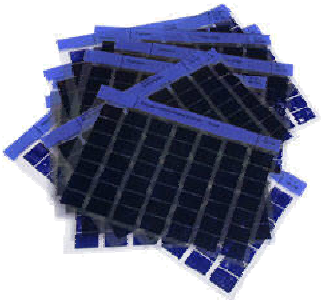


# UNOG Library

## Digitization and Microform Unit (DMU) – December 2009

---



**UNOG**  
THE UNITED NATIONS  
OFFICE AT GENEVA



# Digitization and Microform Unit

## December 2009

---

### Our mandate:

- **Digitization and online dissemination of official documents of the United Nations published before 1993**
- **Targeted preservation of certain UN official documents on microfiche**



# Digitization and Microform Unit

---

- Began digitization in 2005
- Employs six full-time staff, including a supervisor
- Between 2005 and 2009, digitized:
  - 69,563 documents containing 708,192 pages, which represent about 13,913 documents per year, equivalent to 141,600 pages
- It is the only entity in the United Nations system which continues the preservation of official UN documents on microfiche



# Our customers

- The Unit is funded by the Member States of the United Nations through their mandatory contributions
- Worldwide users:
  - Members of Permanent Missions and delegates at the UN and its specialized agencies
  - International organization staff members
  - Staff from the reference libraries of the United Nations system
  - Public users of the UN Official Documents System (ODS) on <http://documents.un.org> and of UN departmental websites



# Our partners

The Digitization Unit at the Dag Hammarskjöld Library at UN Headquarters in New York



# Our goals

---

- “A good document in the right place”
- Documents easily and timely available
- Readable and complete documents
- Good optical character recognition (OCR)
- Digitization and uploading to the ODS of the complete collection of Security Council documents published between 1945 and 1993 in Russian, Chinese and Arabic (general series, agendas and verbatim records) by the end of December 2009



# Digitized documents

---

- Parliamentary documents
- Published by the various organs of the UN (Security Council, General Assembly, Economic and Social Council, Commission on Human Rights, etc.)
- Almost exclusively text, graphics and tables, very few pictures
- In the 6 official languages of the UN: English, French, Spanish, Russian, Chinese, and Arabic



# Digitization equipment available

Combined scanner: flatbed and automatic document feeder (ADF)

- Book scanner
- Microfiche scanner





# Software used

---

- Software for Optical Character Recognition (OCR) working with different languages (English, Spanish, French, Russian and, recently, Chinese)
- Professional software for reading and processing PDF files

# New digitization equipment and software

---

- Acquired a new scanner with an automatic feeder
- Replacing the book scanner with a book scanner with better resolution (300 dpi)
- Acquired server-based OCR software
- Acquisition of an image processing software



# Digitization

- Done systematically or on demand
- At a 300 dpi resolution
- From paper:
  - Loose-leafs using the ADF scanner
  - Bound volumes using the combined scanner flatbed or the book scanner
- From microfiche using the microfiche scanner



# Optical Character Recognition (OCR)

---

- In English, French, Spanish, Russian, and since June 2009, in Chinese, followed by the conversion of TIFF files into PDF text format
- No character recognition in Arabic, conversion of TIFF files into PDF image format
- Obtaining a single format: PDF



# Current workflow

- Each of the five members of the team processes a document or a document series from A to Z, that is to say:
  - Searching for the document or documents in the relevant series in all languages, preferably in second paper copies
  - Preparing lists of documents in a database
  - Preparing the documents physically
  - Scanning all of the documents in the series
  - Processing every document set separately using software to clean, recognize text and convert to PDF
  - Once the entire series is processed, uploading of all documents online to the Official Document System (ODS)



# Digitization process at UNOG Library : 4 phases

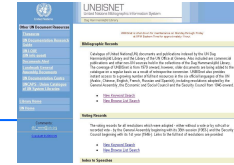


Collect documents...

Input info in e-UNOG database...



Update UNBIS...

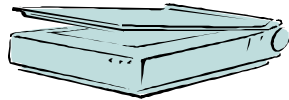
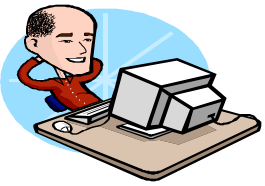


Prepare documents



1. Preparation phase

35%



Paper scanner

or



Microfiche scanner



- TIFF Image e- files
- PDF Image e- files
- PDF Text e- files with OCR

2. Digitization phase

25%



Connect to ODS Mother



Attach UNBIS metadata to e-file



Load to ODS web site

3. Loading phase

25%



Effective loading

OCR random quality

Add to collection

Image quality

Statistics

4. Control phase

15%

# New workflow

---

- Gradual integration of new scanners and software
- Beginning in January 2010, establishing a workflow matrix in which each member of the team will be expert on one of the steps of digitization, while maintaining a variety of tasks
- Quality control for each step of the workflow
- The number and importance of errors are expected to decrease and productivity to increase

