

4 DIGITAL HERITAGE

National Library of the Czech Republic
University Library in Bratislava



Agenda

- ULB projects / activities
- NL CR projects / activities
- ULB + NL CR platforms

Digital Heritage issues

- Acquisition (digitisation, e-publications deposit, web harvesting)
- Standardization
- Long term digital preservation
- Presentation systems (digital libraries)

Digital Resources (ULB)

Digital Resources National Project 2015

- Digital Resources – Webharvesting and E-Born Content Archiving (5.2015 - 12.2015)
- Operational Programme Informatisation of Society (OPIS). EU ERDF
- Complex information system for harvesting, identification, management and long term preservation of web resources and e-Born documents
- Powerful HW and SW infrastructure. Server farm 24x4, archival storage capacity 800 TB.
- Presumptions for the enrichment of the traditional legal deposit.
- The project finished in the fall of 2015.
- Three types: complex, subject based selective website harvest.
- LTP provided via the Central Data Archive

Digital Challenges [www.webdepozit.sk]

- - Digital Object definition
- Digital Object description
- Collection building policy
- Digital Object features
- Digital Object Acquisition
- Digital Object Curation



Central Data Archive (ULB)

CDA National project 2012 - 2014

- Central Data Archive (01.2012,12.2014)
- Operational Programme Informatisation of Society (OPIS). EU ERDF
- Long-time preservation (LTP) of digitized cultural heritage across „the national memory“ institutions: Digital Museum, Digital Gallery, Digital Library and Archive, Digital Audio Vision and Digital Monuments.
- Two active ICT configurations, one passive tape storage in geographically separated localities.
- Follows international standards, ISO 14721 - OAIS, is designed ISO 16363 - LTP.
- HW architecture relies on the powerful midrange servers, 600 TB disc array for information package stream processing and high available tape library with 25 Pbyte capacity using 4 TB cartridges.

Central Archivation - Decentral Digitisation [cda.kulturny.sk]

- Preservation Planning
 - Data formats
 - Data Transfer
 - Data
 - LTP Processes



The image shows a project logo for 'Centrálny dátový archív' (Central Data Archive) with the tagline 'TVORÍME VEDOMOSTNÚ SPOLOČNOSŤ' (We create a knowledge society). The logo includes the OPIS logo and logos for the National OPIS Authority, the Operational Programme Informatisation of Society, and the European Union. Below the logo, it states 'Druh a názov projektu: Centrálny dátový archív', 'Dátum začatia realizácie projektu: január 2011', and 'Dátum skončenia realizácie projektu: december 2014'. At the bottom, it mentions 'Tento projekt je spolufinancovaný Európskou úniou z prostriedkov Európskeho fondu regionálneho rozvoja' (This project is co-financed by the European Union from the resources of the European Regional Development Fund) and the logo for 'univerzitná knižnica v bratislave' (University Library in Bratislava).

LTP Certification (ULB)

LTP Digital Heritage Archive Certification

From start of the CDA project it was necessary to establish a regulatory framework for trusted long-term data storage. Standards ISO 14721 and ISO 16363 provided us with imaginary leading rails. CDA is designed and operates in accordance with the requirements and standards for trusted LTP storage. The essential part and one of the measurable indicators was system certification based on ISO 27001. The process of preparation for certification lasted full calendar year and included the introduction of ISMS in the CDA and the cooperating departments within ULB: accounting, personal, methodical, operations, administration. Preparation consisted of the introduction, of the registration of all processes, issuing directives, forms, training of employees, establishing principles and processes into practice.

LTP = Longtime Responsibility

- - Awareness
- Regulatory framework set-up
- International Standards
- Organisational infrastructure
- Digital Object Management
- Infrastructure
- Security Risk Management
- Self Audit
- Certification Cycles
- Population
- Comparison and Know-How Exchange

Methodological challenges (ULB)

Central Data Archive [cda.kultury.sk]

A **harmonization framework** for all OPIS projects was established. The involved memory institutions followed **current legislation** in the field of copyright, respected existing standards of state and public administration systems, OASIS and LTP and established appropriate policies. Preparation of **methodological manuals** for digitisation, processing and long-term storage of digital content across the national OPIS projects proved to be a strategic act. The CDA as a LTP subject compiled and published methodological materials for cooperating institutions.

Digital Resources [www.webdepozit.sk]

Methodological problems reach from **metadata** up to the **legislation**. We still do not have regulations for **legal deposit** of on-line publications as well as for harvesting, archiving and providing access to archived content. E-content can not be stored and provided to the public without an approval of the providers. Exception represent **public licensed documents** (OA, CC). **Management** of e-Born objects has required creating corresponding MARC metadata structures for websites, e-Born serials and e-Monographs. **Selection** of e-resources for archiving, due to their origin, require an appropriate policy. Web sites are chosen by subject or by events, e-Born serials are selected in cooperation with the ISSN Centre Slovakia.

E-deposit (NL CR)

- The project for establishing system for acquisition of e-publications from publishers (2012-2015)
- Based on voluntary deposit
- The main deliverables:
 - E-publications legal deposit bill
 - Workflow for collecting e-publications
 - Metadata standards for e-publications (for LTP and cataloguing)
- Analysis of the file formats used by Czech e-publishers
- Recommended file formats for LTP



Národní knihovna
České republiky

[Partneři](#)

[Kontakty](#)

[Aplikace](#)

[Akce](#)

[Články a dokumenty](#)

[Pro nakladatele](#)

[Bezpečnost elektronických publikací](#)

[Plán sběru elektronických publikací
na 2013-15](#)

[Ukládání elektronických publikací](#)

[Ukládání publikací na fyzických
nosičích](#)

[Ukládání webových stránek](#)

[Uživatelský přístup](#)

e-deposit

Nacházíte se zde: Úvod

Vítejte na stránkách e-deposit NK ČR

Systém e-deposit je vytvářen v rámci projektu NAKI Správa elektronických publikací v síti knihoven České republiky.

Cílem projektu je:

1. Řešení sběru a zpracování knih a periodik v elektronické podobě a to včetně elektronických předloh tištěných publikací, monografií i periodik.
2. Vytvoření nástrojů pro sdílení a zpřístupnění tohoto typu obsahu v knihovní síti České republiky.
3. Dlouhodobá archivace a ochrana tohoto typu dokumentů v systému budované v projektu Národní digitální knihovna.
4. Návrh metodiky bibliografického popisu tohoto typu dokumentů.
5. Analýza legislativy v oblasti elektronického povinného výtisku v zahraničí, návrh právní úpravy včetně změn zákona ČR v této oblasti.

Řešení projektu je rozděleno do 3 etap. V první etapě realizované od 1.3.2012 do 31.12.2012 proběhla analýza současného stavu. V druhé etapě od 1.1.2013 do 31.12.2014 byl zahájen vývoj systému pro sběr, zpracování a sdílení elektronických publikací, pokračoval výzkum v oblasti bibliografického popisu elektronických publikací a byly navrženy procesy pro přípravu elektronických publikací k archivaci. Ve třetí etapě od 1.1.2015 do 31.12.2015 je spuštěn poloprvok infrastruktury pro sběr, zpracování a sdílení elektronických publikací a je realizován návrh konceptu centrálního bibliografického zpracování.

Projekt navazuje na pilotní projekt Dobrovolné uložení elektronických publikací a je financován z grantového programu NAKI. Systém e-deposit je komponentou Národní digitální knihovny ČR.

Webarchiv (NL CR)

- Webarchive of NL CR has been harvesting Czech web content since 2001.
- Collecting, archiving and providing long-term access to Czech web resources
- The types of harvesting:
 - comprehensive harvesting (the whole national domain) – 1-2 a year
 - selective archiving
 - thematic „event-based“ collections
- All harvested content is available for on-site users only (i.e. from PCs in the library building)
- Access to the selective harvests to all internet users (due to our agreements with publishers).
- Today volume: 220 TB of data, 4 billions objects

The screenshot shows the homepage of webarchiv.cz. At the top, there are navigation links: úvod, o Webarchivu, katalog stránek, and přidat web. In the top right corner is the logo of the National Library of the Czech Republic (NL CR) with the text 'CZ EN'. The main heading is 'Webarchiv' in blue, with the subtitle 'památník českého internetu, více'. Below this is a search bar with the placeholder text 'hledjte „webarchiv.cz“ nebo „webarchiv“, pokročilé'. The page is divided into three columns. The left column features a blue header 'Světový den knihy a autorského práva' and an image of an open book. Below the image is a paragraph of text: 'Psané slovo má nesmírnou moc. Dokáže vykouzlit čarokrásné světy, rozvinout fantazii, sdílet vědecká poučení. Má obrovskou schopnost spojovat nebo rozdělovat společnost. Je proto přitvozené, že si připomínáme 23. duben jako Světový den knihy a Autorský práv. Z našeho archivu jsme vybrali zdroje, které toto téma'. The middle column has a blue header 'Památník národního písemnictví : PNP' and a paragraph: 'Stránky přináší informace o památníku, jeho činnosti, sbírkách, výzkumu, publikacích, pořádaných výstavách a pořadech. Součástí jsou také informace o soutěži Nejkrásnější české knihy roku. Vydavatel: Památník národního písemnictví dějiny české literatury, písemnictví, literární muzea, knižní kultura, Památník národního písemnictví'. The right column has a blue header 'Webarchiv k 6. 5. 2016 obsahuje' and a large number '221 TB'. Below this is the text 'dat. První dokument byl archivován 3. 9. 2001.' and 'Celkem jsme s autory uzavřeli' followed by a large number '4872'. At the bottom of the right column is the text 'smluv. Poslední aktuální smlouvy: Historie c. k. armády v napoleonské epoše, FFUK PRAHA | Ústav pro archeologii, Ústav hospodářských a'.

webarchiv.cz

Metadata (NL CR)

- The national metadata standard for digitisation (still images) in libraries since 2012
- Used for National Digital Library project as well as for digitisation projects carried out by many other libraries
- Developed and managed by NL CR
- Based on international standards:
 - METS, MODS, DC, PREMIS, MIX
- Standardization of production facilitates subsequent LTP in digital repository
- The standard history and changes carefully recorded
- All versions available at www.ndk.cz



The screenshot shows the website for the National Digital Library (Národní digitální knihovně). The navigation menu at the top includes: O projektu, Standardy digitalizace, Digitalizace, Archivace, Zpřístupnění, Formátový úbor NDK, and Otázky. The main content area features a large image of an open book on a scanner with the text "Vítejte v Národní digitální knihovně" and "užívající systém Kramerius". Below the image, there is a section titled "Standardy digitalizace" with a list of standards: Standardy digitalizace, Standardy pro obrazová data, and Standardy pro metadata. The date "26.09.2014" is visible at the bottom left of the page content.

ndk.cz

LTP Certification (NL CR)

- LTP Repository requirements are specified in ISO 14721 and ISO 16363 standards
- Certification of our LTP Repository (assessing as trustworthy) is an important goal of NL CR
- Formal certification is a lengthy process involving both SW/HW and organizational processes (Organizational infrastructure, Digital object management, Infrastructure and security risk management).
- NL CR has to perform several steps
- The prerequisite step is ISO/IEC 27001 certification (Information security management)
- Self-audit and formal audit according to ISO 16363

INTERNATIONAL
STANDARD

ISO
16363

First edition
2012-02-15

**Space data and information transfer
systems — Audit and certification of
trustworthy digital repositories**

Systèmes de transfert des informations et données spatiales — Audit et certification des référentiels numériques de confiance

Cooperation

For enabling the synergy for Digital Heritage UNESCO recommends:

“to intensify national and international cooperation and exchanges, in particular through the pooling of human and material resources to assist research and the protection and preservation of documentary heritage, Member States should support the exchange of research data, publications, and information; the training and exchange of specialist personnel and equipment. They should promote the organization of meetings, study courses and working groups on particular subjects, such as cataloguing, risk management, identification of endangered documentary heritage and modern research”.

Know-How Platform

Main goals

- Platform for sharing knowledge and practical experiences and for collaboration of experts
- Sharing standards, practices and research news
- Cooperative research
- Eg. – ULB / NL CR certification experiences
- Regular meetings of experts

Current issues

- *Social media DH videos archiving*
 - technical difficulties
 - what has the lasting value?
- *Collaborative harvesting of .eu domain*
 - European domain crawling is often omitted by national web archives
- *Documenting repository processes*
 - best practices for writing documentation which is requisite for certification

Synergy Platform

Main goals

- Providing better public access for DH documents
 - Current projects concentrate mostly on collecting and archiving
- Sharing resources
- Encouraging education / trainings in DH topics

Current issues

- *Harmonization of the legislation*
 - towards the exemptions in copyright act for libraries
 - legal deposit for all types of documents
- *Software sharing*
 - eg. Czech resolver software for other countries?
- *Raising a new generation of specialists*
 - internships at V4 memory institutions

Methodology Platform

Main goals

- Platform for the co-operative development of methodologies for all stages of the life cycle of the DH documents (production, collecting, archiving, access)

Current issues

- *Methodology for bibliographical description of crawled web resources*
 - new challenges
 - eg. the definition of intellectual entity for bibliographic record
- *Methodology for publishers*
 - how to create digital documents suitable for digital preservation
 - recommendations of file formats, migration tools and validators
- *Methodology for AIP life-cycle*
 - metadata profile, migration etc.

Usage Platform

Main goals

- Improving access services for users
- Incorporating new digital trends to DH content
 - digital humanities
 - big data

Current issues

- *Development of user friendly interfaces for DH portals*
 - applying state-of-the-art usability standards and optimization methods
 - eg. eye-tracking, usage and behaviour analysis
- *Implementation of data processing, extraction, evaluation and visualization tools, methods and environments*
 - semantical enrichment, full-text processing, data labs

The first steps

- Kick-off meeting
- The formal establishment of the platforms
- Action plan for 2016

Thank you for your attention



Zuzana Kvašová, Jaroslav Kvasnica, Zdeněk Vašek, Iveta Lodrová,
Ladislav Cubr



univerzitná knižnica
v bratislave

Alojz Androvič, Beáta Katrincová, Jana Matúšková, Bibiana Žigová,
Andrea Hrčková (Comenius University in Bratislava)