

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

**Jeff Rothenberg**

Zdeněk Hruška  
zd.hruska@gmail.com



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



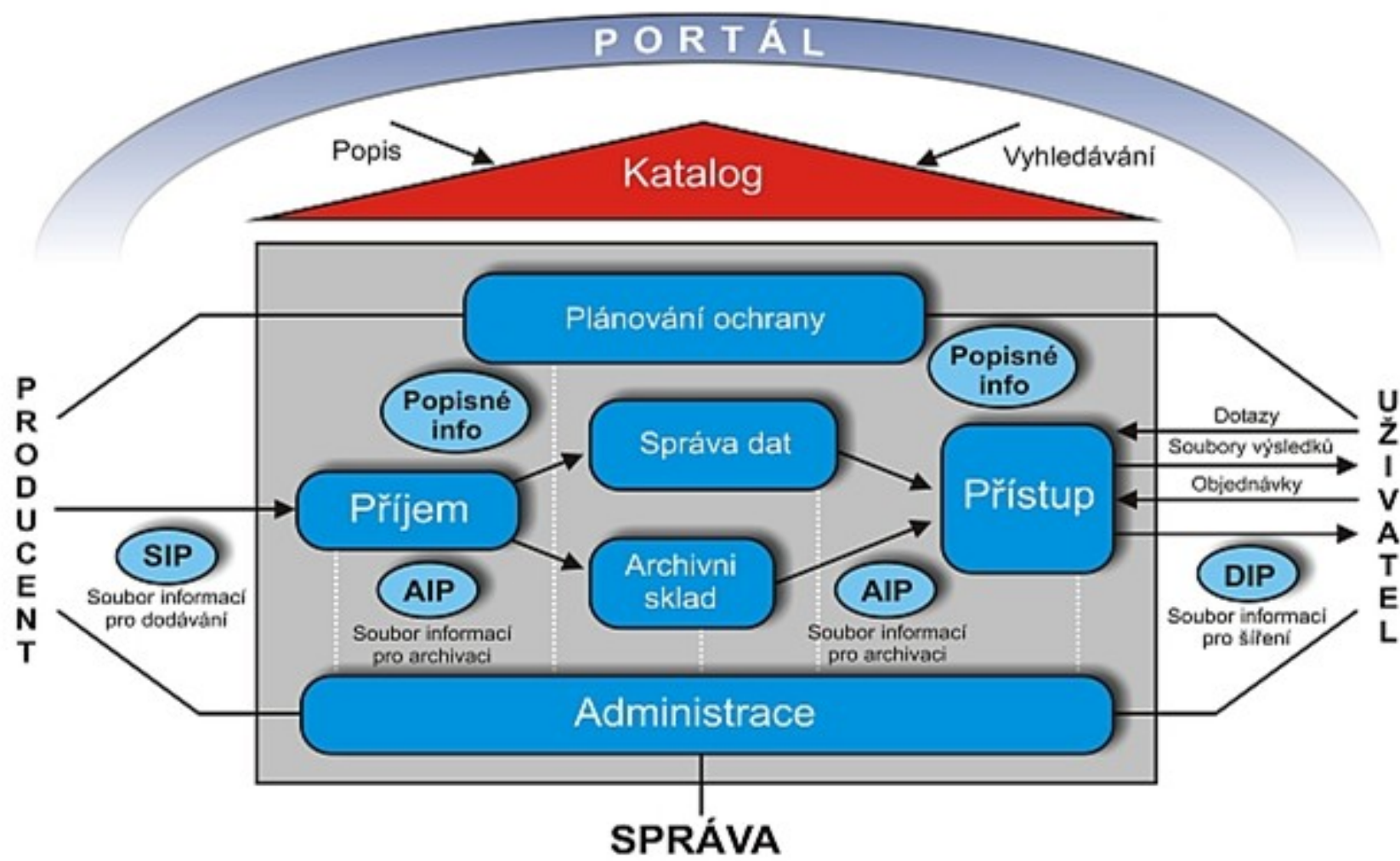
```
<mdWrap MDTYPE="PREMIS:EVENT"><xmlData><event xsi:schemaLocation="info:lc/xmlns/premis-v2
http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd" version="2.2">
<eventIdentifier>
<eventIdentifierType>UUID</eventIdentifierType>
<eventIdentifierValue>2cfc30f8-458d-4e14-9a42-6dc5e1c7e964</eventIdentifierValue>
</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, fonoválečky, (CD)
- ❑ ? kroniky, datová CD a DVD, ...

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





# Datové balíčky v LTP repozitáři



- ❑ SIP = Submission Information Package
  - ❑ Data, která do repozitáře tečou.
  - ❑ Běžně 1:1, ale někdy SIP ≠ AIP
  - ❑ normalizace, dodání metadat
- ❑ AIP = Archival Information Package
  - ❑ Data, uložená v repozitáři
- ❑ DIP = Dissemination Information Package
  - ❑ Data, která jdou z repozitáře
  - ❑ Může být 1:1, ale často AIP ≠ DIP

# 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



- ❑ Zvýšení reputace - vyjádření důvěryhodnosti
  - ❑ dodavatelé dat, uživatelé, zřizovatel
  
- ❑ Pohled dovnitř
  - ❑ ujištění že to funguje/jak by to fungovat mělo
    - ❑ co vylepšit, kde zapracovat

# Důvěryhodný digitální repozitář



## 2010 Memorandum o porozumění

- ❑ Základní certifikace (Basic certification)
  - ❑ CoreTrustSeal/Data Seal of Approval (DSA).
- ❑ Rozšířená certifikace (Extended certification)
  - ❑ self-audit podle ISO 16 363 nebo DIN 31 644, který je zveřejněn a externě ohodnocen
- ❑ Formální certifikace (Formal certification)
  - ❑ externí, nezávislý audit a certifikaci dle ISO 16 363 nebo DIN 31 644

# Nástroje pro audit digitálních repozitářů



Společné požadavky:

- ❑ Dokumentace, dokumentace, dokumentace - smlouvy, strategie, procesy, ...
- ❑ Transparentnost
- ❑ Dlouhodobost
  - ❑ Příprava na audit, ale i akce po něm
  - ❑ Technologické a další změny - potřeba opakovat audity

# Nástroje pro audit digitálních repozitářů



## Okruhy:

- ❑ Repozitář - poslání, strategie, procesy, smlouvy, finance, zaměstnanci, komunita, ...
- ❑ Data management - balíčky (SIP, AIP, DIP), integrita, autenticita, metadata, ...
- ❑ Technické aspekty - vhodný HW a SW, infrastruktura, ...
- ❑ Bezpečnost - útoky, technické havárie, přírodní katastrofy, ...

# Nástroje pro audit digitálních repozitářů



- ❑ CoreTrustSeal/Data Seal of Approval
- ❑ NESTOR (DIN 31644)
- ❑ ISO 16363





Není Seal of  
Approval  
jako  
Seal of  
Approval



# CoreTrustSeal/DSA

❑ Dříve Data Seal of Approval



❑ Pro repozitáře vědeckých dat  
(ale nejen pro ně)



❑ Certifikací DSA prošlo i několik repozitářů vědeckých dat v  
ČR - Univerzita Karlova

- ❑ LINDAT-Clarín - Centre for Language Research Infrastructure in the Czech Republic
- ❑ Repository of Charles University in Prague, Map Collection
- ❑ Český sociálněvědný datový archiv

❑ Překlad DSA do češtiny

- ❑ <https://dsa.cuni.cz/DSA-3.html?look=new>

# CoreTrustSeal



- ❑ Stanovuje 16 kritérií (repozitář musí splnit body 4 - 13)
- ❑ U zbytku musí umožnit jejich splnění producenty a uživateli dat
- ❑ 1. Producent dat vkládá data do digitálního repozitáře společně s informacemi, které ostatním subjektům umožňují zhodnotit jejich **kvalitu** a to, na kolik odpovídají **etickým a jiným normám** platným pro danou disciplínu (3)
- ❑ 2. Producent dat odevzdává data ve **formátech doporučovaných** digitálním repozitářem (3)
- ❑ 3. Producent dat odevzdává **data** společně s **metadaty** vyžadovanými digitálním repozitářem (4)

<https://www.coretrustseal.org>

# CoreTrustSeal



- ❑ 4. Digitální repozitář má jasně stanovené **poslání** (mission) v oblasti digitální archivace a uplatňuje je (4)
- ❑ 5. Digitální repozitář věnuje dostatečnou péči dodržování **právních předpisů a smluv**, a to případně včetně těch, které se vztahují k **ochraně osob** (4)
- ❑ 6. Digitální repozitář aplikuje **zdokumentované procesy a postupy** pro správu ukládání dat (4)
- ❑ 7. Digitální repozitář má **plán dlouhodobé ochrany** digitálního obsahu v něm uloženého (3)

# CoreTrustSeal



- ❑ 8. Archivace probíhá v průběhu celého **životního cyklu dat** a podle jasně **stanovených postupů** (3)
- ❑ 9. Digitální repozitář přebírá od producentů dat **odpovědnost za zpřístupnění** digitálních objektů (4)
- ❑ 10. Digitální repozitář umožňuje uživatelům najít a použít data a trvale na ně odkazovat (4)
- ❑ 11. Digitální repozitář zajišťuje integritu digitálních objektů a metadat (3)
- ❑ 12. Digitální repozitář zajišťuje autenticitu digitálních objektů a metadat (3)

# CoreTrustSeal



- ❑ 13. **Technická infrastruktura** výslovně podporuje úkoly a funkce popsané v **mezinárodně uznávaných archivních standardech** jako je např. OAIS (4)
- ❑ 14. Uživatel dat dodržuje **přístupová pravidla** stanovená digitálním repozitářem (4)
- ❑ 15. Uživatel souhlasí s **pravidly pro sdílení a správné využívání znalostí a informací** obecně uznávanými v dané oblasti a řídí se jimi (3)
- ❑ 16. Uživatel respektuje digitálním repozitářem stanovená **licenční omezení** týkající se užití dat (3)

<b>Úroveň</b>	<b>Význam</b>	<b>Poznámka/URL</b>
<b>0</b>	<b>N/A: nelze aplikovat</b>	<b>Uved'te vysvětlení.</b>
<b>1</b>	<b>Ne, ještě jsme to nezvažovali.</b>	<b>Uved'te vysvětlení.</b>
<b>2</b>	<b>Máme teoretický koncept.</b>	<b>Dodejte URL.</b>
<b>3</b>	<b>Rozpracováno: jsme v implementační fázi.</b>	<b>Dodejte URL.</b>
<b>4</b>	<b>Implementováno: Tuto zásadu jsme plně implementovali pro potřeby našeho repozitáře</b>	<b>Dodejte URL.</b>

# NESTOR

- ❑ Pochází z Německa a reflektuje místní právní, institucionální i finanční prostředí (DIN 31644, 2009)
  - [www.langzeitarchivierung.de](http://www.langzeitarchivierung.de)
- ❑ Primárně pro paměťové instituce, ale i další organizace a LTP komunity (například vývojáři)
- ❑ Nestor je rozdělen do 14 sekcí, které pokrývají v podstatě totožná témata jako ISO 16363
- ❑ V ČR nestor převzal Národní archiv ČR:
  - [https://files.dnb.de/nestor/materialien/nestor\\_mat\\_17\\_cs.pdf](https://files.dnb.de/nestor/materialien/nestor_mat_17_cs.pdf)



# ISO 16 363



## ❑ **Organizační infrastruktura:**

- Řízení a organizační životaschopnost
- Organizační struktura a lidské zdroje
- Procesní odpovědnost a rámec politiky uchovávání
- Finanční udržitelnost
- Smlouvy, licence a závazky

## ❑ **Management digitálních objektů:**

- Příjem: akvizice obsahu
- Příjem: tvorba AIP balíčku
- Plánování dlouhodobé ochrany
- Uchovávání AIP balíčků
- Informační management
- Správa přístupu

# ISO 16 363



- ❑ **Infrastruktura a management bezpečnostních rizik**
  - ❑ Management rizik technické infrastruktury
  - ❑ Management bezpečnostních rizik
  
- ❑ K tomuto dokumentu se velmi těsně váže ISO 16 919
  - ❑ požadavky na certifikační orgány

# Stav LTP v ČR



- ❑ LTP oddělení a LTP repozitář NK ČR (NDK)
- ❑ Webarchiv (NK ČR)
- ❑ Národní digitální archiv (NDA)
- ❑ projekt ARCLib -> provoz KNAV
- ❑ 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

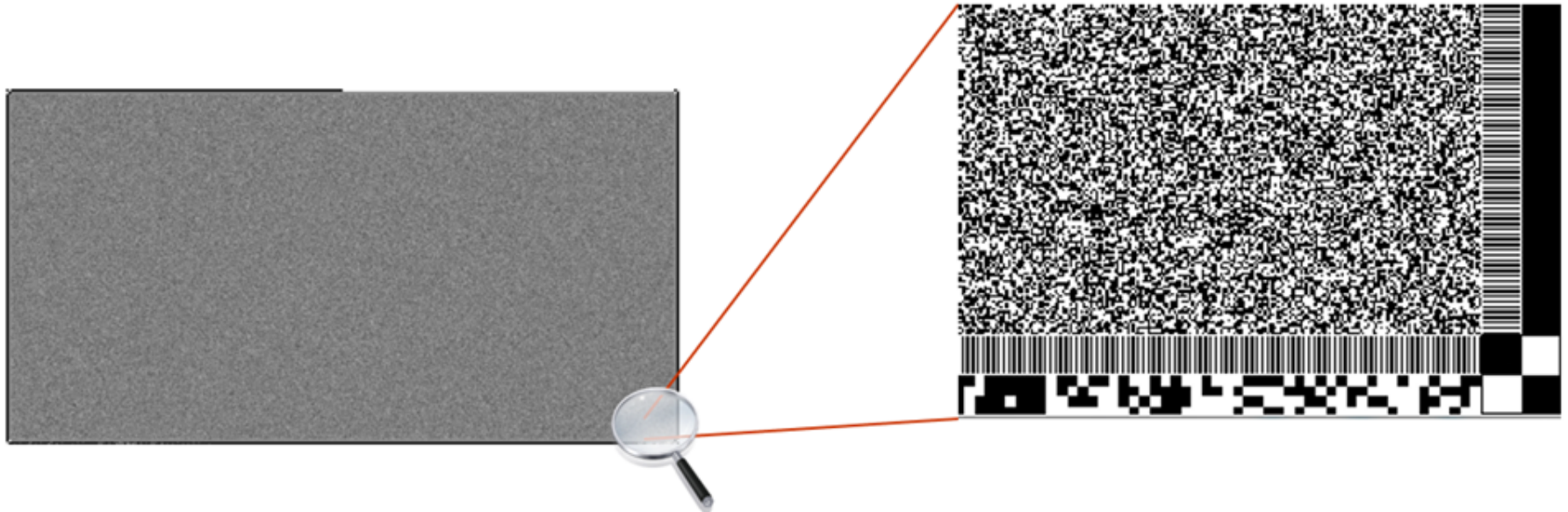
# Stav LTP v zahraničí



- ❑ Centrálny dátový archív (CDA)
  - ❑ cca 5 PB dat (2020)
  - ❑ 3 geograficky oddělené kopie
  - ❑ pro instituce pod MK SR (centralizované řešení)
  - ❑ bit-level ochrana, logickou ochranu neřeší
  - ❑ dark archive

# 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>

# 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.digitalniknihovna.cz/nkp/view/uuid:81e4de40-ca94-11e7-9c14-005056827e51>
- ❑ 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/>
  - ❑ <https://www.coretrustseal.org/>



- ❑ <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**  
zd.hruska@gmail.com