Slide 1

www.ndk.cz

Czech National Digital Library :
Contextual Selection of Holdings and the Need for (Not Only) Organisational Changes

PhDr. Bohdana Stoklasova
bohdana.stoklasova@nkp.cz

Digital Preservation Summit 2011
19.11.2011

1

Slide 2

Content

- **History: Digitisation, Harvesting and Acquisition of Born-Digital Documents without Digital Preservation**
- **National Digital Library Project**
- **The Need for (Not Only) Organisational Changes**
- **Contextual Selection of Holdings**
- **Conclusions**

2

Slide 3

History:
Digitisation, Harvesting and Acquisition of Born-Digital Documents without Digital Preservation

- **Digitisation:** 1992: card catalogues (Retrocon), 1996: historical manuscripts (Manuscriptorium), 2000: endangered newspapers and monographs (Kramerius)
- **Web harvesting:** 2000 (WebArchiv)
- **E-deposit:** Acquisition of born-digital documents + digital preprints of printed documents: 2011 (Pilot project)
- Limited budget
- HW infrastructure, bit stream preservation, no LTP system

3

Slide 4

NDL Project: Goals

- **Digitisation:** Bohemica, 26 million pages by 2014, 50 million pages by 2019, 300,000 volumes
- **LTP:** 300,000 volumes from NDK, 200 000 volumes from Google, WebArchive, e-deposit
- **Access:** integration

4

Slide 5

NDL Project: Budget and Schedule

- **Budget:** 12 million EUR
- **EU Structural Funds, Integrated Operational Programme (85%), state budget (15%)**
- **Schedule:** 2010-2014 (2019)
- **Present situation:** tender

5

Slide 6

NDL system

6

International Context

- **DPE** – 2006: understanding the LTP in its complexity, building staff, meeting experts from other projects (PLANETS, CASPAR)
- Translation and using **tools** like DRAMBORA, PLATTER etc.
- Other tools: PLATO, JHOVE and DROID – used in the NDL workflows
- **Visits:** NL of th Netherlands, New Zealand NL, NL of Australia, German NL, Welcome Trust Library
- **Analysis and testing** of commercial and free SW
- All experience included in the project

7

The Need for (Not Only) Organisational Changes

- Manuscriptorium and Kramerius – **independent projects**, small independent teams
- WebArchive – small reorganisation needed, library not „disturbed"
- National Digital Library Project – **complex project requiring deep reorganisation**, new people, new strategy, different priorities, not very popular in conservative memory institution
- **Digital preservation – financially demanding area, benefits not visible at first sight**
- Printed books and periodicals – protection from water, fire or entry of unauthorised persons well understood
- **Digital preservation – underestimated**

8

Contextual selection of holdings

Two different contexts:

- **Library context** – preservation of our national cultural heritage (most endangered documents)

- **Smart administration context** – access to factographic information, full texts, access for everyone (most popular documents, authors etc.)

Compromise needed

9

Priorites for digitisation

- Documents of the 19th and 20th century which are most threatened by acidic paper degradation
- The most user-interesting and most requested publications from 1901 to the present
- The most user-interesting historical documents until 1801 (including Czech full-text translation)
- Documents freely accessible to the wider public or through Czech libraries
- Other types of documents according to the criteria of significance, uniqueness, physical state etc.

10

Examples

Authors (cooperation with archives, museums etc.)

- Karel Čapek
- Josef Čapek
- Josef Váchal
- Eduard Bass
- Vladislav Vančura
- Ivan Olbracht
- Karel Poláček
- Jaroslav Havlíček
- Jaroslav Hašek
- others...

11

Examples

Book editions:

- Aventinum
- Odeon
- Laichter
- Melantrich
- others...

12

Slide 13 — Examples

NDK — Národní digitální knihovna

Examples

Periodicals (with printed version)

- Cooperation with publishers and aggregators
- NDK – older periodicals
- Publishers and aggregators – 1990-
- Publishers and aggregators – new periodicals
- Aggregators (Newton Media) – metadata – article level + access for libraries

13

Slide 14 — Metadata and snippet

NDK — Národní digitální knihovna

Metadata and snippet

14

Slide 15 — Full text

NDK — Národní digitální knihovna

Full text

15

Slide 16 — Conclusions

NDK — Národní digitální knihovna

Conclusions

LTP priorities:

Priority 1: documents at risk to be irrecoverably lost

- born digital documents (e-books, e-periodicals, WebArchive)
- d-books, d-periodicals (19 and 20th centuries: most threatened by acid-paper degradation)

Priority 2: LTP saving money for digitisation and/or bringing added value for users

- LTP for publishers (no need to digitise books and periodicals)
- LTP for aggregators (article level metadata + full texts access)

16

Slide 17 — Thank you

NDK — Národní digitální knihovna

www.ndk.cz

Thank you for your attention

PhDr. Bohdana Stoklasová
bohdana.stoklasova@nkp.cz

17