Data recording and reuse: Supporting research and digitally augmented learning experiences.

Eric Harris, Sara Price,
Ted Phelps, Hilary Smith
Interact Lab, Department of Informatics
University of Sussex
Brighton, BN1 9QH, UK
{erich, sarap, tedp, hilarys}@sussex.ac.uk

Mark Weal
School of Electronics & Computer Sciences
University of Southampton
Southampton, SO17 1BJ, UK
mjw@ecs.soton.ac.uk

ABSTRACT
In this paper we briefly describe an outdoor learning experience supporting children engaging in scientific enquiry while learning about habitat distributions and interdependencies. To support them in their learning a variety of handheld and wireless devices were used and supported by different infrastructures, enabling learners to collect various forms of information. This paper describes how this data was recorded for later use; to support further learning for participants in the experience, for research analysis purposes and for software development and maintenance. We discuss strengths and weaknesses of our approach and raise issues for various forms of record and re-use that emerge from our research.

INTRODUCTION
Recording of participant action and interaction may be done for several different reasons, such as, research analysis, re-use by the same participants, re-use by different participants, or to build recommender systems that can provide information or guidance for other users (Resnick and Varian 1997).

A primary goal of the Ambient Wood project was to support active learning, which has several aspects including: experienced-based learning where learners actively engage in meaningful activities in the real world; collaboration, taking part by talking about what is being learned, and making it an active social interaction (Chickering and Gamson, 1987); and reflection, where learners need to experience, construct, test, and revise knowledge (Thompson and Jorgensen 1989), which makes the learning process itself an active process. They need also to interpret and transform information (Schomberg, 1986), understand the content in context and create personal meaning (Peterson *et al,* 1996). This involves reflecting in various ways in their activities. Structuring and restructuring information is instrumental in assimilating or internalizing newly acquired knowledge (e.g. Piaget 1972, Vygostsky 1978), and is cited to be an important part of the 'reflective process' (Kolb 1984) as well as playing a role in advancing knowledge (Karmiloff-Smith 1996). A further goal of the ambient wood was to support the process of scientific enquiry (National Curriculum Online 2004), which involves skills such as planning, exploration, collecting data, interpreting and explaining data. This process itself requires the organising structuring and re-representation of information.

Research suggests that recording and re-using information that has been collated during a learning task can provide the facility to support this re-representation and reorganization process (Levene and Peterson 2002). The ambient wood provides the opportunity for students to physically engage in and collect contextually relevant digital information about an environment. This provides them with both concrete and representational data. The data can be recorded and re-used in a variety of ways (e.g. according to location, time, content) providing different ways for learners to manipulate and re-represent their information and, in so doing, to support their thinking at different levels of abstraction. In addition, re-using their data in alternative ways encourages collaboration and discussion, thus supporting the reflective process.

To achieve this kind of data re-use requires an infrastructure to be built that supports recording in different ways (e.g. location, time, content), and requires software infrastructures to manipulate recorded data in appropriate ways for re-representation and re-use. The choice and configuration of infrastructure enables context to be delivered to recorded data. Positional information set along side past and present user interactions can provide meaningful insights into motivation as well as being useful in providing "setting" for users when revisiting their experience

Recording of data during experiences such as the ambient wood is also critical for high-quality research analysis. Recording of this kind of data can take different forms,

which serve different purposes and can contribute to analysis in different ways. Video recording all participant activity enables close examination of verbal interchange, activities undertaken and interaction with the environment and experience as a whole. Logging of digital information accessed or received by learners enables more quantitative data to be examined, for example, numbers of readings taken, comparisons of location data. In addition, the data logging enables verification of digital information accessed or received through cross referencing with video data.

In this paper, we examine not only the strengths and weaknesses of the approach used to record data during the Ambient Wood experience but also issues that are likely to be common to any method of data capture which spans media and which is to support post hoc analyses.

DESCRIPTION OF AMBIENT WOOD

The Ambient Wood is an outdoor learning experience where pupils aged 11-12 years explore and experiment in a digitally augmented woodland, to learn about habitat distributions and interdependencies. To support their learning they are provided with various handheld and wireless devices, which enable access to relevant information that is not normally visible or audible during a visit to a woodland. For example, plant processes and seasonal changes. Some devices enable pupils to actively collect data about their environment, (e.g. a probe tool to collect light and moisture readings), or to receive and store data (e.g. PDA). Others provide pupils with access to enhanced information about their environment (e.g. periscope) and others receive contextually relevant information that is triggered according to pupil location in the wood (e.g. the Ambient Horn).

The ambient wood experience is divided into four different phases of activity, namely; exploration in the wood, hypothesising in the den, experimentation in the wood and reflection in the den. During the experience all data collected and received by pupils, sent to pupils or accessed by pupils was recorded and logged for several purposes,. This paper focuses on data logging for re-use by pupils in supporting them in abstracting from the woodland learning experience; and on data logging to support research analysis of the experience on several different levels; the value of data logging for the design and development process.

RECORD AND RE-USE OF DATA IN AMBIENT WOOD

Record and re-use by learners

All digital data that students collected, received or triggered was logged and recorded for students to re-access and use in different ways. During the outdoor experience students used their data, which was represented in different ways for discussion and reflection in small groups. Back in the classroom this data was again re-represented to support more abstract reflection about interdependencies in woodland habitats.

Den

As part of the woodland experience a purpose built space or 'den' was provided. After each woodland phase (exploring and experimenting) pairs of pupils came together to reflect upon their discoveries, formulate hypotheses and draw conclusions from their experience. All information that was digitally delivered to or triggered by pupils was logged and recorded using GPS and Elvin. This information was re-represented on a bird's eye photograph of the wood using a large screen display. The information on the display consisted of a dot marking the location in the wood where the information was collected. This was to enable pupils to see in a more global space where they had, for example, taken a particular reading or identified a plant, providing the facility for them to map their readings according to their location and make comparisons about the different distributions within their habitat. In addition, pupils could re-access their digitally collected information using the PDA, which served as digital record. This highlighted the corresponding dot on the map with the same image displayed at the side of the map (fig.1). This enabled learners to display specific information to the other pair of pupils, providing the facility for pairs of children to share and discuss information from the two different habitats, and to make comparisons between them.

Fig. 1 Den display

Schools

This same data was also used to support the students learning about food chains and webs in the classroom. The data here was transformed and re-represented again using the map with digital data overlaid. This time however, the display was back projected onto a horizontal surface. The display showed an aerial view of the wood itself, with coloured shapes indicating the learners' location when various plant and animal cards were received. The sets of cards were pre-loaded onto the PDAs, and could be used by the students to highlight various shapes on the shared display. The corresponding card image could also be seen at the side of the shared display (see figure 2). In addition the light and moisture readings recorded during the Ambient Wood experience were used to extrapolate the light and moisture levels throughout the wood using an inverse distance weighting scheme. These visualizations could be

superimposed on the overhead map, enabling students to see collective distributions of their probe readings.

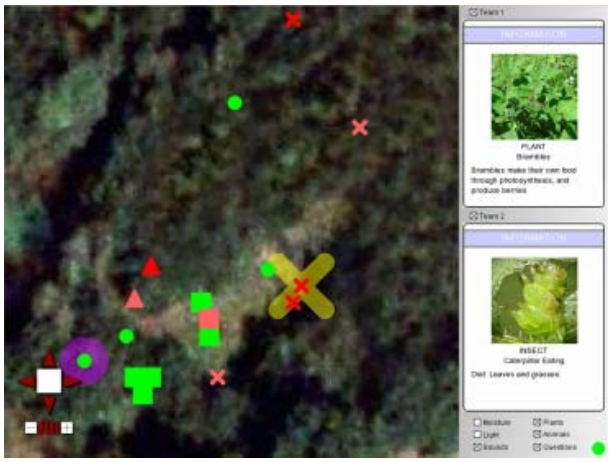

Fig. 2 Classroom table display

The aim here was to help students to reflect about the habitat distributions and interdependencies experienced in the wood at a more conceptual level, thus supporting a move from concrete to abstract understanding.

Record and replay for research analysis

Research analysis requires data gathering that covers many aspects of the event. Video and audio capture provided the core data streams for future analysis. However this needed to be cross referenced to other records kept concerning the "state" of the augmented environment, such as location or time data.

To collect the video data each pair of learners engaged in exploring the wood was accompanied by a roaming camera-person who had a fixed microphone gun attached to the camera. The filming was made as unobtrusively as possible but could not be completely in the background since the focus of interest often demanded closer inspection. Fixed cameras were used where we knew the children would be for a while i.e. the den and in the classroom. Although audio is recorded as part of the video capture we also digitally recorded all of the walkie-talkie traffic. This could then be cross referenced to the video data by applying any timing offset.

To further support the research analysis, use of the Elvin logs provided a record of which pieces of content had been sent, to what and when.

Record and replay for design and development

Recording the sessions on video can support the development process of design, running and managing the Ambient Wood experience over time. Analysis of pilot studies and trials inform subsequent iterations of an experience, for example, Ambient Wood I sessions helped refine and converge ideas on how to scaffold children's learning and activity during the experience.

In addition, record and replay used in this way can help uncover technological and user interaction problems, not always apparent at the time of the study.

INFRASTRUCTURE

Elvin

Elvin is a public/subscribe notification service. It was employed in the Ambient Wood to communicate changes of state among the various software components deployed in the wood; GPS position information, the presence or absence of an RFID tag near a reader's antenna, probe readings and MUD state transitions, are some examples of the notification traffic routed via the Elvin server. A universal consumer, ec, was started before each run and recorded all notification traffic to a text file. The resulting set of notification logs could be replayed using the universal producer, ep, or read into analysis program using an appropriate parser. The logs were used to:

- debug various pieces of software,
- simulate live data during the development phase between Ambient Wood I and Ambient Wood II,
- construct a visualisation of the light and moisture levels in various parts of the wood by extrapolating from probe readings,
- reconstruct the set of information 'cards' each pair of children had collected for use in the classroom session of Ambient Wood II, and
- assist in the videotape analysis of the experience by providing clues about what a PDA was displaying at a given point in time.

The Elvin logs only contain state *change* information, which keeps them reasonably small. In order to reconstruct state for a given point in time, the logs must be searched backwards from that point to locate the last change. For the small amount of data recorded in the Ambient Wood logs, this is not an issue, but it could quickly become problematic with larger logs.

Multi-Media Journal

Using a combination of consolidation techniques, inference tools and storytelling devices, the logs were reused in forms that both recounted events that took place, but also promoted different perspectives, allowing the children to gain new viewpoints on their activities, or alternatively to see their activities relayed as an entirely fictional adventure (Weal, 2003).

The process of building the journal had a number of parts, reflected in the architecture shown in Figure 3. First, a story needed to be selected from the many possible stories that could be told from the raw information available. Then, the events had to be consolidated down to the key events that were to be relayed as part of the story. Finally, the events were turned into a discourse in a particular form, both in terms of media, style, genre, focalisation etc.

A number of record and reuse issues arose from the Multimedia Journals repurposing of the captured event logs. What were perceived as individual events in the wood by the participants, i.e. taking a probe reading, ended up as multiple events in the log. The causality of events was not recorded explicitly and had to be inferred from the chronology of the events and re-composited thereafter. Capturing of broader context rather than just the communication stream might help in this process. In addition, more formal modelling of the underlying information may help in the reuse of the event logs. A more formal structured approach would provide a better base for more complex inference both easing the development of analysis tools and allowing more sophisticated automation of event reuse.

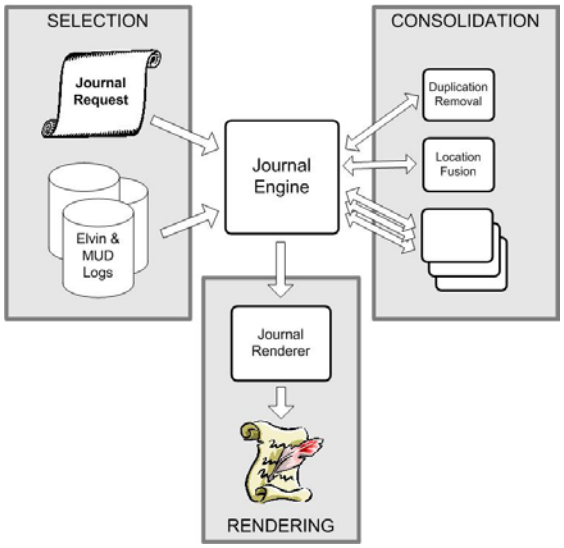

Fig. 3 Journal Architecture

Position (GPS / pingers)

The children's position was used in two ways, to provide location sensitive data in the form of sounds or information on the PDA, and to "locate" the light and moisture readings collected by the children.

For the delivery of information to the children we used pingers with a transmission range of approximately ten metres. This coupled the pinged information to the vicinity of the transmitter but still required the children to explore the area to establish the link. However for the logs we used the GPS position to define the where the children were when they received the pinged information. This was also the case when the children took moisture and light readings, the recorded position for those readings was that of the child who carried the GPS receiver. The GPS in turn transmitted it's position via a pinger to the PDA and thereon to the infrastructure.

DISCUSSION

The ambient wood project enables us to begin to explore a number of issues surrounding the recording and re-use of data from a mixed reality experience. Issues arise from both a learner perspective and from a research perspective, contributing towards design and development as well as research analysis.

Learner Issues;

Active engagement in meaningful activity provides knowledge at a concrete level. Record and re-use is beneficial in supporting the advancement of knowledge by providing the facility for learners to collaboratively discuss their findings, interpreting and transforming their information in personally meaningful ways. This raises a series of interesting questions for analysis and for more specific future research. For example, does personal active engagement affect the way students engage with information later and in more formal settings? Does the contextual emersion of the physical experience provide more meaningful data and facilitate a deeper understanding of the domain being studied? Does the re-representation and reorganization of data facilitate transference or generalization of a hands-on experience to more abstract concepts?

However issues arise that may influence how learners structure their activity and their knowledge by being able to record and re-use data. Whether they are aware or not of data being recorded may influence how they engage in the activity and in the environment. In Ambient Wood learners were unaware that their data was being tracked and recorded until the reflection stage in the den. Awareness of this may have influenced how they structured their activity, what they chose to do, or their level of attention during the activity e.g. to their specific location.

The way in which data is recorded and the redisplayed or re-used may also affect the way in which learners interpret it. For example, there may be instances where technology is unstable resulting in an inaccurate reflection in the logs of learner activity or interaction. For example, in the Ambient Wood there was one occasion where a software bug caused the repeated triggering of one periscope movie. It would be possible to infer from the log that learners had *chosen* to watch the same movie repeatedly. Although in this instance the learners themselves would most likely remember the 'faulty' technology, there may be similar instances where learners may not remember their particular interaction. If the logging data indicates an inaccurate reflection of activity, this may result in misremembering, potentially interfering with appropriate knowledge construction. The way in which this data is recorded and re-used also has implications for quantitative analysis, as well as for re-use by 'other' users. For example, the periscope instance may indicate an importance on a particular video clip that did not really exist.

Therefore, it becomes important to have the selected pieces of information placed in context to prevent misrepresentation of the data. For example a sound delivered on the horn indicating root uptake followed by moisture readings might indicate a casual link between the two, but when cross referenced to the GPS readings might show the moisture probe taken place in the middle of the clearing away from any plant able to engage in root uptake.

Issues for research analysis

To have a record of an event for the purpose of future analysis can pose difficulties in its collection. For example, how to minimize the intrusion from the person filming; where to place the focus of attention; how to dynamically annotate clips of interest; and not least how to ensure the recording includes the focus of interest that have yet to be uncovered.

In the Ambient Wood three quite disparate means of storing data records were used; video with audio, separate audio and Elvin logged data. This proved problematic to cross reference the status of events at any one time even though each record was time stamped. For example, knowing exactly when a card appeared on a PDA can be very important when unraveling a child's motivation in asking a particular question. Often activity about a point of interest can seem quite complex and confused and only becomes clear after careful unscrambling of what happened, when. Here the need is for the recorded data to be synchronised to second precision.

Although collecting data onto different media types can be seen as cumbersome, the use of differing equipment and recording strategies can have benefits. "Observation selection effect" or "Anthropic bias" (a bias introduced by limitations of the data collection process) will taint replayed observations and as such might not tell the whole story. For example when the children took light or moisture readings there are three significant points of interaction, one with one child at the probe, another with the child with the PDA and a third between the children. A wide screen video shot will capture some of the story but not all of it, whereas when the video is placed alongside the Elvin logs a fuller understanding can be made showing the timing of the sent and displayed probe readings.

Attaching context to collected data not only provides clarity in interpreting the data, but also insight into user motivation. Does a stream of logged multiple access to one piece of data mean lots of 'real/meaningful' use or just kids flicking through e.g. PDA data? Are there ways to make these kinds of distinctions? Context can be gleaned through recording the position and orientation of the user when an event happened, but what is more difficult to uncover is the users' mental orientation.

Issues for design and development

Using record and replay was found to be valuable for informing development of design and managing the experience. For example, video analysis showed adult facilitators to have different support-styles particularly when exploring in the wood. By viewing and learning others' styles, a singular approach could then be developed that would best ensure an interaction that would support learners appropriately, for example, to encourage and engender independently exploration and enquiry.

Video-recordings also revealed small glitches that occurred with the technology and showed where there were short breaks in activity, enabling technological glitches to be fixed and/or redesign of the experience to accommodate these shortfalls.

Technical issues;

Within the Ambient Wood, captured data was used for both replay and for reuse. In satisfying each of these requirements demands are made on the supporting technology. These can often be at odds with one another. There may be demands put on the recorded data to be manipulated in a certain way for reuse that is not necessary when just replaying, for example audio captured in mono might have value in stereo reuse. The core requirements of record and reuse need to be understood to best fit the technology (infrastructure and devices) to the task. The value in identifying suitable technology is self evident but in finding the "best fit" care should be taken to provide enough inherent flexibility to the data that the technology can cater for. To address these technical challenges the following aspects should be taken into consideration; *(i)* the core requirements for record and replay/ reuse of data. These might be, for example, accessibility, transferable, ability to be interrogated, referenced and cross synchronized; *(ii)* what affects these core requirements, for example, the media used for storage (digital /analogue), the granularity of the recorded data, the mode of capture (event or time driven), the supporting infrastructure, redundancy and standards used.

In conclusion, this research indicates a number of uses for record and replay, and record and re-use of data. Some particular benefits and disadvantages for particular purposes are beginning to emerge, but most of all it raises a series of questions for further research.

REFERENCES

1. Chickering, A. and Gamson, Z. (1987) Seven principles for good practice. *AAHE Bulletin*, 39, 3-7.
2. Karmilloff-Smith (1992) *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge MIT.
3. Kolb, D. (1984) *Experiential Learning*, Prentice-Hall
4. Levene, M. and Peterson, D. (2002) Trail Record and Ampliative Learning. *Research Report BBKCS-02-10*, School of Computer Science and Information Systems, Birbeck College, University of London.

5. National Curriculum On-line (2004) http://www.nc.uk.net/nc/contents/Sc-2-1-POS.html
6. Peterson, M., Morrison, D., Cram, K., and Misanchuk, E. (1996) CMC: An agent for active learning *Proceedings of 12th Annual Conference on Distance Teaching and Learning: Designing for Active Learning,* August 7-9, Maddison WI..
7. Piaget, J. (1972) *The Principles of Genetic Epistemology*, Routledge and Kegan, London
8. Resnick, P. and Varian, H. (1997) Recommender Systems. *Communications of the ACM*, Vol. 40, 3, March.
9. Schomberg, S. (1986) Strategies for active teaching and learning in university classrooms. Minneapolis: University of Minnesota
10. Thompson, J., and Jorgensen, S. (1989) How Interactive is Instructional Technology? Alternative models for Looking at Interactions between learners and Media. *Educational Technology*, 24
11. Vygotsky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes* (eds) M. Cole, V. John-Steiner, S. Scribner & E. Souberman. Harvard University Press, Cam, Mass.
12. Weal, M. Michaelides, D. Thompson, M. De Roure, D. (2003) *The Ambient Wood Journals - Replaying the Experience* Proceedings of Hypertext'03, The fourteenth conference on Hypertext and Hypermedia. (pages 20-27), 2003, Nottingham, UK.