

Digitální knihovny

Metadata v digitálních knihovnách

Jindřich Marek

Metadata

- „data o datech“
- strukturovaná informace o digitálním objektu a případně jeho fyzické předloze
 - popisují, vysvětlují, lokalizují a jinak pomáhají k vyhledání, použití a správě informačních zdrojů
 - napomáhají archivaci i dlouhodobé ochraně digitálních dat

Typy metadat

- strukturní (vyjádření struktury dokumentu)
 - účelem je zobrazení v rozumné formě (celá kniha)
 - spojení často tisíců objektů do jedné intelektuální entity
- popisná (bibliografická, deskriptivní)
 - popis intelektuálního obsahu jednotky – online pro vyhledávání
- administrativní (nutná ke správě digitálních dat)
 - technická metadata – k uložení a ochraně dat
 - metadata práv – intelektuální a přístupová
 - ochranná metadata

Způsob uložení metadat

1. separátně – nejčastější
 - ulehčuje management metadat i vyhledávání v nich
 - nejčastěji ukládáno „vedle“ digitálních objektů a ideálně v databázi
2. v digitálním objektu
 - např. v hlavičce obrazového souboru
 - výhody: metadata se neztratí, není problém s linkováním dat a relevantních metadat
 - nevýhody: nelze vyjádřit strukturu (popis jen jednoho souboru), obtížněji se aktualizují
 - ne u všech formátů je to možné

Šest zásad pro dobrá metadata

1. Musí být vhodná pro sbírku a materiál v ní, uživatele a použití: robustnost formátu (DC vs. MODS).
2. Musí podporovat interoperabilitu.
3. Musí podporovat dlouhodobou ochranu objektů, které popisují.
4. Musí používat standardní kontrolované slovníky (co, kde, kdy a kdo).
5. Musí být sama digitálním objektem a musí mít vlastnosti pro archivaci, dlouhodobé uchování, jednoznačnou identifikaci apod.
6. Musí být ověřitelná a validovatelná.

Metadata: interoperabilita

METS – Metadata Encoding and Transmission Standard

- funguje jako kontejner (či „pořadač“) pro data v různých formátech

RDF – The Resource Description Framework

- poskytuje mechanismus pro integraci („křížení“) více metadatových formátů
- lze kombinovat různé elementy z různých formátů (pomocí tzv. jmenných prostorů)

Zápis metadat: XML

- metajazyk
 - odvozeno od složitějšího SGML, velmi flexibilní: lze definovat vlastní elementy a komplexní schémata
- jasně definovaná hierarchická (stromová) struktura elementů
 - nemožnost křížení
- elementy: (zpravidla párové) tagy
- atributy

XML Elements vs. Attributes

Take a look at these examples:

```
<person gender="female">
  <firstname>Anna</firstname>
  <lastname>Smith</lastname>
</person>
```

```
<person>
  <gender>female</gender>
  <firstname>Anna</firstname>
  <lastname>Smith</lastname>
</person>
```

In the first example gender is an attribute. In the last, gender is an element. Both examples provide the same information.

There are no rules about when to use attributes or when to use elements in XML.

Schéma pro metadata v XML

- definice elementů, které vytvářejí tagy a pravidla, jak lze elementy použít
- je nutné k vytváření a validaci XML záznamu
- různé druhy: DTD / XSD, RELAX NG apod.

- DTD – Document Type Definition
 - pravidla pro zápis XML – jaké elementy, opakování, jejich obsah, povinné elementy a atributy
- XSD – (XML Schema Definition) - modernější způsob definování/popisu jazyka XML
 - detailnější, flexibilnější
 - je v XML podobě, tj. je snadno strojově zpracovatelné

Typy digitálních dokumentů

- digitalizované dokumenty
- „born-digital“ dokumenty
 - vznikly s předpokladem, že nebudou mít analogový ekvivalent

Digitální dokument/objekt

- podle OAIS (Open Archival Information System)
 - množina bitových posloupností
- podle PREMIS (Preservation Metadata Maintenance Activity)
 - samostatná jednotka informace v digitální podobě
- různé dle různých datových modelů

- jednoduchý objekt
 - ohraničená jednotka – jeden soubor
- komplexní objekt
 - skládá se z digitálního souboru (jednoho nebo více) a metadat

Strukturní metadata

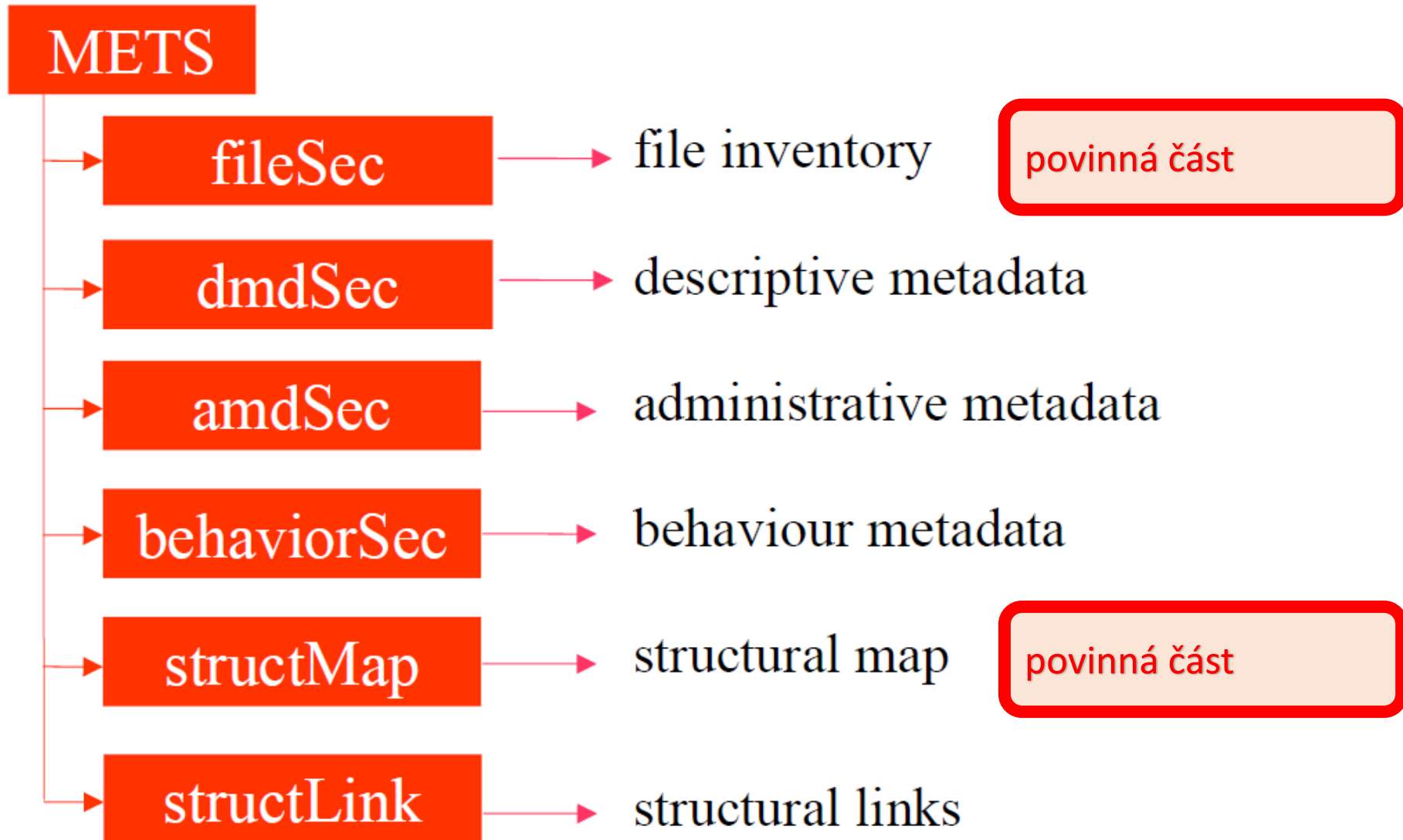
- interní organizace zdroje
 - kniha má 100 stran, pořídí se 100 obrázků ve formátu JPG: nutno vyjádřit logickou strukturu a fyzickou strukturu
 - metadata vyjádří posloupnosti, vztahy mezi objekty
 - mohou obsahovat další popis struktury ve značkovacím jazyce (název kapitoly atp.)
- formáty
 - METS
 - FOXML

METS



- Metadata Encoding and Transmission Standard:
<https://www.loc.gov/standards/mets/>
- kontejnerový formát pro uchování a výměnu metadat v digitálních knihovnách založený na XML
 - provazuje všechny typy metadat v jeden složený metadatový záznam
- některé části se „plní“ externími metadaty, jiné dokáží metadata vyjádřit

Struktura METS



Struktura METS

- nemusí být použity všechny sekce
- sekce pro externí metadata je může obsahovat (mdWrap), nebo na ně odkazovat (mdRef)

```
<dmdSec ID="dmd001">  
  <mdRef LOCTYPE="URN" MIMETYPE="application/xml" MDTYPE="EAD" LABEL="Berol Collection  
  Finding Aid">urn:x-nyu:fales1735</mdRef>  
</dmdSec>
```

```
<dmdSec ID="dmd002">  
  <mdWrap MIMETYPE="text/xml" MDTYPE="DC" LABEL="Dublin Core Metadata">  
    <dc:title>Alice's Adventures in Wonderland</dc:title>  
    <dc:creator>Lewis Carroll</dc:creator>  
  </mdWrap>  
</dmdSec>
```

```
▼<mets:fileSec>
  ▼<mets:fileGrp USE="MASTER">
    ▼<mets:file MIMETYPE="image/tiff" GROUPID="G1" ID="f0178m">
      <mets:FLocat LOCTYPE="URL" xlink:href="http://lcweb4.loc.gov/natlib/ihis/warehouse/afc9999005/AFS_300_A-734_B/0178.tif"/>
    </mets:file>
  </mets:fileGrp>
  ▼<mets:fileGrp USE="SERVICE">
    ▼<mets:file MIMETYPE="image/jpeg" GROUPID="G1" ID="f0178s">
      <mets:FLocat LOCTYPE="URL" xlink:href="http://lcweb4.loc.gov/natlib/ihis/service/afc9999005/AFS_300_A-734_B/0178v.jpg"/>
    </mets:file>
    ▼<mets:file MIMETYPE="image/tiff" GROUPID="G1" ID="f0178z">
      <mets:FLocat LOCTYPE="URL" xlink:href="/media/loc.afc.afc9999005.1153/0178.tif"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
▼<mets:structMap>
  ▼<mets:div DMDID="dmd1" TYPE="bib:modsBibCard">
    ▼<mets:div TYPE="bib:card">
      ▼<mets:div TYPE="lc:image">
        <mets:fptr FILEID="f0178m"/>
        <mets:fptr FILEID="f0178s"/>
        <mets:fptr FILEID="f0178z"/>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
</mets:mets>
```


FOXML

- Fedora Digital Object Model (Fedora v3)
- <https://wiki.lyrasis.org/display/FEDORA38/Fedora+Digital+Object+Model>
 - <https://wiki.lyrasis.org/pages/viewpage.action?pageId=66585857>
 - <https://fedora.info/definitions/1/0/foxml1-1.xsd>
- obsahuje formální (abstraktní) datový model, na rozdíl od METS

Popisná metadata

- pro vyhledávání, popis, identifikaci zdroje
 - jmenný popis - autor, název, vztahy, údaje o předloze aj.
 - věcný popis – MDT, předmětová hesla apod.
 - často kontrolované slovníky, pravidla zápisu, authority
- formáty: MARC21, MARCXML, MODS, DC (na násl. snímcích)
 - MARC21: není určen pro popis digitálních dokumentů, nevyhovuje
 - MODS: pro podrobnější popis a vztahy – 80 elementů
 - DC (Dublin Core): asi nejužívanější, jednoduchý – 15 elementů

ID	Date Added	Date Changed	Author	Title	Copyright Date	Barcode
1	14.3.2019 10:36	21.5.1993 15:51	Sandburg, Carl,	Arithmetic /	1993	

Tag	Ind1	Ind2
LDR		
001		
003		
005		
008		
010		
020		
040		
042		
050	0	0
082	0	0
100	1	
245	1	0
250		
260		
300		
500		
520		
650		0
650		0
650		1
650		1
650		1
700	1	
*		



Data
01142cam 2200301 a 4500
*



```

LDR 01142cam 2200301 a 4500
001      92005291
003      DLC
005      19930521155141.9
008      920219s1993      caua      j      000 0 eng
010      [a]:      92005291
020      [a]: 0152038655 :
          [c]: $15.95
040      [a]: DLC
          [c]: DLC
          [d]: DLC
042      [a]: lcac
050 00 [a]: PS3537.A618
          [b]: A88 1993
082 00 [a]: 811/.52
          [2]: 20
100 1 [a]: Sandburg, Carl,
          [d]: 1878-1967.
245 10 [a]: Arithmetic /
          [c]: Carl Sandburg ; illustrated as an anamorphic
          adventure by Ted Rand.
250      [a]: 1st ed.
260      [a]: San Diego :
          [b]: Harcourt Brace Jovanovich,
          [c]: c1993.
300      [a]: 1 v. (unpaged) :
          [b]: ill. (some col.) ;
          [c]: 26 cm.
500      [a]: One Mylar sheet included in pocket.
520      [a]: A poem about numbers and their characteristics.
          Features anamorphic, or distorted, drawings which can be
          restored to normal by viewing from a particular angle or by
          viewing the image's reflection in the provided Mylar cone.
650 0 [a]: Arithmetic
          [x]: Juvenile poetry.
650 0 [a]: Children's poetry, American.
650 1 [a]: Arithmetic
          [x]: Poetry.
650 1 [a]: American poetry.
650 1 [a]: Visual perception.
700 1 [a]: Rand, Ted,
          [e]: ill.

```

```
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <leader>01142cam 2200301 a 4500</leader>
    <controlfield tag="001"> 92005291 </controlfield>
    <controlfield tag="003">DLC</controlfield>
    <controlfield tag="005">19930521155141.9</controlfield>
    <controlfield tag="008">920219s1993 caua j 000 0 eng </controlfield>
    <datafield tag="010" ind1=" " ind2=" " >
      <subfield code="a"> 92005291 </subfield>
    </datafield>
    <datafield tag="020" ind1=" " ind2=" " >
      <subfield code="a">0152038655 </subfield>
      <subfield code="c">$15.95</subfield>
    </datafield>
    <datafield tag="040" ind1=" " ind2=" " >
      <subfield code="a">DLC</subfield>
      <subfield code="c">DLC</subfield>
      <subfield code="d">DLC</subfield>
    </datafield>
    <datafield tag="042" ind1=" " ind2=" " >
      <subfield code="a">lcac</subfield>
    </datafield>
    <datafield tag="050" ind1="0" ind2="0">
      <subfield code="a">P53537.A618</subfield>
      <subfield code="b">A88 1993</subfield>
    </datafield>
    <datafield tag="082" ind1="0" ind2="0">
      <subfield code="a">811/.52</subfield>
      <subfield code="2">20</subfield>
    </datafield>
    <datafield tag="100" ind1="1" ind2=" " >
      <subfield code="a">Sandburg, Carl,</subfield>
      <subfield code="d">1878-1967.</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
      <subfield code="a">Arithmetic /</subfield>
      <subfield code="c">
        Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
      </subfield>
    </datafield>
    <datafield tag="250" ind1=" " ind2=" " >
      <subfield code="a">1st ed.</subfield>
    </datafield>
    <datafield tag="260" ind1=" " ind2=" " >
      <subfield code="a">San Diego :</subfield>
      <subfield code="b">Harcourt Brace Jovanovich,</subfield>
      <subfield code="c">c1993.</subfield>
    </datafield>
    <datafield tag="300" ind1=" " ind2=" " >
      <subfield code="a">1 v. (unpaged) :</subfield>
      <subfield code="b">ill. (some col.) ;</subfield>
      <subfield code="c">26 cm.</subfield>
    </datafield>
    <datafield tag="500" ind1=" " ind2=" " >
      <subfield code="a">One Mylar sheet included in pocket.</subfield>
    </datafield>
    <datafield tag="520" ind1=" " ind2=" " >
      <subfield code="a">
        A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
      </subfield>
    </datafield>
    <datafield tag="650" ind1=" " ind2="0">
      <subfield code="a">Arithmetic</subfield>
      <subfield code="x">Juvenile poetry.</subfield>
    </datafield>
    <datafield tag="650" ind1=" " ind2="0">
      <subfield code="a">Children's poetry, American.</subfield>
    </datafield>
    <datafield tag="650" ind1=" " ind2="1">
      <subfield code="a">Arithmetic</subfield>
      <subfield code="x">Poetry.</subfield>
    </datafield>
    <datafield tag="650" ind1=" " ind2="1">
      <subfield code="a">American poetry.</subfield>
    </datafield>
    <datafield tag="650" ind1=" " ind2="1">
      <subfield code="a">Visual perception.</subfield>
    </datafield>
    <datafield tag="700" ind1="1" ind2=" " >
      <subfield code="a">Rand, Ted,</subfield>
      <subfield code="e">ill.</subfield>
    </datafield>
  </record>
</collection>
```

```
<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">Sandburg, Carl,</subfield>
  <subfield code="d">1878-1967.</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Arithmetic /</subfield>
  <subfield code="c">
    Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
  </subfield>
</datafield>
<datafield tag="250" ind1=" " ind2=" " >
  <subfield code="a">1st ed.</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" " >
  <subfield code="a">San Diego :</subfield>
  <subfield code="b">Harcourt Brace Jovanovich,</subfield>
  <subfield code="c">c1993.</subfield>
</datafield>
```

MODS



- Metadata Object Description Schema: <http://www.loc.gov/standards/mods/>
- derivát MARCu, sada elementů MODS odpovídá existujícím polím v knihovních databázích (OPAC) – lze mapovat a vyjádřit totéž
- podrobnější elementy než DC
- dobře se spojuje s METS a jeho strukturní mapou (METS: dmdSec)
- lze rozšířit pomocí elementu <extension> – např. jiný metadatový záznam v jiném formátu
- nenárokuje používání konkrétních pravidel pro popis

```
<mods xmlns:xlink="http://www.w3.org/TR/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.loc.gov/mods/" xsi:schemaLocation="http://www.loc.gov/mods/ http://www.loc.gov/standards/mods/mods.xsd">
  <titleInfo>
    <title>Arithmetic </title>
  </titleInfo>
  <name type="personal">
    <namePart>Sandburg, Carl</namePart>
    <namePart type="date">1878-1967</namePart>
    <role>
      <text>creator</text>
    </role>
  </name>
  <name type="personal">
    <namePart>Rand, Ted</namePart>
    <role>
      <text>ill.</text>
    </role>
  </name>
  <typeOfResource>text</typeOfResource>
  <originInfo>
    <place>
      <code authority="marc">cau</code>
      <text>San Diego</text>
    </place>
    <publisher>Harcourt Brace Jovanovich</publisher>
    <dateIssued>c1993</dateIssued>
    <dateIssued encoding="marc">1993</dateIssued>
    <edition>1st ed.</edition>
    <issuance>monographic</issuance>
  </originInfo>
  <language authority="iso639-2b">eng</language>
  <physicalDescription>
    <form authority="marcform">print</form>
    <extent>1 v. (unpaged) : ill. (some col.) ; 26 cm.</extent>
  </physicalDescription>
  <abstract>
    A poem about numbers and their characteristics. Features anamorphic, or distorted,
  </abstract>
  <targetAudience>juvenile</targetAudience>
  <note type="statement of responsibility">
    Carl Sandburg ; illustrated as an anamorphic adventure by Ted Rand.
  </note>
  <note>One Mylar sheet included in pocket.</note>
  <subject authority="lcsh">
    <topic>Arithmetic</topic>
    <topic>Juvenile poetry</topic>
  </subject>
  <subject authority="lcsh">
    <topic>Children's poetry, American</topic>
  </subject>
  <subject authority="lcshac">
    <topic>Arithmetic</topic>
    <topic>Poetry</topic>
  </subject>
  <subject authority="lcshac">
    <topic>American poetry</topic>
  </subject>
  <subject authority="lcshac">
    <topic>Visual perception</topic>
  </subject>
  <classification authority="lcc">PS3537.A618 A88 1993</classification>
  <classification edition="20" authority="ddc">811/.52</classification>
  <identifier type="isbn">0152038655 :</identifier>
  <identifier type="lccn">92005291</identifier>
  <recordInfo>
    <recordContentSource>DLC</recordContentSource>
    <recordCreationDate encoding="marc">920219</recordCreationDate>
    <recordChangeDate encoding="iso8601">19930521155141.9</recordChangeDate>
    <recordIdentifier source="DLC">92005291</recordIdentifier>
  </recordInfo>
</mods>
```

```
<name type="personal">
  <namePart>Sandburg, Carl</namePart>
  <namePart type="date">1878-1967</namePart>
  <role>
    <text>creator</text>
  </role>
</name>
<name type="personal">
  <namePart>Rand, Ted</namePart>
  <role>
    <text>ill.</text>
  </role>
</name>
<typeOfResource>text</typeOfResource>
<originInfo>
  <place>
    <code authority="marc">cau</code>
    <text>San Diego</text>
  </place>
  <publisher>Harcourt Brace Jovanovich</publisher>
  <dateIssued>c1993</dateIssued>
  <dateIssued encoding="marc">1993</dateIssued>
  <edition>1st ed.</edition>
  <issuance>monographic</issuance>
</originInfo>
```

Viewing the image's reflection in the provided Mylar cone.

Hlavní elementy MODS

- Title Info
- Name
- Type of resource
- Genre
- Origin Info
- Language
- Physical description
- Abstract
- Table of contents
- Target audience
- Note
- Subject
- Classification
- Related item
- Identifier
- Location
- Access conditions
- Extension
- Record Info

Dublin Core (DC)

- <http://dublincore.org/>
- na minimální univerzální popis dokumentu
 - dostatečná jednoduchost
 - zároveň dostatečný rozsah
 - možnost popisu prakticky jakéhokoliv zdroje
- 15 základních metadatových prvků
 - [The Simple Dublin Core Generator](#)

Dublin Core: prvky

Obsah:

Název

Předmět a klíčová slova

Popis

Typ zdroje

Zdroj

Vztah

Pokrytí

Intelektuální vlastnictví:

Tvůrce

Vydavatel

Příspěvatel

Autorská práva

Zdroj:

Datum

Formát

Identifikátor

Jazyk

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<dc xmlns="http://purl.org/dc/elements/1.1/">
  <title>Arithmetic / </title>
  <creator> Sandburg, Carl, 1878-1967. </creator>
  <creator> Rand, Ted, ill. </creator>
  <type/>
  <publisher>San Diego :Harcourt Brace Jovanovich,</publisher>
  <date>c1993.</date>
  <language>eng</language>
  ▼<description>
    A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be restored to normal by viewing from a particular angle or by viewing the image's reflection in the provided Mylar cone.
  </description>
  <description>One Mylar sheet included in pocket.</description>
  <subject>Arithmetic</subject>
  <subject>Children's poetry, American.</subject>
  <subject>Arithmetic</subject>
  <subject>American poetry.</subject>
  <subject>Visual perception.</subject>
</dc>
```

```
▼<dc xmlns="http://purl.org/dc/elements/1.1/">
  <title>Arithmetic / </title>
  <creator> Sandburg, Carl, 1878-1967. </creator>
  <creator> Rand, Ted, ill. </creator>
  <type/>
  <publisher>San Diego :Harcourt Brace Jovanovich,</publisher>
  <date>c1993.</date>
  <language>eng</language>
  ▼<description>
    A poem about numbers and their characteristics. Features anamorphic, or distorted, drawings which can be
  </description>
  <description>One Mylar sheet included in pocket.</description>
  <subject>Arithmetic</subject>
  <subject>Children's poetry, American.</subject>
  <subject>Arithmetic</subject>
  <subject>American poetry.</subject>
  <subject>Visual perception.</subject>
</dc>
```

Administrativní metadata

- technická metadata
- metadata práv
- ochranná metadata

- mohou být součástí popisných metadat, ale častěji samostatně

- určena pro správce dat
 - management (správa) dat: strategie uložení, identifikátory, údaje o skenu)
 - obsah závisí na potřebách projektu a plánech do budoucna

Administrativní metadata (METS)

techMD

- formát PREMIS (jeho část Object)
- formát MIX (NISO Metadata for Images in XML)

rightsMD - dva druhy práv k zacházení s objektem:

- administrativní metadata PREMIS Rights
- legislativní práva METS Rights

sourceMD – info o předloze – používá se málo, nutno vyplnit svými údaji, lze užít např. MARCXML

digiprovMD - zaznamenání událostí spojených s objektem

- PREMIS Events

Administrativní metadata: technická metadata

- informace o vlastnostech digitálního objektu
 - lze získat automatickými nástroji
 - informace o objektu (velikost, formát, bitová hloubka, komprese, kodeky apod.)
 - informace nutné k zobrazení a využití
 - informace o procesu skenování (HW a SW)
- MIX (Metadata for Images in XML)

MIX



- Metadata for Images in XML: <http://www.loc.gov/standards/mix/>
- technický popis obrazů
 - vlastnosti digitálního obrazu, jako např. formát souboru, jeho velikost, rozměry, rozlišení, komprese apod.
 - harmonizováno s PREMISem
 - lze použít jako samostatný záznam, nebo jako rozšíření v rámci METS
- služby třetích stran – PRONOM, JHOVE

Administrativní metadata: metadata práv

- správa přístupu a vlastnictví práv k digitálním objektům
- METS Schema for Rights Declaration
 - popis typu práv, vlastníka práv i podmínky přístupu
- PREMIS Rights

Administrativní metadata: ochranná metadata

- logická ochrana dat (nejen ochrana bitstreamu)
 - obsah archivovaných objektů bude možné zobrazit, interpretovat i přes změny technologií v budoucnu
 - nepodporují vyhledávání
 - ne pro analogové dokumenty ani uživatelské kopie (náhledy aj.)
- formát PREMIS
 - části Object, Agents, Rights a Events

PREMIS

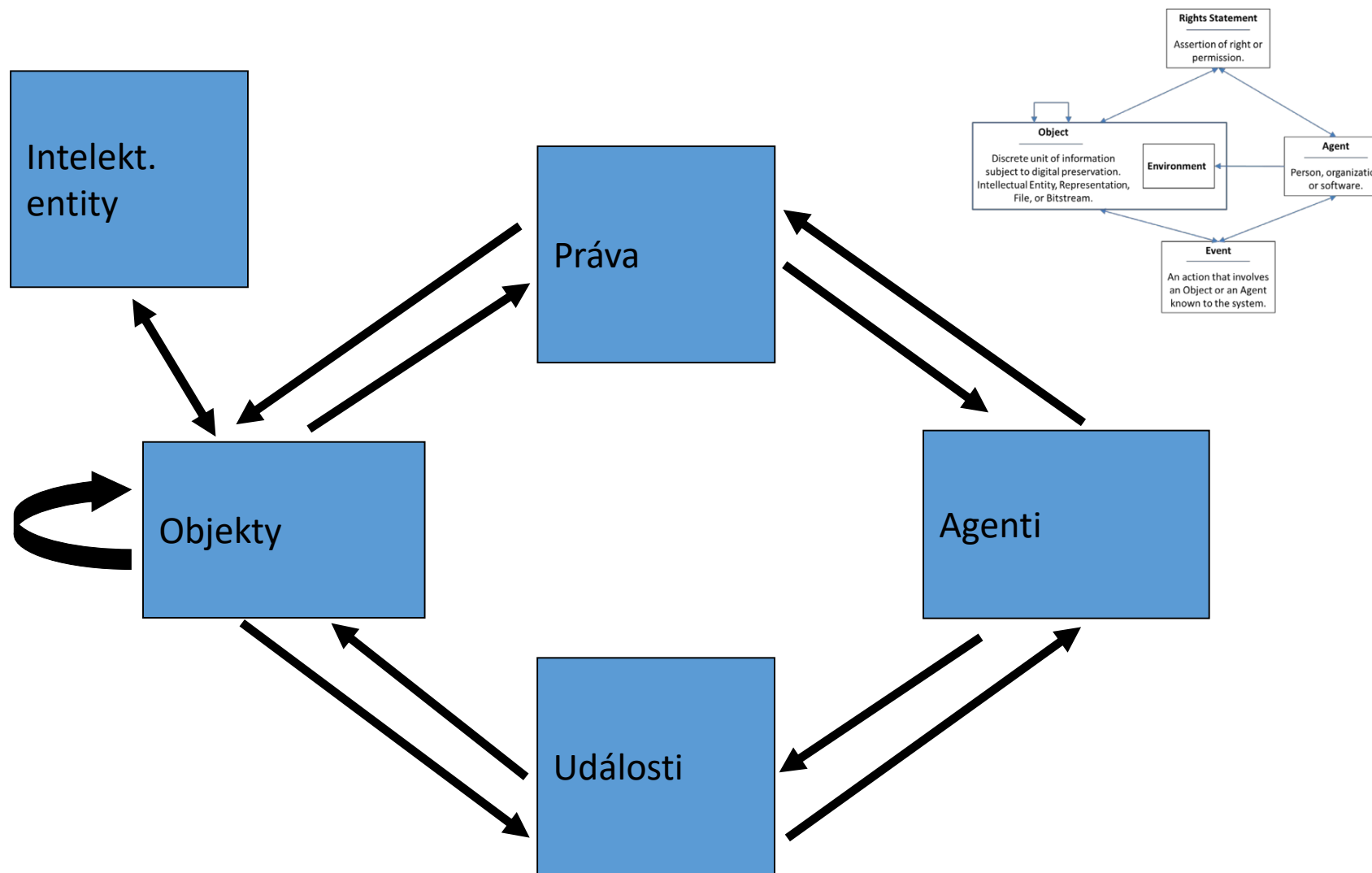


- PREservation Metadata Implementation Strategies
- <http://www.loc.gov/standards/premis/>
- ochranná metadata poskytují informace o:
 - původu
 - struktuře
 - formátech souborů
 - technickém prostředí
 - právech
- v METS – součástí administrativních metadat

Základní principy PREMIS

1. Metadata pro dlouhodobou ochranu.
2. Uchování možnosti zobrazení, porozumění, autenticity a identity objektu.
3. Možnost implementace – jasně definováno, podporováno návody a doporučeními.
4. Technická neutralita.
5. Nepředpokládá specifický systém, technologii, systémovou a databázovou architekturu nebo ochrannou strategii.

Datový model PREMIS



Standards pro digitální knihovny: data

- vlastní obrázky (případně nahrávky apod.)

Obrazová data

Správná volba obrazového formátu je důležitá

1. Kvůli možnosti zobrazení digitalizovaných dat v budoucnu.
2. Kvůli velikosti pořizovaných souborů.

Doporučení Florida Text Archive pro formáty digitálních obrazů

	High Confidence Level	Medium Confidence Level	Low Confidence Level
Raster Image	<ul style="list-style-type: none"> - TIFF (uncompressed) - JPEG2000 (lossless) (*.jp2) - PNG (*.png) 	<ul style="list-style-type: none"> - BMP (*.bmp) - JPEG/JFIF (*.jpg) - JPEG2000 (lossy) (*.jp2) - TIFF (compressed) - GIF (*.gif) - Digital Negative DNG (*.dng) - Google WebP (*.webp) 	<ul style="list-style-type: none"> - MrSID (*.sid) - TIFF (in Planar format) - FlashPix (*.fpx) - PhotoShop (*.psd) - RAW - JPEG 2000 Part 2 (*.jpf, *.jpx) - All other raster image formats not listed here
Vector Graphics	<ul style="list-style-type: none"> - SVG (no Java script binding) (*.svg) 	<ul style="list-style-type: none"> - Computer Graphic Metafile (CGM, WebCGM) (*.cgm) 	<ul style="list-style-type: none"> - Encapsulated Postscript (EPS) - Macromedia Flash (*.swf) - All other vector image formats not listed here

Ad Doporučení

- pro digitalizaci přicházejí v úvahu pouze formáty uvedené v prvním sloupci
- při digitalizaci knih se používají pouze rastrové (nikoliv vektorové) formáty, nejvhodnější jsou tyto:
 - TIFF – Tagged Image Format (případně kvůli úspoře místa na disku možno po digitalizaci konvertovat do JPG)
 - JPEG 2000
 - OSTRÁKOVÁ, Natalie. JPEG 2000 jako archivní formát obrazových dat. *Knihovna: knihovnická revue*, 29(1), 5–26. ISSN 1801-3252. <https://knihovnarevue.nkp.cz/archiv/dokumenty/2018-1/ostrakova.pdf>
- <https://libraries.flvc.org/documents/181844/502298/Recommended+Data+Formats/0b25496f-33ac-4f56-9550-12c34f3d5d7c>

Další rozšířené obrazové formáty

- používány hlavně na webu
- GIF: max. 256 barev (8 bitů)
- JPG (JPEG): široká podpora v programech, 24 bitů
- PNG: náhrada za GIF, ale i některé možnosti TIFFu
- MrSID: komerční formát, vhodné pro mapy (dlaždice)
- RAW: kontejnerový formát, různé specifikace

Použité zdroje

- CUBR, Ladislav. *Dlouhodobá ochrana digitálních dokumentů*. Praha: Národní knihovna ČR, 2010.
- GARTNER, Richard. *An Introduction to METS*. <http://www.odl.ox.ac.uk/projects/docs/mat-intro.ppt>
- GARTNER, Richard. *Metadata for digital libraries: state of the art and future directions*, 2008. <http://www.jisc.ac.uk/whatwedo/services/techwatch/reports/horizonscanning/hs0801.aspx>
- GUENTHER, Rebecca. *The Metadata Object Description Schema (MODS)*. 2004. Dostupné z WWW: <http://www.loc.gov/standards/mods/presentations/niso-mods.ppt>
- *<METS> metadata encoding and transmission standard: primer and reference manual*. Version 1.6, September 2007. <http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf>
- PAPY, Fabrice. *Digital Libraries*. London: ISTE Ltd.; Hoboken: John Wiley & Sons, 2008. ISBN 978-1-84821-042-4.
- *Preservation Metadata for Digital Objects: A Review of the State of the Art. A White Paper by the OCLC/RLG Working Group on Preservation Metadata*, 2001. http://www.oclc.org/research/activities/past/orprojects/pmwg/presmeta_wp.pdf
- SCHREIBMAN, Susan. *Best Practice Guidelines for Digital Collections at University of Maryland Libraries*. 2nd ed., College Park, MD 2007.
- *Standards at the Library of Congress*. <https://www.loc.gov/standards/>

Úkol: Kramerius KNAV pomocí API

- vyhledejte libovolnou knihu, zaznamenejte si její UUID
 - <https://kramerius.lib.cas.cz/view/uuid:e51af7bd-ae80-436d-bbeb-ec4e8bce81a4>
- zobrazte popis v MODS a DC
 - https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:e51af7bd-ae80-436d-bbeb-ec4e8bce81a4/streams/BIBLIO_MODS
 - <https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:e51af7bd-ae80-436d-bbeb-ec4e8bce81a4/streams/DC>
- zobrazte veškerá metadata ve FOXML
 - <https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:e51af7bd-ae80-436d-bbeb-ec4e8bce81a4/foxml>
 - **poznamenejte si UUID obrázku**, který vás zajímá
- dokumentace k API:
<https://github.com/ceskaexpedice/kramerius/wiki/ClientAPIDEV>

Úkol: Kramerius KNAV pomocí API

- zobrazte obrázek v jednotlivých rozlišeních
 - <https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:c69ae36c-478c-4a97-b114-cbd5432b669a/thumb>
 - <https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:c69ae36c-478c-4a97-b114-cbd5432b669a/preview>
 - <https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:c69ae36c-478c-4a97-b114-cbd5432b669a/full>
- zobrazte veškerá data k obrázku ve FOXML
 - <https://kramerius.lib.cas.cz/search/api/v5.0/item/uuid:c69ae36c-478c-4a97-b114-cbd5432b669a/foxml>
 - najděte **text získaný pomocí OCR** (sub: foxml:datastream ID="TEXT_OCR")
- dekódujte získaný text z formátu base64 do čitelného textu
 - použijte <https://www.base64decode.org/>