Národní knihovna

| | |
|---|---|
| **DEFINICE METADATOVÝCH FORMÁTŮ** | 18. července<br>**2017**<br>Dokument verze 0.2 |

# Definice metadatových formátů pro digitalizaci zvukových dokumentů

# Část 1| Gramofonové desky

**Autoři:**
Lenka Damborská, Natalie Ostráková, Zdeněk Vašek, Pavlína Kočišová

# Historie verzí

| jméno | datum | verze dokumentu | provedené změny |
|---|---|---|---|
| Lenka Damborská, Natalie Ostráková, Zdeněk Vašek | prosinec 2016 | 0.1 | První znění. |
| Lenka Damborská, Zdeněk Vašek, Pavlína Kočišová, Natalie Ostráková | květen 2017 | 0.2 | Nová verze na základě vzorových balíčků. |

Tato definice metadatových formátů vychází z Definice metadatových formátů pro monografie ve verzi 1.2 v rámci Standardu Národní digitální knihovny (dále Standard NDK).

# Obsah

1 Popis standardu ..... 5
1.1 Účel standardu ..... 5
1.2 Zodpovědnost ..... 5
1.3 Dohoda mezi dodavatelem dat a zadavatelem ..... 5
1.4 Popis a verze použitých standardů a identifikátorů ..... 6
1.5 Definice pojmů ..... 7
1.6 Terminologie ..... 8
2 Výstupy digitalizace ..... 9
3 Granularita metadatového záznamu ..... 11
4 Identifikátory ..... 12
5 Struktura SIP balíčku ..... 13
5.1 Soubor info.xml ..... 14
5.2 Složka pro původní neupravené zvukové soubory [sourceaudio] ..... 15
5.3 Složka pro zvukové soubory [mastercopyaudio] ..... 15
5.4 Složka pro zvukové soubory [usercopyaudio] ..... 15
5.5 Složka pro obrazové soubory [mastercopy] ..... 15
5.6 Složka pro obrazové soubory [usercopy] ..... 15
5.7 Složka [alto] ..... 16
5.8 Složka [txt] ..... 16
5.9 Složka [amdsec] ..... 16
5.10 Soubor Hlavni_METS.xml ..... 17
5.11 Soubor MD5 ..... 17
6 Názvová konvence složek a souborů ..... 18
7 Metadata ..... 20
7.1 Kořenový element hlavního METS záznamu ..... 20
7.2 METS hlavička <metsHdr> ..... 20
7.3 METS část <dmdSec> - Bibliografická metadata – MODS a Dublin Core ..... 21
7.3.1. Pole MODS a DC pro úroveň kolekce/souboru ..... 24
7.3.2 Pole MODS a DC pro úroveň jedné strany desky zvukového nosiče ..... 35
7.3.3 Pole MODS a DC pro úroveň tracku neboli stopy ..... 37
7.3.4 Pole MODS a DC pro obrazové informace (přílohy) ..... 43

7.4 METS část <amdSec> - Technická a administrativní metadata – formáty AES57, MIX a PREMIS ..... 48
7.4.1 PREMIS Object ..... 50
7.4.2 PREMIS Event ..... 55
7.4.3 PREMIS Agent ..... 57
7.4.4 Technická metadata pro zvukové soubory AES57 ..... 58
7.4.4.1 Pro digitální dokument ..... 59
7.4.4.2 Pro původní zdrojový (analogový) dokument (př. gramodeska), bude v METS v části <sourceMD> ..... 65
7.4.5 Technická metadata MIX ..... 68
7.5 METS část <fileSec> ..... 77
7.5.1 <fileSec> hlavního záznamu METS ..... 77
7.5.2 <fileSec> vedlejšího záznamu amd_mets.xml ..... 79
7.6 METS část <structMap> - Strukturální metadata ..... 81
7.6.1 <structMap> hlavního záznamu METS ..... 81
7.6.1.1 Vyjádření fyzické strukturální mapy ..... 81
7.6.1.2 Vyjádření logické strukturální mapy ..... 82
7.6.1.2.1 Vyjádření logické struktury, s vazbou na ALTO bloky ..... 82
7.7 OCR (ALTO XML a TXT OCR) ..... 83
8 Autorsko-právní metadata ..... 98

# 1 Popis standardu

## 1.1 Účel standardu

Definice metadatových formátů pro zvukové dokumenty (dále DMF) bude sloužit jako předpis pro výsledek procesu digitalizace zvukových dokumentů v digitalizačních projektech knihoven v ČR. V první části bude standard zaměřen na gramofonové desky. Bude sloužit jako jednotný formát pro paměťové instituce, které chtějí svá data dlouhodobě archivovat v některém ze systémů dlouhodobé archivace využívajícím Standard NDK.

**Pro koho není určen?** DMF nebude určen pro malé lokální digitalizační aktivity, které mají jako jediný účel zpřístupnit dokumenty pro omezený okruh uživatelů. Ty si z něj mohou brát inspiraci, ale nemusí ho do důsledku dodržet, pokud nemají v plánu dlouhodobé uložení takto vytvořených dat.

V této specifikaci je blíže definován typ "zvukový dokument".

## 1.2 Zodpovědnost

Za DMF, jeho správnost a rozvoj, je zodpovědná Národní knihovna ČR. Kontaktním oddělením je Odbor digitálních fondů, resp. Oddělení pro standardy. S doplňujícími dotazy lze kontaktovat Zdeňka Vaška (zdenek.vasek@nkp.cz). Garantem obsahu je Moravská zemská knihovna v Brně, kontaktní osobou je Lenka Damborská (lenka.damborska@mzk.cz). Standard vznikl ve spolupráci s Národním muzeem (Filip Šír).

## 1.3 Dohoda mezi dodavatelem dat a zadavatelem

DMF definuje podobu dat a metadat vyrobených dodavatelem a slouží jednak jako závazný dokument pro zadání digitalizace a jednak pro kontrolu dodaných dat. Požadavky definuje zadávající instituce a ošetřuje s dodavatelem dat smlouvou. Jedná se např. o:

- granularita záznamů - zadavatel si musí určit, v jakém rozsahu trvá na vyplnění nepovinných metadatových elementů. DMF obsahuje definici toho, co lze vytvořit, ale je na zadavateli, aby definoval požadovaný rozsah. Z hlediska dlouhodobého uchovávání jsou nezbytné všechny elementy s označením M a jim podřazené elementy.
- podrobnosti skenování obrazů - např. použité rozlišení (300 nebo více)

## 1.4 Popis a verze použitých standardů a identifikátorů

V této verzi specifikace budou použity následující verze standardů:

| Standard | verze | odkaz na webové stránky standardu |
|---|---|---|
| METS | verze 1.11 | http://www.loc.gov/standards/mets/ |
| MODS | verze 3.6 | http://www.loc.gov/standards/mods/ |
| Dublin Core | verze 1.1 | http://dublincore.org/documents/dces/ |
| MIX | verze 2.0 | http://www.loc.gov/standards/mix |
| AES57 | verze 1.0 | http://www.aes.org/publications/standards/ |
| PREMIS | verze 2.2 | http://www.loc.gov/standards/premis/ |
| ALTO | verze 3.1 | http://www.loc.gov/standards/alto/ |
| CopyrightMD | verze 0.91 | http://www.cdlib.org/groups/rmg/ |

Pro převod metadat k titulu zvukového dokumentu z MARC 21 resp. MARCXML do MODS lze využít a dle potřeb knihovny upravit oficiální mapování a konverzní šablonu ze stránek http://www.loc.gov/standards/mods/mods-conversions.html.

Pro vytváření metadat ve formátu Dublin Core lze využít a dle potřeb knihovny upravit oficiální mapování a konverzní šablonu z formátu MODS ze stránek http://www.loc.gov/standards/mods/mods-dcsimple.html.

V této verzi specifikace mohou být použity následující identifikátory:

| | | |
|---|---|---|
| UUID | http://www.ietf.org/rfc/rfc4122.txt | |
| Číslo matrice | číslo vylisované na jedné straně gramofonové desky - na jednom nosiči se může nacházet jedno až dvě matriční čísla | |
| Objednací číslo | katalogové číslo nakladatele - nakladatelské číslo, které je přiřazeno desce nebo kolekci desek (jedno pro desku či kolekci) | |
| URN:NBN | https://resolver.nkp.cz | Pro jednotlivé dokumenty se samostatným bibliografickým záznamem. |
| Další identifikátor | | Je možné použít další standardizovaný identifikátor, v případě nejasností kontaktujte Oddělení pro standardy. |

**Povinné předpoklady:**

- veškerá metadata musí pro zápis používat kódování **UTF-8**
- velikost písmen v názvech souborů a složek - **všechna malá**

**Další doporučení:**

- původní neupravené zvukové soubory by měly mít parametry 96kHz, 24bit, zvukový formát WAV (specifikace: původní z r. 1991 Multimedia Programming Interface and Data Specifications 1.0 s. 56-65 a update z r. 1994 New Multimedia Data Types and Data Techniques s. 12-22) nebo formát BWF (specifikace:Specification of the Broadcast Wave Format (BWF)-Version 1 a Version 2)
- úpravy obrazu, které vedou ke změně rozměrů obrazu, rozlišení apod., se musí dělat před tím, než se vytvoří OCR, tj. budou se dělat zpravidla na TIFF souborech;
- OCR (ALTO XML) bude vznikat z uživatelské kopie - OCR je lepší ze souborů s kompresí (méně šumu)
- je nutné zachovat velikost obrazu uživatelských a archivních kopií stejnou (počet pixelů, rozlišení) tak, aby ALTO XML odpovídalo
- jako výchozí SW pro vytváření souborů JPEG2000 se bude používat Kakadu

**Význam pole "Povinnost":**

Pole "povinnost" uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné. Může nabývat následujících hodnot:

- **M - mandatory** (povinně plnit – element je součástí každého záznamu)
- **MA - mandatory if available** (povinně plnit pokud je to možné, pokud lze apod.)
- **R - recommended** (plnění hodnot elementu je doporučeno, není ovšem povinné)
- **RA - recommended if available** (doporučeno pokud lze plnit)
- **O - optional** (plnění hodnot elementu závisí na konkrétních potřebách instituce)

## 1.5 Definice pojmů

- **SA** = source audio = původní neupravený zvukový soubor
- **MC** = master copy = archivní kopie zvukového nebo obrazového záznamu (MCA, MC)
- **UC** = user copy = uživatelská kopie zvukového nebo obrazového souboru (UCA, UC)
- **PS** = původní sken – obrazový soubor vzniklý při digitalizaci, který se po zpracování (ořez, narovnání apod.) maže a dále se neukládá
- **základní intelektuální entita** = jedna zvuková kolekce (viz dále)
- **balíček** = složka s definovaným názvem a obsahem
- **SIP balíček** – Submission Information Package – je balíček dat a metadat v podobě, ve které je akceptovatelný pro LTP systém nebo pro aplikaci zpřístupnění; balíček bude obsahovat kompletní intelektuální entitu tj. jednu zvukovou kolekci

## 1.6 Terminologie

**Zvukový dokument** = pramen, jehož informačním obsahem je zaznamenaný/nahraný zvuk
**Gramofonová deska** = druh konkrétního zvukového dokumentu/nosiče
**Zvuková kolekce** = v případě gramofonových desek jde o soubor jedné nebo více gramofonových desek v 1 obalu nebo více obalech
**Zvukový záznam** = v katalogizační terminologii: metadatový záznam o zvukovém dokumentu
**Zvuková nahrávka** = záznam zvuku na jakémkoliv nosiči včetně virtuálního
**Zvuková stopa** = podmnožina zvukové nahrávky, nosič může mít více stop-drážek-tracků
**Zvukový soubor** = datový soubor se zvukovou nahrávkou nebo stopou, trackem
**Bibliografický záznam** = užíváno ve významu metadatového popisu obsahu zvukového dokumentu
**Cover** = obal zvukového nosiče
**Booklet** = brožura ke zvukovému nosiči
**IMGdisc** = sken strany desky zvukového nosiče

# 2 Výstupy digitalizace

1. Původní neupravené zvukové soubory (WAV, 96kHz, 24bit) - ***Source Audio (SA)***
   1.1. Jedna či více stop (jedna či více stran desek zvukového nosiče)
   1.2. Bez úprav, bez použití ekvalizačních křivek
   1.3. Parametry vycházejí z metodiky pro digitalizaci https://docs.google.com/document/d/15iRU-77WKsR9jfHinH3s6jPhda_a9BwENnb16beRaQg/edit
   1.4. povinné
   1.5. ekvivalentní pojmy: flat audio, preservation master

2. Archivní kopie zvukových souborů (např. WAV) - ***Master Copy Audio (MCA)***
   2.1. Kopie odvozená ze Source Audio
   2.2. Jedna či více stop (jedna či více stran desek zvukového nosiče)
   2.3. nepovinné - v případě, že se nebude od SA ničím lišit, zachová se pouze SA, MCA je nepovinné (kvůli objemu dat)
   2.4. ekvivalentní pojmy: remaster, production master

3. Uživatelské kopie zvukových souborů (např. MP3) - ***User Copy Audio (UCA)***
   3.1. Kopie odvozené z Master Copy
   3.2. jedná se o jednu stopu (celá strana desky zvukového nosiče) nebo více stop (více stran desek, příp. jedna strana až více stran desek rozdělené na části); preferuje se uchování jedné stopy jakožto celé strany desky (rozdělení na stopy by mělo být záležitostí prezentační vrstvy - např. zabudování tzv. cue sheet do Krameria)
   3.3. povinné
   3.4. ekvivalentní pojmy: access copy

4. Archivní kopie obrazových souborů (JPEG 2000) - ***Master Copy (MC)***
   4.1. desky, booklet, obal
   4.2. doporučené rozlišení je 300 PPI a více, v plné barvě a ve formátu bezztrátový JPEG2000
   4.3. Pro dokumenty (knihy, periodika) … v rozlišení minimálně 300 PPI. Barevná hloubka pro tyto dokumenty bude z důvodu urychlení produkce fixně nastavena na 24 bitů, barva (RGB).

5. Uživatelské kopie obrazových souborů (JPEG 2000) - ***User Copy (UC)***
   5.1. desky, booklet, obal
   5.2. uživatelská kopie se ukládá ve ztrátovém formátu JPEG2000

6. Metadata pro audiosoubory
   6.1. bibliografická metadata – MODS a DC
   6.2. strukturální metadata – METS
   6.3. technická metadata - PREMIS, AES57
   METS nebude obsahovat technická metadata pro UC, jen pro SA a MC.
   6.4. administrativní metadata – PREMIS, METS

7. Metadata obrazových souborů
   7.1. bibliografická metadata – MODS a DC
   7.2. strukturální metadata – METS

7.3. technická  metadata – MIX, PREMIS
METS nebude obsahovat technická metadata pro UC, jen pro MC.
7.4. administrativní metadata – PREMIS, METS

8. OCR - ALTO XML soubor pro každou stránku

9. OCR TXT soubor - pro možnost stáhnout si jen text dokumentu (tam kde kvalita OCR je odpovídající), vyhledávání/indexace.

10. kontrolní metadatové soubory (s kontrolními součty a údaji o vzniku dat apod.)

*Pozn. odkazy vedou na kapitolu s příslušnou problematikou.*

# 3 Granularita metadatového záznamu

Zvukový dokument obsahuje úrovně:

- kolekce, která je tvořena jednou či více stranami desek zvukového nosiče (collection)
- jedna strana desky zvukového nosiče (side)
- jednotlivé stopy, které se nachází na jedné straně desky (track): 1 - libovolný počet
    - v rámci strany budou track(y) neboli stopa(y) rozlišovány primárně na úrovni metadat (data budou tvořena jednou stranou desky)
- související obrazové informace:
    - obal zvukového nosiče (cover)
    - brožura ke zvukovému nosiči (booklet)
    - sken strany desky zvukového nosiče (imgdisc)

- Nevytváří se úroveň desky, protože by se jednalo o duplicitní informaci.
- Základní intelektuální entitou je kolekce, reprezentovaná jedním strojově čitelným jedinečným identifikátorem (např. čárový kód, RFID apod.).
- Metadatový popis je rozepsán až na úroveň stopy.
- Všechny metadatové záznamy jsou zakotveny v METS formátu, 1 kolekce = 1 METS dokument.
- SIP balíček je tvořen pro jednu kolekci (tzn., zahrnuje všechny informace týkající se celku, stran desek, jednotlivých stop a obrazových informací).

# 4 Identifikátory

Producent dat musí dogenerovat do balíčku dle této specifikace:

● identifikátor UUID musí být přidělen ke každé úrovni bibliografických metadat (nenachází se v bibl. záznamu v knihovním katalogu)
● URN:NBN pro úroveň intelektuální entity
● URN:NBN může být přidělováno také nižším logickým úrovním (entitám) – např. textu v bookletu, pokud to má význam, příloha
● syntax URN:NBN musí odpovídat specifikaci identifkátoru URN:NBN pro resolver NK (např. urn:nbn:cz:nk-12345a pro výstupy z projektu NDK)

Doporučené identifikátory jsou následující:

- Strojově čitelný identifikátor (např. RFID, čárový kód EAN apod.) - identifikuje celou zvukovou kolekci dle doporučení metodiky (jedna nebo více desek v 1 obalu)
- Katalogové číslo nakladatele (028 0) - nakladatelské číslo, které je přiřazeno desce nebo kolekci desek
- Číslo matrice (028 1 nebo 2) - číslo vylisované na gramofonové desce
- UUID **-** lze použít pro všechny úrovně digitálního objektu
- URN:NBN pro dokument (desku, kolekci?), který má samostatný bibliografický záznam, bude se přidělovat během digitalizace
- další standardizované identifikátory

Katalogové číslo a matriční číslo se může překrývat.

# 5 Struktura SIP balíčku

**Souborová struktura SIP balíčku:**

| SLOŽKA > | OBSAHUJE >> | OBSAHUJE >>> |
|---|---|---|
| | info.xml | |
| | sourceaudio | složka pro původní neupravené zvukové soubory |
| | mastercopyaudio | složka pro zvukové soubory |
| | usercopyaudio | složka pro zvukové soubory |
| | mastercopy (složka) | složka pro obrazy JPEG 2000 bezztrátové |
| | usercopy (složka) | složka pro obrazy JPEG 2000 ztrátový |
| | alto (složka) | soubory alto.xml pro každou stranu |
| | txt (složka) | soubory ocr.txt pro každou stranu |
| | amdsec (složka) | amd_mets.xml pro každou stranu |
| | hlavní_mets.xml | |
| | soubor.md5 | |

Technická a administrativní metadata nesmí být obsažena v hlavním METS záznamu. Každá stránka musí mít technická a administrativní metadata ve svém souboru s METS záznam (amd_mets.xml), stejně jako každý audio soubor.

**SIP balíček = 1 adresář pro kolekci gramofonových desek. Platí i v případě, že jde o kolekci několika desek společně popsaných.**

## Hlavní složka SIP balíčku obsahuje následující složky a soubory:

### 5.1 Soubor info.xml

Soubor info.xml musí obsahovat každý SIP balíček, budou zde velmi krátce zaznamenány údaje o jeho vzniku.

Aktuální verze XML schématu pro soubor info.xml je k nalezení na stránkách NDK.

| element | atributy | popis | Povinnost |
|---|---|---|---|
| <info> | | kořenový element INFO záznamu | M |
| <created> | | časový údaj o vzniku balíčku ve formátu ISO8601 na úroveň vteřin | M |
| <metadataversion> | | verze metadatové specifikace, podle které byl balíček zpracován | M |
| <packageid> | | název kořenového adresáře balíčku viz kap. 6 | M |
| <mainmets> | | název hlavního METS souboru včetně přípony | M |
| <validation> | | | M |
| | version | verze validačního nástroje (při použití Komplexního validátoru stačí jen verze, v případě použití jiného validačního nástroje nutno vyplnit i jeho název) | M |
| | | výstup validačního nástroje (např. OK; Valid) | R |
| <titleid> | | soupis identifikátorů titulu - vypsat všechny možné (viz type), které titul má přidělen | M |
| | type | možné hodnoty: urnnbn …… | M |
| | | identifikátor | M |
| <collection> | | údaje o větším celku (projektu), pokud do některého balíček patří - např. digitalizace pro ANL | R |
| <institution> | | název instituce, která je zadavatelem digitalizace | R |
| <creator> | | tvůrce balíčku – kód instituce (firmy), která balíček vytvořila | M |
| <size> | | velikost balíčku v kB - bez souboru info.xml | M |

| | | | |
|---|---|---|---|
| <itemlist> | | obsahuje seznam všech souborů v balíčku vč. souborů v podadresářích a souboru info.xml | M |
| | itemtotal | celkový počet souborů | M |
| <item> | | obsahuje cestu k jednomu souboru<br>např. "\txt\txt_nk-00027x_0001.txt"<br><br>element je opakován podle nutnosti - tj. podle počtu souborů | M |
| <checksum> | | odkaz na soubor s MD5 a jeho MD5 kontrolní součet | M |
| | type | bude vždy "MD5" | M |
| | checksum | kontrolní součet souboru - použít stejný algoritmus jako byl použit v MD5 souboru | M |
| | | odkaz na samotný soubor MD5 | M |
| <note> | | poznámka | O |

## 5.2 Složka pro původní neupravené zvukové soubory [sourceaudio]

Složka s původními zvukovými soubory (jedna strana až více stran desek ve formátu WAV), které byly případně poté upraveny.

## 5.3 Složka pro zvukové soubory [mastercopyaudio]

Složka s master kopiemi obsahuje archivní kopie všech zvukových souborů, např. ve formátu WAV. 1 soubor = 1 stopa (jedna strana až více stran desek ve formátu WAV). Bude se uchovávat pouze jeden druh master kopií i přesto, že může vzniknout více druhů master kopií v případě různých druhů úprav. V případě, že se nebude MC nijak lišit od SA, potom tato složka není povinná.

## 5.4 Složka pro zvukové soubory [usercopyaudio]

Složka s uživatelskými kopiemi, všechny uživatelské kopie např. ve formátu MP3. 1 soubor = 1 stopa (jedna strana desky).

## 5.5 Složka pro obrazové soubory [mastercopy]

Složka s archivními kopiemi všech obrazových souborů např. v bezztrátovém JPEG2000 (obal desek, obrazy fyzických nosičů, booklet – jednotlivé stránky). Produkce obrazových dat by se měla řídit doporučenými standardy pro digitalizaci (http://www.ndk.cz/standardy-digitalizace/standardy-pro-obrazova-data).

## 5.6 Složka pro obrazové soubory [usercopy]

Složka s uživatelskými kopiemi všech obrazových souborů

## 5.7 Složka [alto]

Obsahuje ke každé obrazové stránce 1 ALTO XML soubor, tj. tolik ALTO XML souborů, kolik je stránek patřících k jedné kolekci.

## 5.8 Složka [txt]

Obsahuje ke každé obrazové stránce 1 OCR soubor jako čistý text, tj. tolik OCR.TXT souborů, kolik je stránek patřících k jedné kolekci.

## 5.9 Složka [amdsec]

Složka s technickými metadaty musí obsahovat pro každý zvukový soubor i každou naskenovanou stránku 1 METS soubor amd_mets.xml. Každý soubor amd_mets.xml obsahuje následující části METS formátu:

- **amdSec** - administrativní metadata - obsahuje části:
    - technická metadata **techMD**, která:
        - ve formátu **PREMIS Object** popisují vlastnosti:
            - původního zvukového souboru (SA), ze kterého vznikly archivní zvukové kopie, např. soubory WAV
            - původního obrazového souboru (PS), ze kterého vznikly archivní obrazové kopie, např. soubory TIFF
            - archivní kopie zvukového souboru (není povinné v případě, že se MC nijak neliší od SA)
            - archivní kopie obrazového souboru
            - ALTO XML
            - případné jiné reprezentace - obraz i zvuk v případě opravy
        - ve formátu **AES57** popisují vlastnosti:
            - původního zvukového souboru
            - archivní kopie zvukového souboru (není povinné v případě, že se MC nijak neliší od SA)
        - ve formátu **MIX** popisují vlastnosti:
            - archivní kopie obrazu (veškeré generace archivních kopií)
            - případné jiné reprezentace stránky
    - metadata o provenienci digitálních objektů **digiprovMD** - obsahuje části:
        - PREMIS Event
        - PREMIS Agent
- **fileSec -** sekce s výčty jednotlivých souborů a odkazy na ně, povinná část METS záznamu, v případě tohoto METS záznamu pro jednu stránku, který vzniká primárně k zachycení technických a administrativních metadat, bude odkazovat na soubory souvisejícími s tou konkrétní stránkou, tj. archivní kopie zvuku, původní zvukové soubory nebo archivní kopie obrazu, ALTO XML a OCR TXT
- **structMap** - pouze fyzická strukturální mapa, povinná část METS záznamu, bude ukazovat strukturu souborů k dané stopě nebo stránce, tj. source audio, archivní kopie zvuku a obrazu, ALTO XML a OCR TXT
- **copyrightMD** - autorsko-právní metadata

## 5.10 Soubor Hlavni_METS.xml

Hlavní METS záznam bude obsahovat části:

- dmdSec – bibliografická metadata ke každé úrovni zvukového dokumentu ve formátu MODS a DC, včetně obrazových informací
- amdSec - administrativní a technická metadata pro každou zvukovou stopu a obrazovou stránku
- fileSec – výčet všech digitálních objektů s odkazy na ně
- structMap – strukturální mapa pro celý soubor (fyzická i logická část)
- structLink - vazby mezi fyzickou a logickou částí strukturální mapy (tedy mezi jednotlivými zvukovými stopami a stranou desky a mezi obrázky a deskou či kolekcí)

## 5.11 Soubor MD5

Soubor s kontrolním součtem souborů pro 1 SIP balíček. Také obsahuje kontrolní součet pro každý soubor obsažený v SIP balíčku (kromě info.xml a *.md5 souboru samotného).

Obsah souboru je definován následující gramatikou ve formátu ABNF[1]

| | |
|---|---|
| SOUBOR | = *RADEK |
| RADEK | = HODNOTA-MD5 MEZERA NAZEV-SOUBORU KONEC-RADKU |
| HODNOTA-MD5 | = 32HEXDIG |
| MEZERA | = " " / TAB |
| NAZEV-SOUBORU | = *SEGMENT |
| KONEC-RADKU | = ( CR LF ) / LF |
| SEGMENT | = PATH-SEP FILENAME-CHARS |
| PATH-SEP | = *"/" / "\"* |
| TAB | = *%x09* |
| CR | = *%x0D* |
| LF | = *%x0A* |
| FILENAME-CHAR | = ALPHA / DIGIT / *"."* / "_" / "-" |
| FILENAME-CHARS | = 1*FILENAME-CHAR |

kde sémantika je následující:

- HODNOTA-MD5 - produkce tohoto pravidla reprezentuje výsledek výpočtu MD5 hashovací funkce, do které vstupuje obsah souboru označeného jménem souboru NAZEV-SOUBORU
- NAZEV-SOUBORU - produkce pravidla tvoří jméno souboru ve formě absolutní cesty v hierarchické struktuře balíku SIP. Cesta je absolutní vůči kořenové složce SIP balíčku

[1] http://tools.ietf.org/html/rfc4234

# 6 Názvová konvence složek a souborů

V případě použití např. čárového kódu jako identifikátoru bude kód sloužit k jednoznačné identifikaci jedné kolekce desek či desky. Bude také tvořit základ pojmenování souborů. Stejný postup platí i v případě použití jiných strojově čitelných identifikátorů.

**PojmenováníSIP balíčku**

- každý SIP balíček přicházející z digitalizace musí obsahovat pouze jedinou intelektuální entitu (kolekci gramodesek o počtu jedné a více stran).
- název balíčku **musí vycházet** z identifikátoru této entity, tj. URN:NBN nebo UUID
- každá kolekce musí mít svůj jednoznačný identifikátor, tím pádem pak každý SIP balíček a každý soubor v něm má vlastní jednoznačný identifikátor díky využití prefixů
- názvy nesmí obsahovat mezery a diakritiku, doporučené oddělovače jsou podtržítko a pomlčka

**Pojmenování složek**

- viz návrh struktur SIP balíčku (mastercopyaudio, usercopyaudio atp.)
- názvy nesmí obsahovat mezery, diakritiku a dvojtečku, doporučené oddělovače jsou podtržítko a pomlčka.

Př.
Identifikátor URN:NBN přidělený dokumentu vypadá takto: **urn:nbn:cz:nk-00027x.** Z něho se použije pouze sufix **nk-00027x,** takže složka je nazvaná malými písmeny bez diakritiky a dvojteček takto: **nk-00027x.**

**Pojmenování souborů**

V případě použití čárového kódu:

- zvukové soubory - pojmenované jako "druh-souboru_čárový-kód_XXXX(_YYYY).wav", kde druh souboru je “sa”, “mca” nebo “uca”, XXXX je pořadové číslo strany desky (a YYYY je pořadové číslo stopy na desce; toto není povinné, pokud rozdělení desky na stopy nebude aplikováno), př.:
    - sa_1234567890_0001.wav, sa_1234567890_0002.wav
    - mca_1234567890_0001.wav, mca_1234567890_0002.wav
    - uca_1234567890_0001.wav, uca_1234567890_0002.wav
    - anebo v případě rozdělení na stopy: uca_1234567890_0001_0001.mp3, uca_1234567890_0001_0002.mp3, uca_1234567890_0001_0003.mp3, uca_123456789_0001_0004.mp3, uca_1234567890_0002_0001.mp3, uca_1234567890_0002_0002.mp3, uca_1234567890_0002_0003mp3, uca_123456789_0002_0004.mp3

- **obraz strany desky** - skeny označené jako "druh-souboru_čárový_kód_XXXX.jp2", kde druh souboru je “mc” nebo “uc” a XXXX je pořadové číslo skenu; stejné pojmenování ponese i uživatelská kopie
    - mc_1234567890_0001.jp2

    - mc_1234567890_0002.jp2
    - uc_1234567890_0001.jp2
    - uc_1234567890_0002.jp2

- **obraz obalu kolekce** - skeny označené jako "druh-souboru_čárový_kód_XXXX.jp2", kde druh souboru je “mc” nebo “uc”, XXXX je pořadové číslo skenu; stejné pojmenování ponese i uživatelská kopie
    - mc_1234567890_0003.jp2
    - uc_1234567890_0003.jp2

- **digitalizovaný booklet** - bude tvořen jednotlivými skeny označenými jako “druh-souboru_čárový_kód_XXXX.jp2”, kde druh souboru je “mc” nebo “uc”, XXXX je pořadové číslo skenu; stejné pojmenování ponese i uživatelská kopie
    - mc_1234567890_0004.jp2
    - mc_1234567890_0005.jp2
    - uc_1234567890_0004.jp2
    - uc_1234567890_0005.jp2

- soubor ALTO - bude mít označení “alto_čárový_kód_XXXX.xml, kde “alto” určuje druh souboru a XXXX pořadové číslo xml souboru
    - alto_1234567890_0001.xml

- soubor OCR TXT - označení “txt_čárový_kód_XXXX.txt, kde “txt” určuje druh souboru a XXXX pořadové číslo txt souboru
    - txt_1234567890_0001.txt

- soubor MD5 - označení “md5_čárový_kód_XXXX.md5”, kde “md5” určuje druh souboru a XXXX určuje pořadové číslo md5 souboru
    - md5_1234567890_0001.md5

Názvy jakýchkoliv souborů náležející k jedné intelektuální entitě musí být založeny na jednom typu identifikátoru. Názvy musí být pouze malými písmeny a názvy nesmí obsahovat mezery ani diakritiku, doporučené oddělovače jsou podtržítko a pomlčka.

# 7 Metadata

- veškerá metadata budou „zabalena“ pomocí kontejnerového formátu METS
- vložení metadatových formátů do kontejneru METS musí být vždy formou <mdWrap>

## 7.1 Kořenový element hlavního METS záznamu

Kořenový element hlavního METS záznamu musí obsahovat linky na specifikace jednotlivých použitých metadatových schémat (METS, MODS, Dublin Core, CopyrightMD).

| element | atributy | popis | Povinnost |
|---|---|---|---|
| <mets> | | kořenový element METS záznamu | M |
| | LABEL | název zvukové kolekce, včetně roku vydání (název, rok vydání) | M |
| | TYPE | hodnota vždy “sound recording” | M |

## 7.2 METS hlavička <metsHdr>

Dokumentuje vznik a úpravy METS záznamu.

| element | atributy | popis | Povinnost |
|---|---|---|---|
| <metsHdr> | | hlavička METS záznamu | M |
| | LASTMODDATE | LASTMODDATE – datum poslední úpravy záznamu, musí být ve tvaru ISO 8601 (na úroveň vteřin) | M |
| | CREATEDATE | CREATEDATE – datum vytvoření záznamu, musí být ve tvaru ISO 8601 (na úroveň vteřin) | M |
| <agent> | | údaje o tvůrci záznamu METS | M |
| | ROLE | hodnota “creator” | M |
| | TYPE | hodnota “organization” | M |
| <name> | | jméno jednotlivce nebo organizace;<br>tvůrce záznamu, buď dodavatel (firma XY) nebo v případě tvorby záznamu v knihovně bude využita sigla knihoven, tj. pro NK ČR hodnota "ABA001" | M |
| <agent> | | údaje o vlastníkovi METS | M |

| | | | |
|---|---|---|---|
| | ROLE | hodnota “archivist” | M |
| | TYPE | hodnota “organization” | M |
| <name> | | jméno organizace;<br>vlastník záznamu, v případě tvorby záznamu v knihovně bude vždy využita sigla knihoven, tj. pro NK ČR hodnota "ABA001" | M |

## 7.3 METS část <dmdSec> - Bibliografická metadata – MODS a Dublin Core

- MODS a DC budou vloženy v METS části <dmdSec>
- z MARC 21 resp. MARCXML do MODS lze využít a dle potřeb knihovny upravit oficiální mapování a konverzní šablonu ze stránek http://www.loc.gov/standards/mods/mods-conversions.html
- pro vytvoření DC z MODS formátu musí být použito (a dle potřeb knihovny upraveno) oficiální mapování Kongresové knihovny – viz http://www.loc.gov/standards/mods/mods-conversions.html
- U digitalizovaných dokumentů je bibliografický popis vytvářen primárně z pohledu popisu fyzické předlohy, nejde o popis elektronického dokumentu

**Zvukový dokument - gramofonová deska**
**Kolekce gramofonových desek (1 a více desek)**

- základní intelektuální entitou pro popis je kolekce gramofonových desek, tj. v jednom METS záznamu, který bude obsahovat metadata a strukturu jedné a více desky, budou MODS záznamy k této kolekci
- kolekcí se rozumí vrcholná úroveň popisu gramofonové desky, ať již se týká desky jen s jednou stranou nebo souboru mnoha desek. Kolekce odpovídá bibliografickému záznamu v katalogu. V případě kolekce obsahující jen jednu jednostrannou desku bude záznam pro side přebírat z vyšší úrovně vše, co je možné.
- metadata budou popisovat entity[2]:

    - **kolekce (collection)** - soubor jedné či více gramodesek v jednom obalu nebo ve více obalech, které jsou popsány jedním bibliografickým záznamem v katalogu a jejichž obsah spolu souvisí nebo na sebe navazuje. Collection je nadřazená entita pro všechny následující.
    - **strana desky (side)** - popisuje každou stranu desek z kolekce zvlášť. Každá strana se odkazuje pomocí elementu <relatedItem> na disk, na kterém se nachází. Disk sám ovšem netvoří samostatnou úroveň, slouží pouze jako reference k fyzickému nosiči.
    - **zvuková stopa (track)** - popisuje každou zvukovou stopu na straně, každou samostatně. Stopy se odkazují pomocí elementu <relatedItem> ke straně desky, ke které náleží.
    - **příloha (supplement)** - přílohou se rozumí volně vložená entita k intelektuální entitě, např. volně vložený booklet, obsah, reklamní či edukativní příloha apod.

[2] toto pořadí nevyjadřuje logickou strukturu dokumentu, ta je popsána dále ve Specifikaci

Rozlišujeme tři druhy příloh:

1) **obal (cover)** - popis pro fyzický obal kolekce. Pomocí elementu <relatedItem> se odkazuje ke kolekci.
2) **booklet** - popis pro doprovodný text v podobě knížečky či obrazovou přílohu, volně vloženou či sešitou s obalem. Pomocí elementu <relatedItem> se odkazuje k celé kolekci.
3) **obraz desky (imgdisc)** - popis pro sken strany gramodesky. Pomocí elementu <relatedItem> se odkazuje ke straně desky, jejíž je skenem.

Tyto druhy příloh pak budou zohledněny i v rámci fyzické mapy dokumentu v atributu “type”.

Obecná pravidla pro bibliografická metadata:

- pro každou entitu vznikne jeden MODS záznam s vlastním ID, které bude označovat i typ části (např. side, track, supplement). V případě opakování částí stejného typu se bude opakovat tomu odpovídající počet MODS záznamů v jednom SIP balíčku.
- každý MODS záznam má vlastní <dmdSec> část
- každý MODS záznam bude uložen ve vlastní METS části <dmdSec> pomocí mdWrap
- všechny top elementy MODS formátu jsou opakovatelné, kromě <recordInfo>
- všechny elementy Dublin Core jsou opakovatelné
- každá část <dmdSec> musí mít ID a vnořený element <mdWrap> s atributy MDTYPE a MIMETYPE.

| **element** | **atributy** | **popis** | **povinnost** |
|---|---|---|---|
| <dmdSec> | | identifikátor <dmdSec> části METS záznamu | M |
| | ID | pro <dmdSec> s popisem zvukové kolekce hodnota “MODSMD_COLLECTION” a “DCMD_COLLECTION” | M |
| <mdWrap> | | element obsahující vložené záznamy MODS | M |
| | MDTYPE | MDTYPE – hodnota "MODS" pro záznamy v MODS, hodnota "DC" pro záznam v Dublin Core | M |
| | MDTYPEVERSION | číslo verze MODS, hodnota pro záznamy v MODS, např. "3.5" | O |
| | MIMETYPE | MIMETYPE – hodnota "text/xml" | O |
| <xmlData> | | | M |

## Pole MODS a Dublin Core pro jednotlivé úrovně zvukových dokumentů

- ID u elementu <mods>: Identifikátory budou začínat prefixy MODS_COLLECTION, MODS_SIDE, MODS_TRACK, MODS_SUPPL. Obdobně pro DC. Za ty se dále přidá podtržítko a číslo identifikující pořadí identifikátoru, zarovnané a doplněné o nuly na 4 místa (kvůli jednotnosti s ostatními standardy). ID tedy vypadá následovně:

  1) kolekce (v případě jednodiskové i vícediskové kolekce):
     - MODS_COLLECTION_0001
     - DC_COLLECTION_0001

  2) strana:
     - MODS_SIDE_0001
     - DC_SIDE_0001

  3) zvuková stopa:
     - MODS_TRACK_0001
     - DC_TRACK_0001

  4) přílohy (skeny obalu a disků, případně booklet):
     - MODS_SUPPL_XXXX, kde XXXX je pořadové číslo přílohy, např. MODS_SUPPL_0001 je první příloha atd.
     - DS_SUPPL_XXXX, kde XXXX je pořadové číslo přílohy, např. MODS_SUPPL_0001 je první příloha atd.

### 7.3.1. Pole MODS a DC pro úroveň kolekce/souboru

**Obsah sloupce "Popis":**

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné uvést
- povinnost platí pro elementy MODS i pro elementy Dublin Core stejně
- pokud je rodičovský element např. doporučený, a dceřiný element povinný, znamená to, že dceřiný element je povinný pouze tehdy, pokud je použit element rodičovský

| Element MODS | Atributy | Popis | povinnost | Element DC |
|---|---|---|---|---|
| **<mods>** | ID | ID musí vyjadřovat název úrovně, tj. pro soubor gramodesek je to "MODS_COLLECTION_0001" | M | |
| **<titleInfo>** | | Název kolekce, souborný název (pro plnění lze použít katalogizační záznam), samozřejmě lze využít všech prvků a elementů MODS, které názvové informace popisují)<br>opakovatelný element | M | |
| | type | hlavní název bez *type* - pole 245 a $a<br>Může existovat pouze jeden hlavní název bez *type*.<br>type: alternative - pole 246<br>translated - pole 242<br>uniform - pole 130, resp. 240 | MA | |
| <title> | | názvová informace | M | <dc:title> |
| <subTitle> | | podnázev | MA | <dc:title> |
| **<name>** | | údaje o odpovědnosti<br>přebírá se z 1XX a 7XX MARC21<br>opakovatelný element pro více autorů/různé role | MA | |

| | | | | |
|---|---|---|---|---|
| | type | Použít jednu z hodnot:<br>personal<br>corporate<br>conference<br>family | | |
| <namePart> | | Údaje o křestním jméně a příjmení apod. nutno vyjádřit pro křestní jméno i příjmení. Pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě, jaké je, do jednoho elementu <namePart>.<br>Pokud známe datum narození a úmrtí autora, vyplnit ve tvaru RRRR-RRRR s atributem type=”date”. | MA | <dc:creator> nutno do jednoho pole DC spojit jméno i příjmení |
| | type | Použít jednu z hodnot:<br>date<br>family<br>given<br>termsOfAddress | <br>RA<br>MA<br>MA<br>RA | |
| <role> | | specifikace role osoby nebo organizace uvedené v elementu <name> | MA | |
| <roleTerm> | | | MA | |
| | type | hodnota "code" – kód role z kontrolovaného slovníku (http://www.loc.gov/marc/relators/relaterm.html) | | |
| | authority | authority - údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator" | | |
| **<typeOfResource>** | | Přebírá se ze záznamu (Leader/06 - i, j) - "sound recording", "sound recording-musical", "sound recording-nonmusical" | R | <dc:type> |

| | | | | |
|---|---|---|---|---|
| **<genre>** | | bližší údaje o typu dokumentu (dle mapování LoC by zde měla být převedená z pole 655 a 336 v MARC21, ale je možné použít jednu stanovenou hodnotu)<br>stanovená hodnota "sound recording"<br>bez atributu | MA | <dc:type>model:soundrecording</dc:type> |
| **<originInfo>** | | informace o původu předlohy | M | |
| <place> | | údaje o místě spojeném s vydáním, výrobou nebo původem popisovaného dokumentu | MA | <dc:coverage> |
| <placeTerm> | | konkrétní určení místa vydání, např. Praha, resp. "xr" pro ČR<br>odpovídá hodnotám z katalogizačního záznamu, pole 260 $a, resp. pole 008/15-17 | MA | <dc:coverage> |
| | type | "code" pro údaj z pole 008<br>"text" pro údaj z pole 260 | M | |
| | authority | authority - hodnota "marccountry" jen u údaje z pole 008 | MA | |
| <publisher> | | jméno entity, která dokument vydala, vytiskla nebo jinak vyprodukovala, odpovídá poli 260 $b v MARC21<br>pokud má titul více vydavatelů, přebírají se ze záznamu všichni (jsou v jednom poli 260) | MA | |
| <dateIssued> | | datum vydání předlohy, odpovídá hodnotě z katalogizačního záznamu, pole 260 $c a pole 008/07-10=publication | M | <dc:date> |
| | encoding | hodnota "marc" jen u údaje z pole 008 | R | |

| | | | | |
|---|---|---|---|---|
| | point | hodnoty "start", resp. "end" jen u údaje z pole 008, pro rozmezí dat | MA | |
| | qualifier | možnost dalšího upřesnění, hodnota "approximate" pro data, kde nevíme přesný údaj | R | |
| <edition> | | údaj o pořadí vydání,<br><br>odpovídá poli 250 $a katalogizačního záznamu. | R | |
| <issuance> | | údaje o vydávání odpovídá hodnotě uvedené v návěští MARC21 na pozici 07<br>hodnota "monographic" | M | |

**PRO ZÁZNAMY, KTERÉ JSOU ZPRACOVANÉ DLE PRAVIDEL RDA**

Změny v sekci **mods:originInfo**

| | | | | |
|---|---|---|---|---|
| **<originInfo>** | | informace o původu předlohy: odpovídá poli 264 | M | |
| | eventType | hodnoty dle druhého indikátoru pole 264:<br>264_0 "production" (MA)<br>hodnota 0 se uvádí, jestliže pole obsahuje údaje o vytvoření zdroje v nezveřejněné podobě<br>264_1 "publication" (M)<br>hodnota 1 se uvádí, jestliže pole obsahuje údaje o nakladateli zdroje<br>264_2 "distribution" (R)<br>hodnota 2 se uvádí, jestliže pole obsahuje údaje o distribuci zdroje<br>264_3 "manufacture" (R)<br>hodnota 3 se uvádí, jestliže pole obsahuje údaje o tisku, výrobě zdroje ve zveřejněné podobě | | |

| | | | | |
|---|---|---|---|---|
| | | 264_4 "copyright" (R)<br>hodnota 4 se uvádí, jestliže pole obsahuje údaje o ochraně podle autorského práva (copyright)<br>Pro minimální záznam pro zvukové záznamy jsou povinným údajem pouze údaje o nakladateli, tedy pole 264 s druhým indikátorem 1. | | |
| <place> | | údaje o místě spojeném s vytvořením, vydáním, distribucí nebo výrobou popisovaného dokumentu<br>odpovídá hodnotě 264 $a | MA | <dc:coverage> |
| <placeTerm> | | konkrétní určení místa vydání, např. Praha, resp. "xr" pro ČR<br>odpovídá hodnotám z katalogizačního záznamu, pole 264 $a, resp. pole 008/15-17 | MA | <dc:coverage> |
| | type | pokud má dokument více míst vytvoření/vydání/distribuce/výroby v poli 264 $a, přebírají se ze záznamu všechna místa (v jednom poli 264)<br>"code" pro údaj z pole 008<br>"text" pro údaj z pole 264 $a | M | |
| | authority | authority - hodnota "marccountry" jen u údaje z pole 008 | MA | |
| <publisher> | | jméno entity, která dokument vytvořila, vydala, distribuovala nebo vyrobila<br>odpovídá poli 264 $b v MARC21<br>pokud má titul více vydavatelů,/ distributorů/ výrobců, přebírají se ze záznamu všichni (jsou v jednom poli 264) | MA | <dc:publisher> |
| <dateIssued> | | datum vydání předlohy, odpovídá hodnotě z katalogizačního záznamu, pole 264_1 $c a pole 008/07-10<br>!! pro všechny ostatní výskyty v poli 264 $c:<br>264_0 "production" | M | <dc:date> |

| | | | | |
|---|---|---|---|---|
| | | 264_2 "distribution"<br>264_3 "manufacture"<br>264_4 "copyright"<br>využít element <dateOther> s odpovídajícím atributem “type” či element <copyrightDate> | | |
| | encoding | hodnota "marc" jen u údaje z pole 008 | R | |
| | point | hodnoty "start", resp. "end" jen u údaje z pole 008, pro rozmezí dat | M | |
| | qualifier | možnost dalšího upřesnění, hodnota "approximate" pro data, kde nevíme přesný údaj | R | |
| <dateOther> | | datum vytvoření, distribuce, výroby předlohy tento element se využije v případě výskytu $c v:<br>264_0 "production"<br>264_2 "distribution" | R | |
| | type | 264_0: <dateOther type="production"><br>264_2: <dateOther type="distribution"> | M | |
| <copyrightDate> | | využije se pouze v případě výskytu pole 264 s druhým indikátorem 4 a podpolem $c<br>264_4: <copyrightDate> | R | <dc:date> |

| | | | | |
|---|---|---|---|---|
| **<language>** | | údaje o jazyce dokumentu<br>v případě vícenásobného výskytu nutno element <language> opakovat | M | |
| | objectPart | možnost vyjádřit jazyk konkrétní části svazku;<br>možné hodnoty:<br>“summary” - odpovídá poli 041 $b<br>“table of contents” - odpovídá poli 041 $f<br>“accompanying material” - odpovídá poli 041 $g<br>“translation” - odpovídá poli 041 $h | MA | |

| | | | | |
|---|---|---|---|---|
| <languageTerm> | | přesné určení jazyka – kódem; nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php | M | <dc:language> |
| | type | použít hodnotu code | M | |
| | authority | použít hodnotu „iso639-2b“ | M | |
| **<physicalDescription>** | | obsahuje údaje o fyzickém popisu zdroje/předlohy (pole 300, 337, 338) | M | |
| <form> | | údaje o fyzické podobě dokumentu, např. sound recording<br>+<br>Pro záznamy podle RDA:<br>údaje o typu média a typu nosiče zdroje/předlohy<br>odpovídá hodnotám z pole:<br>337 NEPOVINNÉ (hodnota “audio” - viz kontrolovaný slovník pole 337 http://www.loc.gov/standards/valuelist/rdamedia.html)<br>338 POVINNÉ (hodnota “audiodisc” - viz kontrolovaný slovník pole 338 http://www.loc.gov/standards/valuelist/rdacarrier.html) | M | <dc:format> |
| | authority | authority: hodnota “marccategory” nebo “gmd” obsahuje „sound recording“<br>Pro záznamy v RDA, hodnoty:<br>pole 337: authority=”rdamedia” (hodnota audio)<br>pole 338: authority=”rdacarrier” (hodnota audiodisc) | M/MA | |
| | type | pouze pro záznamy v RDA<br>pole 337: type=”media” (audio)<br>pole 338: type=”carrier” (audiodisc) | MA | |

| | | | | |
|---|---|---|---|---|
| <extent> | | údaje o rozsahu<br>odpovídá hodnotám v poli 300 $a, $b, $c, $e, popřípadě 685, resp. 655<br>např.<br>$a 2 LP desky<br>$b 33 ot/min<br>$c 30 cm<br>$e 1 příl. [14 s.].<br>(rozdělit do více opakovaných prvků <extent>)<br>+<br>dle RDA pole 344 (zvukové charakteristiky)<br>$a druh záznamu (digital, analog)<br>$c přehrávací rychlost (počet otáček za minutu)<br>$g konfigurace přehrávacích kanálů (mono, stereo, kvadro, surround)<br>$h speciální přehrávací charakteristiky | M | <dc:coverage> |
| <note> | | poznámka o celkovém fyzickém stavu kolekce; pro každou poznámku je nutno vytvořit nový <note> element.<br><br>případná poškození desek vepisovat do <physicalDescription> na konkrétní úrovni SIDE | RA | |
| **<tableOfContents>** | | seznam názvů desek v kolekci | R | <dc:description> |
| **<note>** | | obecná poznámka k dokumentu jako celku; odpovídá hodnotám v poli 245 $c (statement of responsibility) a v polích 5XX (poznámky) katalogizačního záznamu | RA | <dc:description> |
| **<subject>** | | údaje o věcném třídění, předpokládá se přebírání z katalogizačního záznamu | R | |

| | | | | |
|---|---|---|---|---|
| | authority | vyplnit hodnotu podle zdroje záznamu, např. „czenas“, „eczenas“ nebo nechat prázdné | | |
| <topic> | | libovolný výraz specifikující nebo charakterizující obsah titulu; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) nebo obsah pole 650 záznamu MARC21 nebo obsah pole 072 $x | M | <dc:subject> |
| <geographic> | | geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) nebo obsah pole 651 záznamu MARC21 | R | <dc:subject> |
| <temporal> | | chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) nebo obsah pole 648 záznamu MARC21 | R | <dc:subject> |
| <name> | | jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) nebo obsah pole 600 záznamu MARC21; struktura a atributy stejné jako pro údaje o původcích – viz element <name> | R | <dc:subject> |
| **<classification>** | | klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění odpovídá poli 080 MARC21<br><br>klasifikační údaje věcného třídění podle Konspektu odpovídá poli 072 $a MARC21 | R | <dc:subject> |
| | authority | vyplnit hodnotu “udc” (v případě 072 $a)<br>vyplnit hodnotu “Konspekt” (v případě 072 $9) | M | |
| | edition | vyplnit hodnotu “Konspekt” (v případě 072 $a) | M | |
| **<relatedItem>** | | informace o dalších dokumentech, jejich částech či zdrojích, které jsou ve vztahu k popisované kolekci; | RA | |

| | | | | |
|---|---|---|---|---|
| | | dále lze použít pro vyjádření edice, ve které je dokument vydán (tzn. kolekce je součástí nějaké edice), údaj o edici musí obsahovat minimálně element <title> s jejím názvem<br><br>v případě převzatých písní může obsahovat údaje o originálu.<br><br>poznámka: element <relatedItem> může obsahovat jakýkoliv jiný element MODS – jejich použití se řídí pravidly popsanými pro tyto elementy; | | |
| | type | hodnota „series“ | | |
| **<identifier>** | | údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální – viz přehled typů atributů níže | M | <dc:identifier> |
| | type | budou se povinně vyplňovat následující hodnoty, pokud existují:<br>- „issue number“ – číslo desky, v MARC21 pole 028, např. $a DV 6196, 6228 $b Supraphon, přebírá se z katalogu<br>- uuid (M)<br>- urn:nbn (M)<br>- jiný interní identifikátor, hodnota atributu „local“, lze použít k vyjádření čárového kódu (MA) | MA | |
| **<location>** | | údaje o uložení popisovaného dokumentu, např. signatura, místo uložení apod. | MA | |
| <physicalLocation> | | údaje o instituci, kde je fyzicky uložen popisovaný dokument, např. MZK; nutno použít kontrolovaný slovník – sigly knihoven (BOA001 atd.); odpovídá poli 040 v MARC21<br>pozn. u dokumentů ze soukromých sbírek není možné vyplnit | MA | <dc:source> |
| | authority | Hodnota „sigla ADR“ | | |

| | | | | |
|---|---|---|---|---|
| <shelfLocator> | | signatura nebo lokační údaje o dokumentu, který slouží jako předloha | M | <dc:source> |
| **<recordInfo>** | | údaje o metadatovém záznamu – jeho vzniku, změnách apod. | M | |
| <descriptionStandard> | | popis standardu, ve kterém je přebíraný katalogizační záznam<br>pro záznamy v AACR2:<br>odpovídá hodnotě návěští záznamu MARC21, pozice 18 - hodnota “aacr”, tj. pro LDR/18=”a”<br>pro záznamy v RDA:<br>odpovídá hodnotě záznamu MARC21 pole 040 a podpole $e “rda” - bude použito pro primární rozlišení, zda jde o záznam vytvořený podle AACR2 nebo podle RDA | MA | |
| <recordContentSource> | | kód nebo jméno instituce, která záznam vytvořila nebo změnila | R | |
| <recordCreationDate> | | datum prvního vytvoření záznamu, na úroveň minut podle normy ISO 8601 | M | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordChangeDate> | | datum změny záznamu, podle normy ISO 8601 na úroveň minut | R | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordOrigin> | | údaje o vzniku záznamu; hodnoty: “machine generated” nebo “human prepared” | R | |
| <languageOfCataloging> | | jazyk katalogového záznamu | R | |
| <languageTerm> | | přebírá se z katalogu - pole 40 $b | R | |
| | authority | hodnota “iso639-2b” | R | |

### 7.3.2 Pole MODS a DC pro úroveň jedné strany desky zvukového nosiče

| Element MODS | Atributy | Popis | Povinnost | Element DC |
|---|---|---|---|---|
| **<mods>** | ID | obsahuje název úrovně, pro soubor gramodesek je to "MODS_SIDE_0001", kde číslo je pořadové číslo strany desky. (Pokud je v souboru 5 desek, stran je 10.) | M | |
| **<titleInfo>** | | názvové informace - pokud nemá strana desky svůj samostatný název, kopírují se informace o celé kolekci | M | |
| <title> | | hlavní název | M | <dc:title> |
| <subTitle> | | podnázev | MA | |
| <partNumber> | | číslo strany ve formátu XX, 01, 02 atd. | M | <dc:title> |
| <partName> | | slovní označení desky, jak je vytištěno na desce, např. "side 1", "strana A" apod. | MA | |
| **<genre>** | type | type="model" obsahuje hodnotu pole "side" | M | |
| **<note>** | | obecná poznámka ke straně desky | RA | <dc:description> |
| **<relatedItem>** | | popis desky, jejíž součástí je konkrétní strana desky | M | |
| | type | „host“ | M | |
| <titleInfo> | | názvové informace | M | |

| | | | | |
|---|---|---|---|---|
| <title> | | název desky (případně se přebírají informace z kolekce) | O | |
| <subTitle> | | podnázev desky | O | |
| <partNumber> | | číslo desky v kolekci ve formátu XX<br>01, 02, 03 atd. | M | |
| <partName> | | slovní označení čísla desky, např. "Disc 1", podle toho, jak je vytištěno na desce | MA | |
| <genre> | type | type=”model“ obsahuje hodnotu pole „disc“ | | |
| <identifier> | | identifikátor desky | M | |
| | type | type="issue number" - katalogové číslo desky, v MARC21 pole 028 $a, např. DV 6196, 6228<br>type=“music-publisher“, např. $b Supraphon<br>type=“uuid“ - uuid desky | MA<br>MA<br>R | |
| **<identifier>** | | údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální | M | |
| | type | budou se povinně vyplňovat následující hodnoty, pokud existují:<br>type=“matrix-number“<br>type=“uuid“ | MA<br>M | |
| **<recordInfo>** | | údaje o metadatovém záznamu strany desky – jeho vzniku, změnách apod. | M | |
| <descriptionStandard> | | popis standardu, ve kterém je přebíraný katalogizační záznam<br>pro záznamy v AACR2:<br>odpovídá hodnotě návěští záznamu MARC21, pozice 18 - hodnota “aacr”, tj. pro LDR/18=”a”<br>pro záznamy v RDA: | MA | |

| | | | | |
|---|---|---|---|---|
| | | odpovídá hodnotě záznamu MARC21 pole 040 a podpole $e “rda” - bude použito pro primární rozlišení, zda jde o záznam vytvořený podle AACR2 nebo podle RDA | | |
| <recordContentSource> | | kód nebo jméno instituce, která záznam vytvořila nebo změnila | R | |
| <recordCreationDate> | | datum prvního vytvoření záznamu, na úroveň minut podle normy ISO 8601 | M | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordChangeDate> | | datum změny záznamu, podle normy ISO 8601 na úroveň minut | R | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordOrigin> | | údaje o vzniku záznamu; hodnoty: “machine generated” nebo “human prepared” | R | |
| <languageOfCataloging> | | jazyk katalogového záznamu | R | |
| <languageTerm> | | přebírá se z katalogu - pole 40 $b | R | |
| | authority | hodnota “iso639-2b” | R | |

### 7.3.3 Pole MODS a DC pro úroveň tracku neboli stopy

| Element MODS | Atributy | Popis | Povinnost | Element DC |
|---|---|---|---|---|
| **<mods>** | ID | ID obsahuje název úrovně, pro soubor gramodesek je to "MODS_TRACK_0001", kde | M | |

| | | | | |
|---|---|---|---|---|
| | | čtyřmístné číslo je pořadové číslo tracku na desce | | |
| **<titleInfo>** | | názvové informace o stopě | M | |
| | type | hlavní název bez *type*<br>type:<br>alternative<br>translated<br>uniform | MA | |
| <title> | | hlavní název | M | <dc:title> |
| <subTitle> | | podnázev | MA | |
| <partNumber> | | číslo stopy, 01,02 atd. | MA | <dc:title> |
| <partName> | | slovní označení stopy, např. "track 01" apod. | MA | |
| **<name>** | | autor stopy | RA | |
| | type | Použít jednu z hodnot:<br>personal<br>corporate<br>conference<br>family | | |
| <namePart> | | údaje o křestním jméně a příjmení;<br><br>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se atribut type, a jméno se zaznamená v podobě, jaké je, do jednoho elementu <namePart> | MA | <dc:creator> nutno do jednoho pole DC spojit jméno i příjmení |
| | type | Použít jednu z hodnot: | RA | |

| | | | | |
|---|---|---|---|---|
| | | date<br>family<br>given<br>termsOfAddress | MA<br>MA<br>RA | |
| <role> | | specifikace role osoby nebo organizace uvedené v elementu <name> | MA | |
| <roleTerm> | | | MA | |
| | type | hodnota "code" – kód z kontrolovaného slovníku | M | |
| | authority | hodnota "marcrelator“ – popis role dle slovníku MARC relator | R | |
| **<typeOfResource>** | | vyplnit „sound recording“ | M | <dc:type> |
| **<genre>** | type | type="model" obsahuje hodnotu pole "track" | M | |
| **<language>** | | údaje o jazyce dokumentu | M | |
| <languageTerm> | | přesné určení jazyka – kódem; nutno použít kontrolovaný slovník ISO 639-2,<br>http://www.loc.gov/standards/iso639-2/php/code_list.php | M | <dc:language> |
| | type | použít hodnotu “code” | | |
| | authority | použít hodnotu „iso639-2b“ | | |
| **<physicalDescription>** | | obsahuje údaje o fyzickém popisu zdroje/předlohy | M | |
| <form> | | | M | <dc:format> |

| | | | | |
|---|---|---|---|---|
| | authority | hodnota “marccategory“ obsahuje "sound recording" | | |
| <extent> | | stopáž | M | <dc:coverage> |
| **<relatedItem>** | | popis strany desky, na níž je konkrétní stopa | M | |
| | type | „host“pro stranu desky | M | |
| <titleInfo> | | názvové informace | M | |
| <title> | | název strany desky | O | |
| <subTitle> | | podnázev | O | |
| <partNumber> | | číslo strany desky ve formátu XX<br>01, 02 atd. | M | |
| <partName> | | slovní označení pořadí strany desky, jak je vytištěno na desce<br>např. "side 1", "strana A" apod. | MA | |
| <genre> | type | type=”model“ obsahuje hodnotu pole „side“ | | |
| <identifier> | | identifikátor strany desky | M | |
| | type | type=“uuid“ - uuid desky | M | |
| **<relatedItem>** | | popis předlohy či originálu skladby | RA | |
| | type | “original” pro údaje o původní předloze | M | |
| <titleInfo> | | názvové informace | M | |
| <title> | | název originální skladby | M | |

| | | | | |
|---|---|---|---|---|
| <subTitle> | | podnázev | MA | |
| <partNumber> | | pořadové číslo stopy na straně desky ve formátu XX<br>01, 02 atd. | M | |
| <name> | | údaje o odpovědnosti | RA | |
| | type | Použít jednu z hodnot:<br>personal<br>corporate<br>conference<br>family | | |
| <namePart> | | údaje o křestním jméně a příjmení apod. nutno vyjádřit pro křestní jméno i příjmení.<br>Pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě, jaké je, do jednoho elementu <namePart>.<br>Pokud známe datum narození a úmrtí autora, vyplnit ve tvaru RRRR-RRRR s atributem type=”date”. | M | |
| | type | Použít jednu z hodnot:<br>date<br>family<br>given<br>termsOfAddress | | |
| <role> | | specifikace role osoby nebo organizace uvedené v elementu <name> | MA | |
| <roleTerm> | | | MA | |
| | type | hodnota "code" – kód role z kontrolovaného slovníku | | |

| | | | | |
|---|---|---|---|---|
| | | (http://www.loc.gov/marc/relators/relaterm.html) | | |
| | authority | authority - údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator" | | |
| <identifier> | | identifikátor tracku, ke kterému se popis vztahuje | MA | |
| | type | type="uuid" - uuid tracku | M | |
| **<note>** | | obecná poznámka k tracku | RA | <dc:description> |
| **<identifier>** | | identifikátor tracku, obsahuje unikátní identifikátory mezinárodní nebo lokální | M | <dc:identifier> |
| | type | budou se povinně vyplňovat následující hodnoty, pokud existují:<br>uuid<br>jiný interní identifikátor - hodnota atributu „local“ | M<br>MA | |
| **<recordInfo>** | | údaje o metadatovém záznamu – jeho vzniku, změnách apod. | M | |
| <descriptionStandard> | | popis standardu, ve kterém je přebíraný katalogizační záznam<br>pro záznamy v AACR2:<br>odpovídá hodnotě návěští záznamu MARC21, pozice 18 - hodnota “aacr”, tj. pro LDR/18=”a”<br>pro záznamy v RDA:<br>odpovídá hodnotě záznamu MARC21 pole 040 a podpole $e “rda” - bude použito pro primární rozlišení, zda jde o záznam vytvořený podle AACR2 nebo podle RDA | MA | |

| | | | | |
|---|---|---|---|---|
| <recordContentSource> | | kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník | R | |
| <recordCreationDate> | | datum prvního vytvoření záznamu, na úroveň minut podle normy ISO 8601 | M | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordChangeDate> | | datum změny záznamu, podle normy ISO 8601 na úroveň minut | R | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordOrigin> | | údaje o vzniku záznamu; hodnoty: machine generated nebo human prepared | R | |
| <languageOfCataloging> | | jazyk katalogového záznamu | R | |
| <languageTerm> | | přebírá se z katalogu | R | |
| | authority | hodnota “iso639-2b” | R | |

### 7.3.4 Pole MODS a DC pro obrazové informace (přílohy)

| Element MODS | Atributy | Popis | Povinnost | Element DC |
|---|---|---|---|---|
| **<mods>** | ID | ID obsahuje název úrovně, pro obrazovou přílohu je to "MODS_SUPPL_0001", kde 4-místné číslo je pořadové číslo přílohy | M | |
| **<titleInfo>** | | názvové informace o příloze | M | |

| | | | | |
|---|---|---|---|---|
| | type | hlavní název bez *type* – pole 245 a $a<br>type:<br>alternative – pole 246<br>translated – pole 242<br>uniform – pole 130 resp. 240 | MA | |
| <title> | | hlavní název | MA | <dc:title> |
| <subTitle> | | podnázev | O | |
| <partNumber> | | číslo přílohy<br>01, 02 atd. | M | <dc:title> |
| <partName> | | slovní označení přílohy, označit druh přílohy: obraz desky, obal, booklet | M | |
| **<name>** | | autor přílohy | RA | |
| | type | Použít jednu z hodnot:<br>personal<br>corporate<br>konference<br>family | | |
| <namePart> | | údaje o křestním jméně a příjmení; pokud nelze rozlišit křestní jméno a příjmení, nepoužije se atribut type a jméno se zaznamená v podobě, jaké je, do jednoho elementu <namePart><br>pokud známe datum narození a úmrtí autora, vyplnit ve tvaru RRRR-RRRR s atributem type=”data” | MA | <dc:creator> nutno do jednoho pole DC spojit jméno i příjmení |
| | type | použít jednu z hodnot:<br>date<br>family<br>given | RA<br>MA<br>MA<br>RA | |

| | | | | |
|---|---|---|---|---|
| | | termsOfAddress | | |
| <role> | | specifikace role osoby nebo organizace uvedené v elementu <name> | MA | |
| <roleTerm> | | | MA | |
| | type | hodnota "code" – kód z kontrolovaného slovníku<br><br>(http://www.loc.gov/marc/relators/relaterm.html) | M | |
| | authority | hodnota "marcrelator“ – popis role dle slovníku MARC relator | R | |
| **<typeOfResource>** | | vyplnit „text“ | M | |
| **<genre>** | | hodnota „model" obsahuje druh přílohy, může nabývat hodnot:<br>cover, booklet, imgdisc | M | <dc:type>model:supplement</dc:type> |
| **<language>** | | údaje o jazyce dokumentu<br>v případě vícenásobného výskytu nutno element <language> opakovat | M | |
| <languageTerm> | | přesné určení jazyka – kódem; nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php | M | <dc:language> |
| | type | použít hodnotu code | M | |
| | authority | použít hodnotu „iso639-2b“ | M | |
| **<physicalDescription>** | | obsahuje údaje o fyzickém popisu zdroje/předlohy | M | |

| | | | | |
|---|---|---|---|---|
| <form> | | údaje o fyzické podobě dokumentu | M | <dc:format> |
| | authority | hodnota „marccategory“ obsahuje „text“<br>+<br>Pro záznamy podle RDA:<br>údaje o typu média a typu nosiče zdroje/předlohy<br>odpovídá hodnotám z pole:<br>**337 NEPOVINNÉ** (hodnota např. “bez média” - viz kontrolovaný slovník pole 337) **(R)**<br>**338 POVINNÉ** (hodnota např. “svazek” - viz kontrolovaný slovník pole 338) **(M)** | | |
| | type | pouze pro záznamy v RDA<br>pole 337: type=”media” (bez média)<br>pole 338: type=”carrier” (svazek) | MA | |
| <extent> | | počet stran přílohy<br>brát z 300 $e | M | <dc:format> |
| <note> | | poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element | RA | |
| **<note>** | | obecná poznámka k příloze | RA | <dc:description> |
| **<relatedItem>** | | popis nadřazené úrovně, ke které patří daný druh přílohy (booklet ke kolekci, obal ke kolekci, obraz desky ke straně desky) | M | |
| | type | „host“ | M | |
| <titleInfo> | | názvové informace | M | |
| <title> | | název kolekce (M) nebo strany desky (MA) | M | |

| | | | | |
|---|---|---|---|---|
| <subTitle> | | podnázev | MA | |
| <partNumber> | | POUZE PRO STRANU DESKY:<br>číslo strany desky ve formátu XX<br>01, 02 atd. | M | |
| <partName> | | POUZE PRO STRANU DESKY:<br>slovní označení strany desky, jak je vytištěno na desce, např. "side 1", "strana A" | M | |
| <genre> | type | type=”model“ obsahuje hodnotu pole “collection” nebo “side“ | M | |
| <identifier> | | identifikátor kolekce nebo strany desky | M | |
| | type | type=“uuid“ - uuid kolekce nebo strany desky<br>type=”issue number“ – číslo desky, v MARC21 pole 028, např. $a DV 6196, 6228 $b Supraphon, přebírá se z katalogu<br>jiný interní identifikátor, hodnota atributu „local“, lze použít k vyjádření čárového kódu (MA) | M<br>MA<br>R | |
| **<identifier>** | | údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální – viz přehled typů atributů níže | M | <dc:identifier> |
| | type | budou se povinně vyplňovat následující hodnoty, pokud existují:<br>- uuid (M) | MA | |
| **<recordInfo>** | | údaje o metadatovém záznamu – jeho vzniku, změnách apod. | M | |
| <descriptionStandard> | | popis standardu, ve kterém je přebíraný katalogizační záznam<br>pro záznamy v AACR2: | MA | |

| | | | | |
|---|---|---|---|---|
| | | odpovídá hodnotě návěští záznamu MARC21, pozice 18 - hodnota “aacr”, tj. pro LDR/18=”a”<br>pro záznamy v RDA:<br>odpovídá hodnotě záznamu MARC21 pole 040 a podpole $e “rda” - bude použito pro primární rozlišení, zda jde o záznam vytvořený podle AACR2 nebo podle RDA | | |
| <recordContentSource> | | kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník | R | |
| <recordCreationDate> | | datum prvního vytvoření záznamu, na úroveň minut podle normy ISO 8601 | M | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordChangeDate> | | datum změny záznamu, podle normy ISO 8601 na úroveň minut | R | |
| | encoding | hodnota atributu "iso8601" | | |
| <recordOrigin> | | údaje o vzniku záznamu; hodnoty: machine generated nebo human prepared | R | |
| <languageOfCataloging> | | jazyk katalogového záznamu | R | |
| <languageTerm> | | přebírá se z katalogu - pole 40 $b | R | |
| | authority | hodnota “iso639-2b” | R | |

## 7.4 METS část <amdSec> - Technická a administrativní metadata – formáty AES57, MIX a PREMIS

- pro všechna digitalizovaná data se bude využívat formát PREMIS (jeho části object, event a agent), pro zvukové nahrávky formát AES57 a pro obrazová data náležející k desce formát MIX

- technická a administrativní metadata budou zabalena v části <amdSec> formátu METS ve vlastních formátech (AES57, MIX, PREMIS – části object; events; agent)
- technická a administrativní metadata u obrazových souborů budou vznikat i pro prvotní sken (většinou TIFF), který se po nutných úpravách maže a dále neuchovává
- technická metadata jsou určena primárně pro zachycení technických informací o formátech souborů, o výsledcích validací a kontrol
- administrativní metadata zachycují veškeré změny a procesy, které byly na datech i metadatech provedeny
- technická metadata přicházející z digitalizace jsou dále v maximální míře ukládána v LTP systému (po konverzi do interního formátu LTP systému)
- všechny PREMIS, AES57 či MIX záznamy budou ve vedlejším METS záznamu (amd_mets.xml), který je určen pro administrativní a technická metadata
- tento vedlejší METS záznam (amd_mets.xml) je linkován z hlavního METS záznamu dokumentu (v části <fileSec>)
- plnění technických metadat se předpokládá z výstupů vzniklých využitím služeb třetích stran jakou jsou JHOVE2, PRONOM aj.
- <amdSec> část bude existovat vždy jedna pro všechny reprezentace jedné zvukové stopy či obrazové stránky dokumentu (MC, UC, u obrazu ALTO XML a OCR.txt) a bude obsahovat metadata v <techMD> a <digiprovMD> podčástech

- část <amdSec> musí mít ID a vnořený element <techMD> nebo <digiprovMD>, oba s ID a vnořeným elementem <mdWrap> s atributem MDTYPE

| Element | Atributy | Popis | Povinnost |
|---|---|---|---|
| **<amdSec>** | | element obsahující technická metadata ve formátu PREMIS, AES57 nebo MIX | M |
| | ID | ID – identifikátor konkrétní části <amdSec>, např. pro stranu desky 1 “SIDE01”, pro stranu desky 2 “SIDE02”, obrazovou stránku 1 hodnota “PAGE0001”, pro obrazovou stránku 2 hodnota “PAGE0002” atd. | |
| <techMD> nebo <digiprovMD> | | element rozlišující typy jednotlivých administrativních metadat | M |
| | ID | **ID pro část <techMD>:** | |
| | | pro části obsahující PREMIS Object hodnota: | |

| | | | |
|---|---|---|---|
| | | • "**OBJ_001**" pro původní smazaný soubor TIFF u obrazu nebo WAV u zvuku<br>• "**OBJ_002**" pro MC<br>• "**OBJ_003**" pro ALTO XML u obrazu<br>počet PREMIS Object není omezen, číslování pokračuje OBJ_004 atd.<br>pro části obsahující AES57 hodnota:<br>• "AES57_001" pro původní zvukový soubor<br>• "AES57_002" pro MC<br>• "AES57_003" pro druhou MC<br>pro části obsahující MIX hodnota:<br>• "**MIX_001**" pro původní smazaný obrazový soubor, např. TIFF<br>• "**MIX_002**" pro MC<br>číselná část ID v MIX odpovídá číselné části ID pro PREMIS Object, další záznamy MIX pro novou obrazovou reprezentaci stránky tedy přebírají číslování z techMD pro PREMIS Object (např. pro OBJ_004 v PREMIS Object by to byl MIX_004) | |
| | | **ID pro část <digiprovMD>:** | |
| | | pro části obsahující PREMIS Event hodnota:<br>• "**EVT_001**" atd.<br>pro části obsahující PREMIS Agent hodnota:<br>• "**AGENT_001**" atd. | |
| <mdWrap> | | element obsahující vložené záznamy PREMIS, AES57 či MIX | M |
| | MDTYPE | pro záznamy PREMIS object, event i agent vždy hodnota "PREMIS"<br>pro záznamy AES57 hodnota "AES57"<br>pro záznamy MIX hodnota "NISOIMG" | |
| <xmlData> | | | M |

### 7.4.1 PREMIS Object

- popisovat se pomocí PREMIS Object budou soubory, tj. dle specifikace PREMIS vždy úroveň tzv. **file** (ne reprezentace ani bitstream)
- záznam v PREMIS Object se bude vytvářet pro každý soubor:

        - vzniklý v procesu digitalizace - původní zvukový soubor, který se archivuje, a původní obraz, který se maže
        - archivní kopie zvukové stopy
        - archivní kopie obrazové stránky
        - ALTO XML u obrazu,
        - případné další reprezentace zvukové stopy / obrazové stránky (například nová archivní kopie vytvořená migrací z původní MC)
- PREMIS Object se nebude vytvářet pro UC (zvuk i obraz) ani pro OCR.TXT u obrazových souborů
- pro každý záznam PREMIS Object bude existovat vlastní podčást <techMD>
- záznam PREMIS Object pro jeden soubor bude obsahovat linky na události, které jsou popsány v PREMIS Event ve stejném METS záznamu konkrétního dokumentu v části <digiprovMD> přes <premis:relatedEventIdentification>; to samé platí pro objekty, které budou nalinkovány v případě vztahu (např. MC vzniklá z PS) s popisovaným objektem přes <premis:relatedObjectIdentification>.
    - tj. např. PREMIS Object popisující archivní soubor WAV je tímto způsobem nalinkován na původní nahrávku (resp. na jeho PREMIS Object záznam) – pomocí tagu <relatedObjectIdentification>, který obsahuje ID původního objektu
    - zároveň pomocí tagu <relatedEventIdentification> je záznam PREMIS Object archivního souboru WAV nalinkován na událost, během které vznikl
- POZOR – PREMIS Object bude vznikat a uchovávat se i pro neexistující data (např. původní a posléze smazaný TIFF u obrazových dokumentů)

**Obsah pole “Popis”:**

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné
- výskyt elementu (jak je definováno formátem PREMIS – dle XSD)
    - 0-1 element je nepovinný, neopakovatelný
    - 0-n element je nepovinný, opakovatelný
    - 1-n element je povinný, opakovatelný
    - 1-1 element je povinný, neopakovatelný

**Obsah pole “Použití pro”:**

- použití jednotlivých elementů pro popis MC, PS (původní sken) a XML (ALTO)

| Element | Popis | Použití pro | Povinnost |
|---|---|---|---|
| **<object>** | kořenový element pro PREMIS Object; použít vždy s atributem dle typu objektu<br>xsi:type=“file” – pro soubor<br>1-1 | vše | M |
| **<objectIdentifier>** | identifikátor k jednoznačnému odlišení objektu v určitém kontextu;<br>1-n | MC, XML, PS | M |

| | | | |
|---|---|---|---|
| <objectIdentifierType> | popis kontextu, ve kterém je identifikátor unikátní, např. NDK, ANL nebo název repozitáře; nutno použít kontrolovaný slovník;<br>1-1 | MC, XML, PS | M |
| <objectIdentifierValue> | vlastní hodnota identifikátoru, např. img0001-master, urn.nbn.cz-123465 apod.;<br>1-1 | MC, XML, PS | M |
| **<preservationLevel>** | údaje o úrovni ochrany souboru, která se na něj vztahuje; některé soubory nejsou tak důležité jako jiné, mají menší úroveň ochrany;<br>0-n | MC, XML, PS | MA |
| <preservationLevelValue> | hodnota úrovně ochrany, která je pro soubor relevantní, pro původní sken PS hodnota deleted, pro MC a XML hodnota preservation;<br>1-1 | MC, XML, PS | M |
| <preservationLevelDateAssigned> | datum, kdy byla přiřazena hodnota úrovně ochrany, zápis v ISO 8601, na úroveň dne (DD-MM-RRRR)<br>0-1 | MC, XML, PS | R |
| **<objectCharacteristics>** | technické údaje o souboru<br>1-n | MC, XML, PS | M |
| <compositionLevel> | údaj o tom, zda je nutné digitální objekt rozbalit nebo dekodovat; např. 0 (defaultně pro žádné zabalení nebo kodování); 1 pro jedno zabalení a kódování, podobně pak hodnota 2;<br>1-1 | MC, XML, PS | M |
| <fixity> | údaje o kontrolním součtu<br>1-n | MC, XML, PS | M |
| <messageDigestAlgorithm> | použitý algoritmus kontrolního součtu, např. MD5 aj.<br>1-1 | MC, XML, PS | M |
| <messageDigest> | hodnota kontrolního součtu<br>1-1 | MC, XML, PS | M |
| <messageDigestOriginator> | agent (osoba, instituce, stroj, SW), který kontrolní součet vytvořil (např. JHOVE apod.)<br>0-1 | MC, XML, PS | M |
| <size> | údaje o velikosti souboru v bytech<br>1-1 | MC, XML, PS | M |

| | | | |
|---|---|---|---|
| <format> | údaje o formátu souboru<br>pro soubory ALTO XML je možné vytvořit element dvakrát, jednou popisuje formát XML, podruhé obsahuje informace o použitém standardu ALTO 2.0 (viz příklad 2)<br>1-n | MC, XML, PS | M |
| <formatDesignation> | identifikace formátu souboru, výstup z JHOVE, PRONOM služeb apod.<br>1-1 | MC, XML, PS | M |
| <formatName> | jméno formátu dle MIME type, audio/wav<br><br>1-1 | MC, XML, PS | M |
| <formatVersion> | verze formátu, např. 6.0 nebo jméno verze (př. PCMWAVEFORMATEX--vypíše např. JHOVE)<br>1-1 | MC, XML, PS | M |
| <formatRegistry> | identifikace formátu – dodatečná informace o záznamu formátů v registrech formátů (např. PRONOM aj.)<br>1-1 | MC, XML, PS | M |
| <formatRegistryName> | jméno použitého registru formátů, PRONOM aj.<br>1-1 | MC, XML, PS | M |
| <formatRegistryKey> | unikátní identifikátor (označení) formátu v registru, povinné použít PUID z registru PRONOM, další registry možné<br>1-n | MC, XML, PS | M |
| <creatingApplication> | údaje o aplikaci, ve které byl popisovaný soubor vytvořen, element se používá pro software, tj.SW, kde vzniklo ALTO XML/TXT, SW/kodek pro vytvoření JPEG2000, sw pro úpravu zvuku<br><br>každá aplikace zvlášť v novém elementu<br>1-n | MC, XML, PS | M |
| <creatingApplicationName> | název aplikace, např. WaveLab, ImageGear, Kakadu apod.;<br>1-1 | MC, XML, PS | M |
| <creatingApplicationVersion> | verze aplikace, např. 15.03.000<br>1-1 | MC, XML, PS | M |
| <dateCreatedByApplication> | datum a čas vytvoření, např. 2008-11-10T12:37:46; musí být ve tvaru ISO 8601 (na úroveň vteřin); např. 2008-11-10T12:37:46;<br>1-1 | MC, XML, PS | M |
| **<originalName>** | původní jméno souboru, např. 123456789_01.wav<br>1-1 | MC, XML, PS | M |

| | | | |
|---|---|---|---|
| **<relationship>** | vyjádření vztahu popisovaného souboru k jiným souborům a událostem (eventům)<br>1-n | MC, XML | M |
| <relationshipType> | typ vztahu, doporučené hodnoty: derivation = vztah, kde objekt je výsledkem změny jiného objektu; structural = vztah mezi částmi objektu;<br>tj. např. MC vytvořené z SA bude mít vztah “derivation”;<br>1-1 | MC, XML; | M |
| <relationshipSubType> | upřesnění vztahu, doporučené hodnoty: created from; has source; is source of; has sibling; has part; is part of; has root; includes; is included in apod.;<br>tj. např. MC vytvořené z SA budou mít vztah “created from”;<br>1-1 | MC, XML; | M |
| <relatedObjectIdentification> | identifikace souvisejícího souboru<br>1-n<br>pro MC, XML pro vyjádření vztahu k původnímu objektu | MC, XML | M |
| <relatedObjectIdentifierType> | specifikace kontextu, ve kterém je identifikátor souboru jedinečný, např. URN; temporary filepath; objectID<br>1-1 | MC, XML | M |
| <relatedObjectIdentifierValue> | vlastní řetězec identifikátoru, např. URN:NBN:cz-1301091_011#0001 nebo název souboru, cesta k souboru apod.<br>1-1 | MC, XML | M |
| <relatedEventIdentification> | identifikace s popisovaným souborem související události (eventu); seznam událostí viz PREMIS event<br>1-n | MC, XML | M |
| <relatedEventIdentifierType> | typ události, např. interní číslovací systém událostí jako no.nb.evt; NK repository event ID, UUID apod.<br>1-1 | MC, XML | M |
| <relatedEventIdentifierValue> | hodnota identifikátoru události, např. NK_EVT_005 nebo hodnota UUID aj.<br>1-1 | MC, XML | M |
| <relatedEventSequence> | pořadí události, např. 003; k určení pořadí lze určit datum události<br>0-1 | MC, XML | R |
| **<linkingEventIdentifier>** | identifikátor události týkající původního skenu PS; typy událostí mohou být např. vytvoření, smazání<br>1-n<br>pro PS nutný link na události vytvoření (digitalizace) a jeho vymazání | PS | M |

| | | | |
|---|---|---|---|
| <linkingEventIdentifierType> | typ identifikátoru události, např. UUID, NK_eventID, vlastní číslovací systém apod.<br>1-1 | PS | M |
| <linkingEventIdentifierValue> | hodnota identifikátoru, např. event_01; img0001-master-event001 apod.<br>1-1 | PS | M |

### 7.4.2 PREMIS Event

- PREMIS Event záznamy shromažďují informace o procesech a událostech, které se týkají jednoho nebo více objektů, v našem případě souborů. Primární použití je k zaznamenání událostí, které popisovaný soubor mění nebo upravují.
- bude vznikat pro události, které se prováděly na zvukových a obrazových datech:
    - digitalizace – vytvoření prvního skenu / nahrávky SA
    - vygenerování MC (v případě více MC bude zopakováno)
    - vygenerování UC
    - vytvoření ALTO XML pro obrazová data
    - vymazání PS v případě obrazových dat

- popis událostí bude zachycovat informace o jejich výsledku/výstupu
- záznamy PREMIS Event budou uloženy v METS záznamu určeném pro administrativní a technická metadata (amd_mets.xml) v jeho části <amdSec>, podčást <digiprovMD>
    - amd_mets.xml je linkován z hlavního METS záznamu dokumentu (v částí <fileSec>)
- pro každou událost bude vytvořena jedna <digiprovMD> část
- každý záznam PREMIS Event je linkován na původce aktivity – tj. na PREMIS Agent záznam

**Obsah pole "Popis":**

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné
- výskyt elementu (jak je definováno formátem PREMIS – dle XSD)
    - 0-1 element je nepovinný, neopakovatelný
    - 0-n element je nepovinný, opakovatelný
    - 1-n element je povinný a opakovatelný
    - 1-1 element je povinný a neopakovatelný

**Pole záznamu PREMIS Event**

| Element | Popis | Povinnost |
|---|---|---|

| | | |
|---|---|---|
| **<eventIdentifier>** | údaje o identifikátoru události v kontextu digitalizace nebo repozitáře<br>1-1 | M |
| <eventIdentifierType> | typ identifikátoru, např. no.nb.evt; NK_eventID, UUID apod.<br>1-1 | M |
| <eventIdentifierValue> | hodnota identifikátoru, např. EVT_001; event_019 apod.<br>1-1 | M |
| **<eventType>** | kategorizace události, nutno použít kontrolovaný slovník; typy událostí, které musí být zaznamenány: capture, migration, derivation, deletion<br>1-1 | M |
| **<eventDateTime>** | datum a čas kdy byla událost provedena; nutno zapsat v ISO 8601 na úroveň vteřin<br>1-1 | M |
| **<eventDetail>** | další údaje o události, doporučené hodnoty pro výše uvedené <eventType> následují za /:<br>− capture/digitization – vznik prvního skenu/nahrávky<br>− capture/XML_creation<br>− capture/TXT_creation<br>− migration/MC_creation<br>− derivation/UC_creation<br>− deletion/PS_deletion<br>1-1 | M |
| **<eventOutcomeInformation>** | informace o výsledku události<br>0-n | R |
| <eventOutcome> | kategorizace výsledku události, např. slovy jako successful, failure; možno použít kódy dle kontrolovaného slovníku nebo seznamu kódů;<br>1-n | M |
| **<linkingAgentIdentifier>** | identifikace jednoho nebo více agentů spojených s událostí<br>1-n | M |
| <linkingAgentIdentifierType> | označení typu identifikátoru, např. NK_AgentID, UUID apod.<br>1-1 | M |
| <linkingAgentIdentifierValue> | hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod.<br>1-1 | M |
| <linkingAgentRole> | role agenta ve vztahu k události, např. software; SW component; operator; bude sloužit pro zápis hardware u zvukových dokumentů (tj. role=player nebo AD nebo extraction workstation nebo preamp apod.) nutno používat kontrolovaný slovník<br>0-n | R |
| **<linkingObjectIdentifier>** | informace o objektu/souboru spojeného s událostí, link na něj<br>0-n | M |

| | | |
|---|---|---|
| <linkingObjectIdentifierType> | označení typu identifikátoru,např. hodnota **"file"**<br>1-1 | M |
| <linkingObjectIdentifierValue> | hodnota identifikátoru, např. URN:NBN:cz-_0011#0001 aj. nebo PS_1_0003_R<br>1-1 | M |

### 7.4.3 PREMIS Agent

- využití PREMIS Agent je spíše myšleno pro tzv. ochranné aktivity, které probíhají na archivních datech (AIP balíček) a je nutné pro každou událost na těchto datech mít přesnější informace o tom, kdo ji provedl (osoba administrátora nebo oprávněné osoby)
    - informace v PREMIS Event a PREMIS Object přicházející z procesu digitalizace v SIP balíčku jsou dostačující a dají nám dostatečné informace o události, kdy byla provedena a na jakém SW (PREMIS Object <creatingApplication> + PREMIS Event <eventDetail> - tj. další upřesnění v PREMIS Agent není nutné)
- záznam PREMIS Agent obsahuje charakteristiku tzv. agenta, který je spojen s provedenou a zaznamenanou událostí (PREMIS Event)
    - agent může být osoba, organizace nebo software
- z PREMIS Event je linkováno na agenta, který určitou akci provedl, typ ID agenta a jeho hodnota jsou uvedené v PREMIS Event (<premis:linkingAgentIdentifier>), plný popis agenta je pak v PREMIS Agent
- záznamy PREMIS Agent budou uloženy v METS záznamu určeném pro administrativní a technická metadata (amd_mets.xml) v jeho části <amdSec> a podčásti <digiprovMD>
    - amd_mets.xml je linkován z hlavního METS záznamu dokumentu

- pro každého agenta, tj. jeden PREMIS Agent záznam, bude vytvořena jedna <digiprovMD> část

**Obsah pole „Popis“:**

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné
- výskyt elementu (jak je definováno formátem PREMIS – dle XSD)
    - 0-1 element je nepovinný, neopakovatelný
    - 0-n element je nepovinný, opakovatelný
    - 1-n element je povinný a opakovatelný
    - 1-1 element je povinný a neopakovatelný

| **Element** | **Popis** | **Povinnost** |
|---|---|---|
| **<agentIdentifier>** | popis identifikátoru, který jednoznačně označuje agenta v rámci jednoho kontextu (repozitář např.)<br>1-n | M |

| | | |
|---|---|---|
| <agentIdentifierType> | označení typu identifikátoru, např. NK_AgentID, UUID apod.<br>1-1 | M |
| <agentIdentifierValue> | hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod.<br>1-1 | M |
| **<agentName>** | textové upřesnění agenta, např. přesný název SW, plné jméno osoby apod. - FixImage1.3; Jan Novák; CCS docWorks 6.2.1;<br>0-n | R |
| **<agentType>** | obecné označení agenta – pro osoby např. osoba, pro SW např. software apod.<br>hodnoty: organization; person; software, hardware<br>1-1 | M |
| **<agentNote>** | použití pouze pokud je <agentType> Software a půjde o agenta souvisejícího s migrací původního souboru např. TIFF na JPEG2000 (creation/migration Event);<br>bude obsahovat příkaz k výrobě JPEG2000 souboru v aplikaci Kakadu<br>1-n | MA |
| **<agentExtension>** | pro hardware | MA |
| <NK:manufacturer> | výrobce (př. Mytek, Sony)<br>1-1 | MA |
| <NK:serialNumber> | sériové číslo (př. 01504-1208-043)<br>1-1 | M |
| <NK:settings> | nastavení zařízení<br>0-1 | R |

### 7.4.4 Technická metadata pro zvukové soubory AES57

Pro zvukové soubory se budou uchovávat minimálně následující technická metadata (dle doporučení University Library of Illinois):

1. Formát souboru (MIME type) a informace o (bez)ztrátovosti
2. Velikost souboru v bytech
3. Délka zvukového záznamu ve formátu HH:MM:SS
4. Vzorkovací frekvence
5. Bitová hloubka
6. Počet kanálů
7. Názvy a verze veškerých SW použitých při manipulaci se souborem:
    - nahrávací a editační software - Wavepad, Logic Pro 9, Izotope RX2 (př. MZK)
8. Názvy, výrobce, modely a výrobní čísla veškerého technického vybavení použitého při výrobě souboru:
    - gramofon - Pro-Ject 1-Xpression III Classic, Sumiko Black Pearl 33/45 rpm, Pro-Ject Debut III,...

- přenoska
- hrot
- předzesilovač - Pro-Ject Tube Box S, Pro-Ject Phono Box SE II, CTP-1000 ...
- zvuková karta - Avid Mbox Pro, MOTU Audio Express ,..
- počítačová sestava – Apple iMac Apple OS X

#### *7.4.4.1 Pro digitální dokument*

| **element** | **atributy** | **popis** | **Povinnost** |
|---|---|---|---|
| **<audioObject>** | | | M |
| | *id* | | M |
| | *title* | Např. název dokumentu | |
| | *analogDigitalFlag* | Povolené hodnoty: FILE_DIGITAL | M |
| | *schemaVersion* | Vždy hodnota “1.0.0”, je to verze standardu | |
| <format> | | Oficiální název formátu, tj. „Waveform_Audio“<br>1-1 | M |
| | *specificationVersion* | verze formátu, u wave souborů sem zapsat to, co jhove vypíše u “profiles”, tj. PCMWAVEFORMAT nebo WAVEFORMATEX apod. | MA |
| <appSpecificData> | | Aplikace, která se podílí na tvorbě audia a která do souboru vložila nějaká svá metadata<br>0-n | R |
| | *appVersion* | Verze aplikace | |
| <audioDataEncoding> | | Zatím omezeno pouze na PCM.<br>Hodnotu zapsat tak, jak ji vypíše Jhove v CompressionCode, tj. např.:<br>„PCM audio in integer format“ | M |

| | | | |
|---|---|---|---|
| | | 1-1 | |
| <byteOrder> | | LITTLE_ENDIAN nebo BIG_ENDIAN<br>1-1 | M |
| <firstSampleOffset> | | 0-1 | R |
| <audioDataBlockSize> | | Počet bitů na 1 blok, tj. zapíše se číslo, např. „3“, odpovídá hodnotě BlockAlign ve výpisu z nástroje<br>0-1 | R |
| <use> | | Účel souboru<br>1-n | M |
| | *useType* | Povolené hodnoty: PRESERVATION_MASTER | |
| <primaryIdentifier> | | 1-1 | M |
| | *identifierType* | Může být jen: UMID nebo FILE_NAME nebo SHELF_NUMBER nebo OTHER | MA |
| | *odOtherType* | Vyplníme, pokud u atributu *identifierType* zapíšeme, zde může být cokoliv | MA |
| <fileChecksum> | | Kontrolní součet pro celý soubor<br>1-1 | M |
| <checksumValue> | | Hodnota kontrolního součtu<br>1-1 | M |
| <checksumKind> | | CRC nebo MD5 nebo SHA-1<br>1-1 | M |
| <checksumCreateDate> | | Datum vytvoření kontrolního součtu | M |
| <soundDataChecksum> | | Kontrolní součet pouze pro audio data (tj. bez metadat, hlavičky..)<br>0-1 | O |

| | | | |
|---|---|---|---|
| <face> | | JHOVE vypíše face, region i stream<br>1-n | M |
| | *Id* | Př. Side_1 | M |
| | *direction* | u digitálního souboru vždy hodnota “NONE” | M |
| | *audioObjectRef* | Identifikátor audioObject | M |
| | *label* | Název strany desky, pokud nějaký je (Side_A Rusalka, Půlnoční….) | |
| <timeline> | | 1-1 | M |
| <duration> | | Délka celé jedné strany, zápis asi v sekundách (doplníme z AES60)<br>Př. 55642500<br>1-1 | M |
| | *editRate* | vzorkovací frekvence v herzích<br>hodnota např. “96000” | M |
| <startTime> | | Začátek nahrávky, např. 0, udává se ve frames<br>1-1 | M |
| | *editRate* | | |
| <region> | | Používá se zřejmě pro jednotlivé písně (nutno dořešit)<br>1-n | M |
| | *id* | K id face se přidá 0001, 0002 apod. | M |
| | *formatRef* | Waveform_Audio | M |
| | *faceRef* | Odkaz na id face | M |

| | | | |
|---|---|---|---|
| | *label* | např. Region 1 nebo název písně, proslovu… | M |
| <timeRange> | | 1-1 | M |
| <duration> | | 1-1 | M |
| | *editRate* | | M |
| <startTime> | | 1-1 | M |
| <numChanels> | | Číslo, př. 1<br>1-1 | M |
| <conditionNote> | | Někteří ho používají pro zápis ticha mezi písněmi apod.<br>0-n | R |
| <note> | | Př. „crowd noise“<br>1-1 | M |
| <timeRange> | | 0-1 | R |
| <creationDate> | | 1-1 | M |
| <securityNote> | | 0-n | O |
| <stream> | | Jednotlivé kanály, tj. u stereo nahrávky budou 2 elementy stream<br>1-n | M |
| | *Id* | K id regionu se přidá písmeno, např. u008_20162011_021_0001a | M |
| | *label* | Např. 1 nebo Stream1 | M |

| | | | |
|---|---|---|---|
| | *faceRegionRef* | Id regionu | M |
| <channelAssignment> | | 1-1 | M |
| | *channelNum* | 1 | M |
| | *leftRightPosition* | Př. 0.0 (střed) nebo 100.0 (vpravo) nebo -100.0 (vlevo) | M |
| | | | |
| <conditionNote> | | Poznámka o stavu nahrávky<br>0-n | R |
| <note> | | Text poznámky<br>1-1 | M |
| <timeRange> | | Časový rozsah části nahrávky, které se poznámka týká<br>0-1 | R |
| <creationDate> | | Kdy byla poznámka vytvořena<br>1-1 | M |
| <formatList> | | Obsahuje soupis všech regionů v popisovaném audioObject<br>1-1 | M |
| <formatRegion> | | 0-n | M |
| | *Id* | WAVE nebo Waveform_audio | |
| | *Label* | audio/x-wav | |
| | *ownerRef* | Soupis všech ID regionů | |

| | | | |
|---|---|---|---|
| <bitDepth> | | Bitová hloubka, předepsaná hodnota je 24<br>0-1 | M |
| <sampleRate> | | Vzorkovací frekvence, předepsaná hodnota je 96000 Hz<br>1-1 | M |
| <wordSize> | | Velikost přenášeného vzorku v bytech. Číslo, např. 3 (odpovídá 24 bitové hloubce)<br>0-1 | R |
| <soundField> | | MONO nebo STEREO nebo SURROUND<br>1-1 | R |
| <objectCreationDate> | | 0-1 | R |

### *7.4.4.2 Pro původní zdrojový (analogový) dokument (př. gramodeska), bude v METS v části <sourceMD>*

| element | atributy | popis | Povinnost |
|---|---|---|---|
| **<audioObject>** | | Pro popis původního analogového dokumentu | M |
| | *id* | | |
| | *title* | Např. název dokumentu | |
| | *analogDigitalFlag* | Povolené hodnoty: ANALOG | |
| | *generation* | Kolikátá je generace fyzického nosiče, u původního analogového dokumentu vždy hodnota “Original” | |
| | *disposition* | kde je dokument uložen nyní, př. *returned to stacks* | |
| | *schemaVersion* | | |
| <format> | | název nosiče, např. „open_reel-tape“ (určit zápis), kontrolovaný slovník | M |
| | *specificationVersion* | Pokud existuje | |
| <physicalProperties> | | | O |
| <analogDiscStructure> | | | O |
| | *stockBrand* | Výrobce, jméno produktu, Např. Maxell, Scotch 211, Sony | |
| <substrateMaterialLayer> | | Jednotlivé vrstvy | O |
| <thickness> | | Tloušťka vrstvy | O |
| | *unit* | | |
| | *composition* | Složení vrstvy | |
| | *role* | Role vrstvy, Může být: LABEL_LAYER, PROTECTIVE_LAYER,DATA_LAYER, PROTECTIVE_DATA_LAYER,SUPPORT_LAYER | |

| | | | |
|---|---|---|---|
| | *order* | Pořadí vrstvy | |
| <fillerLayer> | | | O |
| <thickness> | | | O |
| | *unit* | | |
| | *composition* | | |
| | *role* | | |
| | *order* | | |
| <surfaceLayer> | | | O |
| <thickness> | | | O |
| | *unit* | | |
| | *composition* | | |
| | *role* | | |
| | *order* | | |
| <dimensions> | | rozměry | R |
| <shape> | | Když má dokument nezvyklý tvar je to možno uvést sem (př. hranatá deska) | O |
| <diameter> | | průměr | R |
| | *unit* | | |
| <thickness> | | tloušťka desky | R |
| | *unit* | | |
| <shellDimensions> | | Obal audionosiče | O |
| <length> | | | O |

| | | | |
|---|---|---|---|
| | *unit* | | |
| <width> | | | O |
| | *unit* | | |
| <depth> | | | O |
| | *unit* | | |
| <diameter> | | | O |
| | *unit* | | |
| <use> | | Účel souboru | M |
| | *useType* | Povolené hodnoty: ORIGINAL_MASTER, OTHER | |
| <primaryIdentifier> | | | MA |
| | *identifierType* | Může být jen: UMID nebo FILE_NAME nebo SHELF_NUMBER nebo OTHER | |
| | *odOtherType* | Vyplníme, pokud u atributu *identifierType* zapíšeme, zde může být cokoliv | |
| <formatList> | | | O |
| <formatRegion> | | | O |
| | *Id* | WAVE nebo Waveform_audio | |
| | *Label* | Audio/x-wave | |
| | *ownerRef* | Soupis všech ID regionů | |
| <physicalProperties> | | | O |
| <speed> | | | |
| <speedDesignated> | | Rychlost v ips apod. Např. 1.875 | R |
| | *unit* | | |

| | | | |
|---|---|---|---|
| <varispeedAdjustment> | | Pokud byly provedeny nějaké úpravy | R |
| | *unit* | | |
| <speedNote> | | | |
| <grooveOrientation> | | LATERAL nebo VERTICAL nebo STEREO | O |
| <grooveWidth> | | Rozměry drážky | O |
| <min> | | | |
| <max> | | | |
| <grooveCreationMethod> | | DIRECT_CUT nebo PRESS_MOULDED | O |
| <soundField> | | MONO nebo STEREO nebo SURROUND | O |
| <noiseReduction> | | Popisuje akce, které je nutné provést při přehrávání zvuku, aby zněl jak má (tj. netýká se úprav po digitalizaci) | |
| <equalization> | | Úpravy frekvence při přehrávání zvuku | |

### 7.4.5 Technická metadata MIX

- MIX záznam vzniká pouze pro obrazové soubory týkajících se obalu desek, bookletu a strany desek
- bude vznikat pro:
    - archivní obrazovou kopii
    - původní sken vzniklý prvotním skenováním (nejčastěji TIFF) a to i přesto, že tento TIFF se v průběhu výroby maže a není archivován
    - v případě vytvoření nové verze archivní kopie např. formátovou migrací, původní MIX záznam se zachovává a vytváří se navíc záznam nový pro aktuálně platnou verzi archivní kopie
- tyto MIX záznamy budou součástí jednoho METS záznamu amd_mets.xml (v části <amdSec> a podčásti <techMD>) pro administrativní a technická metadata, který vznikne ke každému obrazovému souboru a který je linkován z hlavního METS záznamu
- MIX záznamy jednotlivých obrazových souborů se budou lišit – MIX záznam původního skenu nebude obsahovat např. element ImageProcessing, MIX záznam archivního souboru MC nebude naproti tomu obsahovat informace o procesu skenování, které se váží k původnímu skenu a budou v elementu ImageCaptureMetadata apod.
- pro každý záznam MIX bude vytvořena vlastní část <techMD>
- MIX může být také zapouzdřen v PREMIS Object <premis:objectCharacteristicsExtension>

- externí služby jako JHOVE nebo PRONOM budou využívány k plnění polí formátu MIX
- ve formátu MIX nebude uvedena informace o kontrolních součtech (fixity), která je obsažena v PREMIS Object a není nutno ji opakovat
- <fileSize> je pouze doporučené, údaj je již součástí popisu u PREMIS Object

**Obsah pole "Popis":**

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné
- výskyt elementu (jak je definováno formátem MIX – dle XSD)
    - 0-1 element je nepovinný, neopakovatelný
    - 1-1 element je povinný, neopakovatelný
    - 0-n element je nepovinný, opakovatelný

**Obsah pole "Použití pro"**

- použití jednotlivých elementů pro MC, PS (původní sken) – určuje, který element je a který není součástí MIX záznamu MC nebo MIX záznamu popisujícího původní obrazový dokument ze skeneru

| Element | Popis | Povinnost | Použití pro |
|---|---|---|---|
| **<BasicDigitalObjectInformation>** | | | |
| **<ObjectIdentifier>** | **údaje o identifikátoru obrazového dokumentu, který je formátem MIX popsán;<br>0-n** | R | MC, PS |
| <objectIdentifierType> | např. jméno souboru, nebo jiný identifikátor;<br>1-1 | M | MC, PS |
| <objectIdentifierValue> | hodnota identifikátoru, např. 20110306_001.jp2 nebo urn:nbn:123456;<br>1-1 | M | MC, PS |
| **<fileSize>** | velikost souboru<br>0-1 | R | MC, PS |
| **<FormatDesignation>** | **údaje o formátu obrazového souboru<br>1-1** | **M** | **MC, PS** |
| <formatName> | název formátu, např. lze využít MIME types (Image/jp2 apod.)<br>1-1 | M | MC, PS |
| <formatVersion> | verze formátu, např. 1.0<br>1-1 | M | MC, PS |

| | | | |
|---|---|---|---|
| **<byteOrder>** | **endianita, možnosti jsou little endian, middle (mix) endian a big endian<br>1-1** | **M** | **MC, PS** |
| **<Compression>** | **údaje o kompresi obrazového souboru<br>1-1** | **M** | **MC, PS** |
| <compressionScheme> | informace o kompresním schématu, vyjádřeno číslem (např. 34712 je komprese JPEG2000) nebo slovy (např. JP2 Lossless)<br>1-1 | M | MC, PS |
| **<BasicImageInformation>** | **základní technické údaje o obrazovém dokumentu<br>1-1** | **M** | **MC, PS** |
| **<BasicImageCharacteristics>** | **1-1** | **M** | **MC, PS** |
| <imageWidth> | šířka obrazu v pixelech, např. 3987<br>1-1 | M | MC, PS |
| <imageHeight> | výška obrazu v pixelech, např. 2345<br>1-1 | M | MC, PS |
| <PhotometricInterpretation> | photometrická interpretace<br>0-1 | M | MC, PS |
| <colorSpace> | barevný prostor, např. RGB<br>1-1 | M | MC, PS |
| <ColorProfile> | údaje o barevném profilu<br>0-1<br>povinné pro dokumenty, kde je nutno uchovat přesnou reprezentaci barvy původního dokumentu a používá se ICC profil) | MA | MC, PS |
| <IccProfile> | ICC profil<br>1-1 | M | MC, PS |
| <iccProfileName> | jméno profilu, např. sRGB, Adobe RGB aj.<br>1-1 | M | MC, PS |
| <iccProfileVersion> | verze profilu, např. sRGB IEC61966-2.1<br>1-1 | M | MC, PS |
| <iccProfileURL> | odkaz na profil, např. www.profily.cz/sRGB_v4_ICC_pref.icc;<br>0-1 | R | MC, PS |
| **<SpecialFormatCharacteristics>** | **speciální technické údaje o obrazovém dokumentu, povinné použití pro formát JPEG2000<br>0-1** | **MA** | **MC** |
| **<JPEG2000>** | **1-1** | **M** | **MC** |

| | | | |
|---|---|---|---|
| <CodecCompliance> | údaje o kodeku<br>1-1 | M | MC |
| <codec> | název kodeku, např. Kakadu, LuraWave aj.<br>1-1 | M | MC |
| <codecVersion> | verze kodeku, např. 3.1<br>1-1 | M | MC |
| <codestreamProfile> | popis codestream profilu JPEG2000, např. P0 a P1 (viz ISO/IEC 15444-4);<br>1-1 | M | MC |
| <complianceClass> | specifikace největší výšky, šířky a počtu komponentů, které dekodér dokáže dekódovat, lze použít hodnoty C0, C1 a C2;<br>1-1 | M | MC |
| <EncodingOptions> | obsahuje informace o kodování JPEG2000<br>1-1 | M | MC |
| <Tiles> | popis pixelové velikosti dlaždic formátu JPEG2000<br>1-1 | M | MC |
| <tileWidth> | šířka dlaždice, např. 128<br>1-1 | M | MC |
| <tileHeight> | výška dlaždice, např. 128<br>1-1 | M | MC |
| <qualityLayers> | číselná hodnota počtu vrstev, do kterých byl JPEG2000 rozdělen, např. 12<br>1-1 | M | MC |
| <resolutionLevels> | popis počtu nižších rozlišení, které lze z obrazu získat, např. 6<br>1-1 | M | MC |
| **<ImageCaptureMetadata>** | **popis procesu skenování, je důležité vyplnit, protože tyto údaje nelze zjistit z finálního master/archivního souboru**<br>**1-1** | **M** | **PS** |
| **<SourceInformation>** | **informace o předloze**<br>**1-1** | **R** | **PS** |
| <sourceType> | Booklet, Cover aj.; nutno používat kontrolovaný slovník<br>1-1 | M | PS |
| <SourceID> | identifikátor předlohy<br>0-n | R | PS |
| <sourceIDType> | typ identifikátoru, např.URN:NBN<br>1-1 | M | PS |

| | | | |
|---|---|---|---|
| <sourceIDValue> | vlastní hodnota identifikátoru<br>1-1<br>povinné | M | PS |
| **<GeneralCaptureInformation>** | **základní údaje o skenování**<br>**1-1** | **M** | **PS** |
| <dateTimeCreated> | údaj o datu a čase skenování, např. 2009-01-03T08:25:28; zapsat v ISO 8601 na úroveň vteřin<br>1-1 | M | PS |
| <imageProducer> | entita provádějící skenování, např. The National Library of the Czech Republic, osoba apod.<br>1-1 | M | PS |
| <captureDevice> | typ skenovacího zařízení<br><br>použít jedno z hodnot:<br>− reflection print scanner<br>− transmission scanner<br>− digital still camera<br>− still from video<br><br>1-1 | MA | PS |
| **<scannerCapture>** | **údaje o skeneru**<br>**1-1** | **M** | **PS** |
| <scannerManufacturer> | výrobce skeneru, např. 4DigitalBooks, Treventus, Zeutschel<br>1-1 | M | PS |
| <ScannerModel> | údaje o konkrétním typu skeneru<br>1-1 | M | PS |
| <scannerModelName> | jméno modelové řady skeneru, např. DL<br>1-1 | M | PS |
| <scannerModelNumber> | číslo/označení modelu, např. 3000<br>1-1 | M | PS |
| <scannerModelSerialNo> | výrobní číslo skeneru, např. E4R0003649<br>1-1 | M | PS |
| <MaximumOpticalResolution> | údaje o maximálním optickém rozlišení skeneru<br>1-1 | M | PS |
| <xOpticalResolution> | optické rozlišení na ose x, pouze číselné vyjádření např. 300<br>0-1 | M | PS |
| <yOpticalResolution> | optické rozlišení na ose y, pouze číselné vyjádření např. 300<br>0-1 | M | PS |

| | | | |
|---|---|---|---|
| <opticalResolutionUnit> | jednotka optického rozlišení, jedna z hodnot: no absolute unit; in.; cm<br>0-1 | M | PS |
| <scannerSensor> | popis typu snímacího senzoru skenovacího zařízení, jedna z hodnot:<br>undefined; MonochromeLinear; ColorTriLinear; ColorSequentialLinear; MonochromeArea; OneChipColourArea; TwoChipColorArea; ThreeChipColorArea; ColorSequentialArea;<br><br>1-1 | M | PS |
| <ScanningSystemSoftware> | údaje o softwaru skenovacího zařízení<br>0-1 | M | PS |
| <scanningSoftwareName> | název softwaru, např. Copinet<br>0-1 | M | PS |
| <scanningSoftwareVersionNo> | číslo verze softwaru, např. 3.7<br>0-1 | M | PS |
| **<digitalCameraCapture>** | **údaje o snímacím zařízení (fotoaparát)<br>1-1<br>povinné, pokud je používán fotoaparát a není používán skener** | **MA** | **PS** |
| <digitalCameraManufacturer> | výrobce fotoaparátu, např. Canon<br>0-1 | M | PS |
| <DigitalCameraModel> | popis modelu fotoaparátu<br>0-1 | M | PS |
| <digitalCameraModelName> | název modelové řady, např. EOS<br>0-1 | M | PS |
| <digitalCameraModelNumber> | označení modelu fotoaparátu, např. 1000D<br>0-1 | M | PS |
| <digitalCameraModelSerialNo> | výrobní číslo přístroje, např. E12345<br>0-1 | M | PS |
| <camerarSensor> | typ senzoru fotoaparátu, např. matrix aj.<br>0-1 | M | PS |
| <CameraCaptureSettings> | údaje o nastavení fotoaparátu použitého ke snímání předloh<br>0-1 | M | PS |
| <ImageData> | v rámci tohoto kontejnerového elementu budou použity následující sub-elementy:<br>fNumber<br>exposureTime<br>isoSpeedRatings<br>shutterSpeedValue<br>apertureValue | M | PS |

| | | | |
|---|---|---|---|
| | brightnessValue<br>exposureBiasValue<br>maxApertureValue<br>subjectDistance<br>meteringMode<br>lightSource<br>flash<br>focalLength<br>backLight<br>exposureIndex<br>sensingMethod<br>cfaPattern<br>autoFocus<br>PrintAspectRatio<br><br>všechny hodnoty budou přebrány v případě použití fotoaparátu z údajů Exif | | |
| **<orientation>** | **popis orientace obrazu tak, jak je uložen vzhledem k jeho řádkům a sloupcům, např. normal*; normal, image flipper; normal, rotated 180°; unknown apod.<br>0-1** | **M** | **PS** |
| **<ImageAssessmentMetadata>** | **informace o digitálním obrazu pro jeho hodnocení a využití z hlediska dlouhodobé ochrany apod.<br>0-1** | **M** | **MC, PS** |
| **<SpatialMetrics>** | **rozměry obrázku, 2 rozměrná projekce objektů tak jak ji „vidí“ snímací zařízení<br>0-1** | **M** | **MC, PS** |
| <samplingFrequencyPlane> | popis základní roviny, např. object plane (pro přímo ze předlohy digitalizované dokumenty), source object plane (pro digitalizaci mikrofilmů), camera/scanner focal plane (indikace sampl. frekvence fyzického senzoru);<br>0-1 | R | MC, PS |
| <samplingFrequencyUnit> | jednotka měření sampl. frekvence<br><br>hodnoty: "no absolute unit of measurement"; "in."; "cm";<br>0-1 | M | MC, PS |
| <xSamplingFrequency> | údaje o počtu pixelů na jednotku samplovací frekvence pro šířku obrázku<br>0-1<br>povinné, pokud hodnota samplingFrequencyUnit je "in." nebo "cm" | MA | MC, PS |
| <numerator> | čitatel, číselné vyjádření, např. 300<br>0-1 | M | MC, PS |
| <denominator> | jmenovatel, číselné vyjádření např. 1<br>0-1 | M | MC, PS |

| | | | |
|---|---|---|---|
| <ySamplingFrequency> | údaje o počtu pixelů na jednotku samplovací frekvence pro výšku obrázku<br>0-1<br>povinné, pokud hodnota samplingFrequencyUnit je "in." nebo "cm" | MA | MC, PS |
| <numerator> | čitatel, číselné vyjádření, např. 300<br>0-1 | M | MC, PS |
| <denominator> | jmenovatel, číselné vyjádření např. 1<br>0-1 | M | MC, PS |
| **<ImageColorEncoding>** | **doplňující údaje o barvě obrazu**<br>**1-1** | **M** | **MC, PS** |
| <BitsPerSample> | počet bitů na kanál<br>1-1 | M | MC, PS |
| <bitsPerSampleValue> | hodnota počtu bitů, např. 8, 1, 4 nebo 8,8,8 apod.<br>1-n<br>POZOR – pro každou hodnotu je nutno element opakovat, tj. např. 3x element <bitsPerSampleValue> s hodnotou 8<br><br><mix:BitsPerSample><br><br><mix:bitsPerSampleValue>8</mix:bitsPerSampleValue><br><br><mix:bitsPerSampleValue>8</mix:bitsPerSampleValue><br><br><mix:bitsPerSampleValue>8</mix:bitsPerSampleValue><br></mix:BitsPerSample> | M | MC, PS |
| <bitsPerSampleUnit> | specifikace jednotky, např. "integer" nebo "floating point"<br>1-1 | R | MC, PS |
| <samplesPerPixel> | počet barevných komponentů na pixel, např. 1, 3, 4<br>1-1 | M | MC, PS |
| **<TargetData>** | **informace o kalibračních tabulkách** (calibration target); **povinné pro obrazy, kde se dělá kontrola oproti kalibrační tabulce**<br>**1-1** | **MA** | **MC** |
| <targetType> | typ kalibrační tabulky; 0= external (kalibrační tabulka se neobjeví na dig. obraze, je to oddělený dig. soubor); 1= internal (tabulka je naskenována spolu s přelohou a objeví se na dig. obraze);<br>1-n | M | MC |
| <targetID> | údaje o původu kalibrační tabulky<br>0-n | M | MC |
| <targetManufacturer> | výrobce/původce kalibrační tabulky, např. Eastman Kodak nebo NK ČR, oddělení kontroly kvality apod.<br>1-1 | M | MC |

| | | | |
|---|---|---|---|
| <targetName> | název kalibrační tabulky, např. ColorChecker, MicrofilmScanTarget aj.<br>1-1 | M | MC |
| <targetNo> | číslo nebo verze kalibrační tabulky<br>0-1 | M | MC |
| <targetMedia> | údaj o tom, na jakém médiu je kalibrační tabulka, např. film, paper aj.<br>0-1 | R | MC |
| <externalTarget> | údaje o externí kalibrační tabulce; např. link na http://skenservis.cz/target-00000001 nebo název a cesta ke konkrétnímu souboru<br>0-n<br>povinné v případě, že byla použita externí kalibrační tabulka (targetType = 0) | MA | MC |
| <performaceData> | odkaz na soubor obsahující charakteristiku výkonu systému vzhledem k nastaveným hodnotám rozlišení atd.; možné hodnoty plnění – link URN nebo URL, nebo název souboru<br>0-n | R | MC |
| **<ChangeHistory>** | **dokumentace procesů provedených na obrazovém souboru v jeho životním cyklu**<br>**1-1** | **M** | **MC** |
| **<ImageProcessing>** | **údaje o zpracování obrazového souboru**<br>**1-n** | **M** | **MC** |
| <dateTimeProcessed> | čas v ISO 8601 na úroveň vteřin, př. 2009-01-04T15:12:06<br>0-1 | M | MC |
| <sourceData> | odkaz na původní zdrojová data, ze kterých byl vytvořen finální obrazový soubor; může to být např. URL nebo cesta do složky s původním skenem včetně názvu souboru;<br>0-1 | M | MC |
| <processingAgency> | Instituce, př. The National Library of the Czech Republic<br>0-n | R | MC |

## 7.5 METS část <fileSec>

Část <fileSec> bude obsahovat výčet všech souborů, které patří do zvukové kolekce.

### 7.5.1 <fileSec> hlavního záznamu METS

**file group**

- pro zvuk, obraz i texty budou v hlavním METS záznamu použity elementy <fileGrp>, jeden element <fileGrp> bude existovat vždy pro jednu skupinu souborů stejného formátu a účelu, tedy pro původní zvukové soubory SA, archivní kopie zvukových souborů, uživatelské kopie zvukových souborů, archivní kopie obrazů, uživatelské kopie obrazů, další pro ALTO XML, OCR.TXT soubory a další pro METS záznamy s technickými metadaty (AMD_METS.xml)

  **<fileGrp> pro původní zvukové soubory SA** - atributy ID="SA_AUDIOGRP" USE="source"
  - každý soubor bude mít vlastní element <file> s následujícími atributy:
    - ID - kopíruje název souboru bez přípony
    - MIMETYPE - hodnota audio/wav
    - SIZE - velikost souboru
    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE - hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
  - subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na zvukový soubor (xlink:href) a atribut LOCTYPE

  **<fileGrp> pro zvukové soubory MCA** - atributy ID="MC_AUDIOGRP" USE="master"
  - každý soubor bude mít vlastní element <file> s následujícími atributy:
    - ID – kopíruje název souboru bez přípony
    - MIMETYPE – audio/wav
    - SIZE – velikost souboru
    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE – hodnota MD5
    - CHECKSUM – hodnota kontrolního součtu
    - CREATED – datum vytvoření, ISO8601 na úroveň vteřiny
  - subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na zvukkový soubor (xlink:href) a atribut LOCTYPE

  **<fileGrp> pro zvukové soubory UCA** - atributy ID="UC_AUDIOGRP" USE="user"
  - každý soubor bude mít vlastní element <file> s následujícími atributy:
    - ID - kopíruje název souboru bez přípony
    - MIMETYPE - audio/mp3
    - SIZE - velikost souboru

    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE - hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
- subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na zvukový soubor (xlink:href) a atribut LOCTYPE

**<fileGrp> pro obrazy MC** - atributy ID="MC_IMGGRP" USE="master"

- každý soubor bude mít vlastní element <file> s následujícími atributy:
    - ID - kopíruje název souboru bez přípony
    - MIMETYPE - image/jpeg2000
    - SIZE - velikost souboru
    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE - hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
- subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na obrazový soubor (xlink:href) a atribut LOCTYPE

**<fileGrp> pro obrazy UC** - atributy ID="UC_IMGGRP" USE="user"

- každý soubor bude mít vlastní element <file> s následujícími atributy:
    - ID - kopíruje název souboru bez přípony
    - MIMETYPE - image/jpeg2000
    - SIZE - velikost souboru
    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE - hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
- subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na obrazový soubor (xlink:href) a atribut LOCTYPE

**<fileGrp> pro ALTO XML** bude mít následující atributy: ID="ALTOGRP" USE="layout"

- každý ALTO XML soubor bude mít vlastní element <file> s následujícími atributy:
    - ID – kopíruje název souboru bez přípony
    - MIMETYPE – text/xml
    - SIZE – velikost souboru xml
    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE – hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
- subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na xml soubor obsahující ALTO (xlink:href) a atribut LOCTYPE

**<fileGrp> pro soubory METS s technickými metadaty** amd_mets.xml bude mít následující atributy:

i. ID="TECHMDGRP" USE="technical metadata"

- každý METS xml soubor bude mít vlastní element <file> s následujícími atributy:
    - ID - identifikátor souboru amd_mets.xml jak je používán v METS záznamu; kopíruje název souboru bezp přípony
    - MIMETYPE – text/xml
    - SIZE – velikost souboru xml
    - CHECKSUMTYPE – hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - SEQ – pořadí souboru
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
  - subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na xml soubor amd_mets.xml (xlink:href) a atribut LOCTYPE
- každý element <file> bude mít tedy vnořený subelement <FLocat>, který bude odkazovat na konkrétní soubor a bude obsahovat atributy:
    - LOCTYPE - typ odkazu na soubor, typicky “URL”
    - xlink:href - odkaz na soubor

**<fileGrp> pro soubory OCR.TXT** bude mít následující atributy: ID="TXTGRP" USE=text"

- každý OCR.TXT soubor bude mít vlastní element <file> s následujícími atributy:
    - ID - identifikátor souboru OCR.TXT jak je používán v METS záznamu; kopíruje název souboru bez přípony
    - MIMETYPE – text/plain
    - SIZE - velikost souboru
    - SEQ - pořadí souboru ve fileGrp
    - CHECKSUMTYPE – hodnota MD5
    - CHECKSUM - hodnota kontrolního součtu
    - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
  - subelementem pod <file> je element <Flocat>, který obsahuje link (ideálně v podobě nějakého identifikátoru) na txt soubor (xlink:href) a atribut LOCTYPE

### 7.5.2 <fileSec> vedlejšího záznamu amd_mets.xml

- <fileSec> ve vedlejším METS záznamu AMD_METS.xml bude obsahovat jeden element <fileGrp> s vnořenými elementy <file> pro každou reprezentaci zvukové stopy/stránky, tj. SA, MC pro zvuk, MC pro obraz, ALTO XML a OCR.TXT
- atributy jednotlivých <file> elementů odpovídají atributům pro jednotlivé typy dokumentů uvedených výše pro <fileSec> hlavního METS záznamu

- navíc pro SA a MC pro zvuk, MC pro obraz a ALTO XML (pro každou existující reprezentaci stránky s vlastním záznamem PREMIS Object nebo MIX) bude pro element <file> existovat atribut ADMID s ID těch <techMD> záznamů, které danou reprezentaci stránky popisují

## 7.6 METS část <structMap> - Strukturální metadata

### 7.6.1 <structMap> hlavního záznamu METS

- strukturální mapy v METS záznamu existují dvojího typu - fyzická a logická
- fyzická mapa zaznamenává zejména vazby na fyzické soubory, ze kterých se skládají jednotlivé úrovně dokumentu
- logická mapa zaznamenává hierarchickou strukturu dokumentu bez vazeb na fyzické soubory; popisuje jednu kolekci na úroveň stopy a obrazové přílohy
- vazby mezi konkrétní stranou desky a stopami či konkrétní přílohou a obrázky jsou popsány v METS sekci <structLink>

- strukturální mapa logická i fyzická bude v hlavním záznamu hlavni_METS.xml

#### *7.6.1.1 Vyjádření fyzické strukturální mapy*

- fyzická strukturální mapa bude mít následující atributy: LABEL=”physical_structure”, TYPE=”physical”
- obsahuje element  první úrovně, který popisuje kolekci; v něm jsou dále vnořeny další elementy  druhé úrovně, které odkazují na všechny soubory patřící do kolekce (zvukové stopy, obrázky, alto, ocr)
- ** první úrovně** obsahuje atributy:
    - ID - identifikátor  této úrovně (např. DIV_0000)
    - LABEL - název kolekce (nejvyšší úroveň)
    - TYPE - sound recording
    - DMDID - odkazuje na část bibliografických metadat náležejících ke kolekci
    - ADMID - identifikátor amdsec, ve kterém jsou obsažena (nepovinná) autorsko-právní metadata
- ** druhé úrovně pro zvuk** obsahuje tyto atributy:
    - ID - identifikátor  této úrovně (např. DIV_AUDIO_0000)
    - TYPE - typ objektu, tj. track
    - ORDER - reálné pořadí tracku na desce
    - ORDERLABEL - pořadové číslo tracku buď dle strany desky nebo obalu desek
    -  druhé úrovně obsahuje vždy subelement <fptr>, v kterém je odkaz na soubory související se stopou (tj. SA, MCA, UCA, MC, UC, AMD_METS.xml)
        - link na soubor má v elementu <fptr> atribut FILEID, který obsahuje ID souboru
- ** druhé úrovně** se zopakuje i pro **obraz**, bude mít atributy:
    - ID - identifikátor  této úrovně (např. DIV_PAGE_0000)
    - TYPE - typ stránky (stránka bookletu, strana desky, obal)
    - ORDER - pořadí stránky v příloze
    - ORDERLABEL - pořadové číslo stránky, jak je na ní vytištěno
    - element <fptr> s linkem na konkrétní soubory (SA, MCA, UCA, MC, UC, OCR.TXT, ALTO XML, AMD_METS.xml)
        - link na soubor má v elementu <fptr> atribut FILEID, který obsahuje ID souboru
        - pro element <fptr> obsahující ALTO XML platí, že má navíc tyto atributy: BEGIN=”P1”, kde P1 je ID elementu <fptr> z ALTO XML souboru; a atribut BETYPE=”IDREF”

### *7.6.1.2 Vyjádření logické strukturální mapy*

#### 7.6.1.2.1 Vyjádření logické struktury, s vazbou na ALTO bloky

- bude mít následující atributy: LABEL=”logical_structure” TYPE=”logical”
- logická struktura na úroveň stopy a obrazové přílohy se vyjádří pomocí zanořených elementů 
-  musí obsahovat identifikátor na příslušný blok metadat, př. odkaz na úroveň kolekce DMDID=”MODS_COLLECTION_0001”
- element  první úrovně - kolekce (collection)
- element  druhé úrovně - strana desky (side)
- element  třetí úrovně - zvuková stopa (track)
- element  opět druhé úrovně - obrazové informace (supplement či rozdělit na booklet, cover a imgdisc)

- výčet stop na straně desky a výčet obrázků v příloze bude uveden v části <structLink>

|  type | Atributy | Popis | Povinnost |
|---|---|---|---|
| **COLLECTION** | ID<br>LABEL<br>TYPE<br>DMDID |  obsahuje údaje o nadřazené kolekci<br>ID - identifikátor , hodnota např. COLLECTION_0001<br>LABEL - název celé kolekce<br>TYPE - hodnota COLLECTION<br>DMDID - obsahuje identifikátor popisné části MODS ke kolekci | M |
| **SIDE** | ID<br>LABEL<br>TYPE<br>DMDID |  obsahuje údaje o straně desky<br>ID - identifikátor , hodnota např. SIDE_0001<br>LABEL - název strany desky (číslo)<br>TYPE - hodnota SIDE<br>DMDID - obsahuje identifikátor popisné části MODS ke straně desky | M |
| **TRACK** | ID<br>LABEL<br>TYPE<br>DMDID |  obsahuje údaje o zvukové stopě<br><br>ID - identifikátor , hodnota např. TRACK_0001<br>LABEL - název zvukové stopy (např.: “Stairway to heaven”)<br>TYPE - hodnota TRACK<br>DMDID - obsahuje identifikátor popisné části MODS k tracku | M |
| **SUPPLEMENT** | ID<br>LABEL<br>TYPE<br>DMDID |  obsahuje údaje o obrazové příloze<br>ID - identifikátor , hodnota např. SUPPL_0001<br>LABEL - název přílohy - hodnoty BOOKLET, COVER, IMGDISC<br>TYPE - hodnota SUPPL<br>DMDID - obsahuje identifikátor popisné části MODS k příloze | M |

**<structMap> vedlejšího záznamu METS (AMD_METS.xml)**

- bude obsahovat pouze fyzickou strukturální mapu (TYPE=”PHYSICAL”)
- ta bude obsahovat pouze jeden  element s atributem TYPE=”TRACK” nebo TYPE=”SUPPL”
- do  budou vnořeny odkazy na jednotlivé reprezentace zvukové stopy nebo obrazové přílohy pomocí elementu <ftpr> s atributem FILEID

**METS část <structLink>**

Bude ukazovat vztah mezi fyzickou a logickou strukturou, tedy vztah mezi jednotlivými stopami a stranou desky a jednotlivými obrázky a přílohou.

- element <structLink> obsahuje subelement <smLink> pro každý zvukový soubor nebo obrázek
- bude mít atributy xlink:from (odkaz na stranu desky či přílohu) a xlink:to (odkaz na zvukový soubor či obrázek)

Př.:

```
<mets:structLink>
<mets:smLink xlink:from="SIDE_1" xlink:to="DIV_AUDIO_0001"/>
<mets:smLink xlink:from="SIDE_1" xlink:to="DIV_AUDIO_0002"/>
<mets:smLink xlink:from="SIDE_2" xlink:to="DIV_AUDIO_0003"/>
<mets:smLink xlink:from="SIDE_2" xlink:to="DIV_AUDIO_0004"/>

<mets:smLink xlink:from="SUPPL" xlink:to="DIV_PAGE_0001"/>
<mets:smLink xlink:from="SUPPL" xlink:to="DIV_PAGE_0002"/>
<mets:smLink xlink:from="SUPPL" xlink:to="DIV_PAGE_0003"/>
</mets:structLink>
```

*Otázka, jestli přílohu nerozdělit na tři části pro obaly desek, obrazy stran desek a booklet.*
*Anebo pod jednou přílohou “supplement” budou všechny obrázky nezávisle na typu a typ se popíše v atributu “type” ve fyzické strukturální mapě (booklet/supplement, cover, imgdisc). Bude pak potřeba vytvořit pravidla popisu :)*

## 7.7 OCR (ALTO XML a TXT OCR)

- bude použita poslední verze formátu ALTO XML aktuální v době implementace, nebo verze předchozí (prosinec 2010 verze 2 – viz http://www.loc.gov/standards/alto/)
- níže uvedená specifikace **neobsahuje všechny elementy a atributy formátu ALTO XML, obsahuje pouze ty, které jsou pro tuto konkrétní specifikaci relevantní – každý uvedený element má vyjádřenou míru relevance výrazy: povinné, doporučené a nepovinné**
- elementy a atributy, které v této specifikaci nejsou uvedeny, nepovažujeme pro účely specifikace za důležité
- ALTO XML i OCR TXT vzniknou pro všechny obrazové soubory náležející k jedné intelektuální entitě (kolekce) včetně prázdných stran, fotografií hřbetu, předsádky apod.
- ALTO XML i OCR TXT budou vznikat na úroveň stránky
- ALTO XML soubor pro zcela prázdné stránky bude obsahovat element /alto/Layout/Page/PrintSpace, ten ovšem nebude obsahovat podelementy:

/alto/Layout/Page/PrintSpace/TextBlock; /alto/Layout/Page/PrintSpace/TextBlock/Illustration; /alto/Layout/Page/PrintSpace/TextBlock/GraphicalElement ani /alto/Layout/Page/PrintSpace/TextBlock/ComposedBlock

- struktura ALTO XML bude generovaná na úrovni rozpoznání slova generovaná OCR
- kvalita rozpoznání znaků bude akceptována do určité hranice, výstupy nebudou ručně opravovány
- struktura ALTO umožní vyhledávání textu a jeho zvýraznění na úrovni slova, pokud bude použit odpovídající prohlížeč
- obrazy reprezentující stránku, které budou použity jako UC, musí odpovídat rozměry, orientací a natočením obrazu, který byl použit pro vytvoření OCR
- OCR TXT bude vznikat z hotových ALTO XML během procesu digitalizace
- ALTO XML se bude vytvářet pouze pro novodobé dokumenty, nebo dokumenty s určitou hranicí kvality OCR
- jméno OCR souboru musí odpovídat jménu obrazového souboru, ke kterému náleží; např. pr_0007.jp2 a al_0007.xml nebo např. 123456_006_alto.xml a 123456_006_archiv.jp2
- kódování ALTO XML i TXT OCR musí být v UTF-8
- souřadnice pozic (HPOS, VPOS, WIDTH, HEIGHT) musí být vyjádřeny v pixelech

- v této specifikaci ALTO XML se počítá s OCR i pro text mimo tzv. textové "zrcadlo", tj. mimo hlavní text, jako jsou např. čísla stránek, běžící nadpisy a jiné části vyskytující se na okrajích stránky (top, left, top a bottom margin)
    - elementy topMargin, leftMargin, rightMargin, bottomMargin budou obsahovat elementy <TextBlock>, pro které platí stejná pravidla, jako pro element <textBlock> pro hlavní text stránky
    - pozor: údaje z OCR mimo hlavní text stránky by neměly být vyhledávatelné v aplikaci zpřístupnění, docházelo by ke zmatení uživatele a výsledků (např. při hledání titulu kapitoly by byly zobrazeny výsledky pro každou stránku, která obsahuje běžící nadpis apod.)

- pokud je na konci věty dělící znaménko, ALTO XML i OCR TXT musí obsahovat oba fragmenty slova s dělítkem a současně také kompletní slovo – je vysvětleno dále v tabulce
- ilustrace, reklamy a jiné grafické části stránky nebudou vyjádřeny v tazích /alto/Layout/Page/PrintSpace/Illustration ani Layout/Page/PrintSpace/GraphicalElement, tyto nejsou v popisu/tabulce níže vůbec uvedeny
- ilustrace, reklamy a jiné grafické části stránky budou vyjádřeny v tagu /alto/Layout/Page/PrintSpace/ComposedBlock/ s vyjádřením atributu TYPE, který bude označovat typ bloku (illustration, advertisement aj.)
    - např. ilustrace bude popsána v elementu /alto/Layout/Page/PrintSpace/ComposedBlock/GraphicalElement, kde ComposedBlock TYPE je Illustration
    - reklama s textem v rámečku bude popsána v elementu Layout/Page/PrintSpace/ComposedBlock/TextBlock, kde ComposedBlock TYPE je Advertisement
    - tabulky, grafy obdobně

- elementy /alto/Layout/Page/PrintSpace/ComposedBlock/Illustration a Layout/Page/PrintSpace/ComposedBlock/ComposedBlock také nebudou využity
- /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock a /alto/Layout/Page/PrintSpace/ComposedBlock/GraphicalElement nebudou obsahovat elementy <Shape>; tvar těchto bloků je vyjádřen v elementu <Shape> samotného elementu <ComposedBlock>; logicky pak souřadnice tvaru <TextBlock> nebo <GraphicalElement> obsaženého v /alto/Layout/Page/PrintSpace/ComposedBlock jsou většinou shodné, pokud není tvarů nebo bloků v rámci /alto/Layout/Page/PrintSpace/ComposedBlock více
- všechny vyplněné hodnoty jsou příklady plnění, plnění v konkrétní instituci je nutno specifikovat vlastními pravidly a kontrolovanými slovníky
- ALTO XML bude využíváno pro tzv. pořadí čtení, tj. např. článek vyskytující se na více stránkách nebo na více různých místech jedné stránky bude možné zobrazit celý a ve správném pořadí. K tomu je nutno znát jeho strukturu. Struktura bude vyjádřena v korespondujícím METS záznamu v logické strukturální mapě. Ta bude obsahovat odkazy na jednotlivé textové bloky článku, pomocí ID textových bloků použitých v ALTO XML.

**Obsah pole "Popis":**

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné
- výskyt elementu (jak je definováno formátem ALTO XML – dle XSD)
    - 0-1 element je nepovinný, neopakovatelný
    - 1-1 element je povinný, neopakovatelný
    - 0-n element je nepovinný, opakovatelný

| Element | Atribut | Popis | Povinnost |
|---|---|---|---|
| <Description> | | | |
| <MeasurementUnit> | | měřící jednotka pro souřadnice v ALTO XML; možné hodnoty – dpi, pixel, inch1200 a mm10); inch1200 = 1/1200 inche; doporučené plnění je „mm10“ nebo „pixel“; 1-1 | M |
| <sourceImageInformation> | | informace o obrazovém souboru, ze kterého vzniklo ALTO XML; 1-1 | MA |
| <fileName> | | jméno obrazového souboru, ze kterého bylo ALTO XML vytvářeno; ideálně i s filesystem cestou jeho uložení; např. nlaImageSeq-33386-b.tif//produkce/OCR/digibok_XY/XY_011.tiff 1-1 | M |

| | | | |
|---|---|---|---|
| <fileIdentifier> | | jedinečný identifikátor obrazového souboru;<br>0-n | R |
| <OCRProcessing> | ID | popis procesu vzniku OCR;<br>1-n<br>---<br>ID OCR procesu, např. <OCRProcessing ID="OCRPROCES_1">;<br>povinné | M |
| <preProcessingStep> | | procesy před vznikem OCR, které provádí SW pro OCR (např. natočení obrazu)<br>1-n | MA |
| <processingDateTime> | | určení času procesu, který předcházel samotnému OCR;<br>např. 2008-03-29T19:42:23 dle ISO 8601 na úroveň vteřin;<br>1-1 | O |
| <processingAgency> | | jméno nebo kód instituce, např. NK CZ, název externí firmy apod.;<br>doporučujeme použít kontrolovaný slovník hodnot;<br>0-1 | R |
| <processingStepDescription> | | popis procesu (např. zarovnání, ořez apod.);<br>0-n | O |
| <processingStepSettings> | | nastavení kroku popsaného v <processingStepDescription>, např. CCS OCR Processing Filter<br>0-1 | O |
| <processingSoftware> | | popis SW, který upravoval obrázek před vznikem OCR;<br>1-1 | MA |
| <softwareCreator> | | výrobce softwaru - např. CCS Content Conversion Specialists GmbH, Germany;<br>1-1 | M |
| <softwareName> | | jméno softwaru - např. CCS docWORKS;<br>1-1 | M |
| <softwareVersion> | | verze SW, např. 6.2-1.16;<br>1-1 | M |
| <ocrProcessingStep> | | popis procesu vzniku OCR<br>1-1 – povinné pole | M |
| <processingDateTime> | | okamžik kdy bylo OCR vytvořeno;<br>nutno zapsat v ISO 8601 na úroveň vteřin;<br>0-1 | M |

| | | | |
|---|---|---|---|
| <processingAgency> | | jméno nebo kód instituce, např. NK CZ<br>doporučujeme použít kontrolovaný slovník hodnot;<br>0-1 | M |
| <processingSoftware> | | popis SW, který dělal vlastní OCR;<br>1-1 | M |
| <softwareCreator> | | výrobce softwaru - např. ABBYY, Russia;<br>1-1 | M |
| <softwareName> | | jméno softwaru - např. FineReader;<br>1-1 | M |
| <softwareVersion> | | např. 8.0;<br>1-1 | M |
| <Styles> | | styly definují vlastnosti jednotlivých grafických prvků stránky.<br>styl definovaný v elementu vrchní úrovně je použit jako výchozí pro podřízené elementy;<br>1-1 | MA |
| <TextStyle> | ID<br>FONTSTYLE<br>FONTFAMILY<br>FONTSIZE | definuje font textu;<br>1-n<br>---<br>ID pro každý text style použitý v OCR souboru – povinné<br><br>FONTSTYLE – např. bold, italics apod.; doporučujeme používat kontrolovaný slovník; doporučené<br><br>FONTFAMILY – např. arial, calibri apod.; doporučujeme používat kontrolovaný slovník; povinné<br><br>FONTSIZE – velikost fontu, např. 10, 12 apod.; povinné | MA |
| <ParagraphStyle> | ID<br>ALIGN | definuje formátování textových bloků;<br>1-n<br>---<br>ID pro každý odstavec + zarovnání;<br>např. PAR_01, PAR_02 apod.<br>povinné<br><br>ALIGN – zarovnání; povolené hodnoty: Left, Right, Center, Block aj.;<br>povinné | MA |

| | | | |
|---|---|---|---|
| <Layout> | | layout - rozložení struktur (slov, odstavců apod.) na jedné stránce dokumentu;<br>1-1 povinný výskyt<br>element není opakovatelný | M |
| <Page> | ID<br>ACCURACY<br>POSITION<br>QUALITY<br>PHYSICAL_IMG_NR<br>HEIGHT<br>WIDTH<br>PC | element popisující jednu stránku dokumentu;<br>1-n<br>---<br>ID – vygenerovaný identifikátor stránky, např. PAGE1, nebo P1 apod.;<br>povinné<br><br>ACCURACY – procentuální odhad přesnosti OCR (0-100);<br>doporučené<br><br>POSITION – pozice stránky; hodnoty k plnění: Left, Right, Foldout, Single, Cover;<br>nepovinné<br><br>QUALITY – krátký údaj o kvalitě předlohy stránky; hodnoty k plnění: OK, Missing, Missing in original, Damaged, Retained, Target, As in original;<br>nepovinné<br><br>PHYSICAL_IMG_NR - fyzické (pořadové) číslo stránky v dokumentu; vyjádřeno číslem, např. 1,2,3 apod.;<br>povinné<br><br>WIDTH – šířka stránky vyjádřená v pixelech;<br>povinné<br><br>HEIGHT – výška stránky vyjádřená v pixelech;<br>povinné<br><br>PC - Confidence level OCR souboru – hodnota mezi 0 (nejistá kvalita) a 1 (dobrá kvalita);<br>nepovinné; pokud nevyplníte ACCURACY – tak je vyplnění doporučené | M |

| | | | |
|---|---|---|---|
| <TopMargin> | ID<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | horní okraj – prostor mezi vrchní hranou listu a vrchní linkou textu;<br>0-1<br>---<br>ID: unikátní ID pro element TopMargin, např. P1_TM0001 (page 1, topMargin0001);<br>povinné<br><br>HPOS: horizontální pozice;<br>povinné<br><br>VPOS: vertikální pozice;<br>povinné<br><br>WIDTH – šířka vrchního okraje;<br>povinné<br><br>HEIGHT – výška vrchního okraje;<br>povinné | M |
| <TextBlock> | | stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | MA |
| <LeftMargin> | ID<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | levý okraj – prostor mezi levým okrajem stránky a textem;<br>0-1<br>---<br>ID: unikátní ID pro element LeftMargin, např. P1_LM0001 (page 1, leftMargin0001);<br>povinné<br><br>HPOS: horizontální pozice;<br>povinné<br><br>VPOS: vertikální pozice;<br>povinné<br><br>WIDTH – šířka levého okraje;<br>povinné<br><br>HEIGHT – výška levého okraje;<br>povinné | M |
| <TextBlock> | | stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | MA |

| | | | |
|---|---|---|---|
| <RightMargin> | ID<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | pravý okraj – prostor mezi pravým okrajem stránky a textem;<br>0-1<br>----------<br>ID: unikátní ID pro element RightMargin, např. P1_RM0001 (page 1, rightMargin0001);<br>povinné<br><br>HPOS: horizontální pozice;<br>povinné<br><br>VPOS: vertikální pozice;<br>povinné<br><br>WIDTH – šířka pravého okraje;<br>povinné<br><br>HEIGHT – výška pravého okraje;<br>povinné | M |
| <TextBlock> | | stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | MA |
| <BottomMargin> | ID<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | pravý okraj – prostor mezi spodním okrajem stránky a textem;<br>1-1<br>----------<br>ID: unikátní ID pro element BottomMargin, např. P1_BM0001 (page 1, bottomMargin0001);<br>povinné<br><br>HPOS: horizontální pozice;<br>povinné<br><br>VPOS: vertikální pozice;<br>povinné<br><br>WIDTH – šířka spodního okraje;<br>povinné<br><br>HEIGHT – výška spodního okraje;<br>povinné | M |
| <TextBlock> | | stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | MA |

| | | | |
|---|---|---|---|
| <PrintSpace> | ID<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | popis tvaru pokrývajícího textové pole stránky;<br>1-1<br>----------<br>ID: unikátní ID pro element <printSpace>, např. P1_PS0001 (page 1, printSpace0001); - povinné<br><br>HPOS: horizontální pozice;<br>povinné<br><br>VPOS: vertikální pozice;<br>povinné<br><br>WIDTH – šířka textového pole;<br>povinné<br><br>HEIGHT – výška textového pole;<br>povinné | M |
| <TextBlock> | ID<br>STYLEREFS<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | popisy textových bloků na konkrétní stránce;<br>1-n<br>pokud je stránka prázdná, TextBlock není potřeba uvádět;<br>pokud je na stránce text tak ano<br>----------<br>ID obsahuje identifikátor textového bloku na stránce, např. "BLOCK1" nebo P1_TB0002 (stránka 1, textový blok 2);<br>povinné<br><br>STYLEREFS: reference na ID definice formátování textových bloků <ParagraphStyle>;<br>povinné<br><br>HPOS: horizontální pozice bloku;<br>povinné<br><br>VPOS: vertikální pozice bloku;<br>povinné<br><br>WIDTH – šířka textového bloku;<br>povinné<br><br>HEIGHT – výška textového bloku;<br>povinné | MA |
| <Shape> | | tvar textového bloku;<br>0-1 – pro jeden výskyt <TextBlock> jeden nebo žádný výskyt <Shape>;<br>plnit v případě, že je tvar textového bloku nestandardní (víceúhelník) | RA |

| | | | |
|---|---|---|---|
| <Polygon> | POINTS | popis (souřadnice) tvaru víceúhelníku;<br>0-1<br>---<br>POINTS – vyjádření jednotlivých bodů víceúhelníku;<br>povinné | M |
| <TextLine> | ID<br>STYLEREFS<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | popis jedné řádky textu v rámci textového bloku;<br>1-n<br>nutný alespoň jeden výskyt v rámci textového bloku<br>---<br>ID obsahuje identifikátor řádky textu v textovém bloku, např. "P1_TL0002 (stránka 1, řádka 2);<br>povinné<br><br>STYLEREFS: reference na ID definice formátování textových bloků <ParagraphStyle>;<br>nepovinné<br><br>HPOS: horizontální pozice řádky;<br>povinné<br><br>VPOS: vertikální pozice řádky;<br>povinné<br><br>WIDTH – šířka řádky;<br>povinné<br><br>HEIGHT – výška řádky;<br>povinné | M |
| <String> | ID<br>CONTENT<br>HEIGHT<br>WIDTH<br>HPOS<br>VPOS<br>CC<br>WC<br><br>V případě dělení slov také:<br>SUBS_TYPE<br>SUBS-CONTENT | řetězec znaků – vlastní obsah OCR;<br>znaky tvoří jednotlivá slova a více tagů <String> větu <TextLine>;<br>1-n v rámci <TextLine><br>---<br>ID obsahuje unikátní sekvenční číslo řetězce na stránce, např. "P3_ST0001" (strana 3, řetězec 1);<br>povinné<br><br>CONTENT – ukládá vlastní řetězec znaků (slovo);<br>povinné<br><br>HPOS: horizontální pozice řetězce;<br>povinné<br><br>VPOS: vertikální pozice řetězce;<br>povinné | M |

| | | | |
|---|---|---|---|
| | | WIDTH – šířka řetězce;<br>povinné<br><br>HEIGHT – výška řetězce;<br>povinné<br><br>CC – úroveň důvěry v přesnost OCR rozpoznání každého znaku v řetězci;<br>jde o seznam čísel, každé z nich mezi hodnotami 0 (jistá) a 9 (nejistá) pro každý znak;<br>např. CC="0001" pro CONTENT="TEXT";<br>nepovinné<br><br>WC – úroveň důvěry v přesnost OCR výstupu celého řetězce - slova (word confidence);<br>hodnota mezi 0 (nejistá) a 1 (jistá);<br>např. WC="0,99";<br>nepovinné<br><br>SUBS_CONTENT – obsah chybějící části řetězce v případě, že je slovo na konci řádku rozdělené i do druhého řádku;<br>obsahuje celý řetězec - aby byl vyhledatelný i v případě, že slovo se na stránce vyskytuje, ale je rozděleno;<br>povinné<br><br>SUBS_TYPE – označení typu substituce;<br>možné hodnoty: HypPart1; HypPart2; Abbreviation;<br>povinné - při výskytu SUBS_CONTENT<br><br>HypPart1 se vyskytuje při rozdělení slova u jeho první OCR části (u první části tagu <CONTENT> ve větě (stringu) první;<br>HypPart2 se vyskytuje u následujícího tagu <CONTENT> v následující větě (stringu), který obsahuje druhou část rozděleného slova/řetězce;<br>Abbreviation – typ substituce používaný při rozepisování zkratek v textu na jejich plný text;<br>při dělení slov v textu HypPart1 a HypPart2 povinné, abbreviation nepovinné | |
| <ALTERNATIVE> | | alternativní hodnota OCR řetězce pro jednotlivá slova;<br>0-n<br>lze použít v případě nejistoty rozpoznání řetězce; | O |

| | | | |
|---|---|---|---|
| <HYP> | CONTENT<br>WIDTH<br>HPOS<br>VPOS | zápis znaku rozdělovníku slov<br>0-1 pro jeden výskyt <TextLine>; vždy pro poslední <String>;<br>může se vyskytnout pouze na konci řádku (1x)<br>----------<br>CONTENT – obsahuje řetězec znaků, které jsou v textu použity na rozdělení slova, nejčastěji „-“;<br>povinné<br><br>WIDTH – šířka dělícího znaku;<br>doporučené<br><br>HPOS: horizontální pozice dělícího znaku;<br>doporučené<br><br>VPOS: vertikální pozice dělícího znaku;<br>doporučené | MA |
| <SP> | ID<br>WIDTH<br>HPOS<br>VPOS | prázdný prostor mezi řádky;<br>0-n v rámci jednoho <TextLine>;<br>vždy mezi řádky, tj. mezi tagy <String>;<br>----------<br>ID: unikátní ID pro prázdný prostor mezi řádky, např. P1_SP0001 (stránka 1, prázdný prostor 0001);<br>povinné<br><br>HPOS: horizontální pozice;<br>povinné<br><br>VPOS: vertikální pozice;<br>povinné<br><br>WIDTH – šířka prázdného prostoru;<br>povinné | M |

| | | | |
|---|---|---|---|
| <ComposedBlock> | ID<br>TYPE<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT<br>STYLEREFS | blok sestávající z jiných bloků;<br>může obsahovat<br>PrintSpace/ComposedBlock/TextBlock,<br>PrintSpace/ComposedBlock/Illustration,<br>PrintSpace/ComposedBlock/GraphicalElement,<br>/PrintSpace/ComposedBlock/ComposedBlock, tj. stejné elementy (bloky), které obsahuje samotný element /alto/Layout/Page/PrintSpace;<br>0-n<br>povinné pro vyjádření bloků textu (např. orámovaný text, reklamy), pro vyjádření ilustrací, tabulek a grafik<br>---<br>ID: unikátní ID komponovaný blok, např. P6_CB0001 (stránka 6, komponovaný blok 0001);<br>povinné<br><br>TYPE – označení typu komponovaného bloku; nutné používat kontrolovaný slovník (illustration, Advertisement, apod.);<br>povinné<br><br>HPOS: horizontální pozice bloku;<br>povinné<br><br>VPOS: vertikální pozice bloku;<br>povinné<br><br>WIDTH – šířka komponovaného bloku;<br>povinné<br><br>HEIGHT – výška komponovaného bloku;<br>povinné | MA |
| <Shape> | | tvar komponovaného bloku;<br>0-1 – pro jeden výskyt /alto/Layout/Page/PrintSpace/ComposedBlock jeden nebo žádný výskyt /alto/Layout/Page/PrintSpace/ComposedBlock/Shape;<br>doporučeno – v případě, že je tvar komponovaného bloku nestandardní (víceúhelník) | RA |
| <Polygon> | POINTS | popis tvaru víceúhelníku;<br>1-1<br>---<br>POINTS – vyjádření jednotlivých bodů víceúhelníku<br>povinné | M |

| | | | |
|---|---|---|---|
| <TextBlock> | ID<br>STYLEREFS<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | v případě, že komponovaný blok (např. orámovaný tvar) obsahuje text;<br>platí stejná pravidla jako pro normální element /alto/Layout/Page/PrintSpace/TextBlock;<br>0-n (pro jeden výskyt <ComposedBlock> 0 nebo více elementů /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock>;<br>plnit pokud je v komponovaném bloku text<br>---<br>ID obsahuje identifikátor textového bloku v komponovaném bloku, např. P1_CB0002_SUB (stránka 1, textový blok 2, SUB značí komponovaný blok);<br>povinné<br><br>STYLEREFS: reference na ID definice formátování textových bloků /alto/Styles/ParagraphStyle;<br>povinné<br><br>HPOS: horizontální pozice bloku;<br>povinné<br><br>VPOS: vertikální pozice bloku;<br>povinné<br><br>WIDTH – šířka textového bloku;<br>povinné<br><br>HEIGHT – výška textového bloku;<br>povinné | MA |
| <TextLine> | | /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock/TextLine a ostatní elementy v rámci /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock mají stejná pravidla a výskyty jako jako ve vrchním elementu /alto/Layout/Page/PrintSpace/TextBlock | |

| | | | |
|---|---|---|---|
| <GraphicalElement> | ID<br>HPOS<br>VPOS<br>WIDTH<br>HEIGHT | popis grafického tvaru;<br>v případě využití v rámci /alto/Layout/Page/PrintSpace/ComposedBlock označuje rozměry tvaru v rámci něhož je tabulka, ilustrace, reklama apod.;<br>0-1 - pro jeden výskyt /alto/Layout/Page/PrintSpace/ComposedBlock 0 nebo max. 1 výskyt <GraphicalElement>;<br>plní se, pokud je na stránce a tedy v komponovaném bloku ilustrace, tabulka apod.;<br>---<br>ID – identifikátor grafického tvaru;<br>povinné<br><br>HEIGHT – výška grafického tvaru;<br>povinné<br><br>WIDTH – šířka grafického tvaru;<br>povinné<br><br>HPOS – horizontální pozice grafického tvaru;<br>povinné<br><br>VPOS – vertikální pozice grafického tvaru;<br>povinné | MA |

# 8 Autorsko-právní metadata

- Autorsko-právní metadata jsou kompletně nepovinná.
- V případě rozhodnutí o vytvoření autorsko-právních metadat platí povinnost elementů uvedených ve sloupci povinnost.
- V případě nevytvoření autorsko-právních metadat je status defaultně považován za neznámý (*unknown*).
- Autorsko-právní metadata mohou být vyplněna jak pro celý dokument (hlavní záznam METS) tak i pro konkrétní části dokumentu (vedlejší záznam amd_mets). Platí zde pravidlo hierarchické dědičnosti: v případě, že má vedlejší záznam vyplněné autorko-právní metadata, tak jsou relevantní právě ta. V případě, že je vedlejší záznam neobsahuje, tak je přebírá z hlavní záznamu METS. Naopak to neplatí. Hlavní záznam nepřebírá autorsko-právní metadata z vedlejších záznamů.
- Přítomnost autorsko-právních metadat je třeba zohlednit ve fyzické strukturální mapě.

**Opakovatelnost elementů:**
0 - 1 nepovinný a neopakovatelný
0 - n nepovinný a opakovatelný

**Část METS pro autorsko-právní metadata**

| element | atributy | popis | Povinnost |
|---|---|---|---|
| <amdSec> | | element obsahující autorsko-právní metadata ve formátu ContainerMD | M |
| | ID | Použití jak u hlavního METS, tak vedlejšího záznamu amd_mets<br><br>ID – identifikátor konkrétní části <amdSec>, např. pro svazek "AMD_MONOGRAPH_0001", pro stránku 1 hodnota "PAGE_0001", pro stránku 2 "PAGE_0002" atd. | |
| <rightsMD> | | element pro typ autorsko-právních metadat | M |
| | ID | **ID pro část <rightsMD>: např. "RIGHTS_VOLUME", RIGHTS_0001; …** | |
| <mdWrap> | MDTYPE | MDTYPE – hodnota "OTHER" | M |
| | OTHERMDTYPE | OTHERMDTYPE – hodnota "CopyrightMD" | |
| | MIMETYPE | MIMETYPE – hodnota "text/xml" | |
| <xmlData> | | | M |

**Nepovinné atributy:**
U některých elementů mohou být použity nepovinné atributy, které upřesní informaci v elementu.

1. *iso.code:* elementy: <country.publication>, <country.creation> - do atributu se zapisuje standardizovaný kód země podle ISO3166-1 (alpha 2-letter code list[3]) (viz příklad)
2. *year.type:* elementy: <year.copyright> <year.renewal> <year.publication> <year.creation> <year.birth> <year.death> - do atributu se zapisuje přesnost uvedeného roku. Možné hodnoty *exact* (rok je známý), *approximate* (odhadovaný rok na základě jiných zdrojů), *unknown* (rok není znám a není možné jej odhadnout/dohledat)

| Element | Popis | Povinnost |
|---|---|---|
| **<copyright>** | Povinný kořenový element. Označuje status dokumentu**. Povinné jsou i oba atributy:**<br><br>copyright.status<br>● Copyrighted (autorsky chráněný document)<br>● pd (volně dostupný dokument)<br>● pd_expired (volně dostupný dokument, kterému vypršelo trvání majetkových práv)<br>● unknown (status neznámý)<br><br>publication.status<br>● published (publikovaný dokument)<br>● unpublished (nepublikovaný dokument)<br>● unknown (informace o publikování není známa)<br>**0-1** | **M** |
| <creation> | Informace o vytvoření dokumentu.<br>0-1 | O |
| <year.creation> | Rok vytvoření dokumentu ve formátu YYYY.<br>0-1 | RA |
| <country.creation> | Země, ve které byl dokument vytvořen: vyjádřeno iso kódem (viz Nepovinné atributy) př.: <country.creation iso.code="cz">Česká republika</country.creation><br>0-1 | RA |
| <creator> | Informace o tvůrci dokumentu.<br>0-1 | MA |

[3]Dostupné z: http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html

| | | |
|---|---|---|
| <creator.corporate> | Nátev instituce, pokud je autorem dokumentu.<br>0-n | MA |
| <creator.person> | Informace o autorovi dokumentu.<br>0-n | MA |
| <name> | Jméno a přijmení autora.<br>0-1 | MA |
| <year.birth> | Datum narození autora. Formát YYYY.<br>0-1 | MA |
| <year.death> | Datum úmrtí autora. Formát YYYY.<br>0-1 | MA |
| <note> | Doplňující informace k tvůrci dokumentu.<br>0-n | O |
| <publication> | Informace o vydání dokumentu<br>0-1 | MA |
| <country.publication> | Země vydání – vyjádřeno iso kódem (viz Nepovinné atributy), př:<br><country.publication iso.code="cz">Česká republika</copyrightMD.country.publication><br>0-1 | R |
| <publisher> | Nakladatel.<br>0-1 | R |
| <year.publication> | Rok vydání. Formát YYYY.<br>0-1 | MA |
| <year.copyright> | Rok copyrightu. Formát YYYY.<br>0-1 | RA |
| <year.renewal> | Rok případného obnovení copyrightu. Formát YYYY.<br>0-1 | RA |
| <note> | Doplňující informace k vydání dokumentu.<br>0-n | O |
| <rights.holder> | Informace o držiteli práv. Používá se v případě, že práva drží někdo jiný než autor nebo nakladatel. Např. občanské sdružení autorů apod.<br>0-1 | O |
| <contact> | Kontakt na držitele práv.<br>0-n | O |
| <name> | Jméno nebo název držitele práv.<br>0-1 | O |

| | | |
|---|---|---|
| <note> | Doplňující informace o držiteli práv.<br>0-n | O |
| <notice> | Do pole se vkládá oznámení o copyrightu, tak jak je uvedeno v dokumentu.<br>0-1 | O |
| <general.note> | Pole slouží k vložení jakékoliv další informace, která je relevantní vzhledem k autorským právům dokumentu a nemohla být zařazena do žádného jiného pole.<br>0-n | O |