

Digitální data vydrží navždy
– nebo pět let: podle toho, co
přijde dřív.

Jeff Rothenberg

Zdeněk Hruška

Moravská zemská knihovna v Brně

hruska@mzk.cz



I.

Byl pozdní večer — první máj —
večerní máj — byl lásky čas.
Hrdliččin zval ku lásce hlas,
kde borový zaváněl háj.
O lásce šeptal tichý mech;
kvetoucí strom lhal lásky žel,
svou lásku slavík růži pěl,
růžinu jevil vonný vzdech.
Jezero hladké v křovích stinných
zvučelo temně tajný bol,
břeh je objímal kol a kol;
a slunce jasná světů jiných
bloudila blankytnými pásky,
planoucí tam co slzy lásky.
I světy jich v oblohu skvoucí
co ve chrám věčné lásky vzešly,
až se — milostí k sobě vroucí
změnivše se v jiskry hasnoucí —
bloudíci co milenci sešly.

Významné vlastnosti (significant properties) I

I. Byl pozdní večer — první máj — večerní máj — byl lásky čas. Hrdliččin zval ku lásce hlas, kde borový zaváněl háj. O lásce šeptal tichý mech; kvetoucí strom lhal lásky žel, svou lásku slavík růži pěl, růžinu jevil vonný vzdech. Jezero hladké v křovích stinných zvučelo temně tajný bol, břeh je objímal kol a kol ; a slunce jasná světů jiných bloudila blankytnými pásky, planoucí tam co slzy lásky. I světy jich v oblohu skvoucí co ve chrám věčné lásky vzešly, až se — milostí k sobě vroucí změnivše se v jiskry hasnoucí bloudící co milenci sešly.

Zdroj: K. H. Mácha. Máj. (1896) Dostupné z:

<http://www.digitalniknihovna.cz/mzk/view/uuid:f5a09c95-2fd8-11e0-83a8-0050569d679d>

Významné vlastnosti (significant properties) II



LTP v paměťových a výzkumných institucích



- ❑ Digitalizace a digital born dokumenty
 - ❑ ochrana kulturního dědictví;
 - ❑ data bez analogové verze (vědecká data, hry, ...);
 - ❑ e-maily - historie projektu, historie oboru (dopisy);
 - ❑ ochrana investice = digitalizace (práce, peníze);

LTP je důležité i jinde!



zdroj:
<https://youtu.be/pbBa6Oam7-w>

Formáty vhodné pro dlouhodobé uložení



Formát by měl být:

- ❑ dobře popsáný s veřejně dostupnou dokumentací;
- ❑ multiplatformní, nezávislý (patentově, technologicky);
 - ❑ (x proprietární);
- ❑ široce rozšířený;
- ❑ robustní;
- ❑ perspektivní.

registr formátů PRONOM nebo UDFR

Formáty vhodné pro dlouhodobé uložení

- ❑ obrazová data: TIFF, JPEG2000, PNG, JPG
- ❑ textová data: EPUB, PDF/A, PDF, XML-based, RTF, TXT
- ❑ zvuková data: WAVE, FLAC
- ❑ dataseť/databáze: samopopisné - JSON, XML-based; CSV; otevřené formáty - .sqlite, .db, .db3
- ❑ web: WebARChive (WARC)

Důležitost metadat a jejich druhy

Popisná – popisné údaje o intelektuální entitě - údaje o původu (autor, název); vyhledání a zpřístupnění digitálního objektu. Případně i údaje o originálním (tištěném) objektu či identifikátory (signatura, čár. kód, čČNB, URN:NBN).

- MARCXML, Dublin Core, MODS, nebo EAD (Encoded Archival Description).

Strukturální – zachycují vztahy mezi dílčími digitálními objekty a jak dohromady tvoří jednu intelektuální entitu. Např.:

- kde na webové stránce se nachází obrázek (fyzická struktura)
- jak jdou stránky a kapitoly v knize za sebou (logická struktura)

Důležitost metadat a jejich druhy

Technická – údaje o počtu souborů, velikosti, formátech souborů; dalších vlastnostech (rozlišení, délka audio souboru), upřesňují údaje o hardwaru a softwaru, na nichž mohou být digitální objekty spuštěny.

- textMD (nebo audioMD, videoMD), MIX (Metadata for Images in XML Schema) či ANSI/NISO Z39.87-2006.

METS - spojuje všechny typy MD do jednoho balíku.

Administrativní – info o vzniku entity, o odpovědné osobě za její správu, informace o ochranných činnostech, o právech k přístupu. Též ochranná metadata (PREMIS).

Metadata a jejich druhy - METS

```
<?xml version="1.0" encoding="UTF-8"?><mets:mets xmlns:mets="http://www.loc.gov/METS/"
xmlns:copyrightMD="http://www.cdlib.org/inside/diglib/copyrightMD" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:mods="http://www.loc.gov/mods/v3" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:premis="info:lc/xmlns/premis-v2" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" LABEL="Panenská věž, 1926" TYPE="Monograph"
....
....
  <mets:metsHdr CREATEDATE="2018-09-19T09:50:04" LASTMODDATE="2018-09-19T09:50:04">
    <mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
      <mets:name>BOA001</mets:name>
    </mets:agent>
    <mets:agent ROLE="ARCHIVIST" TYPE="ORGANIZATION">
      <mets:name>BOA001</mets:name>
    </mets:agent>
  </mets:metsHdr>
```

Metadata a jejich druhy - DC

```
<mets:mdWrap MDTYPE="DC" MIMETYPE="text/xml">
  <mets:xmlData>
    <oai_dc:dc>
      <dc:title>Panenská věž : Román na moři a na zemi. [Díl 1-4]</dc:title>
      <dc:creator>Flygare-Carlén, Emilie</dc:creator>
      <dc:publisher>Šolc a Šimáček</dc:publisher>
      <dc:date>1926</dc:date>
      <dc:language>cze</dc:language>
      <dc:language>und</dc:language>
      <dc:format>print</dc:format>
      <dc:description>Sv. 1. [díl 1-2] 536 s. -- sv. 2. [díl 3-4] [3] s., s. 540-1071</dc:description>
      <dc:description>od Emilie Flygaré Carlénové ; podle čtvrtého vydání přeložil Hugo
Kosterka</dc:description>
      <dc:type>model:Monograph</dc:type>
      <dc:identifier>ccnb:cnb000930620</dc:identifier>
      <dc:identifier>barCode:2619286474</dc:identifier>
      <dc:identifier>uuid:9e208580-bbe0-11e8-bc37-005056827e51</dc:identifier>
```

Metadata a jejich druhy - MODS



```
<mets:dmdSec ID="MODSMD_VOLUME_0001">
  <mets:mdWrap MDTYPE="MODS" MDTYPEVERSION="3.5" MIMETYPE="text/xml">
    <mets:xmlData>
      <mods:mods ID="MODS_VOLUME_0001">
        <mods:titleInfo>
          <mods:title>Panenská věž</mods:title>
          <mods:subTitle>Román na moři a na zemi</mods:subTitle>
          <mods:partNumber>2</mods:partNumber>
        </mods:titleInfo>
```

Metadata a jejich druhy - MIX

```
<?xml version="1.0" encoding="UTF-8"?>
  <reportingModule release="1.5" date="2007-10-02">TIFF-hul</reportingModule>
  <lastModified>2018-09-27T09:03:46+02:00</lastModified>
  <size>8096958</size>
  <format>TIFF</format>
  <version>6.0</version>
  <status>Well-Formed and valid</status>
  <sigMatch>
  <module>TIFF-hul</module>
  </sigMatch>
  <mimeType>image/tiff</mimeType>
  <profiles>
    <profile>Baseline RGB (Class R)</profile>
    <profile>DLF Benchmark for Faithful Digital Reproductions of Monographs and Serials: color</profile>
  </profiles>
```

Metadata a jejich druhy - PREMIS



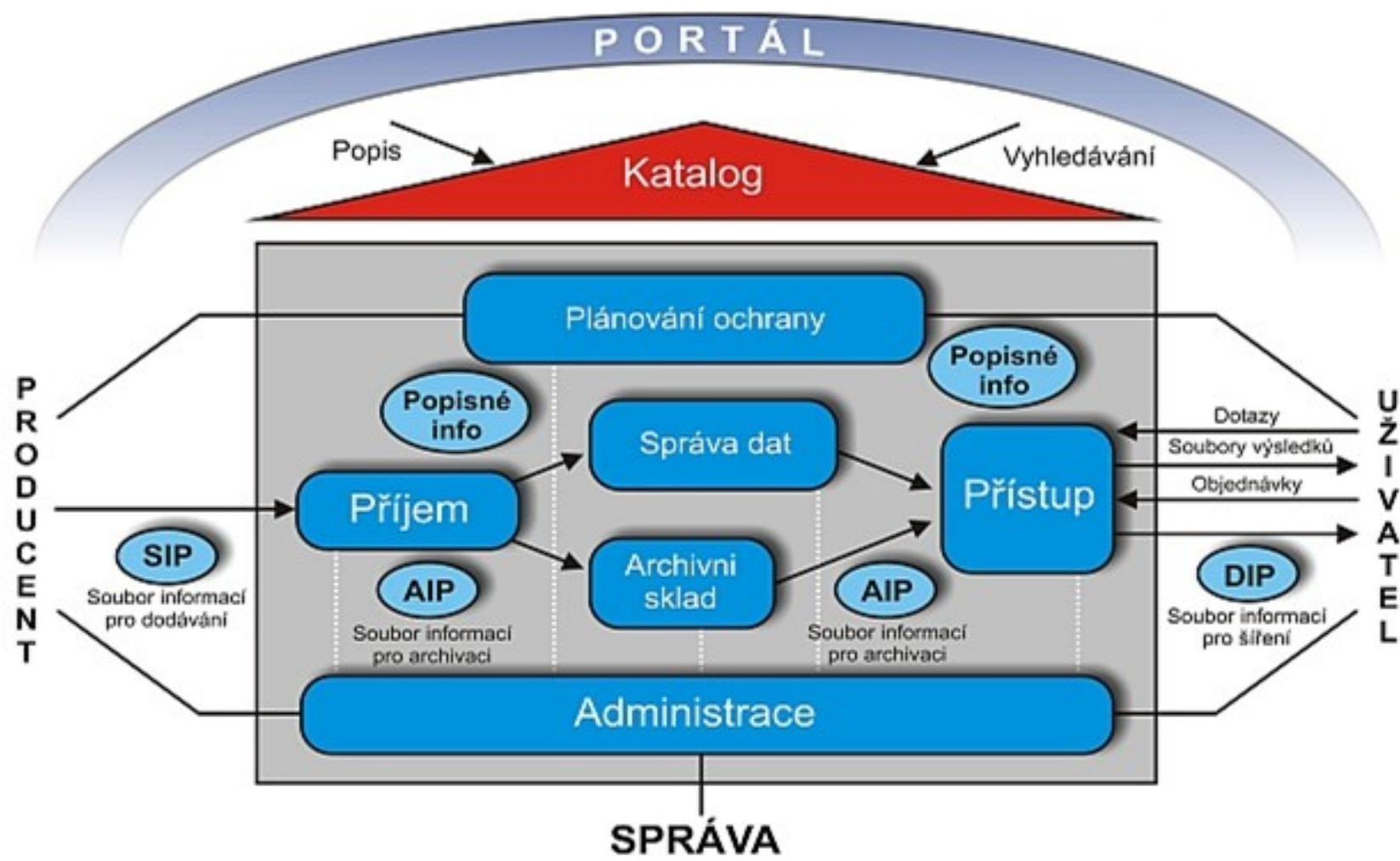
```
<mdWrap MDTYPE="PREMIS:EVENT"><xmlData><event xsi:schemaLocation="info:lc/xmlns/premis-v2
http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd" version="2.2">
<eventIdentifier>
<eventIdentifierType>UUID</eventIdentifierType>
<eventIdentifierValue>2cfc30f8-458d-4e14-9a42-6dc5e1c7e964</eventIdentifierValue>
</eventIdentifier>
<eventType>viruscheck</eventType>
<eventDateTime>2014-01-15T19:34:30</eventDateTime>
<eventDetail>program="Clam AV"; version="ClamAV 0.97.8"; virusDefinitions="18354/Wed Jan 15
04:56:57 2014
"</eventDetail>
<eventOutcomeInformation>
<eventOutcome>Pass</eventOutcome>
```

Standardy NDK



- ❑ standardy digitalizace - formáty + metadata
- ❑ monografie, periodika
- ❑ zvukové dokumenty - LP desky, (CD), (fonoválečky)
- ❑ ? kroniky, datová CD a DVD, ...

Referenční model OAIS (Open Archival Information System) - ISO 14721



Datové balíčky v LTP repozitáři



- ❑ SIP = Submission Information Package
- ❑ AIP = Archival Information Package
- ❑ DIP = Dissemination Information Package

Digitální repozitář, jeho budování a provoz



- ❑ Digitální repozitář = lidé + technologie + postupy
 - ❑ zajištění autenticity a integrity dat
 - ❑ mandát
 - ❑ dlouhodobé financování
 - ❑ politiky a strategie
 - ❑ přijímaná data, exit strategy
 - ❑ cílové skupiny

Digitální repozitář, jeho budování a provoz



- ❑ Důvěryhodný digitální repozitář
- ❑ Hodnocení vůči souboru požadavků
 - ❑ interní audit
 - ❑ externí audit
- ❑ Nástroje na hodnocení:
 - ❑ CoreTrustSeal
 - ❑ ISO 16 363
 - ❑ Nestor

Důvěryhodný digitální repozitář



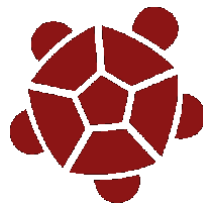
Není Seal of
Approval
jako
Seal of
Approval



Ochranné strategie LTP

- ❑ refreshing (obnovení),
- ❑ technologické muzeum,
- ❑ migrace nosiče,
- ❑ migrace formátu, normalizace,
- ❑ emulace

Lots of Copies Keep Stuff Safe (LOCKSS)



- ❑ vyvinuto na Stanford University Libraries
- ❑ open-source technologie (<https://github.com/lockss>)
- ❑ uchování obsahu od vydavatelů, na který mají knihovny nárok
- ❑ P2P infrastruktura - desítky (stovky) lokálních úložišť
 - ❑ ale není to sharing!
- ❑ kontrola pomocí metadat (kontrolní součty)
- ❑ pokud není obsah dostupný u vydavatele, stáhne se z LOCKSS
- ❑ robustní, komunitní, lokální kontrola



Zdroj: <https://geoffreymultimedia.com/buying-a-new-computer/>

Emulace



- ❑ Emulace jako služba - EaaS (Emulation as a Service)
- ❑ EaaSI - Yale University Library
- ❑ infrastruktura pro emulační prostředí
- ❑ problém s autorskými právy u SW (Apple)

<https://www.softwarepreservationnetwork.org/eaasi/>

Uložení + přístup k datům a metadatům

- ❑ Podle druhu dat a repozitáře
 - ❑ “živá” data (knihovny, galerie, muzea)
 - ❑ “archivní” data (archivy, vědecká data?)
- ❑ média - optická, magnetická
- ❑ rychlost zápisu/čtení
- ❑ cena/výkon
- ❑ spolehlivost/bezpečnost

Technologická bezpečnost uložených dat

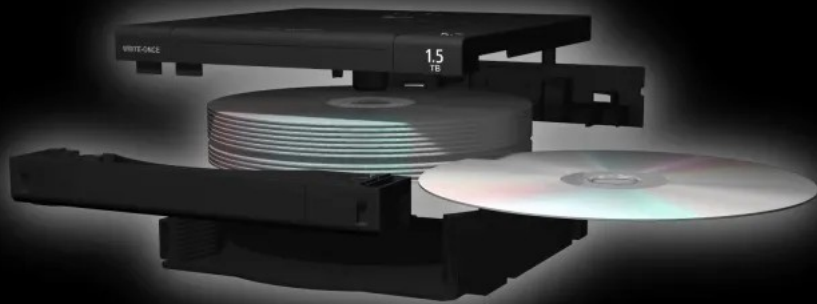
- ❑ LTO (magnetické pásky) - cena/výkon
- ❑ optické disky - odolnost
- ❑ pevné disky - rychlost čtení



zdroj: <http://edpllc-usa.com/lto-8-release-support-engineered-data-products/>



Optical Disc Archive

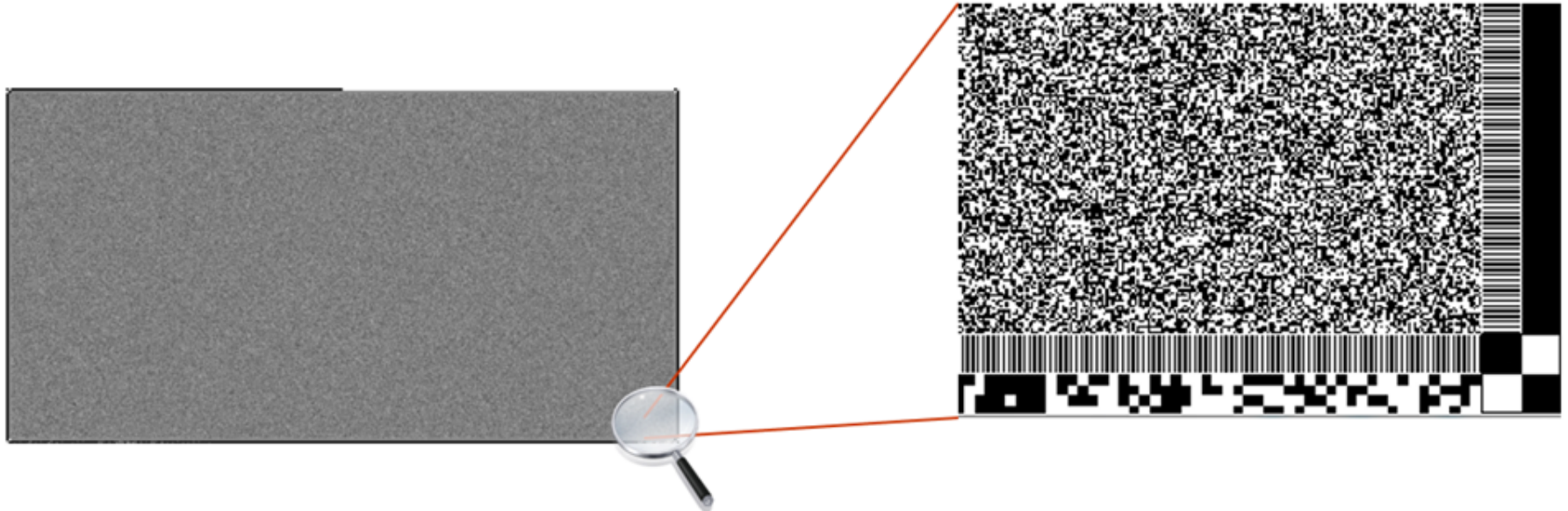


zdroj: <https://5thingsseries.com/archive/>

zdroj: <http://www.techwaredist.com/wordpress/index.php/new-optical-disc-archival-format-unveiled-by-sony/>

Pasivní digitální archivace

- ❑ Piql - uložení dat v podobě “QR” kódu na černobílý filmový pás
- ❑ životnost 100 let, snadná čitelnost (světlo, digi foťák a PC)





Piql - World Arctic Archive



zdroj:
<https://youtu.be/hz170R51w7k>

Stav LTP v ČR a v zahraničí

- ❑ LTP oddělení a LTP repozitář NK ČR (NDK)
- ❑ Webarchiv (NK ČR)
- ❑ Národní digitální archiv (NDA)
- ❑ projekt ARCLib -> pilotní provoz KNAV (2020)
- ❑ repozitáře vědeckých dat UK (CTS pečet'):
 - ❑ Repository of Charles University in Prague, Map Collection
 - ❑ LINDAT-Clarín - Centre for Language Research Infrastructure in the Czech Republic
 - ❑ Český sociálněvědný datový archiv

Stav LTP v ČR a v zahraničí



- ❑ Centrálny dátový archív (CDA)
- ❑ Archives New Zealand
- ❑ National Archives (UK)
- ❑ Library of Congress
- ❑ Die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB)
- ❑ Museum of Modern Art - MOMA (NY, USA)

Použité zdroje + pro zájemce



- ❑ <http://www.digitalpreservation.cz/>
- ❑ <http://ltp-portal.mzk.cz/>
- ❑ <http://ltp.knihovna.cz/>
- ❑ CUBR, Ladislav. Dlouhodobá ochrana digitálních dokumentů. Praha: Národní knihovna České republiky, 2010. ISBN 978-80-7050-588-5.
- ❑ PLATTER: <https://www.ndk.cz/platter-cz/Platter.pdf>
- ❑ ČSN ISO 16363: Audit a certifikace důvěryhodných digitálních úložišť
- ❑ OAIS (též ČSN ISO 14721: Otevřený archivační informační systém - Referenční model)

- 
- ❑ <https://www.ndk.cz/>
 - ❑ <https://www.ndk.cz/archivace/resolver-urn-nbn-sluzba-cidlo>
 - ❑ <https://www.nacr.cz/digitalni-archiv/>
 - ❑ <https://dsa.cuni.cz/>
 - ❑ Kapitola LTP v rámci kurzu Digitalizace v knihovnách na:
<http://kurzy.knihovna.cz/>
 - ❑ <http://cda.kultury.sk/>



- ❑ <http://www.loc.gov/preservation/digital/>
- ❑ <https://www.loc.gov/librarians/standards>
- ❑ <https://www.loc.gov/preservation/resources/rfs/>
- ❑ <http://www.nationalarchives.gov.uk/>
- ❑ <http://preservationmatters.blogspot.com/>
- ❑ <https://www.lib.umich.edu/blogs/bits-and-pieces>
- ❑ <http://digital-archiving.blogspot.com/>
- ❑ <http://www.dcc.ac.uk/news>



Děkuji za pozornost!

Zdeněk Hruška

Moravská zemská knihovna v Brně

hruska@mzk.cz