Digitisation and digital preservation at National Library of the Czech Republic

Monika Holoubková, Jana Hrzinová, Pavlína Kočišová

National Library of the Czech Republic

Digitisation in the Czech Republic

- National Library of the Czech Republic
- regional science libraries
- National Museum of the Czech Republic
- science institutions
- digitisation by service

What we do

- digitising of bohemical related monographic or periodical documents
- digitising documents in danger of physical destruction
- harvest web pages into webarchive
- we also participate in digitising of gramophone records on shellac discs
- we run web application Kramerius for accesing digitised documents for reading
- we strave for a long term archive of digital object either of our production and production of other digitising institutions in Czech Republic

Standartisation

- for all types of digitised documents
- METS 1.9.1
- MODS 3.6
- DC 1.1
- PREMIS 2.2
- MIX 2.0
- ALTO 3.1
- CopyrighMD 0.91
- AS57 1.0 (for sound recording only)

Stardardization

- aim: homogenous data input
- standardised data formats
- standardised description part based on cataloging rules and possibilities of document types

LTP

- project National Digital Library
- 2010 – 2014
- LTP archive as a part of the project
- LTP as a part of deed of foundation
- every new version of standardization = system adaptation
- Format committee for communication with other institutions

LTP

- 260TB
- inputs:
  - own production
  - data from external producents
  - converted data from old digitization
- IBM archive tapes
- 3 localities – 2x Prague, 1x Brno

LTP

- roles at LTP department
  - content administrator
    - data and content administration
    - supervising and revising data
    - replication of digitized documents for other institutions
    - receiving digitized data from other institutions
    - data conversion within old and new digital library
  - curator of digital library
    - content administration
    - revising and opening public domain documents
    - copyright supervising
    - communication and technical assistance for readers

LTP

- data „migration" between old digital library Kramerius 3 and new Kramerius 5
- large amount
  - Periodical (796/5736010)
  - Monographic units (11412/2997698)

- new archive packages
- using archive JPG (instead of DjVu) - higher quality
- original description metadata + adding new administrative metadata
- Long Term Preservation
- data accessible in new digital library

- limitations
  - different data structure – using various scripts - metadata adaptation
  - capacity of infrastructure (10.000 pages/500 intelectual entities per day) = e.g. 1 volume of larger periodical

Webarchiv

- start at 2000, first copy of website at 2001
- 1 head of Webarchiv, 3 curators, 1 IT guy; 1 IT guy and 1 female data analyst on project
- member of IIPC
- 311 TB of data

- **Comprehensive harvest**
  - CZ.NIC, 1 300 000 czech domain, 1-2x per year, visible from National Library
- **Selective harvest**
  - Conspectus Method, bohemical character - territory, authorship, languague, content
  - contract or Creative Commons license, visible from anywhere
- **Topic collection**
  - important events, elections, floods, anniversary, Olympic games (with IIPC), periodic publications, extinct websites etc.

Webarchiv

- cooperate with instituions
  - Institute of Czech Literature (periodical publications and articles)
  - National Archives (websites of ministry, central administration and institutions of justice)

- cooperate with researchers
  - metadata

- czech copyright, autorship law
  - limitations, long way

webarchiv.cz   facebook.com/webarchivcz   slideshare.net/webarchivCZ

Contact informations

Standards

Pavlína Kočišová: pavlina.kocisova@nkp.cz

LTP Storage

Jana Hrzinová: jana.hrzinova@nkp.cz

Webarchiv

Monika Holoubková: monika.holoubkova@nkp.cz