POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

Preparing digital versions of cultural heritage objects

Adam Dudczak
Poznań Supercomputing and Networking Center
maneo@man.poznan.pl

Digitisation lab. on pendrive

- Based on Ubuntu 11.10 (32 bit)
  - User: digitlab
- Created using tool Relinux
  - https://launchpad.net/relinux
- English is a default language but other languages are installed
  - Albanian, Croatian, Greek, Serbian and Turkish
- There are some packages which are not available in public Ubuntu repositories
- Check out `/home/digitlab/apps`

Digitisation lab. on pendrive (2)

- It contains software useful during digitisation:
  - Audio tools:
    - Audacity
  - Video tools:
    - ffmpeg, OpenShot Video Editor, pitivi, VLC
  - PDF:
    - gsView, hocr2pdf, pdf2djvu, gscan2pdf docsplit, pdftk

Digitisation lab. on pendrive (3)

- It contains software useful during digitisation
  - Other tools:
    - Calibre, Libre Office, Firefox,
  - DjVu:
    - DjView, DjVuLibre, gScan2PDF
  - Graphical tools:
    - Gimp, GraphicsMagick and ImageMagick
    - Google Maps image cutter, Hugin
    - Scan Tailor, magicktailor

Digitisation lab. on pendrive (4)

- OCR software
  - Tesseract v3 with support for multiple languages
- Digital library frameworks
  - Dspace
  - Greenstone
  - Dlibra
    - default license allows to create only 3 publication
- Metadata aggregation frameworks
  - Mint

Preparing objects for publication in the digital library

How to digitize and prepare for the web delivery objects with textual content?

http://dl.psnc.pl/moodle/mod/lesson/view.php?id=109

How to digitize photographs/drawings/posters and prepare them for web delivery?

http://dl.psnc.pl/moodle/mod/lesson/view.php?id=108

Preparing objects for publication in the digital library

What is a Digital Master Copy and why it is so important?

http://dl.psnc.pl/moodle/mod/lesson/view.php?id=107

Plan of the practical part

- User Scan Tailor to post-process scans
- Use gScan2Pdf to run Tesseract OCR
- Create PDF and DjVu files
- Optimize PDF file
- How to use Tesseract from command-line
- Start dSpace
    - Login, create a community and collection
    - Add a right to create collection
    - Add both objects to DSpace library

Plan of the practical part (2)

- Stop DSpace
- Start dLibra
  - Launch digital library website
  - Launch Editor application
  - Start submission of the objects
  - Show how do they look in the browser
  - Show how full text search works
- Show how PDF can be rendered using Document Cloud viewer

Plan of the practical part (3)

- Preparation of large images
  - Create Zoomify tiles using magicktiler
  - Publish object in dLibra based library
  - Show what should be do to have similar effect in Dspace
  - Show how OpenLayers viewer works