Integrating Data from Multiple Contexts

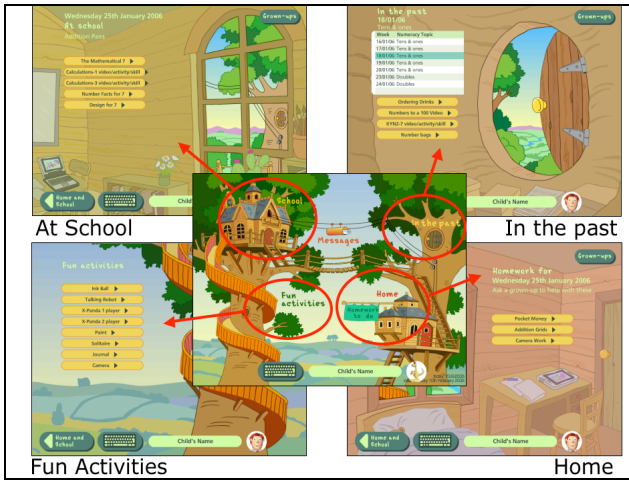
Joshua Underwood[1], Hilary Smith[1], Rosemary Luckin[2], Benedict du Boulay[1], Joe Holmberg[1], Lucinda Kerawalla[1], Hilary Tunley[1]
[1]Human Centred Technology Research Group, School of Science and Technology, University of Sussex, Brighton BN1 9QJ, UK [2]London Knowledge Lab, Institute of Education, University of London, 23-29 Emerald Street, London WC1N 3QS
{ joshuau, hilarys, rosel, bend, J.E.Holmberg, L.J.Kerawalla, hilaryt}@sussex.ac.uk

The HOMEWORK research project provided individual Tablet PCs to a class of primary school children for use in school and at home. Here, we describe our use of both automated and manual data capture and analysis techniques to map out the nature of interaction with the tablets across multiple locations. Outside of school, the tablet PCs offered children access to a variety of multimedia content relevant to current school numeracy work through a custom-built 'home' user interface. We want to know what was used, when, how, where and by whom. Our approach combines usage data captured from multiple sources: the 'home' user interface, system level logging and individual content resources. We then integrate and triangulate with other data sources such as user diaries and interviews to build up a rich picture of technology use across multiple contexts. We also consider the need to transform usage data generating representations appropriate to varied audiences each with very different requirements.

Usage data capture, Usage data analysis, Educational technology.

1. INTRODUCTION

The HOMEWORK project[1] [1] designed and developed a system to support the use and delivery of multimedia content for learning both in school and at home for children aged 6-7, their parents, and teachers. Content included existing broadcast video media and associated interactive resources for primary numeracy as well as custom made (Flash and Journal) content. This material was delivered via Tablet PCs and Interactive Whiteboards at school and via Tablet PCs out of school. When the Tablet PC was taken home children could revisit school 'numeracy hour' work, watch related Number Crew[2] episodes, and complete homework activities, either working alone or with members of their family, in the kitchen, in the car, or elsewhere. At least in theory, access to content at home was controlled through the 'home' user interface (see figure 1). Content resources relating to current and upcoming school numeracy learning objectives were preloaded on the Tablet PCs before they were taken home and linked to from the 'home interface', which updates daily. All content ran locally.



The centre image is the home start-up page. Hotspots (circled in red) link to:

- "At school" (top left) has links to the resources used in today's school numeracy session.
- "In the past" (top right) lists numeracy topics from previous days, selecting one of these reveals links to resources used for that numeracy session.
- "Home" (bottom right) has links to the current homework activities.
- "Fun Activities" (bottom left) has links to a selection of applications and games.
- Clicking "Messages", in the centre of the home page, opens an email client which can be used for parent and/or child to teacher email communication.

The green "Grown-ups" button at top right of all pages reveals extra contextual information relevant to adult carers.

FIGURE 1: HOMEWORK 'home' user interface – displayed at start-up when out of range of school wireless network.

In the longest study, we lent 32 Tablet PCs out for individual use by children in school and at home for 4 weeks. We were interested in capturing home usage data, both to answer focussed questions such as whether our evolving interface was in fact being used as we expected, and more openly to explore how activities were used across multiple contexts and reveal unexpected and interesting patterns of activity. We used both manual data gathering through diaries, observation, focus groups and interviews and automated activity logging. Usage data from various sources was recorded locally on each tablet and retrieved at opportune moments during studies and after studies.

1 For more information about HOMEWORK see http://www.informatics.sussex.ac.uk/users/joshuau/homework/

2 The Number Crew, Open mind for Channel 4 learning, see http://www.channel4.com/learning/microsites/N/numbercrew/frame_home.html

This paper focuses on automated data capture and analysis. We describe: 1) the ways usage was logged at system level, by the home user interface and by individual content items, 2) the retrieval, processing and integration of this usage data and 3) the benefits and limitations of our approach. Finally, we address the need to transform this usage data and represent it for audiences with very different needs, e.g. interaction designers, educational researchers, teachers, child users and their parents.

2. AUTOMATED CAPTURE OF HOME USAGE DATA

2.1 Motivation for Automated Data Capture and Specific Issues

HOMEWORK represents a challenging scenario, we wanted to capture the history of an individual learner's experience with their technology over protracted periods in realistic and varied contexts (the home, in the car, wherever). In these contexts, observational techniques (e.g. video) can be difficult and disruptive. Many of our home users did provide self-report data on usage through diaries and interviews. However, this was often incomplete and retrospective and we had no reliable means of verification. An additional challenge for us is that we are interested not only in the use of the single application that we are designing: "the home interface", but also the use of the various content and applications it links to. We therefore need to integrate data captured by our own interface with the varied data (not under our control) captured by content items and at the operating system level.

2.2 Methods and Tools

2.2.1 What to log? Practical and ethical concerns constrained what we were able to log. Practical concerns included: the effect on Tablet PC performance of intensive logging, concerns about available disc space for storage on individual Tablets (the devices were in use for several weeks with limited opportunities for data recovery) and not least the difficulty of and time required to retrieve and analyse large quantities of data from 32 devices. The Tablet PCs we used had built in microphones and cameras potentially allowing us to automatically capture images and sound from the context of use. This kind of contextual data would be extremely useful, but such data introduces additional ethical concerns and we decided against automated capture. However, children and their parents did take photos and video clips and these can be integrated to add detail to the picture of use.

2.2.2 What was logged? At the system level we used "Activity Monitor[3]", configuring a client application on each Tablet to record start and end times for all processes and to take screenshots every minute during use. The 'home' user interface, developed by us and through which access to other content was controlled, logged button clicks and other user input to a local PostgreSQL database, as did all other content under our control. Other data relating to usage was available to us from some standalone content that logged to text files and also in the form of photos and other content created by our users (mainly MS Windows Journal and Paint files) saved locally.

2.3 Data Retrieval Issues

A particular concern for us was developing efficient mechanisms for rapid and reliable retrieval of large and varied data from all devices used during a study. In order to capture data that might be overwritten or lost during the course of the study (i.e. accidentally or through repeated work on the same file) we aimed to regularly retrieve data from all tablets during the study periods while they were in school and available to us over the wireless network. However, for organisational reasons (e.g. time and access to a space large enough to get out and turn on all 32 tablets) it was rarely possible to spend more than 20 minutes after each school numeracy lesson recovering data. Consequently, we limited data retrieval during the study to running a script from the classroom server to copy files that had changed since the last retrieval from key local directories (e.g. "My Documents"). This meant that if children were off school or did not bring in their tablets we did not recover their data. After the study, when we had recovered all 32 tablets, we used "Activity Monitor" on the server to recover all screenshot and system level logs (+/- 2Gb) and ran scripts to copy all other files from key directories. Even so, this process was time consuming, error prone and we had to recover some data manually. One difficulty we encountered was ensuring data recovered from particular devices were associated to the correct user, this was complicated by the fact that users sometimes changed device because of breakdowns. Another concern was to delete all personal data, particularly photos, from the devices to protect privacy but only after we were certain this data had been successfully retrieved.

3. PROCESSING AND ANALYSIS OF HOME USAGE DATA

Our data was largely pre-processed by hand; by recording and reflecting on this process we hope to identify mechanisms for future automation. Activity Monitor data was retrieved both as html files, interleaving process start times with screenshots and as Excel spreadsheets. We hand coded the spreadsheets to identify 'sessions' (periods of use from start up to shutdown or crash) and we assigned sessions to 'school' or 'home' use using knowledge of the school timetable and absences. We also categorised particular executables as belonging to 'categories of use': fun, numeracy practice, video, homework, other. When an application may have been used for either fun or homework (e.g. Journal) we referred to screenshots to help decide on how to categorise a period of use. Analysis at this level enables us to paint a broad picture of how the tablets were used at home. For example, tablets were initially used out of school very intensively with average use dropping to 1 hour 40 minutes in the first week of a study and 1 hour in the second week, typically divided over 2 or 3 evening sessions. Approximately 30%

3 See http://www.softactivity.com/employee-monitoring.asp for more information about Activity Monitor.

of use was for 'fun activities' not directly related to numeracy learning. For information about how the user interface was actually used we are looking at logs of the buttons clicked. For individuals, we have plotted button clicked against time, this provides a picture of the route users have followed through the home interface in a single session and indicates how long they have spent on particular pages (see figure 2). It also clearly reveals issues such as impatience with application launch time (rapid multiple clicks on the same button). To gain an indication of global 'home interface' use we mapped the total number of users clicking on each button (see figure 3), however this representation loses the sequential information. We are investigating other possible representations, which might provide a global overview of use while retaining sequential information. We are also now looking at integrating log information from individual items of content in this overall picture of use. For, example for some items of content we can see how individuals have gradually built up correct answers through repeated attempts and we would like both to compare this activity across users and to contextualise it within the overall picture of activity in a session built from the data discussed previously.

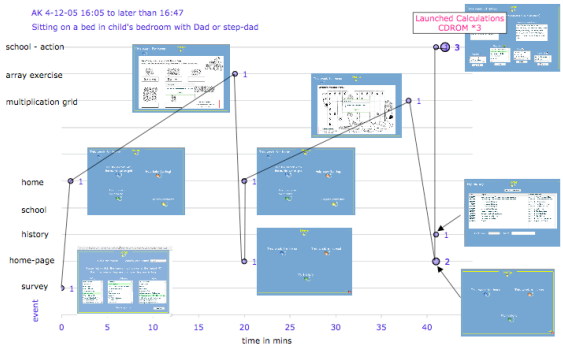


FIGURE 2. Page and button clicked against time, size of dot indicates number of clicks. From early version of home interface.

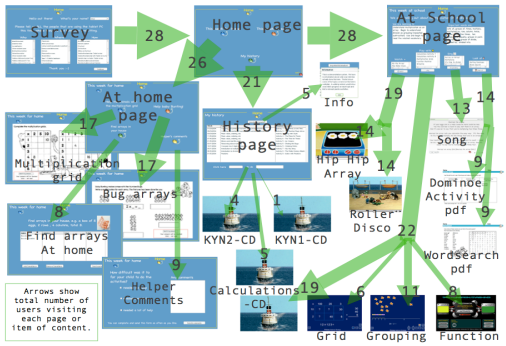


FIGURE 3. Width of green arrow and number indicates number of users visiting each page or piece of content.

We are also working on the integration of multiple sources of data in order to verify and explicate issues highlighted in qualitative data. For example, some parents reported frustration over a task involving using the built in camera. However, they were not able to express clearly which aspects of the activity had caused problems. By integrating log data, screenshots, photos and journal files created during this activity we can build a detailed picture and identify specific issues. In summary, our data processing and analysis has been largely 'manual' and time intensive but through this process we are identifying many opportunities and requirements for new tools.

4. DISCUSSION

To facilitate data retrieval we need easily configurable tools for non-experts that allow: rapid collection of relevant data from multiple devices over wireless networks, provide accurate reporting of retrieval failures and manage archiving of data by users even when users change devices. We also need tools that allow us to identify and retrieve all files and data relating to an activity even when this is spilt across several applications. This is not dissimilar to the requirements Shraefl [2] derives for e-science lab books, although our focus is on learning activities rather than digital experiments. For data analysis we require tools, which enables us to easily associate, inspect and represent data of varied types and from multiple sources, for example diary comments, interview comments and log files. We would wish to be able to define 'user activities' and 'sub-activities' and look at all data from any source associated with these activities. We would want to be able to do this both for a single sequence of activity and across several sequences of the same activity within and across users. However, for large quantities of data even with appropriate tools manual identification of activities would be time consuming and automated techniques might feasibly associate data from different sources and help identify patterns of activity, reducing the time required to process data. While Activity Monitor provides some of the retrieval and integration (html pages combining screenshot and process data) features we require, it does not easily support retrieval and integration of data from other sources, allow us to define user activities or support us in analysing and generating representations of these data. Finally, we require tools that facilitate the automatic generation of representations of the mountain of captured data, tailored to the specific needs of different users.

REFERENCES.

[1] Luckin, R. Underwood, J. du Boulay, B. Holmberg, J. Kerawalla, L. O'Connor, J. Smith, H. and Tunley, H. (In press) Designing Educational Systems Fit for Use: A Case Study in the Application of Human Centred Design for AIED. Accepted for publication in the International Journal of Artificial Intelligence in Education (in press).

[2] schraefel, mc. Brostoff, S. Cooke, R. Stevens, R. Gibson, A. (2005) Transparent interaction; dynamic generation: context histories for shared science. ECHISE 2005 - 1st International Workshop on Exploiting Context Histories in Smart Environments. http://eprints.ecs.soton.ac.uk/10693/01/contextHistoriesMyTea-ECHISE05.pdf

ACKNOWLEDGEMENTS
We would like to thank our schools in Brighton and James Luckin for helping with data analysis.