

# Documenting the Cambodian Genocide on Multimedia

Helen Jarvis  
in collaboration with Nereida Cross  
The University of New South Wales  
Sydney, Australia

Paper presented at the  
Genocide Studies Program Seminar  
October 1, 1998

GS04

YALE UNIVERSITY  
Mellon Sawyer Seminar Series  
Ben Kiernan, Director

Yale Center for International and Area Studies  
Luce Hall, 34 Hillhouse Ave.  
New Haven, CT. 06520-8206



# DOCUMENTING THE CAMBODIAN GENOCIDE ON MULTIMEDIA

paper presented in the Mellon Foundation Sawyer Seminar Series  
Genocide Studies Program  
Yale University, New Haven, Connecticut  
1 October 1998

Helen Jarvis  
in collaboration with Nereida Cross  
School of Information, Library and Archive Studies  
The University of New South Wales  
Sydney, Australia

The two images that dominate Cambodia's representation in book, film and art alike are those of the temples of Angkor and the terrors of Angkar -- "the organisation" of the Khmer Rouge (KR) and its killing fields. And it is of course the latter with which we are concerned in the Cambodian Genocide Program (CGP). Angkar is generally represented in the emotive and dramatic images of skulls and black-clothed ant-like slaves building dykes. Our challenge is to move beyond those authentic yet essentially reductionist images to arrive at a deeper understanding of what took place in Cambodia from 1975 to 1979. So much has been invested by all parties and observers over more than twenty years of political disputation and discord. So much has been lost or destroyed. So much has been forgotten or covered up. How does one begin to provide the documentation for research and rescue of the evidence? This was our task as we embarked upon our program in early 1995.

Our responsibility at the University of New South Wales has been to design the overall structure of the four integrated Cambodian Genocide Data Bases, to develop the methodology to be followed, to select hardware and software, and to advise on all aspects of documentation, including training staff and maintaining quality. This research project has required path-breaking design to enable integration of data found in multiple formats (paper records, photographs and film, oral testimony, physical geographic sites, remote sensing images, computer files) and in multiple languages (principally Khmer, French and English, but also in Vietnamese, Thai, Chinese, Russian) and locations (Cambodia, Vietnam, United States, Australia and elsewhere) with the research team itself been based in three different places (Phnom Penh, New Haven and Sydney).

We have launched investigations for both known and previously unknown evidence and records, and have had to contend with an unexpected plethora of material requiring classification, coding and preservation. Existing international standards (such as Machine-Readable Cataloguing formats and Human Rights Classification Codes) have had to be applied and frequently extended to cope with our unusual range of data, and the software has been pushed into new and challenging areas (such as displaying Khmer script; linking retrieved records to associated image files, and displaying retrieved records and images on the Internet).

In addition to meeting our research objectives, we have needed to have high regard for the integrity of all our data, its provenance and its security, due to the likelihood of its being used in evidence in a future trial. Needless to say, the continuous media spotlight, the intense political interest in the issue, and the continued presence and threat of the Khmer Rouge have demanded constant vigilance regarding the security of both staff and documents, as well as a high degree of responsiveness and sensitivity in presenting our results to the public, particularly as regards respecting the memory of those killed and the privacy and integrity of the survivors.

### **The Cambodian Genocide Data Bases (CGDB)**

To manage the material we selected CDS/ISIS, the database management program developed by Unesco,<sup>1</sup> and its Windows interface Winisis. This is a micro computer based information retrieval software package, used quite widely throughout the world, particularly in the developing countries. It is available from Unesco free of charge which is one of its major attractions. It can run in different languages and indeed in different scripts. CDS/ISIS is a very powerful and flexible package, particularly suited for the complexities posed by a wide and ever growing multiplicity of data types and formats, and for the challenge of handling material in at least two different languages and scripts. We have developed a suite of databases, called CGDB (the Cambodian Genocide Data Bases), within which we manage bibliographic, biographic, geographic and image-based material.

Our work has taken place with the active involvement of staff on three continents who have communicated on a daily basis. For the first few months faxes and international telephone calls were an indispensable aspect of our work, due to the slow and uncertain nature of the postal service to and in Cambodia. And the first transfer of files was implemented by physically carrying a disk drive from Phnom Penh to New Haven! As soon as DC-Cam became a member of Camnet<sup>2</sup>, email transformed our modus operandi and when in 1997 it gained full Internet access through Telstra's Bigpond service, we have had faster and easier email access, and have been able to utilise File Transfer Protocol (FTP) for transferring data files, including GIS and scanned data. As well, the staff in Cambodia have also been able to benefit from using the CGP site's Web browsing capabilities, although this is still costly in Cambodia (US\$7/hour).

### **CGP Bibliographic Database (CBIB)**

For the structure of the bibliographic database we adopted UNIMARC, an international bibliographic data format, supplemented to cater for the wide range of archive and manuscript material we must include -- both print and non-print

---

<sup>1</sup> CDS/ISIS (which is a DOS based program) and Winisis are available from national or regional distributors. For information as to your local distributor contact Unesco, 7 Place de Fontenoy, 75352 Paris 07, France.  
<http://www.unesco.org/webworld/isis/isis.htm>

<sup>2</sup> Camnet (a store and forward email system using Fido technology) was established in 1995 by the Cooperation Committee for Cambodian (CCC) with assistance from Pactok, an Australian NGO assisting development of low-cost computer communications in the Asia Pacific Region. Camnet ceased operations at the end of 1997 when full Internet services became available in Cambodia.

(including articles, handwritten reports, petitions and confessions).<sup>3</sup> We have had to determine codes for the identification of items that are dealt with in the material<sup>4</sup>, such as human rights violations and geographic codes for provinces<sup>5</sup> as well as codes to refer to specific places as developed by the Geographic Department in Cambodia for each and every village in Cambodia.<sup>6</sup>

At the time of its launching in January 1997, CBIB contained 2,000 records covering a wide range of material and it now stands at over 3,000. The first category of material to be included was that of the court documents from the People's Revolutionary Tribunal (PRT) of August 1979, the Cambodian government trial of Pol Pot and Ieng Sary. They were presented to the court in Khmer, French and English. A set of these documents was held in the National Archives of Cambodia in a very sorry state and not very well organised. I was given permission to take a set back to Australia and, with a small grant from the Australian Research Council, they were catalogued, and the different language versions (which had been organised in three quite different sequences) were related to each other and linked to the scanned images of the Khmer (and now English) documents.<sup>7</sup>

The documents collected in Phnom Penh by the Documentation Center since the CGP began are turning out to be of great significance and ever growing dimensions. These consist of such items as confessions, photographs, prison note books, and personnel records from Tuol Sleng and other Khmer Rouge prisons throughout Cambodia. We obtained the first major such collection, referred to as 'the million documents', in late 1995 from what had been the Renakse (United Front for the Defence and Reconstruction of Kampuchea). It turns out that rather than a million documents there collection consists of over 10,000 documents bearing the signatures or fingerprints of perhaps a million people. In 1982/83, following the People's Revolutionary Tribunal, the government established a Research Committee to go around the country to every province and in some provinces right down to the village level to gather evidence on what happened from 1975-79. In addition they asked people to support the decision of the Tribunal to condemn the Khmer Rouge, and also to ask the United Nations to seat the the People's Republic of Kampuchea to represent Cambodia and to oust the Khmer Rouge from that position.

These Renakse documents are very vulnerable and had been lying around in boxes since 1983. To our knowledge the existence of these petitions was never brought to

---

<sup>3</sup> UNIMARC was developed in 1977 by IFLA (the International Federation of Library Associations and Institutions) as an international standard format for exchange of bibliographic data, able to cope with multiple languages and scripts and different cataloguing rules, and compliant with ISO2709. <http://www.nlc-bnc.ca/ifla/VI/3/p1996-1/unimarc.htm>

<sup>4</sup> HURIDOCs Standard Formats for the Recording and Exchange of Bibliographic Information Concerning Human Rights. 1993 <http://homepages.iprolink.ch/~huridocs/standard.htm>

<sup>5</sup> BISA (Bibliographic Information Services on Asia and the Pacific) at the School of Information, Library and Archive Studies at the University of New South Wales has been training information workers from the region since 1981. BISA has held three workshops in Cambodia, and at the Third BISA Workshop in January 1995 the list of codes was developed by Ms Mang Thula.

<sup>6</sup> *Gazetteer of Cambodia*. 2nd ed. Phnom Penh : Geographic Department of the Council of Ministers, 1996.

<sup>7</sup> The complete set of PRT documents is to be presented in the edited book on the PRT currently in preparation by John Quigley and Howard DeNike.

the attention of the United Nations, and they have until now never been analysed or summarised. Most of them seem to be general statements or petitions appealing for the United Nations to take action. Some of them go on to state "in our village or our province so many people were killed and so many Buddhist wats were burned down, schools were burned down," giving rather general figures, but some of them go down to specifics, such as "in my family these people were killed on such and such a date", so there is a huge discrepancy in the importance and significance of the documents and value of them to any court of law. In any event, this is a very important collection that needs careful attention and research. The complete set of documents from Siem Reap province has been scanned as an example, showing one province in depth in order to indicate the range of materials in the collection. The documents from other provinces have been categorised as to their district and content, and the documents considered to be more significant, in the sense of providing concrete data, have been scanned and some translated. One of the documents in the collection was a table that gives the figures from each province of the number of reported deaths and the number of petitioners, and this is where this million figure comes from, as it reports that 1,166,307 petitioners had signed all these documents, reporting the deaths of 3,314,000. This table, compiled in July 1982, appears to be the source for the figure used officially by the PRK for the number of deaths caused by the Khmer Rouge when in government. It gives as the source of its figures various telexes and documents from provincial authorities, but these have so far proved elusive. It should be noted then that the Research Committee's province by province data gathering took place after the compilation of the table, and is not the source of the data for the table, as we erroneously suggested in 1996.<sup>8</sup>

Other documents have been provided to the Documentation Center by government bodies like the National Archives of Cambodia, and various Ministries as well as by private individuals. A range of primary material such as personal autobiographies, transcripts of interviews, collections of photographs, tapes etc are being included, in particular the material that Ben Kiernan has collected over the years, including interview transcripts and a diary from Ieng Sary's Ministry of Foreign Affairs, just published in full by CGP on the Internet in both Khmer and English. The Tuol Sleng Genocide Museum in Phnom Penh, the school that was used as the S-21 prison and torture centre, has provided a wealth of material. In the early 1990s Cornell University led an effort to microfilm the confessions held there but, in addition to the material microfilmed, quite a number of other important documents, have been found and are now being included in our bibliographic database. These comprise personnel records and notebooks maintained by the prison staff, and their biographical questionnaires, which provided many items for the biographical database discussed below.

In 1996/7 perhaps the most valuable collection was acquired – over 100,000 pages from the Santebal, or Security Office, the nerve centre of the Khmer Rouge security

---

<sup>8</sup> Helen Jarvis and Nereida Cross, "Documenting the genocide in Democratic Kampuchea : the Cambodian Genocide Program", presented at the *Cambodia : power, myth and memory* conference, held at Monash University, December 1996.

apparatus. Over 10,000 biographies and 11,000 confessions, letters and other documents are now being catalogued, summarised and copied at DC-Cam, and the biographies are being further analysed by Toni Shapiro, a research affiliate at the CGP. These items accounted for some 810 of the records recently added to CBIB.

Documents are also being located outside Cambodia for inclusion in the database. This includes both primary and secondary literature (journal articles, books and films) Added to CBIB in September 1998 are 96 records on articles on the Khmer Rouge published in the Bangkok Post during 1975-79, catalogued by Puangthong Rungswadisab, a research affiliate at the CGP, who is now moving on to add similar Thai language material.

### **CGP Image Database (CIMG)**

A virtual database of scanned images links back into the searchable databases.<sup>9</sup> So that the full text of a document or a photograph may be viewed while displaying a record about a document retrieved from the bibliographic database, and later from the biographic database also,<sup>10</sup> A large number of documents already scanned include the People's Revolutionary Tribunal documents and significant or key documents, such as the Santebal collection, particularly those bearing handwriting and even signatures of officials, showing that these individuals were at least aware of, and in some cases, actually directed the committal of specific crimes.

A specific subset of the image database relates to over 5,000 photographs from Tuol Sleng prison -- from one quarter to one third of the people who were held there, most of whom are believed to have been executed at Choeung Ek on the outskirts of Phnom Penh. Scanned images have been made from photographic prints made from the negative film restored and printed by the Tuol Sleng Museum staff and the Cambodian Photo Archive Group, led by Chris Riley and Doug Niven.<sup>11</sup> The prisoners were photographed, in standard ID photo style, presumably as they arrived at the prison. A small subset of these photographs shows prisoners during or even after torture and death, though we have deliberately avoided mounting these gruesome images on the Internet, and we reserve access to them to serious researchers visiting our offices.

We have developed a physical database pertaining to this subset of CIMG, to form the Cambodian Tuol Sleng Image Database (CTS). Each photograph has a record giving salient details such as gender, age, clothing, and whether a name or number

---

<sup>9</sup> Most scanning has been done by DC-Cam in Phnom Penh using Corel Photo Paint version 5.0 with documents scanned as black and white drawings, and photographs as such. The images are saved as GIF files, selected because of its compact storage algorithm and widespread acceptance for transmission of images over the Internet. We originally intended to utilise OCR to gain access to the textual content of the English language documents, but testing revealed that the documents were of such poor quality that this was not feasible.

<sup>10</sup> The SHOWP (show picture) routine, distributed by Unesco with CDS/ISIS, was adapted on behalf of BISA as SHOWB to meet our needs where one bibliographic record might relate to multiple versions of each document, each in multiple language editions with multiple pages

<sup>11</sup> See *Facing death* (edited by Chris Riley and Doug Niven. Santa Fe : Twin Palms, 1996 and a special issue of *Photographers International*, no.9 (April 1995)

in the province of Svay Rieng alone, one such written document suggest that 94,000 people were killed and that there are over 1,000 mass graves.

Just physically getting the opportunity to visit all of the genocide sites is beyond what we can do, so our priority has been to map the major sites in each province and district. However, some areas have been inaccessible due to security or transport considerations, and in some provinces we have only been able to make a preliminary survey, with sites selected on the basis of their accessibility.

The maps generated from ArcInfo and its PC interrogation and presentation package ArcView show locations of Mass Graves (giving the estimated number of victims per site ranging from 2 to 36,000); DK prisons; Mass Graves and DK prisons; and Memorials. Roads, watercourses and district borders are displayed.

The process of mapping the genocide sites has involved observing fragile evidence and interviewing aging informants in the field. The vulnerability of both physical and personal records makes compelling a research program on the genocide sites themselves. Proposals for physical and social research (involving exhumation and forensic examination and exploration of the sites' place in cultural memory) have been have been outlined but as yet remain unfunded.<sup>15</sup>

#### **Accessing the Cambodian Genocide Data Bases**

All material collected by CGP and DC-Cam is publicly available, whether in original form<sup>16</sup> or in electronic format as presented via the Cambodian Genocide Data Bases on the Internet or in CD-ROM. The bibliographic, biographic and Tuol Sleng photographic databases are all searchable directly over the Internet<sup>17</sup>, while the individual province maps for the geographic database have been generated and printed with help from Yale University's Center for Earth Observation and Institute for Biospheric Studies<sup>18</sup> and are loaded onto CGP's Internet site as static images (pending the imminent installation of the ArcView Internet Map Server at UNSW's School of Geomatic Engineering, which will use CGEO as its pilot dynamically interrogatable database).

We are continuing to add data to all the databases, and the output from our three CDS/ISIS databases is periodically converted to WAIS (Wide Area Information Service) format and then made searchable over the World Wide Web by using CGI scripting and SFGate.<sup>19</sup> We have also produced a CD-ROM version of the

---

<sup>15</sup> Details of such proposals are available from the author of this paper. One aspect of this research -- the Cambodian genocide sites as objects of cultural tourism -- is being commenced by Rachel Hughes, a doctoral research student from the University of Melbourne.

<sup>16</sup> Application for consultation of documents may be directed to Mr Youk Chhang, Director, Documentation Center of Cambodia (DC-Cam), PO Box 1110, Phnom Penh, Cambodia.

<sup>17</sup> CGDB and other findings are displayed on the CGP Internet site available via two Internet addresses: CGP at Yale: <http://www.yale.edu/cgp/> and CGP at UNSW: <http://www-cgp.silas.unsw.edu.au/>

<sup>18</sup> Thanks to Jodi Rowlands, Art Gleeson and Paul Boon.

<sup>19</sup> For more details on the technical aspects of this process, see Helen Jarvis and Robert Loomans, "The Cambodian Genocide Program (CGP)", *Proceedings of the Second New South Wales Symposium on Information*



databases<sup>20</sup>, particularly for those who do not have Internet access. The cost is US\$100, but it is made available at no charge to Cambodian government departments and non-government organisations, as well as to donors to the work of CGP.

### **Future Plans for the CGDB**

The suite of databases developed for CGP has proved to be adequate to the task of coping with the highly varied material included in CGP's documentation component. The presentation of our material on the Internet has attracted attention and commendation<sup>21</sup>. Nevertheless, we are looking towards improvement in a number of areas.

#### *a) Database management*

We intend to explore the possibility of moving the CDS/ISIS databases into a more modern, capable, widely used and better supported environment (such as using Oracle with a search engine such as Infoseek) and to provide direct interrogation over the Internet without our current idiosyncratic and somewhat cumbersome conversion and scripting routine. However, we need earmarked funding to undertake this definitely non-trivial exercise and, as long as the current arrangement remains robust, this is not an urgent matter. It would also be necessary to ensure that all three locations of the CGP had the hardware and software enhancements that would be required as a concomitant of such a development.

#### *b) Further integration*

We are currently investigating the possibility of arranging an integrated search facility across all the CDS/ISIS databases, so that users may enter a single search statement without being required to decide in advance whether they wish to search for bibliographic, biographic or photographic data. In addition, it would be desirable to provide a more direct means of searching image data than the current CDS/ISIS database developed to allow searching of the Tuol Sleng photographs. And we look forward to the day when optical character recognition (OCR) may be carried out on Khmer script material, although it must be noted that much of the material we deal with is in the form of handwritten manuscript or poor typescript that has proved difficult to interpret by computer even when in English.

During 1998 DC-Cam has been granted specific funding from the Sterling Memorial Library at Yale University, Cornell University Libraries and the Southeast Asia Microfilms Project to undertake microfilming of a substantial portion of the documents it has uncovered in the interests of ensuring long term preservation of their content. At present this is a stand-alone operation, but it would seem highly

---

*Technology and Information Systems (SITIS '97)*, 21 February 1997. Sydney : School of Information Systems, UNSW, 1997, [7p.]

<sup>20</sup> The *Cambodian Genocide Data Bases on CD-ROM* is available for purchase from SILAS, UNSW, Sydney 2052, Australia.

<sup>21</sup> The CGP site has received several awards, being named Internet Site of the Day on 29 January 1997 by *Academe Today*, the online journal of the *Chronicle of Higher Education*; and History Site of the Week on 16 March 1997 by *World History Compass*. It continues to inspire viewers (particularly Cambodians) to send letters of appreciation for the work done.

desirable that funding be sought to digitise these images and then to integrate them into the bibliographic database as has been done with the scanned material.

*c) Geo-referencing*

Of course the CGEO database is fully geo-referenced, as it derives from GPS and GIS data. A start has been made on linking bibliographic data to the individual genocide sites, with the site reports as the first items to be so linked. As all the bibliographic records are being assigned their relevant geographic codes, they have the potential to be linked, as do biographic and photographic records given appropriate codes.

One outcome of this geo-referencing approach is that the CGDB are appropriately coded to become a part of the Electronic Cultural Atlas Initiative (ECAI) an international research project aimed at the creation of spatially referenced, GIS-style cultural databases which can be accessed seamlessly across the Internet from a common front-end. <sup>22</sup>

*d) Unicode*

We commenced the CGP documentation in 1995 well before the Khmer character set was defined for Unicode, even in draft form. As a result, we decided to adopt the Khek Brothers' Anlongvill font widely used in Cambodia and preferred by DC-Cam. This means that users must first install the Anlongvill Khmer font, and some also must install Adobe Type Manager (ATM)<sup>23</sup> at their own work station before being able to see the Khmer script parts of our data. We are currently working on making this font available for users to download from our web site.

In the interests of facilitating access to all our data, and of keeping to international standards, we would like to switch to using Unicode's Khmer coding, as soon as it is endorsed, and as soon as we can plan for conversion of data entered to date, and ensure that all parts of the CGP operations are able to obtain the higher-level software and hardware required. Utilisation of Unicode would also facilitate the inclusion of further scripts, such as Thai and Chinese, into CGDB.

*e) Research Reports*

In addition to the provision of access to the databases, the CGP intends to embark upon a publication program to make various aspects of its findings available in a more synthesised and analysed form as research monographs in hard-copy print format and/or on the Internet. Several items have already been published on the Internet<sup>24</sup> and several are in press.<sup>25</sup>

---

<sup>22</sup> )The Electronic Cultural Atlas Initiative (ECAI) is conceived and coordinated by Prof. Lewis R. Lancaster (buddhst@socrates.berkeley.edu), and Richard Henderson (richard@haas.berkeley.edu), University of California, Berkeley, USA.

<http://www.ciolek.com/ECAI/ecai-home.html>

<sup>23</sup> ATM comes bundled with Windows 95 and above, as well as with Windows NT, but Windows 3.1 users must purchase it separately. <http://www.adobe.co.uk/products/prodtbl.html>

The Anlongvill font is prepared by Khek Brothers <http://www.khekbros.com>

<sup>24</sup> See CGP Translations on the site.

Despite the fact that CGP and the Documentation Center of Cambodia have been fortunate to receive a number of grants, including major ongoing funding from the US Department of State, we are still seeking funding to continue and extend the program. Huge numbers of documents are being uncovered in Cambodia, as well, to a lesser extent, in private and government archives and databases around the world. We want to do considerably more imaging and cataloguing to be able to analyse the documents in more detail and to make them more readily accessible. And, as mentioned above, we wish to carry out further research on the genocide sites themselves. In addition, we want to continue the training aspects of the program, developing a core of information specialists and documentalists set up with all the necessary equipment and skills to manage their own national historical documents, in a country of severely impaired education infrastructure.

### **Conclusion**

The CGP is a child of the Internet Age. It began its work at the time of the creation of the World Wide Web. Within the first month of operation -- even before an office was established in Phnom Penh -- the initial bibliographic database was designed and Cambodian staff with computer experience were recruited. The program has continued to expand its use of information technology, with scanners, GPS recorders and microfilm facilities complementing the ever growing number of PCs in DC-Cam, and with Internet servers and a range of high-end GIS hardware and software utilised at UNSW and at Yale.

The emphasis on documentation described in this paper is one of the distinctive features of the CGP. From its inception, the CGP has devoted considerable resources to the systematic recording of all its findings, in a wide range of media, and to harnessing new information technologies in an effort to making our results publicly available in a form that facilitates access and retrieval of crucially important evidence of one of the darkest episode of our time. In this way it may serve as a model for documenting other genocides and systematic human rights abuses as well as a broader range of social phenomena.

---

<sup>25</sup>A volume reporting the mapping of the Cambodian genocide sites is now in preparation, and this paper outlining the policies and practices of the documentation component of the CGP will also form one of the series. The research paper by Sara Colm on The Northeast Zone in DK is to be published by White Lotus in 1999.

