core feature of our ongoing work.

4. Conclusion

The emergence of Grid-based computing and e-Social Science opens up new and exciting possibilities for qualitative research. Our own involvement in the enterprise is concerned with the development of new resources and tools that support qualitative research in the digital age. Computing permeates everyday life in contemporary society, cutting across work, education, leisure and play. Understanding the impact of computing on everyday life has been a longstanding concern in the social sciences, yet remarkably few resources and tools exist to support qualitative research in this area. We have presented the development of *digital records* as new resources that might promote the enterprise. Digital records combine system-based recordings of interaction in computational environments with external recordings of interaction (such as video and audio). *Replay Tool* allow the component parts of digital records to be replayed side-by-side to provide a much more comprehensive view on interaction in existing and emerging computational environments. *Replay Tool* also supports the production of representational accounts of collaboration and interaction. Version 1 of the tool supports the production of naturalistic accounts and may be downloaded at www.ncess.ac.uk/nodes/digitalrecord. Version 2 is currently under construction and aims to support more structured forms of qualitative analysis and the production of quantitative representations. We welcome enquiries, requests for support, and feedback from qualitative researchers interested in using *Replay Tool* so that we might develop resources and tools that are of value to the broad research community.

Acknowledgement

ESRC e-Social Science Research Node *Understanding New Forms of Digital Record for e-Social Science* (DReSS), www.ncess.ac.uk/nodes/digitalrecord

References

1. Benford, S., Seager, 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 the 6th International Conference on Ubiquitous Computing*, pp. 70-87, Nottingham: Springer.
2. Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M. and Row Farr, J. (2006) "Can You See Me Now?", *Association of Computing Machinery's Transactions on Computer-Human Interaction*, vol. 13 (1), pp. 100-133.
3. Chalmers, M. and Areti, G. (2004) "Seamful interweaving: heterogeneity in the theory and design of interactive systems", *Proceedings of the 2004 ACM Symposium on Designing Interactive Systems*, pp. 243 - 252, Cambridge, Massachusetts: Association of Computing Machinery.
4. Crabtree, A. (2003) *Designing Collaborative Systems: A Practical Guide to Ethnography*, London: Springer.
5. Crabtree, A. (2004) "Design in the absence of practice: breaching experiments", *Proceedings of the 2004 ACM Symposium on Designing Interactive Systems*, pp. 59-68, Cambridge, Massachusetts: Association of Computing Machinery.
6. 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 SIGCHI Conference on Human Factors in Computing Systems*, pp. 391-398, Vienna: Association of Computing Machinery.
7. Crabtree, A. and Rouncefield, M. (2005) "Working with text logs: some early experiences of e-social science in the field", *Proceedings of the 1st International Conference on e-Social Science*, Manchester, UK: ESRC.
8. Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M. and Brown, B. (2006a) "Supporting ethnographic studies of ubiquitous computing in the wild", *Proceedings of the 2006 ACM Symposium on Designing Interactive Systems*, Penn State, Pennsylvania: Association of Computing Machinery.
9. Crabtree, A., French, A., Greenhalgh, C., Rodden, T. and Benford, S. (2006b) "Working with digital records: developing tool support", *Proceedings of the 2nd International Conference on e-Social Science*, Manchester: ESRC.
10. Crabtree, A., French, A., Greenhalgh, C., Benford, S., Cheverst, K., Fitton, D., Rouncefield, M. and Graham, C. (2006c) "Developing digital records: early experiences of record and replay", *Computer Supported Cooperative Work: The Journal of Collaborative Computing* (special issue on e-Research), Springer.
11. Dicks, B., Mason, B., Coffey, A. and Atkinson, P. (2005) *Qualitative Research and Hypermedia Ethnography for the Digital Age*, London: Sage.
12. Equator (2006), www.equator.ac.uk
13. Flintham, M., Anastasi, R., Benford, S., Hemmings, T., Crabtree, A., Greenhalgh, C., Rodden, T., Tandavanitj, N., Adams, M. and Row-Farr, J. (2003) "Where online meets on-the-streets: experiences with mobile mixed reality games", *Proceed-*

ings of the 2003 ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 569-576, Florida: Association of Computing Machinery.

14. French, A., Greenhalgh, C., Crabtree, A., Wright, M., Brundell, P. Hampshire, A. and Rodden, T. (2006) "Software replay tools for time-based social science data", *Proceedings of the 2nd International Conference on e-Social Science*, Manchester: ESRC. www.ncess.ac.uk/events/conference/2006/papers/papers/FrenchSoftware ReplayTools.pdf
15. Foster, I., Kesselman, C. and Tuecke, S. (2001). "The anatomy of the Grid", *International Journal of High Performance Computing Applications*, vol. 15 (3).
16. Garfinkel, H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, New Jersey: Prentice-Hall.
17. Heath, C. and Button, G. (eds.) (2002) *Special Issue on Workplace Studies*, British Journal of Sociology, vol. 53 (2).
18. Hine, C. (1998) *Virtual Ethnography*, London: Sage.
19. Hughes, J. and Sharrock, W. (1997) *The Philosophy of Social Research* (3rd Edition), London: Longman.
20. Jordan, B. and Henderson, A. (1995" "Interaction analysis: foundations and practice", *The Journal of the Learning Sciences*, vol. 4 (1), pp. 39-103.
21. OWL (2006) *Web Ontology Language*, www.w3.org/2004/OWL
22. Sacks, H., Schegloff, E., and Jefferson, G. (1974) "A simplest systematics for the organization of turn-taking in conversation", *Language*, vol. 50, pp. 696-735.
23. Weiser, M. (1991) "The computer for the 21st Century", *Scientific American,* vol. 265 (3), pp. 94-104.
24. Wolcott, H.F. (1999) *Ethnography: A Way of Seeing*, London: Altamira Press.