



JISC Final Report

Project Information			
Project Acronym	ClimbieDataCorpus@hud.ac.uk		
Project Title	Climbié Inquiry Data Corpus Online		
Start Date	October 2008	End Date	September 2009
Lead Institution	University of Huddersfield		
Project Director	Sue White		
Project Manager & contact details	Graham Stone CSB 5/23, Computing and Library Services, University of Huddersfield, Queensgate, Huddersfield HD1 3DH Tel. 01484 472042 Fax. 01484 473167 Email: g.stone@hud.ac.uk		
Partner Institutions	n/a		
Project Web URL	http://victoriaclimbie.hud.ac.uk/		
Programme Name (and number)	Enriching Digital Resources		
Programme Manager	Ben Showers		

Document Name			
Document Title	Final Report		
Reporting Period	Total Project		
Author(s) & project role	Graham Stone (Project Manager) and Dawn Clarke (Research Assistant)		
Date	16 July 2010	Filename	ClimbieDataCorpus_finalreport_vers3.0
URL	<i>if document is posted on project web site</i>		
Access	<input checked="" type="checkbox"/> Project and JISC internal	<input type="checkbox"/> General dissemination	

Document History		
Version	Date	Comments
Draft Ver. 2.0	16 July 2010	
Final Ver. 3.0	22 July 2010	After comments by BS



JISC Final Report (*Public Report*)

Victoria Climbié Inquiry Data Corpus Online

Table of Contents

Acknowledgements	3
Executive Summary - The Climbié online corpus	4
Background - The Victoria Climbié Inquiry Data Corpus Project	5
Aims and Objectives	6
Methodology	6
Implementation	7
Outputs and Results	7
Outcomes	11
Evaluation of learning resources	11
Conclusions	12
Implications	12
Further work	13
References	13
Appendices	14

Acknowledgements

The project team would like to thank Ben Showers at the JISC for his support and assistance throughout the project.

Acknowledgements are also due to:

Project Team: Graham Stone (Project Manager), Graham Gibbs, Christopher Hall, Antony Osborne, Andrew Teal and Dawn Clarke.

Steering Group: Sue White, Nigel Parton, Helen Masson, Rachel Balen, Sue Peckover, Graham Stone, Graham Gibbs, Dawn Clarke and Antony Osborne.

We would also like to thank the Victoria Climbié Foundation (<http://vcf-uk.org/>) for their support.

The project team would also like to acknowledge the National Archives for the PSI click-use licence (PSI Licence number C2008002563) for the reuse of Crown copyright and Public Sector Information (<http://www.opsi.gov.uk/click-use/psi-licence-information/index>) and the Guardian newspaper for permission granted to the University Of Huddersfield for a 5 year licence for the reproduction of content regarding the Inquiry.

Executive Summary

The Climbié online corpus

The Victoria Climbié Inquiry Data Corpus Project at the University of Huddersfield has undertaken work coding and annotating the witness statements from the Victoria Climbié Inquiry. The project took this large, analysed data set (in the form of an Atlas.ti project) and made it available via the University Repository (<http://eprints.hud.ac.uk/4842/>). The data set was converted to XML for deposit in the repository. Functions were developed to allow users not only to search the data in the normal fashion but to retrieve 'tagged' passages of text in the fashion common to qualitative data analysis and to link these retrievals with other metadata and contextual information. These data are of central interest to researchers in child welfare, professional and legal studies, public administration and politics as well as teachers and students of a range of subjects, such as health care and social work, who deal with child welfare and to professionals needing to develop management and administrative skills in child welfare. The project investigated the technical issues involved in depositing and making available a data set such as this and evaluated its utility and utilisation by researchers, learners and professionals.

Web access to the corpus is available at <http://victoriaclimbie.hud.ac.uk/>. There are two main sections to the website:

- A set of pages with information about the inquiry, its background, the people involved and the research literature it has generated
- Suggestions, in both a teaching and a research context of how data might be retrieved and what kinds of teaching exercise and what kinds of research projects this might be used to support

Further publicity articles will be available via the website and the Repository at a future date.

The conclusions drawn from the project are:

- The Repository is not the best way to archive and retrieve the coded data set due to usability and robustness problems. In short the data set was so large, retrieval of the sub sets was taking too long to make it practical. Hence the Repository now operates as gateway for storing the full data set
- There is a difference between coding for research and teaching and learning. The coding for research would be in depth, interpretative, analytical coding for specific use by the particular researcher. For teaching, training and learning the coding has been applied in order for the user(s) to gain access to the different themed subsets which then can be analysed in much the same manner used by a research focus
- This would appear to be the best/most appropriate way forward for coding of data because the data set is accessible, usable and can be tailored to meet the specific requirements of the different users

There are several exciting possibilities for use and development of the materials produced. These are:

- additional coding on the data set
- development for teaching application
- development for training application within the care sector
- evaluation of the use of the data set by both academics and practitioners
- update the data set held in the Repository
- coding of other similar data sets for use in academia and the relevant community

Background

The Victoria Climbié Inquiry Data Corpus Project

The Victoria Climbié Inquiry Data Corpus is an archive of transcriptions of the examination of witnesses at a child protection inquiry. The inquiry, chaired by Lord Laming, was established by the Department of Health, Department of Education and Skills and the Home Office after an 8-year-old girl from the Ivory Coast, Victoria Climbié, was murdered by her great aunt and her partner in February 2000. The inquiry was held in public in London from September 2001 to July 2002 and a report (Laming 2003) was presented to the Government in January 2003.

The Corpus Project was established in September 2003, supported by the University of Huddersfield and followed from earlier work commissioned by the NSPCC (Hall 2003). The initial aim was to make the inquiry evidence accessible to researchers by organising the oral evidence into a data corpus. The transcriptions of the witnesses have been coded and categorised, and organised in such a way that users can search for topics across the entire verbal evidence. The examination of witnesses provides a large body of data about the state of child protection services in the late 1990s. There is detailed testimony about professional practice, decision-making, inter-agency working, the context of service delivery and policy-making across all agencies. Witnesses were questioned about the day-to-day contact with Victoria and her great aunt but also policy and practice more generally. There was questioning about training, supervision, work allocation and atmosphere, resources, procedures and guidance, local authority policy and management, and Government audit. Witnesses included social workers, police officers, doctors, health workers and hospital staff, housing officials, voluntary agencies, team managers, advisory staff, senior managers, local politicians and government inspectors.

Work undertaken by the Corpus Project prior to the JISC funded project

The Delphi consultation with potential users

The first task was to consult with experts, using the Delphi method (Delbecq et al., 1995), on the appropriate themes around which to organise the coding frame. Delphi is 'a method for structuring a group communication process' which facilitates a panel of experts to deal with a complex problem by focusing on a set of predetermined tasks. A group of experts was contacted including some from social work, health, education, legal studies and voluntary agencies and asked to complete a series of sequential questionnaires, which were interspersed with summaries of respondents' comments (Gabb et al., 2006). They were asked to rate and comment on a series of statements regarding aspects of the Inquiry and child protection in general. For example, "Managers did not take adequate responsibility for the handling of the Victoria Climbié case". They were also asked whether they had accessed the Victoria Climbié website and how they might make use of the data corpus. There were three rounds of consultation and 50 experts returned questionnaires. The findings of this exercise were analysed and themes emerged from the consultation, many of which had not been identified by the researchers.

Thematic coding

Stage 2 of the project coded a subset of the data using more than 45 major themes identified in the Delphi investigation. Atlas.ti, a computer assisted qualitative data analysis (CAQDAS) program was used to do this. The coded data have been made available to professionals and researchers as an Atlas.ti electronic file.

The University Repository

The project to set up a Repository began in January 2006 with the support of the Chair of the University Research Committee, the chairs of the School Research Committees and the Director of Computing and Library Services. The Repository was formally launched in May 2007 by the Vice Chancellor, who has continued to support the objective of embedding the Repository in the research and scholarly communication of the University, and was initially used for the RAE submission in 2008.

The University Research Committee subsequently suggested that the Repository should record details of all University research output from 2000 to date. In direct response to this the Repository now contains approximately 7200 records of research output, including nearly 1900 full text research items and 200 theses dating back as far as 1966.

Phase II of the Repository project is to build upon the initial success and goodwill by ensuring that the Repository becomes central to the University research strategy and workflows. The Repository will play a central role in the preparation for the REF and will include raw data as well as research outputs on open access.

Aims and Objectives

The main aim of the project was to make the coded and categorised Climbié corpus available to users via the University of Huddersfield Repository. The Climbié data corpus includes all the transcriptions from the 68 days of oral evidence which was made available to the Project by the Department of Education and Skills. There is testimony from 168 witnesses, which produced over 5,200, single-spaced, A4 pages of evidence (approx. 2 million words).

Core aims and objectives of the project were to:

- Complete qualitative coding of corpus
- Develop Repository using API to
 - accept data in DExT format
 - enable code Searches alongside plain text searches
 - enable biographical/cover sheet data retrieval
 - enable critical incident material retrieval
- Evaluate technical issues
- Assess the of usability of outputs using Delphi team
- Develop style sheets for educational use
- Disseminate the outcomes of the project via journal articles and conference presentations
- Produce technical and supporting documentation

Methodology

Relationship with research data deposited in the ESDS Data Archive

Unlike most qualitative data coordinated or held by the ESDS Qualidata, the data set (though not the analysis) is already publicly available through a website rather than having been collected for a specific research project. There are considerable advantages to using publicly available data, reducing ethical problems of confidentiality and accessibility associated with gaining 'informed consent' (Hammersley 1997, Fielding 2000) making it particularly good for teaching use. All the witnesses and barristers will have been aware that their evidence was being made available on the internet; hence there is no need to gain permissions for use or re-use of the data. The company which managed the website, which might be considered to have copyright of the data, has also indicated that the material is publicly available (personal communication) and the Department for Education and Skills has provided the project with a complete version of the website.

ESDS Qualidata have developed a proposed standard for representing qualitative data and its coding in archives. This is the DExT standard. Atlas.ti, the software we have used for coding the data set produces XML output of the data set and coding. This was used for importing the data into the Repository and for delivering requested copies of the data set. It is a subset of the DExT standard and hence is compliant with the standard.

Proposed methods for retrieval and categorisation

As the Climbié data corpus is not just a collection of records, but has been coded and annotated, the project undertook development work of the Repository API to create functions that allowed the retrieval of the data using the codes and annotations.

Implementation

Retrieval based on thematic coding

Users can use the coding to pinpoint particular themes which interest them and retrieve all the relevant text. For example a central theme is inter-agency communication. For other users such themes are too broad and they may wish to identify more specific data, for example cross-referencing inter agency communication and nursing in order to study how nurses communicate with different professionals.

Although we intended to develop the means of retrieving relevant data using the Repository API, the response time to a query done in this way turned out to be far too slow as it involved the parsing of the very large XML file for each query. Therefore, we established a version of the parsed file on a separate server where queries could be delivered to the used at reasonable speed.

The original Dephi consultation established a set of themes that we used to code the data. Further work on this project and formative evaluation suggested further themes that would make it easier to create the subset of the data that researchers and educationalists would need. A full list is in Appendix 1.

Users can use a menu system to select Boolean combination of codes for which they wish to retrieve text. The codes used have been extended so that they include biographical data, for instance identifying the speaker by job and employer (local authority). Users are helped in this by guidance notes that indicated possible teaching and research questions and which retrievals should be used to get the requisite data. In addition, these notes then suggest further questions and exercises that will guide students using the data in a teaching context

Users who wish to, can also download the entire coded data file that can be used in any software that can read XML files and DEX standard files. (At the moment this is just Atlas.ti)

Outputs and Results

Formative evaluation with researchers, teachers and potential professional trainers suggested that most had only a limited understanding of the process of coding retrieval at best. It was therefore important to develop description and guidance for the snapshots and subsets of data that particular code retrievals would produce. These linked learning resources have been produced and are on the website and linked to the code retrieval.

In 2010 we ran a dissemination session coordinated by the HEA centre for Social Work and Social Policy (SWAP), we will follow this by writing a series of short publicity articles for centre newsletters, copies of which will also be available through the website and Repository.

The Victoria Climbié Inquiry Data Corpus Online website

Web access to the corpus (and hence to the data on the Repository) has been implemented. The website is at <http://victoriaclimbie.hud.ac.uk/>

The data from the Victoria Climbié Inquiry and Laming Report forms a very powerful source of data about child protection practice in the UK and some of the key issues that affect that practice. The project developed a mechanism for obtaining 'snapshots' of the data set that can be used by students in projects and studies about child protection. These used the coding and linked data in the

Repository. They are, in effect, themed subsets of whole corpus that will be of sufficient size to support research projects and contain key relevant information, yet not be too large for college and university undergraduate students to use in project work.

There are two main sections to the website, a set of pages with information about the inquiry, its background, the people involved and the research literature it has generated. We have licensed two files of information from the Guardian Website for five years. This information now appears on the website with some colour coding (see Fig 1.)

FAMILY

Victoria Climbié
After suffering months of horrific abuse, Victoria Climbié died aged just eight on February 25 2000. Nearly 18 months earlier her parents had sent her from her home in a shanty town in the Ivory Coast to live with her great-aunt in the hope she would get a good education and enjoy a better life. Instead she was **beaten with bicycle chains and kept trussed up in a plastic sack** in an unlit, unheated bathroom.

Berthe Amoissi and Francis Climbié
Victoria's parents. Berthe and her husband Francis, say Marie Thérèse Kouao duped them into believing she would take good care of their daughter. The couple, who have six other children, live in the poverty-stricken suburbs of Abidjan, capital of the Ivory Coast. They were the first witnesses to give evidence at the inquiry on Friday September 28. Mr Climbié broke down after being shown pictures of Victoria with a tooth missing and scald marks on her face. The couple's solicitor has warned they might bring private prosecutions against **police officers and care workers whose negligence contributed to their daughter's murder.**


KILLERS

Marie Thérèse Kouao
Victoria's great-aunt was found guilty of her murder at the Old Bailey in January 2001. **The girl lived with Kouao** and her boyfriend in his tiny studio flat in Tottenham, north London. Kouao, a French citizen, always denied mistreating Victoria - whom she called Anna - and **claimed the girl's horrendous injuries were self-inflicted.** She also **dragged her niece to a church to be exorcised,** claiming the eight-year-old was possessed by evil spirits. Giving evidence on Tuesday 8 January this year, Kouao - the first convicted murderer to give evidence in person at a public inquiry - ranted and raved that she was innocent, claimed she was the victim of a conspiracy and criticised Victoria's parents.

Carl Manning
Kouao's boyfriend. Manning was also **found guilty of Victoria's murder.** In a chilling confession read in court, the bus driver told police: **"You could beat her and she would not cry at all. She could take the beatings and pain like anything."** In video-recorded evidence played at the inquiry on Wednesday January 30, he apologised for his part in Victoria's "sickening" death and **claimed that child protection agencies could not be blamed for her suffering.**

HARINGEY COUNCIL

Lisa Arthurworrey



Victoria's allocated **social worker** was suspended by **Haringey council** after Kouao and Manning were found guilty of the girl's murder in January 2001. Ms Arthurworrey, who worked in the child protection team at the council's north Tottenham office, is accused by her employers of **negligence and failing to safeguard and promote the interests** of a **child in her care.** The social worker, who told the inquiry she has suffered from depression since Victoria's death, claims she has been the victim of a witch hunt, and was poorly **supervised** and **advised** by her seniors.

Angella Mairs

A **social work manager** and Lisa Arthurworrey's second supervisor, Ms Mairs has also been suspended by **Haringey council.** She is alleged to have removed a **key document** from Victoria's file after her murder, made false or misleading statements and failed to **maintain childcare standards.** She denied these charges at the inquiry on Friday December 7.

Figure 1. Sample information page from the Victoria Climbié website

The colour coding here and elsewhere on the site is to indicate examples of the kind of categories that can be used to define the kinds of subsets of data that can be retrieved from the full corpus.

The second main part of the website consists of suggestions, in both a teaching and a research context of how data might be retrieved and what kinds of teaching exercise and what kinds of research projects this might be used to support. An example of this can be seen in Fig. 2.

Using the code 'workplace practices'

This code can be used to retrieve information about each professional involved in the court case, their individual role, workplace activities and perspectives. This can be particularly useful because it gives an insight into what was taking place in all the different organisations. Hence the user can begin to gain an understanding of the different models, attitudes, tensions, problems and successes encountered by the different professionals involved in the case. These data can then be used in many ways such as helping the researcher/student analyse what works in practice and what does not, and highlight the potential for missed opportunities in a child protection case.

E.g. What workplace practices affect the actions of a SW compared with a paediatrician or a nurse working on a child protection case?

Do the following:

1. If anything is still selected (text in the lists are highlighted) then click on the 'Reset Form' button at the bottom of the page.
2. In the area where you can select **the role or occupation of the witness:** - Select 'Social Worker' (click once with the mouse)
3. In the area where you can select **whether the witness was discussing workplace practices:** - Select 'Workplace practices' (click once with the mouse)
4. Click on the 'Show Quotations' button at the bottom of the page
5. A new page opens showing all the data you have retrieved. This can be cut and pasted into a word processing file for later use.
6. Repeat this for other occupations such as paediatrician or nurse.

Teaching use

Exercise: in groups retrieve and analyse data about a specific group of people and their workplace practices. Compare and contrast the different roles. Question: What are the different issues that each group has to consider when working on a child protection case? How can these issues impact on other professional groups working on the case? What issues can affect communication between the different groups?

Figure 2. An example of a way to retrieve data from the corpus and a sample of teaching use

These suggestion pages also contain ideas for further refinement of the retrieval that can address additional teaching and research issues. An example is given in Fig. 3.

Refining the retrieval

Combine workplace practices, job/status and contact with VC to answer the question: What was different about the actions of the people in the group who had contact with VC compare to those who did not?

E.g. What workplace practices affect the actions of a SW who had contact with VC compared with a paediatrician who had no contact?

Do the following:

1. If anything is still selected (text in the lists are highlighted) then click on the 'Reset Form' button at the bottom of the page.
2. In the area where you can select **the role or occupation of the witness:** - Select 'Social Worker' (click once with the mouse)
3. In the area where you can select **whether the witness speaking had contact with Victoria Climbié:** - Select 'Contact with VC' (click once with the mouse)
- 4.
5. In the area where you can select **whether the witness was discussing workplace practices:** - Select 'Workplace practices' (click once with the mouse)
6. Click on the 'Show Quotations' button at the bottom of the page
7. A new page opens showing all the data you have retrieved. Cut and paste this into a word processing file for later use.
8. Click on the 'Reset Form' button at the bottom of the page.
9. In the area where you can select **the role or occupation of the witness:** - Select 'paediatrician' (click once with the mouse)
10. In the area where you can select **whether the witness speaking had contact with Victoria Climbié:** - Select 'Contact with VC' (click once with the mouse)
- 11.
12. In the area where you can select **whether the witness was discussing workplace practices:** - Select 'Workplace practices' (click once with the mouse)
13. Click on the 'Show Quotations' button at the bottom of the page
14. A new page opens showing all the data you have retrieved. Cut and paste this into a word processing file for later use.

Figure 3. Instructions about how to refine the retrieval

These pages also include the access to the data in the corpus itself and the options available for retrieving subsets of the data. Figure 4 shows this data retrieval page.

The screenshot shows the 'Choose statements to display' page of the Victoria Climbié Inquiry Data Corpus Online. The page features a navigation menu on the left with links for 'University homepage', 'Project homepage', 'The Inquiry', 'Background', 'Witnesses', 'Research Literature', 'The Data Corpus', 'How to use', 'Learning Resources', 'Research Resources', 'Retrieve Data', 'Links', 'Project team', and 'Contact details'. The main content area is titled 'Choose statements to display' and includes instructions: 'The figures in parentheses in the menus below indicate the number of statements classified under each heading. See the [How to use this corpus](#) page'. Below this, there are four filter sections, each with a dropdown menu and a highlighted instruction: 1. 'Narrow your choice by selecting **the role or occupation of the witness**:' with a dropdown list including 'Witnesses' Job/Status: Admin Officer (227)', 'Witnesses' Job/Status: Ambulance Technician (79)', 'Witnesses' Job/Status: Child Protection Advisor (55)', 'Witnesses' Job/Status: Childminder+childminder's family (554)', 'Witnesses' Job/Status: Clergy (362)', 'Witnesses' Job/Status: Councillor (1481)', 'Witnesses' Job/Status: Education Officer (36)', 'Witnesses' Job/Status: General Practitioner (1)', 'Witnesses' Job/Status: Hospital doctor (1143)', and 'Witnesses' Job/Status: Housing Officer (1045)'. 2. 'Narrow your choice by selecting **the borough or the hospital where the witness worked**:' with a dropdown list including 'Hospitals: Central Middlesex Hospital (2749)', 'Hospitals: North Middlesex Hospital (3415)', 'SS Borough: Brent (8263)', 'SS Borough: Griffith (18)', 'SS Borough: Haringey (11031)', and 'SS Boroughs: Ealing (2460)'. 3. 'Narrow your choice by selecting **whether the witness speaking had contact with Victoria Climbié**:' with a dropdown list including 'Contact with VC (13011)' and 'No contact with VC'. 4. 'Narrow your choice by selecting **whether the witness was discussing workplace practices**:' with a dropdown list including 'Workplace practices (22007)' and 'No discussion of workplace practices'. At the bottom, there is a fifth filter section: 'Narrow your choice by selecting **the topic being discussed by the witness**:' with a dropdown list including 'Assessment: Action/inaction (1806)', 'Assessment: Decision plan of action (964)', 'Assessment: Exchanging information (428)', 'Assessment: General (554)', 'Blame/Mistakes (9433)', 'Categorisation: Child in need case (298)', 'Categorisation: Child protection case (520)', 'Categorisation: Emotional abuse case (111)', 'Categorisation: Housing/homeless/subsistence case (119)', and 'Categorisation: Sick Child case (139)'. Below the dropdowns are 'Show Quotations' and 'Reset Form' buttons, and a timer showing 'Elapsed time: Less than one second'. On the right side, there is a 'Partners' section with the 'JISC' logo and links to 'The Victoria Climbié Foundation' and 'Open Access and Institutional Repositories with Eprints'.

Figure 4. The page where the user specifies what data they would like retrieved from the Corpus

Again this is colour coded to match that used in the background pages and in the suggestions for retrieval and educational and research use. The idea is to make it easier for those unfamiliar with multi-level retrieval to understand the logic of what they are doing.

Learners and researchers are able to search the data corpus for those occasions where certain subjects, professions or instances in the life of Victoria were examined. For example, a search might identify all those occasions in the corpus where staff supervision was examined, or all those occasions where inter-agency communication took place. It provides an important resource of accessible data for researchers, practitioners, trainers and managers, who can use it to identify research topics and data, case studies for training material, teaching risk assessment and decision-making, or processes of policymaking.

Outcomes

The principal outcome of the project was the retrieval of subsets of the data using the Repository API, and indeed this was done initially, however, it turned out that each query required the eprints software to re-parse the whole data set. This introduced an unacceptably long wait (more than a minute) whilst seemingly nothing happened. So it was decided to establish a parsed version of the data set on a local server from which retrievals could be done in a much faster time.

This is a consequence of the difference between the mode of operation of a Repository, which focuses on the retrieval of documents and what the project required, which was the retrieval of selected subsections of the document. This was exacerbated by the size of the document.

Nevertheless, the Repository remains the principal source of the document so that at any time (e.g. after the repository documents have been updated) the local server version can be recreated.

A second major outcome is the recognition that the kind of coding needed to prepare such a data set for secondary use is different from the kind of coding that might be created by a social researcher using the data set in a research project. The needs of the secondary user (or the users seeking just a subset of the data) are different from those of the primary researcher. We are currently preparing a publication on this issue. We believe it will have important implications for those who are promoting the qualitative secondary use of research and other large data sets from data archives. For example, it may mean the at least some of the coding undertaken by principle researchers for their own analysis will be of little use to users coming to do secondary analysis.

On the other hand our project does demonstrate a way in which other publicly available data sets, which have not been subject to any kind of research analysis, might be made accessible to both researchers and to teachers. There are now a number of inquiries and commissions which are making full transcripts of witness testimony (along with other evidence) publicly available on the web. In the UK these include the Bichard Inquiry and the Hutton Inquiry. There is great potential for the use of these data sets in research (Gibbs and Hall, 2007) and in teaching. This project has demonstrated how with simple coding and with linked guidance notes such data sets can be made accessible to users.

During 2010 we ran dissemination sessions at the appropriate HEA centre (SWAP) and will write a short publicity article for the centre newsletter.

In October and November 2010 we will evaluate them in a teaching context by using some with classes of students. Whilst students will not use the guidance notes and the menu based retrieval system directly (we envisage this will be done by their teachers), feedback from them will enable us to improve the content of the guidance notes.

Evaluation of the learning resources

A task of the project was to evaluate the learning resources and the possible uses of the Corpus. However, because the interface was only completed towards the end of the project, there has not been the opportunity to use the corpus with students. However, we have gathered feedback from academics and practitioners testing out the website and from those who attended the launch event. This led to some suggestions both for improving the website design and for different ways of categorising and presenting the data for retrieval.

There was some positive feedback from all parties in the launch event evaluation which has led to some interesting, exciting uses of the database. The academics will be using the material in both undergraduate and postgraduate teaching during 2010 – 2011. The general opinion about the strength of the materials produced by the research team is that they give access to the daily activities involved in social work practice, especially the relevance to serious case reviews. This was reiterated by the practitioners who felt that there is a definite gap in the training available to them across several

local authorities. With this in mind the team would like to develop the materials further and evaluate the use of them within both the academic and professional practice arena.

Conclusions

The conclusions drawn from the project are:

- The Repository is not the best way to archive and retrieve the coded data set due to usability and robustness problems. In short the data set was so large, retrieval of the sub sets was taking too long to make it practical. Hence the Repository now operates as gateway for storing the full data set
- There is a difference between coding for research and teaching and learning. The coding for research would be in depth, interpretative, analytical coding for specific use by the particular researcher. For teaching, training and learning the coding has been applied in order for the user(s) to gain access to the different themed subsets which then can be analysed in much the same manner used by a research focus
- This would appear to be the best/most appropriate way forward for coding of data because the data set is accessible, usable and can be tailored to meet the specific requirements of the different users

Implications

Despite the small amount of evaluation carried out on the project and the coding of the data set there are several exciting possibilities for use and development of the materials produced. These are:

- In a teaching and learning environment the coded data set gives access to detailed information about the daily activities involved in a care practitioner's workload. This was specifically related to serious case reviews which are often not accessed or understood in detail because of the complexity and size of the data sets which are publically available. However, from these data sets there is the potential to explore issues such as: personal responsibility, accountability, court processes/procedures etc., in great detail. This potential has been recognised by academics within the University and as a result used to underpin teaching and assessment for undergraduate and postgraduate students. However, the data set requires further coding in order to fulfil its full potential which will require several months work. Furthermore an in depth evaluation of the use of the coded data set in a teaching context would enable the team to develop the project in the future
- Amongst the practitioners involved in the evaluation there was a great deal of interest and enthusiasm about the use of coded data sets. Once again the potential is for teaching and training care practitioners to develop their professional practice. The data set offers an insight into the daily activities of not only social workers but all other practitioners who could be involved the care sector – and in this instance in a serious case review. This is particularly relevant given the political and social interest in child protection, for example the death of 'Baby P' from which a report was produced that highlighted the same mistakes and problems that were evident in the Victoria Climbié case. Hence the data set is relevant now and can be used to inform policy and practice in many ways such as: to develop individual professionalism; to understand the roles, activities and interactions of care providers, managers and practitioners; to inform/develop policy and practice within the sector and so on
- There is developmental work required on the project in order to meet the pragmatic learning and training needs of the care sector; however, there is the potential to provide materials which can be used to fill an existing gap in the training provided across the sector

- The project has been a test bed for future coding projects; what has become evident is that publically held data sets contain a wealth of information which, if made more accessible, can be of great value for developing policies and practices within the community – in this case the care sector - and academia

Further work

The evaluation process has highlighted several areas for further development on the project, these are:

- additional coding on the data set
- development for teaching application
- development for training application within the care sector
- evaluation of the use of the data set by both academics and practitioners
- update the data set held in the Repository
- coding of other similar data sets for use in academia and the relevant community

References

Delbecq A. L., Van de Ven A. H., Gustafson D. H., (1975) *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Glenview, IL: Scott Foreman & Co

Fielding, Nigel (2000, December). The Shared Fate of Two Innovations in Qualitative Methodology: The Relationship of Qualitative Software and Secondary Analysis of Archived Qualitative Data [43 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 1(3). Available at: <http://qualitative-research.net/fqs/fqs-eng.htm> [Date of access: Sept. 20th 2004]

Gabb, J., Balen, R., Gibbs, G.R., Hall, C. & Teal, A. (2006) "Victoria Climbié Inquiry Data Corpus Project: Using the Delphi Method in Multidisciplinary Child Protection Research" *British Journal of Social Work* Vol.36 (4) pp. 577-596 (<http://eprints.hud.ac.uk/329/>)

Gibbs, G.R. and Hall, C. (2007) "The Research Potential of Testimony from Public Inquiry Websites" *Children & Society* Vol. 21, pp. 69–79 (<http://eprints.hud.ac.uk/345/>)

Hall, C (2003) *Themes In the Part II seminars of the Victoria Climbié Inquiry: Report to the NSPCC* University of Huddersfield

Hammersley, M. (1997) 'Qualitative data archiving: some reflections on its prospects and problems'. *Sociology*, 31(1), 131-42

Laming (2003) Report of the Inquiry into the death of Victoria Climbié, London: Department of Health

Appendix 1. Codes used

Assessment: Action/inaction
Assessment: Decision plan of action
Assessment: Exchanging information
Assessment: General
Blame/Mistakes
Categorisation: Child in need case
Categorisation: Child protection case
Categorisation: Emotional abuse case
Categorisation: Housing/homeless/subsistence case
Categorisation: Sick Child case
Chronology: Phase 1
Chronology: Phase 2
Chronology: Phase 3
Chronology: Phase 4
Chronology: Phase 5
Chronology: Phase 6
Class
Communication between agencies
Communication within agencies
Contact with VC
Family Status
Files/Records
Gender
Good practice
Guidelines and Procedures
Hierarchical exchanges
Hospitals: Central Middlesex Hospital
Hospitals: North Middlesex Hospital
Hospitals: St Mary's
Immediate Family
Inquiry Procedure
Locale
Locum staff
Lying/Discrepancies
Management: Departmental audits
Management: Policy making
Management: Resources
Management: Responsibilities and direction
Management: Roles
Management: Targets
MR GARNHAM: You appeared to be..
Organisational Issues
Prof/ Job discussed: Interpreter
Prof/Job discussed: Administration
Prof/Job discussed: Ambulance
Prof/Job discussed: auditors
Prof/Job discussed: Child protection
Prof/Job discussed: Childminder
Prof/Job discussed: Clergy
Prof/Job discussed: Councillors
Prof/Job discussed: Customer Service/Reception
Prof/Job discussed: Education
Prof/Job discussed: Friends (of aunt + boyfriend)
Prof/Job discussed: GP's
Prof/Job discussed: Health Visitor
Prof/Job discussed: Hospital doctors

Prof/Job discussed: Housing
Prof/Job discussed: Interpreter
Prof/Job discussed: LA Directorate
Prof/Job discussed: Lawyers
Prof/Job discussed: NSPCC family centre
Prof/Job discussed: Nurses
Prof/Job discussed: Paediatricians
Prof/Job discussed: Pathologist
Prof/Job discussed: PCT directorate
Prof/Job discussed: Police Officers
Prof/Job discussed: Schooling
Prof/Job discussed: Social work
Prof/Job discussed: SS management
Prof/Jobs discussed: Unions
Race
Resources
Role/responsibility
Significant Incidents
Skills, Expertise, experience
Somerset Gardens
SS Borough: Brent
SS Borough: Enfield
SS Borough: Haringey
SS Boroughs: Ealing
Supervision
Training
Witnesses' Job/Status: Admin Officer
Witnesses' Job/Status: Ambulance Technician
Witnesses' Job/Status: Audit Commissioner
Witnesses' Job/Status: Child Protection Advisor
Witnesses' Job/Status: Childminder+childminder's family
Witnesses' Job/Status: Clergy
Witnesses' Job/Status: Councillor
Witnesses' Job/Status: Customer Services Officer
Witnesses' Job/Status: Education Officer
Witnesses' Job/Status: Family
Witnesses' Job/Status: Friends
Witnesses' Job/Status: General Practitioner
Witnesses' Job/Status: Health Visitor
Witnesses' Job/Status: Hospital doctor
Witnesses' Job/Status: Housing Officer
Witnesses' Job/Status: Lawyers + Solicitors
Witnesses' Job/Status: Local Authority Directorate
Witnesses' Job/Status: Nurse
Witnesses' Job/Status: Paediatrician
Witnesses' Job/Status: Pathologist/Forensic Scientist
Witnesses' Job/Status: PCT directorate
Witnesses' Job/Status: Police Officer
Witnesses' Job/Status: School Teacher
Witnesses' Job/Status: SS Directorate
Witnesses' Job/Status: Union Reps
Witnesses' Job/Status: Social Worker
Witnesses' Jobs/Status: NSPCC
Work atmosphere
Workplace practices



JISC Completion Report (*JISC Internal only*)

Lessons Learned

Aims and Objectives

The main aim of the project was to make the coded and categorised Climbié corpus available to users via the University of Huddersfield repository. The Climbié data corpus includes all the transcriptions from the 68 days of oral evidence which was made available to the Project by the Department of Education and Skills. There is testimony from 168 witnesses, which produced over 5,200, single-spaced, A4 pages of evidence (approx. 2 million words).

Core aims and objectives of the project were to:

- Complete qualitative coding of corpus
- Develop Repository using API to
 - accept data in DExT format
 - enable code Searches alongside plain text searches
 - enable biographical/cover sheet data retrieval
 - enable critical incident material retrieval
- Evaluate technical issues
- Assess the of usability of outputs using Delphi team
- Develop style sheets for educational use
- Disseminate the outcomes of the project via journal articles and conference presentations
- Produce technical and supporting documentation

In the main the aims and objectives were achieved with the exception of: developing style sheets for educational use; disseminating the outcomes of the project via journal articles and conference presentations and evaluation of the use of the data set within and educational and practice setting. Some evaluation was undertaken by both practitioners and academics which highlighted the need for further development – coding of the data, to refine the project outcomes so that they can be used for teaching, training and learning for practitioners and academics. This would then facilitate an in depth evaluation for the project. This was considered a necessary and exciting development process given the value of the data set for research to be used to inform policy and practice in the care sector. This type of coded data was also regarded as valuable if applied to other publicly held data sets for example in adult mental health.

Overall Approach

The time scale for this type of project was too short; in order for effective evaluation and development to be carried out the time allowed needed to be extended by a further 12 months. The project was fortunate in that it recruited the RA immediately and she stayed in post to the end of the project; had this not been the case then once again the time scale would have been too short for recruitment and/or replacement of a key member of the team. This would have prevented the project from meeting its aims and objectives or indeed producing many useful outcomes.

Project Outcomes

- The principal outcome of this project is rather negative. Although we thought that retrievals of subsets of the data could be implemented using the Repository API, and indeed this was done initially, it turned out that each query required the repository software to re-parse the

whole data set. This introduced an unacceptably long wait (more than a minute) whilst seemingly nothing happened. Nevertheless, the repository remains the principal source of the document so that at any time (e.g. after the repository documents have been updated) the local server version can be recreated.

- A second major outcome is the recognition that the kind of coding needed to prepare such a data set for secondary use is different from the kind of coding that might be created by a social researcher using the data set in a research project. The needs of the secondary user (or the users seeking just a subset of the data) are different from those of the primary researcher. We are currently preparing a publication on this issue.
- Our project does demonstrate a way in which other publicly available data sets, which have not been subject to any kind of research analysis, might be made accessible to both researchers and to teachers. There are now a number of inquiries and commissions which are making full transcripts of witness testimony (along with other evidence) publicly available on the Web. In the UK these include the Bichard Inquiry and the Hutton Inquiry. There is great potential for the use of these data sets in research (Gibbs and Hall, 2007) and in teaching. This project has demonstrated how with simple coding and with linked guidance notes such data sets can be made accessible to users.
- An unexpected outcome was the reaction of both the academic and the practitioner community towards the product. Whilst this was initially a project about the teaching and learning of qualitative data analysis what became evident was that the product was of great value on a subjective level for the users. As a result the data set is to be used to underpin teaching for undergraduate and postgraduate students and also for training within the care sector. It was considered that there is much scope for coded data sets in that they enable the user to access a wealth of information which can be used to inform teaching, learning and policy and practice.

Project Partners

We did not have any other project partners other than eprints who we consulted over for the storing of the data set. However, we worked collaboratively with the Social Work and Social Policy council (SWAP) to launch the project which proved very successful. We also have paid the Guardian newspaper for the use of the material developed by the journalists which will appear on the website for the next 5 years.

Project Management

It would have been useful to have a steering group from the onset of the project.

Programme Support

Support was be provided by the JISC Programme Manager in the first instance, including notification of relevant events or developments.

Future Work

The evaluation process has highlighted several areas for further development on the project, these are:

- additional coding on the data set;
- development for teaching application;
- development for training application within the care sector;
- evaluation of the use of the data set by both academics and practitioners;
- update the data set held in the repository;
- coding of other similar data sets for use in academia and the relevant community.

Sustainability Plan

The work of adding content to the Repository covered by the scope of this bid has a finite goal.

Planned outputs of the project will be taken further as follows

- Report from the Climbié user group on appropriate forms of data presentation, themes needed and the use of archive in management and teaching contexts;
- An online, DExT-METS compliant, XML-based data corpus with coded and annotated data,
- Report on the appropriateness of the work on the Climbié corpus to other public inquiry archives and recommendations for the design of such developments;
- Use by other projects;
- Development of teaching and learning content through JORUM/Open Research Initiative.

Budget

A surplus in the budget has been identified to fund a research assistant for an extra 3 months (0.5 FTE). This would be to develop the future work described above.

It is intended to detail the outcomes of this further work on the website (<http://victoriaclimbie.hud.ac.uk/>) in the latter part of 2010.



JISC Project Final Reports Budget Template

Directly Incurred Staff	TOTAL BUDGET £	Year <Oct 08 – Mar 09> Actual Expenditure	Year <Apr 09 – Mar 10> Actual Expenditure	TOTAL EXPENDITURE £	TOTAL VARIANCE
Research Assistant, 0.5 FTE	£15,721	£0	£14,901.26	£14,901.26	£819.74
Total Directly Incurred Staff (A)	£15,721	£0	£14,901.26	£14,901.26	£819.74
Non-Staff					
Travel and expenses	£600	£91.50	£734.65	£826.15	-£226.15
Hardware/software	£3,000	£0	£0	£0	£3,000.00
Dissemination	£500	£0	£308.85	£308.85	£191.15
Evaluation	£500	£0	£0	£0	£500.00
Other	£7,500	£0	£8,361.88	£8,361.88	-£861.88
Total Directly Incurred Non-Staff (B)	£12,100	£91.50	£9,405.38	£9,496.88	£2,603.12
Directly Incurred Total (A+B=C) (C)	£27,821	£91.50	£24,306.64	£24,398.14	£3,422.86
Directly Allocated					
Staff	£43,573	£22,949.87	£30,994.75	£53,944.62	-£10,371.62

Estates	£2,629	£1,871.01	£2,520.15	£4,391.16	-£1,762.16
Other	£	£	£	£	£
Directly Allocated Total (D)	£46,202	£24,820.88	£33,514.90	£58,335.78	-£12,133.78
Indirect Costs (E)	£24,948	£13,909.33	£18,735.03	£32,644.36	-£7,696.36
Total Project Cost (C+D+E)	£98,971	£38,821.71	£76,556.57	£115,378.28	-£16,407.28
Funds Received from JISC	£49,486	£16,650.71	£32,835.29	£49,486.00	£0
Institutional Contributions	£49,485	£22,171.00	£43,721.28	£65,892.28	-£16,407.28

Nature of Institutional Contributions

Directly Incurred Staff					
Post, Grade & % FTE	£	£	£	£	£
Directly Incurred Non Staff					
Hardware/Software etc.	£	£	£	£	£
Directly Allocated					
Staff, Estates etc.	£24,537	£8,261.67	£24,986.25	£33,247.92	-£8,710.92
Indirect Costs					
Indirect Costs	£24,948	£13,909.33	£18,735.03	£32,644.36	-£7,696.36
Total Institutional Contributions	£49,485	£22,171.00	£43,721.28	£65,892.28	-£16,407.28