Satellite Meeting
"Conservation and preservation of library material in a cultural-heritage oriented context"
31 August - 1 September 2009
Rome, Italy

Organized by IFLA Core Activity on Preservation and Conservation and IFLA Preservation and Conservation Section

Thanks to the support of:

75th IFLA General Conference, Milan, 23-27 August 2009
Worldwide Leader in Book Digitization

“Moving Knowledge From Books to Bytes”

NICOLA ROSSI
nrossi@nicait.com

NICA srl

IFLA satellite - Rome, Sept 1° 2009
Why Libraries Are Prioritizing Mass Digitization

- **Increase Access**
  - Online with advanced search capability
  - Knowledge is power to the holder
  - Reprints, Large Print, E Books, Braille Books, Audio

- **Preservation**
  - Reduce wear and tear on valuable materials
  - Available for future generations
  - Protect from natural and man-made disasters

- **Reduce cost**
  - De-acidification
  - Rebinding costs
  - Inter-library loans
  - Facility costs for books/materials that are seldom used

- **Increase Revenue**
  - Grow revenue by making these books and materials available for sales through online retail channels
  - Increase ability to attract new donors and donations
Concerns Prioritizing Collections & Considerations

Concerns:
• Too many books / materials to choose from
• Rare books are usually not removed from the library so it is hard to track interest.
• Limited data for what is being searched on-line
• Many are not even catalogued
• Cost of creating metadata for each object

Considerations:
• Select books and materials that are important to the institution regardless of demand
• Books and materials that are at risk of being ruined because of aging (Kirtas)
• Materials that must be saved in the event of a natural or man-made disaster
• Digitize what someone else is willing to pay for
  – Donor
  – Kirtasbooks.com
Large Scale Digitization Success Factors

In large scale digitization, Quality means *Longevity*
- *Page Quality*
- *Volume Quality*

Quality digitization requires:
- Understanding Standards
- Faithful Images
- Complete Metadata
- Accurate OCR

Quality digitization results from Rigorous Discipline & Control
What is the difference between Scanning and Digitization?

3 Pillars of Digitization

It's Not Scanning... It's Digitization!
Capture

- TIFF and JPEG
- Industry Standards and Common

TIFF File
Large for Color 22.5 MB
Lossless Compression
Not Web Friendly

JPEG
Smaller for Color 6.6MB
Lossie Compression
Web Friendly
Large Scale Digitization
Success Factors

Images

Capture - High Quality Images
In Full Color w/Color Profile

Processing - Contrast & brightness
De-speckling, rotate, etc.
Text block Alignment
Grayscale, bi-tonal reduction
Image Segmentation
Background removal

Output - Color, Grayscale, Bitonal
TIFF, JPG, JPG2000, PDF

3 Pillars of Digitization
3 Pillars of Digitization

Standard Formats

Color
5.7MB

Bitonal
1.8MB

Segmented
4.3MB

Once upon a time, a lion and a donkey went to hunt. They agreed to share all the game they might catch.

In the evening, the three hunters met at the edge of the forest.

THE LION'S SHARE

Once upon a time, a lion, a fox and a donkey went to hunt together. They agreed to share all the game they might catch.

In the evening, the three hunters met at the edge of the forest.
Large Scale Digitization

3 Pillars of Digitization

OCR

High Level of Accuracy
- Clear background
- No Skew
- Multiple Languages

Repurposing (where OCR matters)
- Online search
- Flowing Text for e-Books, Mobile devices
- Audio (text to speech), Braille
- Other emerging & future formats
Large Scale Digitization
Success Factors

Why is Metadata Important?

Bibliographic
Administration and Management of content
Search across the digital books by bibliographic parameters.

Technical
Authentication trail of what and how the image was acquired, processed, stored

Structural
Tagging images with descriptors for Title Page, Forward, TOC, Index, etc.
Navigate within the books to section desired

OCR
Text extracted from the page images to search within the book key words/phrases
Today’s output may not be suitable to meet tomorrow’s needs without the “key”
3 Pillars of Digitization

Imaging Workflow
Mass Digitization

Dublin Core
MARC
MARC21

Internet Access to Your Digital Books
and Print on Demand.
Kirtas Solutions

- **Systems**
  - 5 fully integrated Solutions for digitizing Books and Materials

- **Software**
  - Image Processing
  - OCR
  - Metadata and xml output

- **Services**
  - On site Services (Facilities Management)
    - Labor, systems, processing
  - Service bureau
    - Full Suite of Services
  - Shared Services
    - Equipment on site- customer scans and Kirtas/Nica complete the rest of the work remotely
Partial List of Customers

- University of Bologna
- British National Library
- National University of Ireland
- Boris Yeltsin Presidential Library
- Kazakhstan National Library
- King Faisal University
- National Library Republic of Tartarstan
- NACESTI Vietnam
- Cambridge University Press
- Japan Science and Technology
- Hong Kong University Library
- Queensland University
- Caval for New South Wales Universities
KABIS I
Non-Destructive & Fast!

KABIS II

KABIS III

1-shift, 3-year Page Volume

7.5 M

5.7 M

4.5 M

Book Solutions
Digitization — KABIS III
Key Advantages

- **SmartCradle™**
  - 110° V-cradle keeps spines safe
  - Sensors automatically adjust cradle
  - Page edge sensor detects multiple-page and no-page pickups
  - Automatic centering improves processing

- **Gentler than a human hand**
  - Safe for Brittle pages...No bending pages!
  - Safe for loose & torn pages

- Up to 3000 pages per hour
- Two Canon EOS 1Ds Mark III cameras
- 21.1 Megapixel images
- Up to 600 dpi with multiple camera positions and lenses
- High powered server computer for processing
Image Processing

**BookScan Editor ™ (BSE)**
- Auto-cropping, deskewing, despeckling
- Background removal
- Text enhancements
- Batch processing

**BookScan Editor PRO (BSE)**
- OCR searchable text
  - Customized ABBYY® solution
  - 180+ languages
  - Searchable PDF
  - Word, Txt or XML output

**BookScan Editor for Arabic**
- Arabic, Farsi, Pashto, Urdu, Jawi
- Searchable PDF
- Word or XML output

**BookScan Editor METS ALTO**
- METS/ALTO for MARC, DC, Other, ONIX

**JPEG 2000** (lossless format)

**Braille Book Formatting**

**e-book Formatting**

**Print On Demand**
Summary

Key Points
• Superior Image Quality & Segmentation
• Comprehensive Metadata in XML Schemas
• High Speed Productivity
• Safe Handling
• OCR Processing for 190+ Languages
• JPEG2000 Option
• Print-on-Demand, Digitize on Demand
• KirtasBooks.com

Kirtas has over 500 customers in 30 countries

We have digitized over 15,000 books a month in our service bureau

Delighted to be in partnership with the some of the world’s most prestigious Libraries, Universities and Corporations