Kontras archive of human rights documents
Notes on a short assessment, 20 January 2012

Background on document

This document was based upon a meeting between Daniel D'Esposito participated in part of the sessions. This needs assessment meeting was agreed upon spontaneously during a visit of D'Esposito and Verstappen to Kontras the previous day. Because of the limited time available, the report is not exhaustive. Only the parts of the archive containing text documents were taken into account.

General characteristics

The Kontras archive has been built up since the establishment of the organisation in 1999. The large majority of the material was provided by survivors and family members of victims. Further material has come from other NGOs as well as from Komnas Ham. There are different types of material: primary material (testimonies of witnesses, police reports, court cases) as well as secondary material (chronologies of events, research reports).

The documents were provided to Kontras on basis of trust and good faith, without detailed provisions regarding the accessibility and the confidentiality of the material.

The archive is used by the staff of Kontras, according to their needs. They can take material out of the archive; there is a system in place for registering loans. Victims also have access to the documents with which they are concerned, and can request photocopies.

About 2/3 of the documents are kept in files on shelves. The files are marked with a variety of codes: region, type of perpetrator (Police, TNI), type of violation. There is no list of the files and their contents.

Preservation

On first sight, a number of documents appear to be in bad state due to moisture. Therefore, a main purpose of the project is to conserve the material in an appropriate way. The books and envelops in which documents are maintained should be opened as soon as possible, independent of whether this project comes off the ground.

Size of the collection and estimated amount of time required

A preliminary estimate indicates that the archive room contains about 54 metres of documents on shelves and about 27 metres of documents in boxes. The estimated size of the collection of paper documents is therefore 81 meters. As the average number of pages per metre is 10,000, this would give a total of 810,000 pages of documents.

It should be verified whether all documents in this room, in particular in the boxes, are to be scanned, and whether there are documents in other rooms which should also be scanned. Therefore, the error margin in the estimate would be around 25%.

Assuming that the collection is around 810,000 documents and that the scanner will handle 20 pages per minute, the total time required only for the physical scanning of the documents would be 675 hours or 84 working days of 8 hours. This figure should be multiplied by three, so as to take into account of the additional time required for taking documents, removing staples or binding, naming, modifying scanner settings, stapling and filing of documents. This gives a total of 252 working days or 12.5 months of 20 working days.

Also the character and size of non-text documents, including audio-types, should be estimated and decisions made with regard to making them accessible.

Estimating roughly that the average size of a document is 20 pages, there would be around 40,500 documents in the collection. The average time for making the description for each document according to the format provided below is estimated at 5 minutes. Therefore, this part of the work would require 3375 hours or 422 working days of 8 hours or 21 months of 20 working days.

Qualifications for staff

A considerable part of the work could be carried out by volunteers or students. However, they would need coordination and supervision by professional staff, both with regard to the physical archiving as well as with regard to the descriptions.

Accessibility

One possible way of providing access to the collection would be through a search engine, such as Solr http://lucene.apache.org/solr/, with sorting by the following facets: region, year, type of document. For an example of how this could work, see the website of the World Organisation against Torture http://www.omct.org. Only a small format would be needed for recording information on each document: time, place, type of document.

The search engine would be hosted on an internal server, or alternatively on an external virtual server.

Selection of scanner

The scanner(s) to be obtained should fulfil the following requirements:
1. Robust to perform heavy-duty scanning during 40+ hours per week
2. Ability to handle Bahasa Indonesia for automatic corrections in PDF files, as the source files are often of modest quality (photocopies, large variety of fonts).
3. Have a feeder tray which would allow to feed in about 50 pages without jamming
4. Allow recto-verso scanning when required

The HP Scanjet Enterprise 7500 Flatbed Scanner appears to be a good choice.

Start of project

The project is to start with a pilot which should aim to:
1. Review and modify the draft format for recording information about each document and the terms proposed
2. Estimate the total time required for scanning the documents
3. Estimate the total time required for annotating the documents.
4. Estimate the total amount of disk space required.

Procedures and good practices

Naming and filing of documents

1. Each document is to be given a systematic name, consisting of region – date – unique number from 1 to x. For larger collections (comprising more than 100 documents) the names are to contain region – date – name of victim – unique number from 1 to x. The names are to be attributed at the time of scanning the document.

2. Develop a numbering system for documents, either based only on numbers or on numbers and regions. If the latter option is chosen, the list of regions used for filing should be the same as the list used for naming of documents (see below)

Before scanning

1. Check whether there are pages with text on recto and verso
2. Check whether there are pages on particularly thin paper – these should be scanned one by one
3. Check whether there are pages in format bigger than A4 – reduce the format (either the scanner has this option or elsewise use a photocopier)

After scanning

1. Ensure that the documents are re-stapled correctly, immediately after scanning
2. Ensure that documents are returned shortly after scanning in the appropriate folder
3. Ensure that folders are returned shortly after scanning to the correct shelf

Format for recording information about documents

1. *Human rights issue*

Arbitrary arrest
Corruption
Death penalty
Destruction and loss of property
Extrajudicial killings
Involuntary disappearances
Religious freedom
Torture

2. *Case name*

List of large-scale human rights violations in Indonesia since the 1960s for which there is a substantial amount of documents present in the Kontras archive.

3. *Date of event*

Exact date in the form DDMMYYYY. If the exact date is not mentioned in the document, browse through it and establish the most appropriate date.

4. *Type of victim*

General list of common occupations in Indonesia plus list of occupations commonly held by vulnerable persons, such as human rights defender, journalist.

5. *Type of perpetrator*

List of perpetrators including TNI (army), police, and vigilantes.

6. *Geographical region*

Term and code from list with three levels 33 provinces, regencies and cities, and districts of Indonesia (this list should be extracted from the one already available in Kontras' application of OpenEvsys).

7. *Type of document*

Including the following:
Audio tape
CD or DVD
Court document
Government report
Letter by government – initial event
Letter by victim – initial event
Letter by government – follow-up
Letter by victim – follow-up
Newspaper article
Police report
Video

The distinction between initial event and follow-up is to allow indications of impunity. In some cases, it can also allow identifying incidents for which there are no documents on the initial violation but only documents on follow-up (usually pointing to non-response of the part of the government).

8. *Date of document*

Exact date on which the document was written, in the form DDMMYYYY

9. *Confidentiality*

Tag "Y" for confidential documents.

Budget

3 scanners x 1500 USD	4500 USD
3 computers x 1200 USD	3600 USD
Scanning by student volunteers 100 USD per month x 12.5 months	1250 USD
Entry of basic information on each document by student volunteers 100 USD per month 21 x months	2100 USD
Creation and hosting of search engine	1000 USD
Equipment for cutting and rebinding of documents	500 USD
Coordination salary eight months x 600 per month	4800 USD
Total	17750 USD

Bert Verstappen, HURIDOCS
info@huridocs.org
23 January 2012