

Definice metadatových formátů pro digitalizaci monografií

- vychází z obecné specifikace metadat v projektu NDK

Jan Hutař, NK ČR

VERZE 0.3 - 24.11.2011

| jméno | datum | verze dokumentu | provedené změny |
|--------------------------|--------------------|-----------------|--|
| Jan Hutař, Přemysl Tvrdý | 22.- 25.11.2011 | 0.3 | další úpravy |
| Jan Hutař – NK | 1.- 21.11.2011 | 0.3 | úpravy celého textu, logické struktury a specifikace jednotlivých metadatových schémat |
| Jan Hutař – NK | únor-říjen 2011 | 0.2 | změny v závislosti na vývoji specifikace metadat pro periodika |
| Jan Hutař – NK | 1.2.2011 | 0.1 | první verze pro zadávací dokumentaci NDK projektu |

| | | |
|-----|--|----|
| 1. | Východiska | 3 |
| 2. | Výstupy digitalizace..... | 4 |
| 3. | Granularita metadatového záznamu | 4 |
| 4. | Identifikátory..... | 5 |
| 5. | Struktura PSP balíčku | 6 |
| 6. | Názvová konvence složek a souborů | 9 |
| 7. | Transportní balík pro jeden nebo více PSP balíčků | 11 |
| 8. | Metadata..... | 11 |
| 8.1 | Kořenový element hlavního METS záznamu | 12 |
| 8.2 | METS hlavička <metsHdr>..... | 12 |
| 8.3 | METS část <dmdSec> - Bibliografická metadata – formát MODS a Dublin Core | 13 |
| 8.4 | METS část <amdSec> - Technická a administrativní metadata – formáty MIX a PREMIS | 38 |
| 8.5 | METS část <fileSec>..... | 55 |
| 8.6 | METS část <structMap> - Strukturální metadata a ALTO XML | 57 |
| 8.7 | OCR (ALTO XML a TXT OCR) | 65 |

1. Východiska

- uživatelské kopie = UC
- archivní kopie = MC
- původní sken – PS – obrazový soubor vzniklý při digitalizaci, který se po zpracování (ořez, narovnání apod.), maže a dále se neukládá
- u všech metadatových formátů budou použity verze aktuální v době implementace projektu NDK, nebo verze předchozí v případě, že nová verze je nová min. 3 měsíce
- základní intelektuální entita ve workflow digitalizace a následně i v LTP systému = svazek monografie (viz dále).
- PSP balíček – producer submission package
 - o balíček dat a metadat, který přichází od producenta dat (tedy např. z workflow digitalizace
 - o PSP balíček bude obsahovat kompletní intelektuální entitu tj. svazek monografie
 - o z workflow digitalizace lze poslat více PSP balíčků v balíku např. [.tar] apod.
 - o pokud má dvousvazkové dílo v katalogu knihovny bibliografický záznam pro každý svazek, vznikne pro každý svazek PSP balíček a každý svazek bude brán jako jedna intelektuální entita; to samé platí i pro případ, že vícesvazkové dílo má pouze jeden záznam
- SIP balíček – submission information package – je balíček dat a metadat v podobě, ve které je akceptovatelný pro LTP systém nebo pro aplikaci zpřístupnění. Může vznikat v transformačním modulu z PSP balíčku po jeho kontrolách. SIP by měl obsahovat jen jednu intelektuální entitu.
- primárně se u monografií nebude dělat členění na vnitřní části (kapitoly apod.) – pouze u některých zvláště důležitých monografií > tj. musí existovat možnost vyjádřit popis částí (např. kapitoly, přílohy apod.) v metadatach
- základní bibliografická metadata budou stahována přímo z knihovních katalogů do workflow digitalizace
- převod stávajících dat, která jsou v proprietárním DTD (periodika a monografie) do vnitřního formátu LTP systému bude probíhat v transformačním modulu
- i u nových dat z digitalizace bude nutný převod do interních formátů LTP systému a aplikací zpřístupnění – opět bude probíhat v modulu transformace
- úpravy obrazu, které vedou ke změně rozměrů obrazu, rozlišení apod., se musí dělat před tím, než se udělá OCR, tj. budou se dělat 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 pasovalo
- jako výchozí SW pro vytváření souborů JPEG2000 se bude používat Kakadu
- veškerá metadata musí pro zápis používat kodování UTF-8

2. Výstupy digitalizace

1. archivní kopie (1 MC pro každou stránku)
 2. uživatelské kopie (1 UC pro každou vzniklou MC, tedy stránku)
 3. OCR - ALTO XML soubor pro každou stránku
 4. OCR TXT soubor - pro možnost stáhnout si jen text dokumentu (tam kde kvalita OCR je odpovídající), vyhledávání/indexace.
 5. metadata pro MC
 - a. bibliografická metadata – MODS a DC
 - b. strukturální metadata – METS
 - c. technická metadata – MIX, PREMIS
 - d. administrativní metadata – PREMIS, METS
 6. kontrolní metadatové soubory (s kontrolními součty a údaji o vzniku dat apod.)
- Pozn.
 - Záznam METS nebude obsahovat žádná metadata pro uživatelské kopie. Tj. METS neobsahuje popisná, ani technická metadata pro UC. Obrazové soubory UC nejsou ani součástí strukturální mapy <structMap> ani <fileSec>. Součástí PSP balíčku budou jen obrazy UC ve složce [userCopy]. Důvodem je to, že metadata pro UC budou vytvářena na vstupu do Krameria4 ve formátu FOXML (Fedora Object XML). Budou se vyrábět z METS záznamu pro MC, jehož specifikace je níže.

3. Granularita metadatového záznamu

Monografie

- základní intelektuální entitou pro monografie je 1 svazek
- pokud má monografie pouze jeden svazek, vznikne jeden metadatový popis (=METS záznam)
- pokud má monografie svazky dva, např. dvousvazkový slovník, jedná se o dvě intelektuální entity (svazek první a svazek druhý) a vzniknou tedy dva metadatové záznamy, ke každému svazku jeden METS záznam a tedy dva PSP balíčky
- v knihovních katalozích jsou někdy vícesvazkové monografie katalogizovány jako jeden soubor, tj. mají jeden záznam v katalogu, někdy jsou jednotlivé díly vedeny jako jednotlivé záznamy v katalogu; obou případech musí vzniknout metadatový popis ke každému svazku jako základní intelektuální entitě a také PSP balíček pro každý svazek
- každý METS záznam musí obsahovat metadata o nadřazené intelektuální entitě (např. soubor monografie) tak, aby bylo možné obě entity virtuálně spojit a bylo jasné uživateli, že se jedná o jeden soubor/titul, který má dva svazky.

4. Identifikátory

Do workflow digitalizace budou přicházet bibliografická metadata, která již budou obsahovat následující identifikátory vrchních úrovní intelektuálních entit (úroveň titulu):

- ISBN – pouze pro titul monografie (jednosvazkové), nebo pro soubor monografií, které mají pouze jeden souborný záznam, ISBN není přiděleno vždy
- ISSN
- ČČNB – identifikátor entity tak jak odpovídá katalogizačnímu záznamu, tj. každá entita se záznamem v katalogu NK/MZK má tento identifikátor
- URN:NBN pro svazek monografie, které bude přidělováno během digitalizace¹.
 - o URN:NBN může být přidělováno také nižším logickým úrovním (entitám) – tedy vnitřní část (zejména např. článek ve sborníku), příloha
 - o syntax URN:NBN musí odpovídat specifikaci identifikátoru URN:NBN pro resolver NK (např. urn:nbn:cz:ndk-123456 pro výstupy z projektu NDK)
- další možností je využití UUID

¹ buď přímo v SW pro workflow digitalizace, nebo za pomoci aplikace Resolver URN:NBN

5. Struktura PSP balíčku

| složka | obsahuje > | obsahuje > | obsahuje> |
|---------------------------------------|--|--|--|
| svazek monografie /číslo periodika | | | |
| | info.xml | údaje o vzniku balíku | |
| složka [masterCopy] | | obrazy JPEG2000 lossless | |
| složka [userCopy] | | obrazy JPEG2000 lossy | |
| složka [ALTO] | | soubory ALTO XML | |
| složka [TXT] | | soubory OCR.TXT | |
| složka [amdSec] | AMD_METS.xml soubor pro každou stránku obsahuje> | amdSec | techMD = PREMISobject pro MC, původní TIFF, ALTO XML) + MIX pro MC, původní TIFF) digiprovMD = PREMISevent + PREMISagent |
| | | fileSec | odkazuje na MC, ALTO XML, OCR TXT soubor popisované 1 stránky |
| | | StructMap | pouze fyzická - pro soubory popisované stránky (MC a ALTO XML, OCR TXT) |
| Hlavní_METS.xml | dmdSec | MODS a DC pro jednotlivé úrovně dokumentu | |
| | fileSec | obsahuje linky na MC, ALTO XML, OCR TXT a technická metadata ve složce [amdSec] | |
| | structMap (včetně ALTO odkazů) | logická a fyzická pro MC, ALTO XML areas, OCR TXT a AMD_METS.xml | |
| MD5 | | kontrolní součty všech souborů v PSP balíku | |

Jedná se o variantu, kdy technická a administrativní metadata nejsou obsažena v hlavním METS záznamu, ale pro každou stránku v jiném dalším METS záznamu (AMD_METS.xml). Důvodem je to, že pokud by bylo vše v hlavním METSu, byl by neúměrně dlouhý. Takto je z hlavního záznamu nalinkováno.

Hlavní metadatový záznam METS obsahuje metadata pro archivní kopie obrazových dat, neobsahuje však žádná metadata pro uživatelské kopie.

PSP balíček = 1 složka pro svazek monografie. V případě, že má monografie 2 svazky/díly, tak 1 svazek = 1 PSP.

Hlavní složka PSP balíčku obsahuje následující složky a soubory:

soubor info.xml

velmi krátce tu budou zaznamenány údaje o vzniku celého PSP balíčku – kdo, kdy ho vytvořil, jakou měl velikost, odkud kam byl nakopírován apod. Obsahovat by také měl informaci o stavu zpracování balíčku.

Zaznamenány by také měly být údaje o obsahu PSP balíčku – počet a názvy souborů apod. Soubor info.xml by také mohl být vedle hlavního PSP balíčku. Údaje a struktura info.xml souboru:

1. vznik balíčku – datum dle ISO8601 na úroveň vteřin
2. ID balíčku – použit identifikátor svazku monografie (URN:NBN) – viz názvová konvence v kap. 6
3. ID titulu - čČNB, ISBN (bez pomlček), URN:NBN nebo ISSN
4. údaje o větším celku, do kterého balíček patří - např. digitalizace pro ANL
5. název instituce, která je zadavatelem digitalizace
6. tvůrce balíčku – kód instituce (firmy), která balíček vytvořila
7. velikost balíčku – v kB
8. z jakého serveru bylo nahráno – URL
9. obsah balíčku
 - názvy souborů včetně directory path a koncovky (mime type)
 - počet souborů v balíčku celkem
10. stav zpracování – možné hodnoty
 - hotovo
 - opraveno
 - added OCR
 - added titles
 - added logical parts (issues, years)
 - updated xml (MODS, DC, identifikátory),
11. poznámka – např. o tom, že balíček neobsahuje OCR apod.

příklad balíčku, který obsahuje 2 soubory, jeden v rootu a druhý ve složce:

1. CREATED=2009-11-10T12:37:46
2. PACKAGEID=NDK_123456
3. TITLEID=ISBN1234567890
4. COLLECTION=NDK
5. INSTITUTION=NKP
6. CREATOR=NazevFirmy
7. PACKAGESIZE=36000155kb
8. SOURCELOCATION= server123.firma.cz/baliky_hotovo/01/2011/12/000025456
9. ITEMLIST=scan01.jp2
ITEMLIST=slozka/hotovo/27.9.2011/scan02.jp2
ITEMTOTAL=2
10. STATUS=hotovo
11. NOTE=noOCR

složka [masterCopy]

složka s master kopiemi, obsahuje soubory JPEG2000 v neztrátové kompresi, 1 soubor = 1 stránka, tj. obsahuje všechny naskenované stránky monografie

složka [userCopy]

složka s uživatelskými kopiemi, pro každou naskenovanou stránku monografie obsahuje jeden JPEG2000 soubor se ztrátovou kompresí

složka [ALTO]

obsahuje ke každé stránce 1 ALTO XML soubor, tj. tolik ALTO XML souborů kolik je stránek svazku monografie.

složka [TXT]

obsahuje ke každé stránce 1 OCR soubor jako čistý text. Tj. tolik OCR.TXT souborů kolik je stránek svazku monografie.

složka [amdSec]

složka s technickými metadaty – **obsahuje pro každou naskenovanou stránku monografie 1 METS soubor (AMD_METS.xml)**. Záměrně nejsou tato metadata v hlavním METS záznamu (hlavni_METS.xml), protože ten by neúměrně narostl a bylo by obtížné s ním pracovat. Musí z něj být ovšem nalinkována (z části fileSec). Každý METS soubor AMD_METS.xml obsahuje následující části METS formátu:

- amdSec – administrativní metadata – obsahuje část
 - o technických metadat (techMD), která ve formátu PREMISobject popisuje vlastnosti archivních kopií, uživatelských kopií, ALTO XML, původního TIFF souboru, ze kterého vznikly archivní kopie. Dále je přítomen záznam technických metadat v MIX formátu pro archivní kopie a pro původní TIFF.
 - o metadat o provenienci digitálních objektů (digiProvMD) – v této části je využit formát PREMISevent a PREMISagent. Zachyceny budou následující události (vytvoření skenu, uložení obrazu, různé migrace v digitalizaci, ořez apod. – informace o datu, místě, osobě nebo SW, který tyto úkony udělal).
 - o fileSec- sekce s odkazy na soubory – 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, které jsou s tou konkrétní stránkou spojeny, tj. archivní kopie, ALTO XML a OCR TXT. Jde o povinnou sekci METS záznamu, pro další mapování do LTP systému nebude potřeba.
 - o structMap – **pouze fyzická** strukturální mapa, povinná část METS záznamu. Bude ukazovat strukturu souborů k dané stránce, tj. opět archivní kopie, ALTO XML a OCR TXT. Pro další mapování do LTP systému nebude potřeba.

soubor Hlavni_METS.xml

další částí PSP balíčku je hlavní METS dokument. Hlavní METS záznam tedy obsahuje:

- dmdSec – bibliografická metadata k svazku monografie včetně popisu nadřazených entit (např. titul) nebo naopak částí (např. kapitola). Základ bude z katalogu, případný další popis částí bude z digitalizace. Formátem hlavním bude MODS, nutná pro LTP je i přítomnost zkráceného záznamu v Dublin Core.

- fileSec – hlavní část s linky na všechny digitální objekty (archivní kopie, ALTO XML a OCR TXT), které se váží k jednomu svazku monografie. Obsahuje také linky na administrativní metadata AMD_METS.xml do složky [amdSec].
- structMap – strukturální mapa pro celý dokument, tj. pro jeden svazek monografie. Obsahuje:
 - o logickou část – vyjadřuje logickou strukturu svazku s odkazy na ALTO XML
 - o fyzickou část obsahující informace o všech reprezentacích konkrétní stránky (archivní kopie, ALTO XML, OCR TXT a AMD_METS.xml)
 - o mapování na ALTO XML areas

soubor MD5

Poslední částí PSP balíčku je soubor s kontrolními součty pro všechny soubory balíčku (kromě info.xml a .md5 souboru samotného). Soubor .md5 je jeden pro 1 celý balíček SPS (balíček se svazkem monografie). Tento soubor .md5 obsahuje kontrolní součet pro každý soubor obsažený v PSP balíčku. Z tohoto důvodu nejsou samostatné kontrolní součty součástí podsložek balíčku. Kontrolní součty jsou také samozřejmě v technických metadatech.

6. Názvová konvence složek a souborů

pojmenování PSP balíčku

- každý PSP balíček přicházející z digitalizace by měl obsahovat pouze jedinou intelektuální entitu (svazek monografie). **Pak musí název balíčku vycházet z identifikátoru této entity, např. URN:NBN, číslo čárového kódu použitého na fyzické jednotce apod.**
- **každý svazek monografie musí mít svůj jednoznačný identifikátor, tím pádem pak každý PSP balíček a každý soubor v něm má vlastní jednoznačný identifikátor**

pojmenování složek

- viz návrh struktur PSP balíčku (kap. 5)

pojmenování souborů

- názvy jakýchkoliv souborů náležejících k jedné základní entitě (svazek) musí být založeny na jednom typu identifikátoru
- pro svazek monografie by takovým identifikátorem mohlo být URN:NBN, čČNB, ISBN nebo ISSN titulu
- podobně využitelným identifikátorem by mohlo být generované číslo UUID, které by se generovalo pro každý soubor. Tím by se ovšem ztratila (i vizuální) vazba na vrchní úroveň titulu i vazba na související soubory (stránka v jp2 a k ní náležející soubor ALTO XML apod.).

S využitím URN:NBN to může vypadat následovně (použit příklad pojmenování pro projekt NDK – digitalizace monografií):

| typ souboru | název souboru | vysvětlení |
|-----------------------------|------------------------------|--|
| PSP balíček (číslo, svazek) | NDK_123456 | název celé složky PSP balíčku, u základních int. entit bude v názvu využito vždy URN:NBN |
| archivní kopie | MC_NDK_123456_0013.jp2 | archivní JPEG2000 stránky 13 svazku monografie s urn:nbn:cz:ndk-123456 |
| uživatelská kopie | UC_NDK_123456_0013.jp2 | uživatelská kopie ve formátu JPEG2000 stránky 13 svazku monografie s urn:nbn:cz:ndk-123456 |
| ALTO XML | ALTO_NDK_123456_0013.xml | ALTO soubor náležející ke 13té stránce ze svazku monografií urn:nbn:cz:ndk-123456 |
| OCR TXT | TXT_NDK_123456_0013.txt | TXT soubor s OCR náležející ke 13té stránce ze svazku monografií urn:nbn:cz:ndk-123456 |
| info.xml | INFO_NDK_123456.xml | info xml k celému PSP balíčku svazku monografie |
| MD5 | NDK_123456.md5 | soubor s kontrolními součty k celému PSP balíčku svazku monografie |
| Hlavni_METS.xml | METS_NDK_123456.xml | hlavní METS záznam k celému svazku monografie s urn:nbn:cz:ndk-123456 |
| AMD_METS.xml | AMD_METS_NDK_123456_0013.xml | METS záznam s technickými metadaty pro stránku 13 ze svazku monografie s urn:nbn:cz:ndk-123456 |

Složka jednoho balíčku PSP, který obsahuje jen jeden obrazový soubor k první stránce svazku monografie, pak může vypadat následovně (příklad balíčku z digitalizace NDK):

| | |
|--------------|------------------------------|
| NDK_123456 | |
| | info.xml |
| [masterCopy] | MC_NDK_123456_0001.jp2 |
| [userCopy] | UC_NDK_123456_0001.jp2 |
| [ALTO] | ALTO_NDK_123456_0001.xml |
| [TXT] | TXT_NDK_123456_0013.txt |
| [amdSec] | AMD_METS_NDK_123456_0001.xml |
| | METS_NDK_123456.xml |
| | NDK_123456.md5 |

7. Transportní balík pro jeden nebo více PSP balíčků

Pokud bude jeden PSP balík obsahující 1 základní intelektuální entitu (svazek monografie) přemísťován např. jako tar, měl by název souboru tar odpovídat názvu PSP balíčku (tedy vycházet z použitého identifikátoru pro entitu svazku).

Výstupem workflow digitalizace ale může také být balík (např. tar), který obsahuje více PSP balíčků - toto sdružování bude omezeno jen kapacitou HW. Takovýto sdružený balík by měl být pojmenován na základě již užívaného identifikátoru.

- v případě, že balík obsahuje svazky jednoho vícesvazkového díla, měl by název balíku vycházet z ČČNB nebo ISBN
- typ identifikátoru musí být vyjádřen v názvu souboru – např. ISBN_1234567890.tar nebo CCNB_12345678910.tar apod.
- lze počítat s tím, že bude docházet k tomu, že sdružený balík nebude obsahovat např. všechny svazky titulu monografie – tato skutečnost musí být patrná z názvu balíku (např. ISBN_1234567890_YYYY kde YYYY může být pořadové číslo, datum, doba vzniku jednoho z více balíčků obsahujících svazky určitého titulu/souboru s identifikátorem ISBN 1234567890).

Transportní balík by měl obsahovat následující části:

- balíčky PSP (svazků)
- informační soubor, který odpovídá specifikaci info.xml
- kontrolní součty všech PSP balíčků
- seznam balíčků v transportním balíku

8. Metadata

- veškerá metadata budou „zabalena“ pomocí kontejnerového formátu METS
- formát METS bude v aktuální verzi v době implementace nebo verzi předchozí (prosinec 2010 verze 1.9- <http://www.loc.gov/standards/mets/mets-schemadocs.html>)
- veškerá metadata ve všech formátech musí být zapsána pomocí XML za použití kodování UTF-8
- **vložení metadatových formátů do kontejneru METS bude vždy formou <mdWrap>, tj. ne odkazováním z METS záznamu ven**

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 je zcela dle konkrétních potřeb)

8.1 Kořenový element hlavního METS záznamu

| element | atributy | popis | Povinnost |
|---------|---------------|---|-----------|
| <mets> | LABEL TYPE | kořenový element METS záznamu ----- LABEL – název titulu monografie, včetně roku vydání, např. Honzíkova cesta, 1979 TYPE – hodnota vždy „Monograph“ | M |

- kořenový element hlavního METS záznamu k jednomu svazku monografie musí obsahovat linky na specifikace jednotlivých použitých metadatových schémat (METS, MODS, Dublin Core)

8.2 METS hlavička <metsHdr>

- dokumentuje vznik a úpravy METS záznamu

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

8.3 METS část <dmdSec> - Bibliografická metadata – formát MODS a Dublin Core

- každá část <dmdSec> musí mít ID a vnořený element <mdWrap> s atributy MDTYPE, MIMETYPE

| element | atributy | popis | povinnost |
|----------|--------------------|---|-----------|
| <dmdSec> | ID | identifikátor <dmdSec> části METS záznamu ----- ID: pro <dmdSec> s popisem svazku (titulu) monografie hodnota „MODSMD_VOLUME“ a „DCMD_VOLUME“ pro <dmdSec> s popisem vnitřní části monografie hodnota dle typů vnitřní části (oddíl ² , obraz) - hodnoty „MODSMD_CHAP“ a „DCMD_CHAP“ pro článek a hodnoty „MODSMD_PICT“ a „DCMD_PICT“ pro obraz pro <dmdSec> s popisem přílohy monografie hodnota „MODSMD_SUPPL“ a „DCMD_SUPPL“ | M |
| <mdWrap> | MDTYPE MIMETYPE | element obsahující vložené záznamy MODS ----- MDTYPE – hodnota „MODS“ pro záznamy v MODS, hodnota „DC“ pro záznam v Dublin Core MIMETYPE – hodnota „text/xml“ | M |

- na samotný bibliografický popis bude použit formát MODS, aktuální verze v době implementace, nebo verze předchozí (prosinec 2010 verze 3.4 viz <http://www.loc.gov/standards/mods/>) a formát Dublin Core (dále DC) kvalifikovaný (<http://dublincore.org/documents/dcmi-terms/>)
- DC je primárně určeno na poskytnutí dat přes OAI-PMH, bude odpovídat OAI XSD (viz http://www.openarchives.org/OAI/2.0/oai_dc.xsd) a bude se jednat o nekvalifikovaný Dublin Core
- DC bude použito, uloženo v METS apod. stejným způsobem jako formát MODS – viz možnosti struktury PSP balíčku výše
- pro vytvoření DC z MODS formátu může být použito oficiální mapování Kongresové knihovny – viz <http://www.loc.gov/standards/mods/mods-conversions.html>
- DC a MODS bude vložen v METS části dmdSec – viz možnosti struktur PSP balíčku v kap. 7

² pozor výraz „kapitola“ je v tomto kontextu obecný a může vyjadřovat nejen kapitolu, ale také např. předmluvu, obsah apod.

- základním zdrojem pro popisná metadata je katalog NK a
- u digitalizovaných dokumentů je bibliografický popis vytvářen primárně z pohledu popisu fyzické předlohy, nejde o popis elektronického dokumentu

Monografie

- základní intelektuální entitou pro popis je svazek monografie, tj. v jednom METS záznamu, který bude obsahovat metadata a strukturu jednoho svazku, budou MODS záznamy k tomuto svazku
- metadata budou popisovat entity³:
 - o **svazek (Volume)**
 - o **vnitřní část (InternalPart) – typy „textový oddíl“ (Chapter) a „obraz“ (Picture)**
 - o **příloha (Supplement)**
-
- pozn: nepočítá se s úrovní titulu, ani pro vícedílné publikace, které mají pouze jeden katalogizační záznam; u jednosvazkových monografií titul splývá s popisem svazku (MODS záznam popisující svazek je záznam titulu z katalogu NK/MZK); pokud přeci jen existuje souhrnný název pro více svazků (např. sebrané spisy), je řešeno plněním souhrnného názvu do údajů o edici
- ad svazek (Volume) – popis svazku u klasické monografie (1 svazek = 1 záznam) odpovídá záznamu v katalogu
- ad vnitřní část - bližší určení typů „kapitol“ a „obrazu“ (fotografie, tabulka, ilustrace, graf apod.) bude možné vyjádřit pomocí atributů a výrazů kontrolovaného slovníku v elementu <genre>
 - o u popisu vnitřní části je i element <recordInfo> - důvodem je to, že popis bude vznikat manuálně a je žádoucí o vzniku záznamu držet kontrolní údaje; element je volitelný
- ad příloha (Supplement) - přílohou se rozumí volně vložená entita do jednotlivého svazku, např. mapa, klíč (řešení úloh), pracovní sešit, CD/DVD apod.

Rozlišujeme 3 druhy příloh monografie:

- o příloha, která **se neskenuje**, ale chceme o ní vytvořit bibliografický záznam, dát najevo čtenáři, že existuje – to může být např. CD/DVD apod.
 - digitální podoba přílohy (pokud existuje) není součástí balíčku PSP svazku
 - popis lze udělat v rámci popisu přílohy (Supplement) v MODS – viz specifikace níže
 - pokud existuje záznam v katalogu k této příloze (např. CD/DVD, mapa apod.), bude využit pro generování MODS záznamu přílohy
 - taková příloha není součástí logické strukturální mapy formátu METS
- o příloha podobného typu, tvaru a velikosti jako je popisovaný svazek monografie, která se spolu s číslem **skenuje**
 - digitální podoba přílohy je, spolu s se svazkem (Volume), součástí PSP balíčku svazku a je součástí hlavního METS záznamu

³ toto pořadí nevyjadřuje logickou strukturu dokumentu, ta je popsána jinde

- popis lze udělat v rámci popisu přílohy (Supplement) v MODS – viz specifikace níže
 - taková příloha může mít vnitřní části (InternalPart) stejně jako svazek (Volume) a jejich text je součástí ALTO XML, které je společné pro svazek (Volume) i přílohu (Supplement)
 - **taková příloha je součástí logické strukturální mapy formátu METS**
 - **taková příloha je součástí fyzické strukturální mapy formátu METS (linky mezi jednotlivými soubory reprezentujícími stránky a popisnými metadaty)**
- příloha odlišného typu tvaru a velikosti než je popisovaný svazek monografie, která **se skenuje zvlášť** na svazku nezávisle
 - může jít např. o mapu apod.
 - k těmto přílohám vznikají metadata podobně jako pro jednotlivé svazky monografií, ovšem na původním svazku, ke kterému příloha patřila, nezávisle - tj. pro „původní“ svazek, u kterého byla příloha, vznikne 1 popis (PSP balíček s jedním hlavním METS záznamem a ALTO XML souborem) a pro přílohu je vytvořen další 1 popis (a PSP balíček s METS záznamem)
- záznamy monografie mohou být v katalozích následující
 - monografie má jen jeden svazek – existuje jeden záznam v katalogu a jedno čČNB
 - monografie má více svazků – pak existuje buď
 - 1) jeden záznam pro soubor, pokud jednotlivé svazky/díly nejsou od sebe příliš odlišné (např. slovník a-k, l-z), k jednomu záznamu existuje jedno čČNB; nebo
 - 2) v případě, že jednotlivé díly/svazky souboru jsou odlišné (např. Vlastivěda česká – díl flora, díl fauna, atd.), tak má každý svazek svůj záznam v katalogu a své čČNB, souborný záznam v tomto případě neexistuje
- popis nadřazené entity, kde tedy existuje pouze 1 katalogizační záznam pro více svazků monografie, nebude součástí metadat popisujících svazek.
- stránka se nebude popisovat, její logické i fyzické číslování i typ stránky je obsaženo ve struktuře METS dokumentu (část structMap)
 - typ stránky (Advertisement, Blank, Content, Index aj.) budou odpovídat přesně seznamu typů z DTD monografie – viz <http://digit.nkp.cz/Monographs/DTD/2.10/Monograph.xsd>
- v katalozích NK a MZK neexistují údaje o kapitolách monografií - tj. vnitřní členění a popis musí vzniknout v digitalizaci, popis titulu/svazku monografie musí být stažen z katalogu do workflow digitalizace
- pro každou entitu vznikne jeden MODS záznam s vlastním ID, které bude označovat i typ části (např. oddíl, ilustrace apod.) v případě opakování částí se bude opakovat odpovídající počet MODS záznamů v jednom PSP 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**

8.3.1 Navrhovaná pole MODS a Dublin Core pro jednotlivé části monografie

- mapování jednotlivých polí z MARC21 lze nalézt např. na jedné z oficiálních stránek MODS formátu u jednotlivých top-level elementů
- <http://www.loc.gov/standards/mods/userguide/generalapp.html>

Obsah sloupce „Popis“:

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

význam sloupce „povinnost“

- pole „povinnost“ uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné
- může nabývat následujících hodnot
 - o M - mandatory (povinně plnit – element je součástí každého záznamu)
 - o MA - mandatory if available (povinně plnit pokud je to možné, pokud lze apod.)
 - o R - recommended (plnění hodnot elementu je doporučeno, není ovšem povinné)
 - o RA - recommended if available (doporučeno pokud lze plnit)
 - o O - optional (plnění hodnot elementu je zcela dle konkrétních potřeb)

Pole MODS a DC pro svazek monografie

| Element MODS | Atributy | Popis | povinnost | Element DC |
|--------------|------------|--|-----------|------------|
| <titleInfo> | ID type | název svazku monografie pro plnění použít katalogizační záznam ----- ID musí vyjadřovat název úrovně, tj. např. „MODS_VOLUME“ type: hodnota „alternative“ pro paralelní a jiné názvy (odpovídají poli 245 podpoli „b“) | M | |
| <title> | | názvová informace – název svazku monografie hodnoty převzít z katalogu, odpovídá poli 245, podpoli „a“ pro | M | <dc.title> |

| | | | | |
|--------------|------|---|----|--|
| | | hlavní název | | |
| <subTitle> | | podnázev svazku monografie | MA | <dc.title> |
| <partNumber> | | číslo části, např. určité řady/edice (část 1, řada B) | R | <dc.description> |
| <partName> | | jméno edice nebo speciální ediční řady, např. Knihy odvahy a dobrodružství | R | <dc.description> |
| <name> | type | <p>údaje o odpovědnosti za svazek</p> <p>-----</p> <p>type: použit jeden z typů</p> <ul style="list-style-type: none"> - personal - corporate - conference - family <p>pokud má monografie autora a ilustrátora, element <name> se opakuje s různými rolemi</p> <p>POZOR – údaje o odpovědnosti nutno přebírat z polí 1XX a 7XX MARCu21</p> | M | |
| <namePart> | type | <p>údaje o křestním jméně a příjmení apod.</p> <p>nutno vyjádřit pro křestní jméno i příjmení</p> <p>-----</p> <p>type: použít jednu z hodnot:</p> <ul style="list-style-type: none"> - date – doporučené pokud lze uvést - family – povinné pokud lze uvést - given – povinné pokud lze uvést - termsOfAddress – doporučené pokud lze uvést <p>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké</p> | M | <dc.creator> nutno do jednoho pole DC spojit jméno i příjmení |

| | | | | |
|------------------|-------------------|---|---|-----------|
| | | je do jednoho elementu <namePart> | | |
| <role> | | specifikace role osoby nebo organizace uvedené v elementu <name> | M | |
| <roleTerm> | type authority | popis role nutno použít kontrol. slovník např. z MARC21 ----- type: code – kód role z kontrolovaného slovníku rolí http://www.loc.gov/marc/relators/ relaterm.html) authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"; | M | |
| <typeOfResource> | | popis charakteristiky typu nebo obsahu zdroje jedna z hodnot: - text - cartographic - notated music - sound recording-musical - sound recording-nonmusical - sound recording - still image - moving image - three dimensional object - software, multimedia - mixed material pro monografie hodnota text; mělo by se vyčítat z MARC21 katalogizačního záznamu z pozice 06 návěští | R | <dc:type> |
| <genre> | | blíže údaje o typu dokumentu | M | <dc:type> |

| | | | | |
|--------------|-----------|--|----|----------------|
| | | hodnota: volume | | |
| <originInfo> | | informace o původu předlohy Poznámka: Jeden nebo více výskytů elementů se předpokládá pro vydavatele, další výskyt v případě nutnosti popsat tiskaře. Pokud je nutno vyjádřit tiskaře (pole 260 podpole „f“ a „e“ a „g“ v MARC21), je nutno element <originInfo> opakovat s atributem transliteration=“printer“ a elementy <place>, <publisher>, <dateCreated>, které budou obsahovat údaje o tiskaři. | M | |
| <place> | | údaje o místě spojeném s vydáním, výrobou nebo původem popisovaného dokumentu | MA | <dc:coverage> |
| <placeTerm> | type | konkrétní určení místa, např. Praha odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „a“ ----- type – bude vždy text | MA | <dc:coverage> |
| <publisher> | | jméno entity, která dokument vydala, vytiskla nebo jinak vyprodukovala odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21; Pokud má monografie více vydavatelů, přebírají se za záznamu všichni (jsou v jednom poli 260). | MA | <dc:publisher> |
| <dateIssued> | qualifier | datum vydání předlohy, | M | <dc:date> |

| | | | | |
|---------------|-----------|---|---|--|
| | | <p>přebírat z katalogu; odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „c“</p> <p>jiná data než rok možno zapsat v následujících podobách:</p> <ul style="list-style-type: none"> - DD.MM.RRRR – pokud víme den, měsíc i rok vydání - MM.RRRR – pokud víme jen měsíc a rok vydání - RRRR – pokud víme pouze rok - DD.-DD.MM.RRRR – vydání pro více dní - MM.-MM.RRRR – vydání pro více měsíců <p>-----</p> <p>qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj</p> | | |
| <dateCreated> | qualifier | <p>datum vytvoření předlohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo></p> <p>odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“</p> <p>-----</p> <p>qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj</p> | R | |
| <issuance> | | <p>údaje o vydávání</p> <p>hodnota monographic odpovídá hodnotě uvedené návěští MARC21 na pozici 07</p> | M | |
| <language> | | údaje o jazyce dokumentu; | M | |

| | | | | |
|-----------------------|---------------------------------|---|----|---------------|
| | | v případě vícenásobného výskytu nutno element <language> opakovat | | |
| <languageTerm> | type authority objectPart | přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php ----- type: použít hodnotu code authority: použít hodnotu „iso639-2b“; odpovídá poli 041 MARC21, podpoli „a“ objectPart: možnost vyjádřit jazyk konkrétní části svazku; možné hodnoty např.: summary (pro shrnutí), original (pro předlohu u překladu) aj. – nutno vytvořit kontrolovaný slovník; jazyk resumé lze přebírat z pole 041, podpole „b“ jazyk předlohy u překladu lze přebírat z pole 041, podpole „h“ | M | <dc:language> |
| <physicalDescription> | | obsahuje údaje o fyzickém popisu zdroje/předlohy | M | |
| <form> | authority | údaje o fyzické podobě dokumentu, např. print, electronic apod. pro monografie hodnota print odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21 ----- authority: hodnota „marcform“ | M | <dc:format> |
| <extent> | | údaje o rozsahu (stran, svazků nebo rozměrů) odpovídá hodnotám v poli 300 | RA | <dc:format> |

| | | | | |
|--------------|-----------|--|----|------------------|
| | | podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat; počet stránek bude vyjádřen ve fyzické strukturální mapě a bude tak vidět v aplikaci zpřístupnění i bez vyplnění tohoto pole | | |
| <note> | | poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element | RA | |
| <abstract> | | shrnutí obsahu jako celku odpovídá poli 520 MARC21 | R | <dc:description> |
| <note> | | obecná poznámka ke svazku monografie jako celku odpovídá poli 500 v MARC21 | RA | <dc:description> |
| <subject> | authority | údaje o věcném třídění předpokládá se přebírání z katalogizačního záznamu ----- authority: vyplnit hodnotu „czenas“ | R | |
| <topic> | | libovolný výraz specifikující nebo charakterizující obsah svazku monografie; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) nebo obsah pole 650 záznamu MARC21 | 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 použit kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) nebo obsah pole 600 záznamu MARC21 | R | <dc:subject> |
| <classification> | authority | klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění odpovídá poli 080 MARC21 ----- authority: vyplnit hodnotu „udc“ | M | <dc:subject> |
| <relatedItem> | type | informace o dalších dokumentech/částech/zdrojích, které jsou ve vztahu k popisovanému dokumentu; Poznámka: element <relatedItem> může obsahovat jakýkoliv jiný element MODS – jejich použití se řídí pravidly popsányými pro tyto elementy; ----- type: hodnota „series“ | RA | |
| <identifier> | type | údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které svazek monografie má – viz přehled typů atributů níže ----- type: budou se povinně vyplňovat následující hodnoty, pokud existují: - doi - hdl - handle - issn - převzít z katalogizačního záznamu NK ČR - isbn - převzít z katalogizačního záznamu NK ČR | M | <dc:identifier> |

| | | | | |
|--------------------|-----------|---|----|-------------|
| | | <ul style="list-style-type: none"> - ccnb – čČNB - převzít z katalogizačního záznamu NK ČR - permalink záznamu z katalogu NK ČR, např. http://aleph.nkp.cz/F/?func=direct&doc_number=002186258&local_base=NKC - urnnbn - pro URN:NBN, např. zápis ve tvaru urn:nbn:cz:ndk-123456 pro projekt NDK; pozor, musí odpovídat URN:NBN, podle kterého je pojmenovaný PSP balíček a jeho jednotlivé soubory - uuid - jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu | | |
| <location> | | údaje o uložení popisovaného dokumentu, např. signatura, místo uložení apod. | MA | |
| <url> | note | pro uvedení lokace elektronického dokumentu ----- note: pro poznámku o typu URL (na plný text, abstrakt apod.) | O | <dc:source> |
| <physicalLocation> | authority | údaje o instituci, kde je fyzicky uložen popisovaný dokument, např. NK ČR nutno použít kontrolovaný slovník – sigly knihoven (ABA001 atd.) odpovídá poli 040 v MARC21 | M | <dc:source> |

| | | | | |
|-----------------------|----------|--|----|-------------|
| | | pozn. u dokumentů v digitální podobě není možné vyplnit ----- authority: hodnota „siglaADR“ | | |
| <shelfLocator> | | signatura nebo lokační údaje o dokumentu | M | <dc:source> |
| <part> | type | popis části, pokud je svazek částí souboru, element může být využit jen na zaznamenání <caption> ----- type: hodnota bude vždy „volume“ | O | |
| <caption> | | text před označením čísla, např. „č.“, „část“, „No.“ apod. | RA | |
| <recordInfo> | | údaje o metadatovém záznamu – jeho vzniku, změnách apod. | M | |
| <recordContentSource> | | kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník | R | |
| <recordCreationDate> | encoding | datum prvního vytvoření záznamu, na úroveň minut ----- encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601 | M | |
| <recordChangeDate> | encoding | datum změny záznamu ----- encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601 | R | |
| <recordOrigin> | | údaje o vzniku záznamu hodnoty: machine generated nebo human prepared | R | |

Pole MODS a DC pro vnitřní část monografie (textový oddíl a obraz)

| Element MODS | Atributy | Popis | Povinnost | Element DC |
|--------------|----------|---|-----------|--|
| <titleInfo> | ID | názvová informace vnitřní části ----- ID musí vyjadřovat název úrovně, tj. např. „MODS_PICTURE“ pro obrázek v textu, „MODS_CHAPTER“ pro textový oddíl apod. | M | |
| <title> | | vlastní název vnitřní části (oddílu, obrazu); u obrazu brát případně z popisku obrazu; pokud není titul, nutno vyplnit hodnotu „untitled“ | M | <dc:title> |
| <subTitle> | | podnázev vnitřní části (oddílu); např. podnázev kapitoly | MA | <dc:title> |
| <partNumber> | | číslo vnitřní části | RA | <dc:title> |
| <partName> | | název vnitřní části | RA | <dc:title> |
| <name> | type | údaje o odpovědnosti za vnitřní část (oddíl i obraz) ----- type: použít jeden z typů - personal - corporate - conference - family | MA | |
| <namePart> | type | údaje o křestním jméně a příjmení apod. nutno vyjádřit pro křestní jméno i příjmení ----- type: použít jednu z hodnot: - date – doporučené pokud lze uvést - family – povinné pokud lze uvést - given – povinné pokud lze | MA | <dc:creator> nutno do jednoho pole DC spojit jméno i příjmení |

| | | | | |
|------------|-------------------|--|----|-----------|
| | | <p>uvést</p> <ul style="list-style-type: none"> - termsOfAddress – doporučené pokud lze uvést <p>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu <namePart></p> | | |
| <role> | | specifikace role osoby nebo organizace uvedené v elementu <name> | MA | |
| <roleTerm> | type authority | <p>popis role nutno použít kontrol. slovník např. z MARC21</p> <p>-----</p> <p>type: code – kód role z kontrolovaného slovníku rolí (http://www.loc.gov/marc/relators/relaterm.html)</p> <p>authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"</p> | MA | |
| <genre> | type | <p>bližší údaje o typu vnitřní části povinné hodnota: chapter nebo picture</p> <p>-----</p> <p>type: doporučené</p> <p>hodnota pro chapter – možnost vyplnit bližší určení typu oddílu (možnost použít DTD monografie, MonographComponentPart Types)</p> <ul style="list-style-type: none"> - table of content - advertisement - abstract | M | <dc:type> |

| | | | | |
|----------------|-------------------|--|----|---------------|
| | | <ul style="list-style-type: none"> - introduction - review - dedication - bibliography - editorsNote - preface - chapter - article - index (použije se pro všechny typy seznamů mimo hlavní obsah; např. seznam obrazů, tabulek apod.) - unspecified – pokud nepatří ani do jedné z výše uvedených kategorií - aj. <p>hodnota pro picture – možnost vyplnit další určení typu obrazu</p> <ul style="list-style-type: none"> - table - illustration - chart - photograph - graphic - map - advertisement - cover - unspecified – pokud nepatří ani do jedné z výše uvedených kategorií - aj. | | |
| <language> | | údaje o jazyce vnitřní části nelze plnit u obrazu; v případě vícenásobného výskytu nutno element <language> opakovat | MA | |
| <languageTerm> | type authority | přesné určení jazyka – kódem nutno použít kontrolovaný slovník | M | <dc:language> |

| | | | | |
|-----------------------|-----------------------|---|----|------------------|
| | | ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php nelze plnit u obrazu ----- type: použít hodnotu code authority: použít hodnotu „iso639-2b“ | | |
| <physicalDescription> | | obsahuje údaje o fyzickém popisu vnitřní části; určeno spíše pro oddíly než pro obrazy | R | |
| <form> | authority | údaje o fyzické podobě vnitřní části, např. print, electronic apod. ----- authority: hodnota „marcform“ | R | <dc:format> |
| <abstract> | | shrnutí obsahu vnitřní části | R | <dc:description> |
| <note> | | obecná poznámka k vnitřní části do poznámky by se měla dávat šifra autora vnitřní části, která se vyskytuje pod vnitřní částí | RA | <dc:description> |
| <subject> | | údaje o věcném třídění | R | |
| <topic> | authority (volitelné) | libovolný výraz specifikující nebo charakterizující obsah vnitřní části; lze (není ovšem nutno) použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) ----- při použití autoritních záznamů použít AUT NK ČR a atribut authority: vyplnit hodnotu „czenas“; při použití volných klíčových slov atribut authority nepoužívat | M | <dc:subject> |
| <geographic> | authority | geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) | R | <dc:subject> |

| | | | | |
|------------------|-----------|---|----|----------------------------|
| | | ----- authority: vyplnit hodnotu „czenas“ | | |
| <temporal> | authority | chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) ----- authority: vyplnit hodnotu „czenas“ | R | <dc:subject> |
| <name> | authority | jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) ----- authority: vyplnit hodnotu „czenas“ | R | <dc:subject> |
| <classification> | authority | klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění plnit pouze pro oddíl odpovídá poli 080 MARC21 ----- authority: vyplnit hodnotu „udc“ | RA | <dc:subject> |
| <identifier> | type | údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které vnitřní část má – viz přehled typů atributů níže ----- type: budou se povinně vyplňovat následující hodnoty, pokud existují pro oddíl nebo obraz: - doi - hdl - handle - urnnbn - pro URN:NBN, u vnitřních částí monografií se s URN:NBN počítá primárně pro články ve sborníku, ne pro „obyčejné“ kapitoly - uuid | M | <dc:identifier> povinné |

| | | | | |
|-----------------------|----------|---|----|---------------|
| | | - jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu | | |
| <part> | | vrchní element, který bude použit pouze na záznam rozsahu vnitřní části; nelze u obrazu | RA | |
| <extent> | | upřesnění popisu části – rozsah na stránkách | MA | <dc:format> |
| <start> | | první stránka, na které vnitřní část začíná | MA | <dc:coverage> |
| <end> | | poslední stránka, na které vnitřní část končí | MA | <dc:coverage> |
| <recordInfo> | | údaje o metadatovém záznamu vnitřní části – jeho vzniku, změnách apod. | M | |
| <recordContentSource> | | kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník | R | |
| <recordCreationDate> | encoding | datum prvního vytvoření záznamu vnitřní části ----- encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601 | M | |
| <recordChangeDate> | encoding | datum změny záznamu vnitřní části ----- encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601 | R | |
| <recordOrigin> | | údaje o vzniku záznamu vnitřní části hodnoty: machine generated nebo human prepared | R | |

Pole MODS a DC pro přílohu

| Element MODS | Atributy | Popis | Povinnost | Element DC |
|--------------|----------|---------------------------|-----------|------------|
| <titleInfo> | ID | názvová informace přílohy | M | |

| | | | | |
|--------------|------|---|----|--|
| | | <p>použít názvové autority nebo katalogizační záznam</p> <p>-----</p> <p>ID musí vyjadřovat název úrovně, tj. „MODS_SUPPLEMENT“</p> | | |
| <title> | | <p>názvová informace – název svazku monografie, jehož součástí příloha je;</p> <p>převzít z katalogu</p> | M | <dc:title> |
| <partNumber> | | <p>číslo přílohy, pokud nějaké má doporučené pokud lze vyplnit</p> | MA | <dc:description> |
| <partName> | | název přílohy | MA | <dc:title> |
| <name> | type | <p>údaje o odpovědnosti za přílohu</p> <p>-----</p> <p>type: použít jeden z typů</p> <ul style="list-style-type: none"> - personal - corporate - conference - family | MA | |
| <namePart> | type | <p>údaje o křestním jméně a příjmení apod.</p> <p>nutno vyjádřit pro křestní jméno i příjmení</p> <p>-----</p> <p>type: použít jednu z hodnot:</p> <ul style="list-style-type: none"> - date – doporučené pokud lze uvést - family – povinné pokud lze uvést - given – povinné pokud lze uvést - termsOfAddress – doporučené pokud lze uvést <p>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu <namePart></p> | MA | <dc:creator> nutno do jednoho pole DC spojit jméno i příjmení |
| <role> | | specifikace role osoby nebo | MA | |

| | | | | |
|------------------|-------------------|---|----|-----------|
| | | organizace uvedené v elementu <name> | | |
| <roleTerm> | type authority | popis role nutno použít kontrol. slovník např. z MARC21 ----- type: code – kód role z kontrolovaného slovníku rolí (http://www.loc.gov/marc/relators/relaterm.html) authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator" | MA | |
| <typeOfResource> | | popis charakteristiky typu nebo obsahu přílohy jedna z hodnot: - text – např. pro přílohu typu časopis, kniha, brožura apod. - cartographic – pro mapy - notated music - sound recording-musical - pro hudební CD/DVD - sound recording-nonmusical - sound recording - still image – fotografie, plakáty apod. - moving image – pro filmová DVD - three dimensional object - software, multimedia – pro CD/DVD se SW - mixed material | R | <dc:type> |
| <genre> | | bližší údaje o typu dokumentu hodnota: supplement | M | <dc:type> |

| | | | | |
|--------------|-----------|--|----|----------------|
| <originInfo> | | <p>informace o původu přílohy <i>plnit pokud se liší od údajů v popisu svazku monografie (platí i pro jednotlivé sub-elementy)</i></p> <p>Poznámka: Jeden nebo více výskytů elementů se předpokládá pro vydavatele, další výskyt v případě nutnosti popsat tiskaře. Pokud je nutno vyjádřit tiskaře (pole 260 podpole „f“ a „e“ a „g“ v MARC21), je nutno element <originInfo> opakovat s atributem transliteration=“printer“ a elementy <place>, <publisher>, <dateCreated>, které budou obsahovat údaje o tiskaři.</p> | MA | |
| <place> | | údaje o místě spojeném s vydáním, výrobou nebo původem přílohy | MA | <dc:coverage> |
| <placeTerm> | type | <p>konkrétní určení místa, např. Praha odpovídá hodnotě katalogizačního záznamu, pole 260, podpole „a“</p> <p>-----</p> <p>type – bude vždy text</p> | MA | <dc:coverage> |
| <publisher> | | <p>jméno entity, která přílohu vydala, vytiskla nebo jinak vyprodukovala</p> <p>odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21</p> | MA | <dc:publisher> |
| <dateIssued> | qualifier | <p>datum vydání přílohy, dle toho jaké údaje jsou k dispozici</p> <p>jiná data než rok možno zapsat v následujících podobách:</p> <ul style="list-style-type: none"> - DD.MM.RRRR – pokud víme den, měsíc i rok vydání - MM.RRRR – pokud víme jen měsíc a rok vydání | MA | <dc:date> |

| | | | | |
|----------------|-------------------|--|----|---------------|
| | | <ul style="list-style-type: none"> - RRRR – pokud víme pouze rok - DD.-DD.MM.RRRR – vydání pro více dní - MM.-MM.RRRR – vydání pro více měsíců <p>možno použít hodnotu z katalogizačního záznamu, pole 260, podpole „c“</p> <p>-----</p> <p>qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj</p> | | |
| <dateCreated> | qualifier | <p>datum vytvoření přílohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo> nebo např. u popisu CD/DVD apod.</p> <p>odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“</p> <p>-----</p> <p>qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj</p> | R | |
| <frequency> | | <p>údaje o pravidelnosti vydávání</p> <p>odpovídá údajům MARC21 v poli 310 nebo pozici 18 v poli 008</p> | RA | |
| <language> | | údaje o jazyce dokumentu | M | |
| <languageTerm> | type authority | <p>přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php</p> <p>-----</p> <p>type: použít hodnotu code</p> <p>authority: použít hodnotu „iso639-</p> | M | <dc:language> |

| | | | | |
|-----------------------|-----------|---|----|------------------|
| | | 2b“ | | |
| <physicalDescription> | | obsahuje údaje o fyzickém popisu | M | |
| <form> | authority | údaje o fyzické podobě dokumentu, např. print, electronic apod. povinné pro tištěné předlohy hodnota „print“, pro elektronické přílohy „electronic“ odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21 ----- authority: hodnota „marcform“ | M | <dc:format> |
| <extent> | | údaje o rozsahu (stran, svazků nebo rozměrů) odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat | RA | <dc:format> |
| <note> | | poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element | RA | |
| <abstract> | | shrnutí obsahu dokumentu odpovídá poli 520 MARC21 | RA | <dc:description> |
| <note> | | obecná poznámka k dokumentu odpovídá poli 500 v MARC21 | RA | <dc:description> |
| <subject> | authority | údaje o věcném třídění ----- authority: vyplnit hodnotu „czenas“ | R | |
| <topic> | | libovolný výraz specifikující nebo charakterizující obsah přílohy; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) | M | <dc:subject> |
| <geographic> | | geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR | R | <dc:subject> |

| | | | | |
|------------------|-----------|---|----|-----------------|
| | | (geografický termín) | | |
| <temporal> | | chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) | R | <dc:subject> |
| <name> | | jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) | R | <dc:subject> |
| <classification> | authority | klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění odpovídá poli 080 MARC21 ----- authority: vyplnit hodnotu „udc“ | M | <dc:subject> |
| <identifier> | type | údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které příloha má – viz přehled typů atributů níže ----- type: budou se povinně vyplňovat následující hodnoty, pokud existují: - doi - hdl - handle - issn - převzít z katalogizačního záznam NK ČR - isbn - převzít z katalogizačního záznam NK ČR - ccnb – čČNB - převzít z katalogizačního záznam NK ČR - permalink záznamu z katalogu NK ČR, např. http://aleph.nkp.cz/F/?func=direct&doc_number=002186258&local_base=NKC - urnnbn - pro URN:NBN - uuid - jiný interní identifikátor, | MA | <dc:identifier> |

| | | | | |
|--|--|--|--|--|
| | | hodnota atributu „local“, lze použít např. k vyjádření čárového kódu | | |
|--|--|--|--|--|

8.4 METS část <amdSec> - Technická a administrativní metadata – formáty MIX a PREMIS

- pro všechna digitalizovaná data se budou využívat formát PREMIS (jeho části object, event a agent), pro obrazová data dále i formát MIX
- technická a administrativní metadata budou zabalena v části <amdSec> formátu METS ve vlastních formátech (MIX, PREMIS – části object; events; agent)
- technická a administrativní metadata budou vznikat i pro prvotní sken (většinou TIFF), který se po nutných úpravách maže a dále neuchovává
 - o technická metadata jsou určena primárně pro zachycení technických informací o formátech souborů, o výsledcích validací a kontrol
 - o administrativní metadata zachycují veškeré změny, procesy apod., které byly na datech i metadatach provedeny
- technická metadata přicházející z digitalizace jsou dále v maximální míře ukládána v LTP systému (po namapování do interního formátu LTP systému)
- všechny PREMIS záznamy budou ve vedlejší METS záznamu (AMD_METS.xml), který je určen pro administrativní a technická metadata (spolu s MIX záznamy).
 - o celý METS záznam (AMD_METS.xml) a je linkován z hlavního METS záznamu dokumentu
- **plnění technických metadat se předpokládá z výstupů vzniklých využitím služeb třetích stran jako jsou JHOVE2, PRONOM aj.)**
- **<amdSec> část bude existovat vždy jedna pro všechny reprezentace jedné stránky dokumentu (MC, ALTO XML, 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> | ID | element obsahující technická metadata ve formátu PREMIS nebo MIX ----- ID – identifikátor konkrétní části <amdSec>, např. pro stránku 1 by hodnota mohla být „PAGE0001“ | M |
| <techMD> nebo <digiprovMD> | ID | element rozlišující typy jednotlivých administrativních metadat ----- ID pro část <techMD>: | M |

| | | | |
|----------|--------|--|---|
| | | <ul style="list-style-type: none"> - pro části obsahující PREMIS-object hodnota „OBJ_001“ – objekt 1 (PREMIS object pro smazaný TIFF, OBJ_002 by bylo pro MC, OBJ_003 pro ALTO XML - pro části obsahující MIX hodnota „MIX_001“ = MIX metadata pro původní TIFF, „MIX_002“ pro MC <p>pro část <digiprovMD>:</p> <ul style="list-style-type: none"> - pro části obsahující PREMIS-event hodnota „EVT_001“ apod. - pro části obsahující PREMIS-agent hodnota „AGENT_001“ apod. | |
| <mdWrap> | MDTYPE | <p>element obsahující vložené záznamy PREMIS, MIX</p> <p>-----</p> <p>MDTYPE</p> <ul style="list-style-type: none"> - pro záznamy PREMIS object, event i agent vždy hodnota „PREMIS“ - pro záznamy MIX hodnota „NISOIMG“ | M |

8.4.1 PREMIS Objects

- bude odpovídat poslední aktuální verzi v době implementace (leden 2011 - PREMIS data dictionary v. 2.1), nebo verzi předchozí
- 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 1) vzniklý v procesu digitalizace (původní sken, který se dále maže); 2) archivní obrazové kopie, 3) ALTO XML, OCR.TXT, 4) uživatelská kopie)
- PREMIS object se nebude vytvářet pro OCR.TXT soubory
- 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 Events ve stejném METS metadatovém záznamu konkrétního dokumentu (svazku monografie) v části <digiprovMD>; přes <premis:relatedEventIdentification>, to samé platí pro objekty, které budou nalinkovány v případě vztahu (např. UC vznikla z MC) s popisovaným objektem přes <premis:relatedObjectIdentification>.
 - o tj. např. PREMIS object popisující archivní soubor JPEG2000 je tímto způsobem nalinkován na původní sken ve formátu TIFF (resp. na jeho PREMIS object záznam) – pomocí tagu <relatedObjectIdentification>, který obsahuje ID původního objektu (např. TIFF)
 - o zároveň pomocí tagu <relatedEventIdentification> je záznam PREMIS object archivního souboru JPEG2000 nalinkován na událost, během které vznikl
- **POZOR – Premis Object bude vznikat a uchovávat se i pro neexistující data (původní a posléze smazaný TIFF)**

Pole záznamu PREMIS Object

Obsah pole „Popis“:

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

význam pole „povinnost“

- pole „povinnost“ uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné
- může nabývat následujících hodnot
 - o M - mandatory (povinně plnit – element je součástí každého záznamu)
 - o MA - mandatory if available (povinně plnit pokud je to možné, pokud lze apod.)
 - o R - recommended (plnění hodnot elementu je doporučeno, není ovšem povinné)
 - o RA - recommended if available (doporučeno pokud lze plnit)
 - o O - optional (plnění hodnot elementu je zcela dle konkrétních potřeb)

Obsah pole „Použití pro“

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

| Element | Popis | Použití pro | Povinnost |
|-------------------------|---|-------------------|-----------|
| <objectIdentifier> | identifikátor k jednoznačnému odlišení objektu v určitém kontextu; 1-n | MC, XML, PS | M |
| <objectIdentifierType> | popis kontextu, ve kterém je identifikátor unikátní, např. NDK, ANL nebo název repozitáře; nutno použít kontrolovaný slovník; 1-1 | MC, XML, PS | M |
| <objectIdentifierValue> | vlastní hodnota identifikátoru, např. img0001-master, urn.nbn.cz-123465 apod.; 1-1 | MC, XML, PS | M |
| <objectCategory> | typ objektu, ke kterým se metadata (PREMIS object) vztahuje, např. file pro soubor, representation pro dig. reprezentaci, bitstream pro bitstream; 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; | MC, XML, PS | M |

| | | | |
|---------------------------------|---|-------------------|---|
| | 0-n | | |
| <preservationLevelValue> | hodnota úrovně ochrany, která je pro soubor relevantní, pro původní sken PS hodnota deleted, pro MC a XML hodnota preservation; 1-1 | MC, XML, PS | M |
| <preservationLevelDateAssigned> | datum, kdy byla přiřazena hodnota úrovně ochrany, zápis v ISO 8601, na úroveň dne (DD-MM-RRRR) 0-1 | MC, XML, PS | R |
| <objectCharacteristics> | technické údaje o souboru 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 kodování, podobně pak hodnota 2; 1-1 | MC, XML, PS | M |
| <fixity> | údaje o kontrolním součtu 0-n | MC, XML, PS | M |
| <messageDigestAlgorithm> | použitý algoritmus kontrolního součtu, např. MD5 aj. 1-1 | MC, XML, PS | M |
| <messageDigest> | hodnota kontrolního součtu 1-1 | MC, XML, PS | M |
| <messageDigestOriginator> | agent (osoba, instituce, stroj, SW), který kontrolní součet vytvořil (např. JHOVE apod.) 0-1 | MC, XML, PS | M |
| <size> | údaje o velikosti souboru v bytech 0-1 | MC, XML, PS | M |
| <format> | údaje o formátu souboru 1-n | MC, XML, PS | M |
| <formatDesignation> | identifikace formátu souboru, výstup z JHOVE, PRONOM služeb apod. 0-1 | MC, XML, PS | M |
| <formatName> | jméno formátu, např. image/tiff nebo Adobe PDF 1-1 | MC, XML, PS | M |
| <formatVersion> | verze formátu, např. 6.0 0-1 | MC, XML, PS | M |
| <formatRegistry> | identifikace formátu – dodatečná informace o záznamu formátů v registrech formátů (např. PRONOM aj.) | MC, XML, PS | M |

| | | | |
|-------------------------------|---|-------------------|---|
| | 0-1 | | |
| <formatRegistryName> | jméno použitého registru formátů, např. UDFR, PRONOM aj. 1-1 | MC, XML, PS | M |
| <formatRegistryKey> | unikátní identifikátor (označení) formátu v registru, např. fmt/155 z PRONOM 1-1 | MC, XML, PS | M |
| <creatingApplication> | údaje o aplikaci, ve které byl popisovaný soubor vytvořen; nutno popsat skener, SW kde vzniklo ALTO XML/TXT, SW/kodek pro vytvoření JPEG2000 MC 0-n | MC, XML, PS | M |
| <creatingApplicationName> | název aplikace, např. ImageGear, Kakadu apod.; 0-1 | MC, XML, PS | M |
| <creatingApplicationVersion> | verze aplikace, např. 15.03.000 0-1 | MC, XML, PS | M |
| <dateCreatedByApplication> | datum a čas vytvoření, např. 2008-11-10T12:37:46; musí být ve tvaru ISO 8601 (na úrovni vteřin); 0-1 | MC, XML, PS | M |
| <originalName> | původní jméno souboru , např. digibok_2007081301091_0011.jp2 0-1 | MC, XML, PS | M |
| <relationship> | vyjádření vztahu popisovaného souboru k jiným souborům a událostem (eventům) 0-n | MC, XML | M |
| <relationshipType> | typ vztahu, doporučené hodnoty: derivation= vztah kde objekt je výsledkem změny jiného objektu; structural= vztah mezi částmi objektu; tj. např. ALTO vytvořené z TIFFU bude mít vztah derivation, podobně jako JPEG2000 z TIFFu vytvořený; 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.; tj. např. ALTO nebo JPEG2000 vytvořený z původního TIFFu budou mít vztah „created from“ 1-1 | MC, XML; | M |
| <relatedObjectIdentification> | identifikace souvisejícího souboru 1-n pro MC, XML pro vyjádření vztahu k původnímu objektu (skenu) | MC, XML | M |

| | | | |
|--------------------------------|---|------------|---|
| <relatedObjectIdentifierType> | specifikace kontextu, ve kterém je identifikátor souboru jedinečný, např. URN; temporary filepath; objectID 1-1 | MC, XML | M |
| <relatedObjectIdentifierValue> | vlastní řetězec identifikátoru, např. URN:NBN:cz-1301091_011#0001 nebo název souboru, cesta k souboru apod. 1-1 | MC, XML | M |
| <relatedEventIdentification> | identifikace s popisovaným souborem související události (eventu); seznam událostí viz PREMIS event 0-n | MC, XML | M |
| <relatedEventIdentifierType> | typ události, např. interní číslovací systém událostí jako no.nb.evt; NK repository event ID, UUID apod. 1-1 | MC, XML | M |
| <relatedEventIdentifierValue> | hodnota identifikátoru události, např. NK_EVT_005 nebo hodnota UUID aj. 1-1 | MC, XML | M |
| <relatedEventSequence> | pořadí události, např. 003; k určení pořadí lze určit datum události 0-1 | MC, XML | R |
| <linkingEventIdentifier> | identifikátor události týkající původního skenu PS; typy událostí mohou být např. vytvoření, smazání 0-n 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. 1-1 | PS | M |
| <linkingEventIdentifierValue> | hodnota identifikátoru, např. event_01; img0001-master-event001 apod. 1-1 | PS | M |

8.4.2 PREMIS Event

- bude odpovídat poslední aktuální verzi v době implementace (leden 2011 - PREMIS data dictionary v. 2.1), nebo verzi předchozí
- 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 dělaly na obrazových datech
 - o digitalizace – vytvoření prvního skenu (např. do TIFF)
 - o vytvoření ALTO XML
 - o vygenerování MC

- vygenerování UC
- vymazání PS
- 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
- 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
 - element je povinný a neopakovatelný

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 je zcela dle konkrétních potřeb)

Pole záznamu PREMIS Event

| Element | Popis | Povinnost |
|------------------------|--|-----------|
| <eventIdentifier> | údaje o identifikátoru události v kontextu digitalizace nebo repozitáře 1-1 | M |
| <eventIdentifierType> | typ identifikátoru, např. no.nb.evt; NK_eventID, UUID apod. 1-1 | M |
| <eventIdentifierValue> | hodnota identifikátoru, např. EVT_001; event_019 apod. 1-1 | M |
| <eventType> | kategorizace události, nutno použít kontrolovaný slovník; typy událostí, které musí být zaznamenány: capture, migration, derivation, deletion 1-1 | M |
| <eventDateTime> | datum a čas kdy byla událost provedena; nutno | M |

| | | |
|--------------------------------|---|---|
| | zapsat v ISO 8601 na úroveň vteřin 1-1 | |
| <eventDetail> | další údaje o události, doporučené hodnoty pro výše uvedené <eventType> následují za /: <ul style="list-style-type: none"> - capture/digitization – vznik prvního skenu - capture/XML_creation - capture/TXT_creation - migration/MC_creation - derivation/UC_creation - deletion/PS_deletion 0-1 | M |
| <eventOutcomeInformation> | informace o výsledku události 0-n | R |
| <eventOutcome> | kategorizace výsledku události, např. slovy jako successful nebo failure, možno použít kódy – nutno používat kontrolovaný slovník nebo seznam kódů 0-1 | M |
| <linkingAgentIdentifier> | identifikace jednoho nebo více agentů spojených s událostí 0-n | M |
| <linkingAgentIdentifierType> | označení typu identifikátoru, např. NK_AgentID, UUID apod. 1-1 | M |
| <linkingAgentIdentifierValue> | hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod. 1-1 | M |
| <linkingAgentRole> | role agenta ve vztahu k události, např. software; SW component; operator; nutno používat kontrolovaný slovník 0-n | R |
| <linkingObjectIdentifier> | informace o objektu/souboru spojeného s událostí, link na něj 0-n | M |
| <linkingObjectIdentifierType> | označení typu identifikátoru, např. PhysUnitID; URN, NK_OBJ, OBJ_001 apod.; hodnoty by se měly brát z kontrolovaného slovníku 1-1 | M |
| <linkingObjectIdentifierValue> | hodnota identifikátoru, např. URN:NBN:cz- _0011#0001 aj. 1-1 | M |

8.4.3 PREMIS Agent

- bude odpovídat poslední aktuální verzi v době implementace (leden 2011 - PREMIS data dictionary v. 2.1), nebo verzi předchozí
- **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)**
 - o informace v PREMIS event a PREMIS object přicházející z procesu digitalizace v PSP balíčku jsou dostačující a dají nám dostatečné informace o události, kdy byla provedena, na jakém SW byla provedena (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)
 - o 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 Events (<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>, podčást <digiprovmD>
 - o AMD_METS.xml je linkován z hlavního METS záznamu dokumentu
- pro každého agenta, tj. jeden PREMIS agent záznam, bude vytvořena jedna <digiprovmD> část

Pole záznamu PREMIS Agent

Obsah pole „Popis“:

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

význam pole „povinnost“

- pole „povinnost“ uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné
- může nabývat následujících hodnot
 - o M - mandatory (povinně plnit – element je součástí každého záznamu)
 - o MA - mandatory if available (povinně plnit pokud je to možné, pokud lze apod.)
 - o R - recommended (plnění hodnot elementu je doporučeno, není ovšem povinné)
 - o RA - recommended if available (doporučeno pokud lze plnit)
 - o O - optional (plnění hodnot elementu je zcela dle konkrétních potřeb)

| Element | Popis | Povinnost |
|------------------------|--|-----------|
| <agentIdentifier> | popis identifikátoru, který jednoznačně označuje agenta v rámci jednoho kontextu (repozitář např.) 1-n | M |
| <agentIdentifierType> | označení typu identifikátoru, např. NK_AgentID, UUID apod. 1-1 | M |
| <agentIdentifierValue> | hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod. 1-1 | M |
| <agentName> | textové upřesnění agenta, např. přesný název SW, plné jméno osoby apod. - FixImage1.3; Jan Novák; CCS docWorks 6.2.1; 0-n | R |
| <agentType> | obecné označení agenta – pro osoby např. osoba, pro SW např. software apod. hodnoty: organization; person; software 0-1 | M |
| <agentNote> | použití pouze pokud je <agentType> Software a půjde o agenta souvisejícího s migrací TIFF na JPEG2000 (creation/migration Event); bude obsahovat příkaz k výrobě JPEG2000 souboru v aplikaci Kakadu 0-n | MA |

8.4.4 Technická metadata MIX

- Bude využit formát MIX, verze aktuální v době implementace projektu, nebo verze předchozí (prosinec 2010 verze 2 – viz <http://www.loc.gov/standards/mix/>)
- **MIX záznam vzniká pouze pro obrazové soubory!**
 - o tj. bude vznikat 1) pro archivní kopii, 2) další MIX záznam bude vznikat pro původní soubor vzniklý prvotním skenováním (nejčastěji TIFF) a to i přesto, že tento TIFF se v průběhu výroby maže a není archivován
 - o tyto dva MIX záznamy budou součástí jednoho METS záznamu AMD_METS.xml (v části <amdSec>, podčást <techMD>) pro administrativní a technická metadata, který vznikne ke každému obrazovému souboru a který je linkován z hlavního METS záznamu svazku monografie
- **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. – podrobnosti viz tabulka níže, sloupec „užití pro MC a PS“**

- pro každý záznam MIX bude vytvořena vlastní část <techMD>
- MIX může být také zapouzdřen v PREMIS Object <premis:objectCharacteristicsExtension>
- externí služby, jako např. JHOVE a 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 (viz MIX profily Nizozemí, Finska a Norska)
- <fileSize> je pouze doporučené, údaj o velikosti souboru je součástí popisu PREMIS object

Pole formátu MIX pro popis archivní kopie a původního skenu

Obsah pole „Popis“:

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

význam pole „povinnost“

- pole „povinnost“ uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné
- může nabývat následujících hodnot
 - o M - mandatory (povinně plnit – element je součástí každého záznamu)
 - o MA - mandatory if available (povinně plnit pokud je to možné, pokud lze apod.)
 - o R - recommended (plnění hodnot elementu je doporučeno, není ovšem povinné)
 - o RA - recommended if available (doporučeno pokud lze plnit)
 - o O - optional (plnění hodnot elementu je zcela dle konkrétních potřeb)

Obsah pole „Použití pro“

- použití jednotlivých elementů pro MC, PS (původní sken) a UC – 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; 0-n | R | MC, PS |
| <objectIdentifierType> | např. jméno souboru, nebo jiný identifikátor; 0-1 | M | MC, PS |
| <objectIdentifierValue> | hodnota identifikátoru, např. 20110306_001.jp2 nebo urn:nbn:123456; 0-1 | M | MC, PS |
| <fileSize> | velikost souboru | R | MC + PS |

| | | | |
|--|---|-----------|----------------|
| | 0-1 | | |
| <FormatDesignation> | údaje o formátu obrazového souboru 0-1 | M | MC, PS |
| <formatName> | název formátu, např. lze využít MIME types ⁴ (Image/jp2 apod.) 0-1 | M | MC, PS |
| <formatVersion> | verze formátu, např. 1.0 0-1 | M | MC, PS |
| <byteOrder> | endianita, možnosti jsou little endian, middle (mix) endian a big endian 0-1 | M | MC + PS |
| <Compression> | údaje o kompresi obrazového souboru (pokud 0-n | M | MC, PS |
| <compressionScheme> | informace o kompresním schématu, vyjádřeno číslem (např. 34712 je komprese JPEG2000) nebo slovy (např. JP2 Lossless) 0-1 | M | MC, PS |
| <BasicImageInformation> | základní technické údaje o obrazovém dokumentu 0-1 | M | MC, PS |
| <BasicImageCharacteristics> | 0-1 | M | MC, PS |
| <imageWidth> | šířka obrazu v pixelech, např. 3987 0-1 | M | MC, PS |
| <imageHeight> | výška obrazu v pixelech, např. 2345 0-1 | M | MC, PS |
| <PhotometricInterpretation> | photometrická interpretace 0-1 | M | MC, PS |
| <colorSpace> | barevný prostor, např. RGB 0-1 | M | MC, PS |
| <ColorProfile> | údaje o barevném profilu 0-1 povinné pro dokumenty, kde je nutno uchovat přesnou reprezentaci barvy původního dokumentu a používá se ICC profil) | MA | MC + PS |
| <IccProfile> | ICC profil 0-1 | M | MC + PS |
| <iccProfileName> | jméno profilu, např. sRGB, Adobe RGB aj. 0-1 | M | MC + PS |
| <iccProfileVersion> | verze profilu, např. sRGB IEC61966-2.1 0-1 | M | MC + PS |
| <iccProfileURI> | odkaz na profil, např. www.profil.cz/sRGB_v4_ICC_pref.icc ; 0-1 | R | MC + PS |

⁴ <http://www.iana.org/assignments/media-types/index.html>

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

| | | | |
|--|--|----------|-----------|
| | 0-1 povinné | | |
| <GeneralCaptureInformation> | základní údaje o skenování 0-1 | M | PS |
| <dateTimeCreated> | údaj o datu a čase skenování, např. 2009-01-03T08:25:28; zapsat v ISO 8601 na úroveň vteřin 0-1 | M | PS |
| <imageProducer> | entita provádějící skenování, např. The National Library of the Czech Republic, osoba apod. 0-1 | M | PS |
| <captureDevice> | typ skenovacího zařízení, např. reflection print scanner; doporučené využívání hodnot z kontrovaného slovníku 0-1 | M | PS |
| < ScannerCapture> | údaje o skeneru 0-1 | M | PS |
| <scannerManufacturer> | výrobce skeneru, např. 4DigitalBooks, Treventus, Zeutschel 0-1 | M | PS |
| <ScannerModel> | údaje o konkrétním typu skeneru 0-1 | M | PS |
| <scannerModelName> | jméno modelové řady skeneru, např. DL 0-1 | M | PS |
| <scannerModelNumber> | číslo/označení modelu, např. 3000 0-1 | M | PS |
| <scannerModelSerialNo> | výrobní číslo skeneru, např. E4R0003649 0-1 | M | PS |
| <MaximumOpticalResolution> | údaje o maximálním optickém rozlišení skeneru 0-1 | M | PS |
| < xOpticalResolution> | optické rozlišení na ose x, např. 300 0-1 | M | PS |
| < yOpticalResolution> | optické rozlišení na ose y, např. 300 0-1 | M | PS |
| < opticalResolutionUnit> | jednotka optického rozlišení, např. inch (in.) 0-1 | M | PS |
| <scannerSensor> | popis typu snímacího senzoru skenovacího zařízení, např. matrix, linear, undefined aj. 0-1 | M | PS |
| <ScanningSystemSoftware> | údaje o softwaru skenovacího zařízení 0-1 | M | PS |
| <scanningSoftwareName> | název softwaru, např. Copinet 0-1 | M | PS |
| <scanningSoftwareVersionNo> | číslo verze softwaru, např. 3.7 0-1 | M | PS |

| | | | |
|-------------------------------|---|----|----|
| <DigitalCameraCapture> | údaje o snímacím zařízení (fotoaparát) 0-1 povinné, pokud je používán fotoaparát a není používán skener | MA | PS |
| <digitalCameraManufacturer> | výrobce fotoaparátu, např. Canon 0-1 | M | PS |
| <DigitalCameraModel> | popis modelu fotoaparátu 0-1 | M | PS |
| <digitalCameraModelName> | název modelové řady, např. EOS 0-1 | M | PS |
| < digitalCameraModelNumber> | označení modelu fotoaparátu, např. 1000D 0-1 | M | PS |
| < digitalCameraModelSerialNo> | výrobní číslo přístroje, např. E12345 0-1 | M | PS |
| <camerarSensor> | typ senzoru fotoaparátu, např. matrix aj. 0-1 | M | PS |
| <CameraCaptureSettings> | údaje o nastavení fotoaparátu použitého ke snímání předloh 0-1 | M | PS |
| <ImageData> | v rámci tohoto kontejnerového elementu budou použity následující sub-elementy: fNumber exposureTime isoSpeedRatings shutterSpeedValue apertureValue brightnessValue exposureBiasValue maxApertureValue subjectDistance meteringMode lightSource flash focalLength backLight exposureIndex sensingMethod cfaPattern autoFocus PrintAspectRatio všechny hodnoty budou přebrány v případě použití fotoaparátu z údajů Exif | M | PS |
| <orientation> | popis orientace obrazu tak, jak je uložen vzhledem k jeho řádkům a sloupcům, např. normal*; normal, image flipper; normal, rotated 180°; unknown apod. | M | PS |

| | | | |
|--|---|-----------|---------------|
| | 0-1 | | |
| <ImageAssessmentMetadata> | informace o digitálním obrazu pro jeho hodnocení a využití z hlediska dlouhodobé ochrany apod. 0-1 | M | MC, PS |
| <SpatialMetrics> | rozměry obrázku, 2 rozměrná projekce objektů tak jak ji „vidí“ snímací zařízení 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); 0-1 | R | MC + PS |
| <samplingFrequencyUnit> | jednotka měření sampl. frekvence, např. hodnoty 1= žádná pevná jednotka ; 2= inch, 3=centimetr; 0-1 | M | MC, PS |
| <xSamplingFrequency> | údaje o počtu pixelů na jednotku smplovací frekvence pro šířku obrázku 0-1 povinné, pokud hodnota samplingFrequencyUnit je 2 nebo 3 | MA | MC, PS |
| <numerator> | čitatel, číselné vyjádření, např. 300 0-1 | M | MC, PS |
| <denominator> | jmenovatel, číselné vyjádření např. 1 0-1 | M | MC, PS |
| <ySamplingFrequency> | údaje o počtu pixelů na jednotku smplovací frekvence pro výšku obrázku 0-1 povinné, pokud hodnota samplingFrequencyUnit je 2 nebo 3 | MA | MC, PS |
| <numerator> | čitatel, číselné vyjádření, např. 300 0-1 | M | MC, PS |
| <denominator> | jmenovatel, číselné vyjádření např. 1 0-1 | M | MC, PS |
| <ImageColorEncoding> | doplňující údaje o barvě obrazu 0-1 | M | MC, PS |
| <BitsPerSample> | počet bitů na kanál 0-1 | M | MC, PS |
| <bitsPerSampleValue> | hodnota počtu bitů, např. 8, 1, 4 nebo 8,8,8 apod. 0-n POZOR – pro každou hodnotu je nutno element opakovat, tj. např. 3x element <bitsPerSampleValue> s hodnotou 8 | M | MC, PS |

| | | | |
|----------------------|--|-----------|-----------|
| | <mix:BitsPerSample> <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue> <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue> <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue> </mix:BitsPerSample> | | |
| <bitsPerSampleUnit> | specifikace jednotky, např. integer nebo floating point 0-1 | R | MC, PS |
| <samplesPerPixel> | počet barevných komponentů na pixel, např. 1, 3, 4 0-1 | M | MC, PS |
| <TargetData> | informace o kalibračních tabulkách 0-1 povinné pro obrazy, kde se dělá kontrola proti kalibrační tabulce | 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); 0-n | M | MC |
| <targetID> | údaje o původu kalibrační tabulky 0-n | M | MC |
| <targetManufacturer> | výrobce/původce kalibrační tabulky, např. Eastman Kodak nebo NK ČR, oddělení kontroly kvality apod. 0-1 | M | MC |
| <targetName> | název kalibrační tabulky, např. ColorChecker, MicrofilmScanTarget aj. 0-1 | M | MC |
| <targetNo> | číslo nebo verze kalibrační tabulky 0-1 | M | MC |
| <targetMedia> | údaj o tom, na jakém médiu je kalibrační tabulka, např. film, paper aj. 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 0-n 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 0-n | R | MC |
| <ChangeHistory> | dokumentace procesů provedených na | M | MC |

| | | | |
|--------------------------------|--|---|----|
| | obrazovém souboru v jeho životním cyklu 0-1 | | |
| <ImageProcessing> | údaje o zpracování obrazového souboru 0-n | M | MC |
| <dateTimeProcessed> | 2009-01-04T15:12:06; zapsat v ISO 8601 na úroveň vteřin 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; 0-1 | M | MC |
| <processingAgency> | The National Library of the Czech Republic 0-n | R | MC |

8.5 METS část <fileSec>

8.5.1 <fileSec> hlavního záznamu METS

file group

- pro obrazy i texty (ALTO XML/OCR.TXT) budou v hlavním METS záznamu použity elementy <fileGrp>, jeden element <fileGrp> bude existovat pro obrazy archivních kopií, další pro ALTO XML, další pro OCR.TXT soubory a další pro METS záznamy s technickými metadaty (AMD_METS.xml)
1. <fileGrp> pro obrazy archivních kopií, bude mít tyto atributy: ID="MC_IMGGRP" USE="Images"
 - každý soubor bude mít vlastní element <file> s následujícími atributy:
 - ID – identifikátor souboru jp2 jak je používán v METS záznamu
 - MIMETYPE – hodnota image/jp2
 - SIZE – velikost souboru jp2
 - 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 na obrazový soubor (xlink:href) a atribut LOCTYPE
 2. <fileGrp> pro ALTO XML bude mít následující atributy: ID="ALTOGRP" USE="Text"
 - každý ALTO XML soubor bude mít vlastní element <file> s následujícími atributy:
 - ID – identifikátor souboru ALTO XML jak je používán v METS záznamu
 - MIMETYPE – text/xml
 - SIZE – velikost souboru xml

- CHECKSUMTYPE – hodnota MD5
 - CHECKSUM - hodnota kontrolního součtu
 - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
 - subelementem pod <file> je element <Flocat>, který obsahuje link na xml soubor obsahující ALTO (xlink:href) a atribut LOCTYPE
3. <fileGrp> pro soubory METS s technickými metadaty AMD_METS.xml bude mít následující atributy: 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
 - 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 na xml soubor AMD_METS.xml (xlink:href) a atribut LOCTYPE
4. <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
 - MIMETYPE – text/plain
 - SIZE - velikost souboru
 - CHECKSUMTYPE – hodnota MD5
 - CHECKSUM - hodnota kontrolního součtu
 - CREATED - datum vytvoření, ISO8601 na úroveň vteřiny
 - subelementem pod <file> je element <Flocat>, který obsahuje link na txt soubor (xlink:href) a atribut LOCTYPE

8.5.2 <fileSec> vedlejšího METS záznam AMD_METS.xml

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

8.6 METS část <structMap> - Strukturální metadata a ALTO XML

8.6.1 <structMap> hlavního záznamu METS

- strukturální mapy v METS záznamu existují dvojího typu, fyzická a logická; fyzická zaznamenává hierarchické informace o dokumentu, včetně vazeb na fyzické soubory, ze kterých se skládají jednotlivé úrovně dokumentu
- 1 logická strukturální mapa v hlavním METS záznamu popisuje 1 svazek monografie a musí popisovat strukturu až na úroveň vnitřních částí (např. kapitol, nebo článků) apod.
 - součástí svazku monografie mohou být přílohy – pokud se skenují spolu se svazkem, popisuje strukturální mapa METS záznamu svazek včetně přílohy (bere se jako jeden svazek)
- strukturální mapa logická i fyzická včetně odkazů na ALTO XML bude v hlavním záznamu hlavni_METS.xml
- pro každou stránku seskupuje METS logická strukturální mapa odkazy na textové bloky (nebo ilustrace), které jsou součástí té stránky. Informace o blocích textu nebo ilustracích na stránce jsou uloženy v 1 ALTO XML souboru, který stránce odpovídá. Každý blok a každá ilustrace má unikátní identifikátor, který je použit jako odkaz v METS strukturální mapě.

Vyjádření fyzické strukturální mapy

- bude mít následující atributy <structMap LABEL="Physical_Structure" TYPE="PHYSICAL">
- fyzická strukturální mapa obsahuje rodičovský <div>, který obsahuje tyto atributy:
 - o LABEL- může obsahovat titul svazku monografie
 - o TYPE – např. monograph
 - o ID – identifikátor div
 - o DMDID – identifikátor části popisných metadat
- jednotlivé stránky jsou zanořeny do rodičovského elementu <div> jako dceřiné <div> elementy
 - <div> pro soubory stránky bude mít tyto atributy:
 - TYPE – bude se plnit typem stránky (viz typy stránek v DTD periodika http://digit.nkp.cz/Monographs/DTD/2.10/DocumentationMonograph/Monograph.html#element_MonographPage_Link032CD908)
 - ID – identifikátor div
 - ORDERLABEL – pořadové číslo stránky, jak je na ní vytištěno
 - ORDER – pořadí stránky ve svazku monografie
 - <div> pro soubory stránky vždy obsahují odkaz <ftpr> na soubor obrazu archivní, na ALTO XML, na OCR.TXT a na AMD_METS.xml pomocí elementu <par>
 - odkaz na obrazový soubor archivní kopie má v elementu <area> následující atributy: FILEID, který obsahuje ID souboru archivní kopie
 - odkaz na ALTO XML má v elementu <area> následující atributy: FILEID, který obsahuje ID ALTO XML souboru, dále BEGIN="P1" kde P1 je ID elementu <page> z ALTO XML souboru; a atribut BETYPE="IDREF"

- link na OCR.TXT soubor má v elementu <area> následující atributy: FILEID, který obsahuje ID souboru OCR.TXT
- link na AMD_METS.xml soubor má v elementu <area> následující atributy: FILEID, který obsahuje ID souboru AMD_METS.xml

Vyjádření logické strukturální mapy

- bude mít následující atributy <structMap LABEL="Logical_Structure" TYPE="LOGICAL">
- logická struktura na úroveň oddílů nebo např. ilustrací se popisuje pomocí do sebe zanořených elementů <div>
- pokud stránka obsahuje jen obraz a žádný text, pak je popsána jedním elementem <div> a link do souboru ALTO XML vede přímo na element <ComposedBlock>
- stránky obsahující více logických oblastí jsou popsány jedním <div> elementem, který má vnořené <div> elementy pro každou logickou oblast, která odpovídá např. textovému oddílu (např. kapitola, článek) nebo obraz.
 - a. pokud se jedná o jednoduchý, celistvý text na jedné straně, tak je popsán jen jedním <div> elementem s atributem TYPE="chapter"
 - v tomto <div> jsou dále jako další <div> elementy zanořeny jednotlivé textové bloky (odstavce, nadpisy, obrazy apod.)
 - u každého bloku je odkaz do ALTO XML souboru na příslušný textový blok <TextBlock> – pomocí tohoto odkazu se v ALTO XML souboru nalezne jak text, tak i informace o jeho umístění na stránce (souřadnice), toto je realizováno pomocí struktury <area> v elementu <ftpr>
 - u bloku tvořeného obrazem je odkaz do ALTO XML na příslušný komponovaný blok <ComposedBlock>; je realizováno pomocí struktury <area> v elementu <ftpr>
 - v případě použití atributu ORDER umožňuje tento princip u oddílů vyjádřit i tzv. pořadí čtení jeho částí, jako jsou např. nadpis, autor, obrázek apod.
 - b. výjimečně, pokud textový oddíl není celistvý a je rozdělen na více částí, které se vyskytují na jedné nebo více stránkách, které nemusejí jít za sebou, je možné určit pořadí čtení těchto částí, opět pomocí atributu ORDER
 - pro každou část oddílu existuje vlastní <div> element, podřízený hlavnímu <div> elementu oddílu
 - element <div> každé části má atribut TYPE hodnotu „chapter-part“ a atribut ID musí vyjadřovat o jakou z částí se jedná, tj. např. ID="chapter5-1" odpovídá první části oddílu číslo pět
- **POZOR – u monografie se dělení oddílů běžně nepředpokládá (kapitoly jsou běžně na více stránkách, většinou po sobě jdoucích)**
 - **tj. dělení oddílů není povinné a lze využít pouze struktury odstavců jak je naznačeno v první části příkladu, tj. pokud kapitola/odstavec pokračuje na další straně, logická mapa uvádí, že poslední odstavec (NORMAL_TEXT) např. na stránce 5 odkazuje na ALTO náležející ke stránce 5 a v něm na poslední textový blok; následující odstavec v logické mapě bude odkazovat na ALTO náležející ke stráně 6 a v něm na první textový blok (viz příklad)**

- do logické struktury PSP balíčku může být v případě její existence zakomponována i příloha (Supplement), která má vlastní <div> element s atributem TYPE="SUPPLEMENT"
 - o vnořené <div> elementy pro obraz a textové oddíly i jejich použití je shodné se způsobem popisu logické struktury u elementu <div> s atributem TYPE="VOLUME"

Příklad

Logická mapa obsahující svazek monografie se 5 textovými oddíly (chapter) a 1 přílohou (supplement). První oddíl (kapitola první) je na stránkách 1-2, na straně 1 je titul (název) kapitoly a odstavec normálního textu, na straně 2 je odstavec s normálním textem (může jít také o navazující odstavec – pokračující z předchozí stránky) a obrázek s popiskem i uvedeným autorem). Následují další tři textové oddíly, které nejsou součástí příkladu. Poslední textový oddíl (závěrečná kapitola pátá – doslov) začíná na straně 20 odstavcem s běžným textem (NORMAL_TEXT) a pokračuje na straně 21 dalším odstavcem. Na kapitole páté je ukázána možnost vyjádření návazností jednotlivých rozdělených částí textového oddílu (podobně jako u dělených článků periodik). Tato možnost může být využitelná v případě článků ve sborníku, nebo např. pokud část oddílu (např. kapitoly) pokračuje jinde ve svazku apod.

```

<structMap LABEL="Logical Structure" TYPE="LOGICAL">
  <div LABEL="Honzíkova cesta" TYPE="VOLUME" ID="VOLUME_1" DMDID="XY">
    <div LABEL="Kapitola první - O Honzíkovi" TYPE="CHAPTER" ID="CHAPTER_1" DMDID="XY" ORDER="0">
      <div TYPE="TITLE" ID="CHAPTER_PART_1" ORDER="1">
        <fptr>
          <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK1"/>
        </fptr>
      </div>
      <div TYPE="NORMAL_TEXT" ID="CHAPTER_PART_2" ORDER="2">
        <fptr>
          <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK2"/>
        </fptr>
      </div>
      <div TYPE="NORMAL_TEXT" ID="CHAPTER_PART_3" ORDER="3">
        <fptr>
          <area FILEID="ALTO_PAGE_2" BETYPE="IDREF" BEGIN="BLOCK1"/>
        </fptr>
      </div>
      <div LABEL="Vesnické nádraží" TYPE="PICTURE" ID="CHAPTER_PART_4" DMDID="XY" ORDER="4">
        <div TYPE="CAPTION" ID="CHAPTER_PART_5">
          <fptr>
            <area FILEID="ALTO_PAGE_2" BETYPE="IDREF" BEGIN="BLOCK2"/>
          </fptr>
        </div>
        <div TYPE="PICT_AUTHOR" ID="CHAPTER_PART_6">
          <fptr>
            <area FILEID="ALTO_PAGE_2" BETYPE="IDREF" BEGIN="BLOCK3"/>
          </fptr>
        </div>
        <div TYPE="IMAGE" ID="CHAPTER_PART_7">
          <fptr>
            <area FILEID="ALTO_PAGE_2" BETYPE="IDREF" BEGIN="COMPOSED_BLOCK1"/>
          </fptr>
        </div>
      </div>
    </div>
    ... následují např. 3 další textové oddíly (kapitoly) až po poslední kapitolu pátou ...
    <div LABEL="Doslov" TYPE="CHAPTER" ID="CHAPTER_5" DMDID="XY" ORDER="1">
      <div TYPE="CHAPTER_PART" ID="CHAPTER_2-1" ORDER="1">
        <div TYPE="NORMAL_TEXT" ID="CHAPTER_PART_1" ORDER="1">
          <fptr>
            <area FILEID="ALTO_PAGE_20" BETYPE="IDREF" BEGIN="BLOCK1"/>
          </fptr>
        </div>
      </div>
      <div TYPE="CHAPTER_PART" ID="CHAPTER_2-2" ORDER="2">
        <div TYPE="NORMAL_TEXT" ID="CHAPTER_PART_1" ORDER="1">
          <fptr>
            <area FILEID="ALTO_PAGE_21" BETYPE="IDREF" BEGIN="BLOCK1"/>
          </fptr>
        </div>
      </div>
    </div>
  </div>
  <div LABEL="Honzíkova cesta" TYPE="SUPPLEMENT" ID="SUPPL_1" DMDID="XY">
    ... popis článků a obrazů stejně jako u TYPE="VOLUME"
  </div>
</structMap>

```

kde jednotlivé části obsahují a popisují...

význam pole „povinnost“

- pole „povinnost“ uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné
- může nabývat následujících hodnot
 - o M - mandatory (povinně plnit – element je součástí každého záznamu)
 - o MA - mandatory if available (povinně plnit pokud je to možné, pokud lze apod.)
 - o R - recommended (plnění hodnot elementu je doporučeno, není ovšem povinné)
 - o RA - recommended if available (doporučeno pokud lze plnit)
 - o O - optional (plnění hodnot elementu je zcela dle konkrétních potřeb)

| <div> type | Atributy | Popis | Povinnost |
|--|---------------------------------------|---|-----------|
| VOLUME nebo SUPPLEMENT | LABEL TYPE ID DMDID | <div> obsahuje údaje o svazku monografie nebo o jeho příloze ----- LABEL – název (titul) svazku monografie, tedy např. „Honzíkova cesta“ TYPE- hodnota VOLUME nebo SUPPLEMENT ID – identifikátor <div>, např. hodnota „VOLUME_1“ nebo „SUPPL_1“ DMDID – obsahuje identifikátor DMD popisné části MODS svazku/přílohy | M |
| CHAPTER | LABEL TYPE ID DMDID ORDER | <div> obsahující údaje o jednom textovém oddílu a jeho částech ----- LABEL – název textového oddílu (např. kapitola, článek ve sborníku apod.) TYPE – hodnota CHAPTER s pořadovým číslem, např. CHAPTER_1 ID – identifikátor <div> elementu DMDID – identifikátor popisných metadat ORDER – pořadí oddílu | M |
| <div> TYPE=“CHAPTER“ může obsahovat další vnořený <div> různých typů popisující různé části textového oddílu, rozlišujeme tyto části (typy): <ul style="list-style-type: none"> - TITLE - SUBTITLE - AUTHOR - TRANSLATOR - NORMAL_TEXT – běžný text bez dalšího upřesnění | | | |

| | | | |
|---|---------------------------|---|----|
| <ul style="list-style-type: none"> - PICTURE - CHAPTER_PART - u oddílů, které jsou rozděleny na více míst na jedné stránce nebo více stránkách (v případě článků ve sborníku např.) <ul style="list-style-type: none"> o tento <div> pro jednu součást rozděleného článku pak může obsahovat stejné části jako <div> pro oddíl, tj. (TITLE, SUBTITLE, AUTHOR, TRANSLATOR, NORMAL_TEXT, PICTURE) | | | |
| TITLE | TYPE ID ORDER | <div> obsahující link na textový blok s nadpisem oddílu (tedy např. kapitoly) ----- TYPE – hodnota „TITLE“ ID – identifikátor <div> elementu, který popisuje jednu část oddílu (nadpis), např. hodnota „CHAPTER_PART_1“ ORDER – pořadí části oddílu | MA |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“ BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| SUBTITLE | TYPE ID ORDER | <div> obsahující link na textový blok s podnadpisem ----- TYPE – hodnota „SUBTITLE“ ID – identifikátor <div> elementu, který popisuje jednu část oddílu (podnadpis), např. hodnota „CHAPTER_PART_2“ ORDER – pořadí části oddílu | MA |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“ BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| AUTHOR | TYPE ID ORDER | <div> obsahující link na textový blok se jménem autora ----- TYPE – hodnota „AUTHOR“ ID – identifikátor <div> elementu, který popisuje jednu část oddílu (autor), např. hodnota „CHAPTER_PART_3“ ORDER – pořadí části oddílu | MA |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“ BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| TRANSLATOR | TYPE ID ORDER | <div> obsahující link na textový blok se jménem překladatele ----- TYPE – hodnota „TRANSLATOR“ ID – identifikátor <div> elementu, který popisuje jednu část oddílu (překladatel), např. hodnota | MA |

| | | | |
|--|---------------------------------------|---|----|
| | | „CHAPTER_PART_3“ ORDER – pořadí části oddílu | |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“ BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| NORMAL_TEXT | TYPE ID ORDER | <div> obsahující link na textový blok (nejčastěji odstavec) s běžným textem ----- TYPE – hodnota „NORMAL_TEXT“ ID – identifikátor <div> elementu, který popisuje jednu část oddílu (běžný text), např. hodnota „CHAPTER_PART_4“ ORDER – pořadí části oddílu | M |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“ BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| PICTURE | LABEL TYPE ID DMDID ORDER | <div> pro obraz náležející k textovému oddílu; plní se pokud se obraz vyskytuje ----- LABEL – název obrazu pokud existuje TYPE - PICTURE ID – identifikátor <div> elementu, který popisuje jednu část oddílu (běžný text), např. hodnota „CHAPTER_PART_3“ DMDID – link na bibliogr. popis obrazu ORDER – pořadí obrazu | MA |
| <div> element s typem PICTURE může obsahovat další <div> elementy s typy CAPTION, PICT_AUTHOR, PICT_TITLE a IMAGE; <ul style="list-style-type: none"> - CAPTION obsahuje text případného popisku k obrazu - PICT_AUTHOR obsahuje text se jménem případného autora obrazu - PICT_TITLE obsahuje text názvu obrazu, pokud nějaký název existuje - IMAGE – obsahuje link do souboru ALTO XML na blok popisující vlastní obraz | | | |
| CAPTION | TYPE ID | <div> obsahující link na textový blok s popisem obrazu ----- TYPE – hodnota CAPTION ID – identifikátor <div> elementu, např. „CHAPTER_PART_4“ | MA |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| PICT_AUTHOR | TYPE | <div> obsahující link na textový blok s autorem obrazu | MA |

| | | | |
|------------------|---------------------------|---|----|
| | ID | ----- TYPE – hodnota PIT_AUTHOR ID – identifikátor <div> elementu, např. „CHAPTER_PART_5“ | |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| PICT_TITLE | TYPE ID | <div> obsahující link na textový blok s názvem obrazu ----- TYPE – hodnota PICT_TITLE ID – identifikátor <div> elementu, např. „CHAPTER_PART_6“ | MA |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| IMAGE | TYPE ID | <div> obsahující link na komponovaný blok ALTO XML obsahující souřadnice vlastního obrazu ----- TYPE – hodnota IMAGE ID – identifikátor <div> elementu, např. „CHAPTER_PART_7“ | MA |
| <fptr> <area> | FILEID BEGIN BETYPE | FILEID – ID ALTO XML souboru BEGIN – ID komponovaného bloku v ALTO XML souboru BETYPE – hodnota IDREF | |
| CHAPTER_PART | TYPE ID ORDER | <div> obsahující další vnořené <div> odkazující na jednotlivé části konkrétní části rozděleného textového oddílu; možnos použít pro dělený oddíl (typu články např. ve sborníku) Pozn: pod <div> TYPE=“CHAPTER_PART“ lze vnořit všechny typy <div> jako pod <div> TYPE=“CHAPTER“ ----- TYPE – hodnota „CHAPTER_PART“ ID – identifikátor <div> konkrétní části, pro první část děleného oddílu např. „CHAPTER_2-1“, tj. první část oddílu 2 ORDER – pořadí konkrétní části děleného oddílu | MA |

POZOR - jednotlivé <div> elementy lze kombinovat a vytvářet nové struktury.

8.6.2 <fileSec> vedlejšího záznamu METS (AMD_METS.xml)

- bude obsahovat pouze fyzickou strukturální mapu (TYPE="PHYSICAL")
- ta bude obsahovat pouze jeden <div> element s atributem TYPE="MONOGRAPH_PAGE"
- do <div> budou vnořeny odkazy na jednotlivé reprezentace stránky svazku (MC, ALTO XML a OCR.TXT) pomocí elementu <fptr> s atributem FILEID

```
<structMap TYPE="PHYSICAL">
  <div TYPE="MONOGRAPH_PAGE">
    <fptr FILEID="JP2_0001"/>
    <fptr FILEID="ALTOXML_0001"/>
    <fptr FILEID="OCRTEXT_0001"/>
  </div>
</structMap>
```

8.7 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 (svazku monografie) včetně prázdných stran, fotografií hřbetu, předšátky apod.
- ALTO XML i OCR TXT budou vznikat na úrovni 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 ani jiné části vyskytující se na okrajích stránky (top, left, top a bottom margin)
 - o elementy topMargin, leftMargin, rightMargin, bottomMargin budou obsahovat elementy <TextBlock>, pro které platí stejná pravidla, jako pro element <textBlock> pro hlavní text stránky
 - o 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 dělicí 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.)
 - o např. ilustrace bude popsána v elementu /alto/Layout/Page/PrintSpace/ComposedBlock/GraphicalElement, kde ComposedBlock TYPE je Illustration
 - o reklama s textem v rámečku bude popsána v elementu Layout/Page/PrintSpace/ComposedBlock/TextBlock, kde ComposedBlock TYPE je Advertisement
 - o 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)
 - o 0-1 element je nepovinný, neopakovatelný
 - o 0-n element je nepovinný, opakovatelný

význam pole „povinnost“

- pole „povinnost“ uvádí, zda je plnění jednotlivých elementů povinné, doporučené nebo volitelné
- může nabývat následujících hodnot
 - o M - mandatory (povinně plnit – element je součástí každého záznamu)
 - o MA - mandatory if available (povinně plnit pokud je to možné, pokud lze apod.)
 - o R - recommended (plnění hodnot elementu je doporučeno, není ovšem povinné)
 - o RA - recommended if available (doporučeno pokud lze plnit)
 - o O - optional (plnění hodnot elementu je zcela dle konkrétních potřeb)
- POZOR - u elementů, které obsahují atributy, jsou atributy rozepsány pod čarou (vysvětlení, povinnost uvádění apod.)

| Element | Atribut | Popis | Povinnost |
|---------------------------------------|---------|---|-----------|
| <Description> | | | |
| <MeasurementUnit> | | měřicí jednotka pro souřadnice v ALTO XML; možné hodnoty – dpi, pixel, inch1200 a mm10); inch1200 = 1/1200 inche; doporučené plnění je „mm10“ nebo „pixel“; 0-1 | M |
| <sourceImageInformation> | | | |
| | | informace o obrazovém souboru, ze kterého vzniklo ALTO XML; 0-1 | M |
| <fileName> | | jméno obrazového souboru, ze kterého bylo ALTO XML vytvářeno; ideálně i s filesystem cestou jeho uložení; např. n1almageSeq-33386-b.tif//produkce/OCR/digibok_XY/XY_011.tiff | M |

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

| | | | |
|----------------------|---|---|---|
| <softwareVersion> | | verze SW, např. 6.2-1.16; 0-1 | M |
| <ocrProcessingStep> | | popis procesu vzniku OCR 1-1 – povinné pole | M |
| <processingDateTime> | | okamžik kdy bylo OCR vytvořeno; nutno zapsat v ISO 8601 na úroveň vteřin; 0-1 | M |
| <processingAgency> | | jméno nebo kód instituce, např. NK CZ doporučujeme použít kontrolovaný slovník hodnot; 0-1 | M |
| <processingSoftware> | | popis SW, který dělal vlastní OCR; 0-1 | M |
| <softwareCreator> | | výrobce softwaru - např. ABBYY, Russia; 0-1 | M |
| <softwareName> | | jméno softwaru - např. FineReader; 0-1 | M |
| <softwareVersion> | | např. 8.0; 0-1 | M |
| <Styles> | | styly definují vlastnosti jednotlivých grafických prvků stránky. styl definovaný v elementu vrchní úrovně je použit jako výchozí pro podřízené elementy; 0-1 | M |
| <TextStyle> | ID FONTSTYLE FONTFAMILY FONTSIZE | definuje font textu; 0-n ----- ID pro každý text style použitý v OCR souboru – povinné FONTSTYLE – např. bold, italics apod.; doporučujeme používat kontrolovaný slovník; doporučené FONTFAMILY – např. arial, calibri apod.; doporučujeme používat kontrolovaný slovník; povinné | M |

| | | | |
|------------------|---|---|---|
| | | <p>FONTSIZE – velikost fontu, např. 10, 12 apod.; povinné</p> | |
| <ParagraphStyle> | <p>ID ALIGN</p> | <p>definuje formátování textových bloků; 0-n</p> <p>-----</p> <p>ID pro každý odstavec + zarovnání; např. PAR_01, PAR_02 apod. povinné</p> <p>ALIGN – zarovnání; povolené hodnoty: Left, Right, Center, Block aj.; povinné</p> | M |
| <Layout> | | <p>layout - rozložení struktur (slov, odstavců apod.) na jedné stránce dokumentu; 1-1 povinný výskyt element není opakovací</p> | M |
| <Page> | <p>ID ACCURACY POSITION QUALITY PHYSICAL_IMG_NR HEIGHT WIDTH PC</p> | <p>element popisující jednu stránku dokumentu; 1-n</p> <p>-----</p> <p>ID – vygenerovaný identifikátor stránky, např. PAGE1, nebo P1 apod.; povinné</p> <p>ACCURACY – procentuální odhad přesnosti OCR (0-100); doporučené</p> <p>POSITION – pozice stránky; hodnoty k plnění: Left, Right, Foldout, Single, Cover; nepovinné</p> <p>QUALITY – krátký údaj o kvalitě předlohy stránky; hodnoty k plnění: OK, Missing, Missing in original, Damaged, Retained, Target, As in original; nepovinné</p> | M |

| | | | |
|--------------|--|---|----|
| | | <p>PHYSICAL_IMG_NR - fyzické (pořadové) číslo stránky v dokumentu; vyjádřeno číslem, např. 1,2,3 apod.; povinné</p> <p>WIDTH – šířka stránky vyjádřená v pixelech; povinné</p> <p>HEIGHT – výška stránky vyjádřená v pixelech; povinné</p> <p>PC = Confidence level OCR souboru – hodnota mezi 0 (nejistá kvalita) a 1 (dobrá kvalita); nepovinné; pokud nevyplníte ACCURACY – tak je vyplnění doporučené</p> | |
| <TopMargin> | <p>ID HPOS VPOS WIDTH HEIGHT</p> | <p>horní okraj – prostor mezi vrchní hranou listu a vrchní linkou textu; 0-1</p> <p>-----</p> <p>ID: unikátní ID pro element TopMargin, např. P1_TM0001 (page 1, topMargin0001); povinné</p> <p>HPOS: horizontální pozice; povinné</p> <p>VPOS: vertikální pozice; povinné</p> <p>WIDTH – šířka vrchního okraje; povinné</p> <p>HEIGHT – výška vrchního okraje; povinné</p> | M |
| <TextBlock> | stejně plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | | MA |
| <LeftMargin> | <p>ID HPOS VPOS</p> | <p>levý okraj – prostor mezi levým okrajem stránky a textem; 0-1</p> <p>-----</p> | M |

| | | | |
|----------------|--|---|----|
| | WIDTH HEIGHT | ID: unikátní ID pro element LeftMargin, např. P1_LM0001 (page 1, leftMargin0001); povinné HPOS: horizontální pozice; povinné VPOS: vertikální pozice; povinné WIDTH – šířka levého okraje; povinné HEIGHT – výška levého okraje; povinné | |
| <TextBlock> | stejně plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | | MA |
| <RightMargin> | ID HPOS VPOS WIDTH HEIGHT | pravý okraj – prostor mezi pravým okrajem stránky a textem; 0-1 ----- ID: unikátní ID pro element RightMargin, např. P1_RM0001 (page 1, rightMargin0001); povinné HPOS: horizontální pozice; povinné VPOS: vertikální pozice; povinné WIDTH – šířka pravého okraje; povinné HEIGHT – výška pravého okraje; povinné | M |
| <TextBlock> | stejně plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | | MA |
| <BottomMargin> | ID HPOS VPOS WIDTH HEIGHT | pravý okraj – prostor mezi spodním okrajem stránky a textem; 0-1 ----- ID: unikátní ID pro element | M |

| | | | |
|--------------|--|---|----|
| | | <p>BottomMargin, např. P1_BM0001 (page 1, bottomMargin0001); povinné</p> <p>HPOS: horizontální pozice; povinné</p> <p>VPOS: vertikální pozice; povinné</p> <p>WIDTH – šířka spodního okraje; povinné</p> <p>HEIGHT – výška spodního okraje; povinné</p> | |
| <TextBlock> | stejně plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace> | | MA |
| <PrintSpace> | <p>ID</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p> | <p>popis tvaru pokrývajícího textové pole stránky; 0-1</p> <p>-----</p> <p>ID: unikátní ID pro element <printSpace>, např. P1_PS0001 (page 1, printSpace0001); - povinné</p> <p>HPOS: horizontální pozice; povinné</p> <p>VPOS: vertikální pozice; povinné</p> <p>WIDTH – šířka textového pole; povinné</p> <p>HEIGHT – výška textového pole; povinné</p> | M |
| <TextBlock> | <p>ID</p> <p>STYLEREFS</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p> | <p>popisy textových bloků na konkrétní stránce; 0-n</p> <p>pokud je stránka prázdná, TextBlock není potřeba uvádět; pokud je na stránce text tak ano</p> <p>-----</p> <p>ID obsahuje identifikátor textového bloku na stránce, např. "BLOCK1" nebo P1_TB0002</p> | MA |

| | | | |
|------------|--|--|----|
| | | <p>(stránka 1, textový blok 2); povinné</p> <p>STYLEREFS: reference na ID definice formátování textových bloků <ParagraphStyle>; povinné</p> <p>HPOS: horizontální pozice bloku; povinné</p> <p>VPOS: vertikální pozice bloku; povinné</p> <p>WIDTH – šířka textového bloku; povinné</p> <p>HEIGHT – výška textového bloku; povinné</p> | |
| <Shape> | | <p> tvar textového bloku; 0-1 – pro jeden výskyt <TextBlock> jeden nebo žádný výskyt <Shape>; plnit v případě, že je tvar textového bloku nestandardní (víceúhelník)</p> | RA |
| <Polygon> | POINTS | <p>popis (souřadnice) tvaru víceúhelníku; 0-1</p> <p>-----</p> <p>POINTS – vyjádření jednotlivých bodů víceúhelníku; povinné</p> | M |
| <TextLine> | ID STYLEREFS HPOS VPOS WIDTH HEIGHT | <p>popis jedné řádky textu v rámci textového bloku; 1-n nutný alespoň jeden výskyt v rámci textového bloku</p> <p>-----</p> <p>ID obsahuje identifikátor řádky textu v textovém bloku, např. "P1_TL0002 (stránka 1, řádka 2); povinné</p> <p>STYLEREFS: reference na ID definice formátování textových bloků <ParagraphStyle>; nepovinné</p> | M |

| | | | |
|----------|--|--|---|
| | | <p>HPOS: horizontální pozice řádky; povinné</p> <p>VPOS: vertikální pozice řádky; povinné</p> <p>WIDTH – šířka řádky; povinné</p> <p>HEIGHT – výška řádky; povinné</p> | |
| <String> | <p>ID CONTENT HEIGHT WIDTH HPOS VPOS ----- CC WC</p> <p>V případě dělení slov také: SUBS_TYPE SUBS-CONTENT</p> | <p>řetězec znaků – vlastní obsah OCR; znaky tvoří jednotlivá slova a více tagů <String> větu <TextLine>; 1-n v rámci <TextLine></p> <p>-----</p> <p>ID obsahuje unikátní sekvenční číslo řetězce na stránce, např. "P3_ST0001" (strana 3, řetězec 1); povinné</p> <p>CONTENT – ukládá vlastní řetězec znaků (slovo); povinné</p> <p>HPOS: horizontální pozice řetězce; povinné</p> <p>VPOS: vertikální pozice řetězce; povinné</p> <p>WIDTH – šířka řetězce; povinné</p> <p>HEIGHT – výška řetězce; povinné</p> <p>CC – úroveň důvěry v přesnost OCR rozpoznání každého znaku v řetězci; jde o seznam čísel, každé z nich mezi hodnotami 0 (jistá) a 9 (nejistá) pro každý znak; např. CC="0001" pro</p> | M |

| | | | |
|---------------|--|---|---|
| | | <p>CONTENT="TEXT"; povinné</p> <p>WC – úroveň důvěry v přesnost OCR výstupu celého řetězce - slova (word confidence); hodnota mezi 0 (nejistá) a 1 (jistá); např. WC="0,99"; povinné</p> <p>SUBS_CONTENT – obsah chybějící části řetězce v případě, že je slovo na konci řádku rozdělené i do druhého řádku; obsahuje celý řetězec - aby byl vyhledatelný i v případě, že slovo se na stránce vyskytuje, ale je rozděleno; povinné</p> <p>SUBS_TYPE – označení typu substituce; možné hodnoty: HypPart1; HypPart2; Abbreviation; povinné - při výskytu SUBS_CONTENT</p> <p><i>HypPart1</i> se vyskytuje při rozdělení slova u jeho první OCR části (u první části tagu <CONTENT> ve větě (stringu) první; <i>HypPart2</i> se vyskytuje u následujícího tagu <CONTENT> v následující větě (stringu), který obsahuje druhou část rozděleného slova/řetězce; <i>Abbreviation</i> – typ substituce používaný při rozepisování zkratk v textu na jejich plný text; při dělení slov v textu HypPart1 a HypPart2 povinné, abbreviation nepovinné</p> | |
| <ALTERNATIVE> | | <p>alternativní hodnota OCR řetězce pro jednotlivá slova; 0-n</p> | 0 |

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

| | | | |
|-----------------|--|---|----|
| | | povinné | |
| <ComposedBlock> | ID TYPE HPOS VPOS WIDTH HEIGHT STYLEREFS | <p>blok sestávající z jiných bloků; může obsahovat PrintSpace/ComposedBlock/Text Block, PrintSpace/ComposedBlock/Illust ration, PrintSpace/ComposedBlock/Grap hicalElement, /PrintSpace/ComposedBlock/Co mposedBlock, tj. stejné elementy (bloky), které obsahuje samotný element /alto/Layout/Page/PrintSpace; 0-n povinné pro vyjádření bloků textu (např. orámovaný text, reklamy), pro vyjádření ilustrací, tabulek a grafik</p> <p>----- ID: unikátní ID komponovaný blok, např. P6_CB0001 (stránka 6, komponovaný blok 0001); povinné</p> <p>TYPE – označení typu komponovaného bloku; nutné používat kontrolovaný slovník (illustration, Advertisement, apod.); povinné</p> <p>HPOS: horizontální pozice bloku; povinné</p> <p>VPOS: vertikální pozice bloku; povinné</p> <p>WIDTH – šířka komponovaného bloku; povinné</p> <p>HEIGHT – výška komponovaného bloku;</p> | MA |

| | | | |
|-------------|--|--|----|
| | | povinné | |
| <Shape> | | <p>tvár komponovaného bloku; 0-1 – pro jeden výskyt /alto/Layout/Page/PrintSpace/ComposedBlock jeden nebo žádný výskyt /alto/Layout/Page/PrintSpace/ComposedBlock/Shape; doporučeno – v případě, že je tvár komponovaného bloku nestandardní (víceúhelník)</p> | RA |
| <Polygon> | POINTS | <p>popis tvaru víceúhelníku; 0-1</p> <p>-----</p> <p>POINTS – vyjádření jednotlivých bodů víceúhelníku povinné</p> | M |
| <TextBlock> | ID STYLEREFS HPOS VPOS WIDTH HEIGHT | <p>v případě, že komponovaný blok (např. orámovaný tvar) obsahuje text; platí stejná pravidla jako pro normální element /alto/Layout/Page/PrintSpace/TextBlock; 0-n (pro jeden výskyt <ComposedBlock> 0 nebo více elementů /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock>; plnit pokud je v komponovaném bloku text</p> <p>-----</p> <p>ID obsahuje identifikátor textového bloku v komponovaném bloku, např. P1_CB0002_SUB (stránka 1, textový blok 2, SUB značí komponovaný blok); povinné</p> <p>STYLEREFS: reference na ID definice formátování textových bloků /alto/Styles/ParagraphStyle; povinné</p> | MA |

| | | | |
|--------------------|---|--|----|
| | | <p>HPOS: horizontální pozice bloku; povinné</p> <p>VPOS: vertikální pozice bloku; povinné</p> <p>WIDTH – šířka textového bloku; povinné</p> <p>HEIGHT – výška textového bloku; povinné</p> | |
| <TextLine> | <p>/alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock/TextLine a ostatní elementy v rámci</p> <p>/alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock mají stejná pravidla a výskyty jako jako ve vrchním elementu /alto/Layout/Page/PrintSpace/TextBlock</p> | | |
| <GraphicalElement> | <p>ID</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p> | <p>popis grafického tvaru; v případě využití v rámci</p> <p>/alto/Layout/Page/PrintSpace/ComposedBlock označuje rozměry tvaru v rámci něhož je tabulka, ilustrace, reklama apod.;</p> <p>0-1 - pro jeden výskyt</p> <p>/alto/Layout/Page/PrintSpace/ComposedBlock 0 nebo max. 1 výskyt <GraphicalElement>; plní se pokud je na stránce a tedy v komponovaném bloku ilustrace, tabulka apod.;</p> <p>-----</p> <p>ID – identifikátor grafického tvaru; povinné</p> <p>HEIGHT – výška grafického tvaru; povinné</p> <p>WIDTH – šířka grafického tvaru; povinné</p> <p>HPOS – horizontální pozice grafického tvaru;</p> | MA |

| | | | |
|--|--|---|--|
| | | povinné VPOS – vertikální pozice grafického tvaru; povinné | |
|--|--|---|--|