

# Definice metadatových formátů pro digitalizaci periodik pro ANL

Jan Hutař, NK ČR

VERZE 1.0 - 7.9.2011

jméno	datum	verze dokumentu	provedené změny
Jan Hutař - NK	25.-7.9.2011	verze 1.0	kompletní úprava logické strukturální mapy; doplnění detailů o METS záznamu
Jan Hutař - NK	22-24.8.2011	draft 0.2	vymazání UC z celého dokumentu, úpravy textu, výměna obrázku; k vnitřní části přidán element recordInfo – nutno kvůli údajům o popisu článku; element <subject> přidán k popisu titulu, čísla, vnitřní části a přílohy; k číslu přidán abstrakt
Jan Hutař - NK	15.8.2011		vymazán element <name> na úrovni titulu; drobné úpravy
Jan Hutař - NK	21.7.2011	draft 0.1	první znění

1.	VÝCHODISKA .....	2
2.	VÝSTUPY DIGITALIZACE .....	2
3.	GRANULARITA METADATOVÉHO ZÁZNAMU .....	3
4.	NÁZVOVÁ KONVENCE SLOŽEK A SOUBORŮ .....	3
5.	TRANSPORTNÍ BALÍK PRO JEDEN NEBO VÍCE PSP BALÍČKŮ.....	4
6.	IDENTIFIKÁTORY .....	5
7.	STRUKTURA PSP BALÍČKU .....	5
8.	METADATA .....	8
8.1	METS hlavička <metsHdr> .....	8
8.2	METS část <dmdSec> - Bibliografická metadata – formát MODS a Dublin Core.....	9
8.3	METS část <amdSec> - Technická a administrativní metadata – formáty MIX a PREMIS.....	33
8.4	METS část <fileSec> .....	55
8.5	METS část <structMap> - Strukturální metadata a ALTO XML.....	56
8.6	OCR (ALTO XML a TXT OCR).....	64

## 1. Východiska

- uživatelské kopie = UC
- archivní kopie = MC
- původní sken – PS – obrazový soubor vzniklý při digitalizaci, který se po zpracování (ořez, narovnání apod.), maže se 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 = číslo periodikaPSP 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. číslo periodika
  - o z workflow digitalizace lze poslat více PSP balíčků v balíku např. [.zip], [.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
- základní bibliografická metadata budou stahována přímo z knihovních katalogů do workflow digitalizace
- 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

Periodika

- základní intelektuální entitou periodik je 1 číslo
- každé číslo periodika má svůj vlastní metadatový záznam (=METS), který obsahuje údaje o nadřazených entitách čísla jako jsou ročník, titul periodika, tj. je pro uživatele i pro systém možné spojit jednotlivá čísla do ročníků a titulů

### 4. Názvová konvence složek a souborů

pojmenování PSP balíčku

- každý PSP balíček přicházející z digitalizace by měl obsahovat pouze jedinou intelektuální entitu (číslo periodika). Pak by název balíčku měl vycházet z identifikátoru této entity, např. URN:NBN, číslo čárového kódu použitého na fyzické jednotce apod.

pojmenování složek

- viz návrh struktur PSP balíčku (kap. 7)

pojmenování souborů

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

S využitím URN:NBN mohlo by to vypadat následovně (podobný princip může být použit s využitím čárového kódu nebo jiného identifikátoru):

typ souboru	název souboru	vysvětlení
PSP balíček (číslo, svazek)	URN_123456	název celé složky PSP balíčku, u základních int. entit bude v názvu využito vždy URN:NBN

archivní kopie	MC_URN_123456_0013.jp2	archivní JPEG2000 stránky 13 čísla periodika s URN:NBN 123456
uživatelská kopie	UC_URN_123456_0013.jp2	uživatelská kopie ve formátu JPEG2000 stránky 13 čísla periodika s URN:NBN 123456
ALTO XML	ALTO_URN_123456_0013.xml	ALTO soubor náležející ke 13té stránce z čísla periodika s URN:NBN 123456
OCR TXT	TXT_URN_123465_0013.txt	TXT soubor s OCR náležející ke 13té stránce z čísla periodika s URN:NBN 123456
info.xml	INFO_URN_123456.xml	info xml k celému PSP balíčku čísla periodika
MD5.xml	MD5_URN_123456.xml	xml s kontrolními součty k celému PSP balíčku čísla periodika
Hlavni_METS.xml	METS_URN_123456.xml	hlavní METS záznam k celému číslu periodika s URN:NBN 123456
AMD_METS.xml	AMD_METS_URN_123456_0013.xml	METS záznam s technickými metadaty pro stránku 13 z čísla periodika s URN:NBN 123456

## 5. Transportní balík pro jeden nebo více PSP balíčků

Pokud bude jeden PSP balík obsahující 1 základní intelektuální entitu (číslo periodika) přemísťován např. jako zip nebo tar, měl by název souboru zip/tar odpovídat názvu PSP balíčku (tedy z čČNB, ISSN nebo ISBN).

Výstupem workflow digitalizace ale může také být balík (např. zip nebo 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 čísla titulu periodika, měl by název balíku vycházet z čČNB nebo z ISSN
- v případě, že balík obsahuje svazky vícesvazkového díla, měl by název balíku vycházet z čČNB nebo ISBN
- typ identifikátoru musí být vyjádřen v názvu souboru – např. ISSN\_123456.zip nebo CCNB\_123456.zip apod.
- lze počítat s tím, že bude docházet k tomu, že sdružený balík nebude obsahovat např. všechny čísla určitého titulu periodika – tato skutečnost musí být patrná z názvu balíku (např. ISSN\_123456\_YYYY kde YYYY může být pořadové číslo, datum, doba vzniku jednoho z více balíčků obsahujících čísla určitého titulu s identifikátorem ISSN 123456).

Transportní balík by měl obsahovat následující části:

- balíčky PSP (svazků nebo čísel)
- kontrolní součty všech PSP balíčků
- seznam entit, které balík má obsahovat

Do úvahy mohou přijít balicí metody jako BagIt<sup>1</sup>, tar, zip apod.

## 6. 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 – pouze pro titul periodika, ISSN není přiděleno vždy (chybí např. u starých titulů z 19. století)
- ČČNB – identifikátor entity tak jak odpovídá katalogizačnímu záznamu, tj. každá entita se záznamem v katalogu NK/MZK má tento identifikátor

Bylo by ideální, aby během digitalizace byl přidělován (generován přímo nebo vyžádán z aplikace Resolver URN:NBN) identifikátor URN:NBN:

- přidělován bude logickým úrovním (entitám)
  - o u periodik tedy: číslo, případně ročník a titul a vnitřním částem
- další možností je využití UUID

## 7. Struktura PSP balíčku

V kapitole je návrh struktury balení dat a metadat v jednom PSP balíčku na výstupu z workflow digitalizace.

---

<sup>1</sup> <https://confluence.ucop.edu/display/Curation/BagIt>

složka	obsahuje >	obsahuje >	obsahuje>								
svazek monografie /číslo periodika											
	info.xml	údaje o vzniku balíku (PREMIS nebo DC)									
	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>	<table border="1"> <tr> <td>amdSec</td> <td>techMD = PREMISobject pro MC, původní TIFF, ALTO XML) + MIX pro MC, původní TIFF)</td> </tr> <tr> <td></td> <td>digiprovMD = PREMISevent + PREMISagent</td> </tr> <tr> <td>fileSec</td> <td>odkazuje na MC, ALTO XML, OCR TXT soubor popisované 1 stránky</td> </tr> <tr> <td>StructMap</td> <td>pouze fyzická - pro soubory popisované stránky (MC a ALTO XML, OCR TXT)</td> </tr> </table>	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)
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)										
	Hlavni_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.XML		kontrolní součet pro celý PSP balík = pro všechny soubory a složky v 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.

PSP balíček = 1 složka pro 1 číslo periodika.

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.

#### **složka [masterCopy]**

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

#### **složka [userCopy]**

složka s uživatelskými kopiemi, pro každou naskenovanou stránku čísla periodika 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 čísla periodika.

#### **složka [TXT]**

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

#### **složka [amdSec]**

složka s technickými metadaty – **obsahuje pro každou naskenovanou stránku čísla časopisu 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, 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.
  - 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.
  - 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 číslu periodika včetně popisu nadřazených entit (např. ročník, 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 číslu periodika. Obsahuje také linky na administrativní metadata AMD\_METS.xml do složky [amdSec].
- structMap – strukturální mapa pro celý dokument, tj. pro jedno číslo periodika. Obsahuje:
  - o logickou část pro archivní, ALTO XML, OCR TXT a AMD\_METS.xml
  - o fyzickou část pro archivní, ALTO XML, OCR TXT a AMD\_METS.xml
  - o mapování na ALTO XML areas

#### soubor MD5.XML

poslední částí PSP balíčku je xml soubor s kontrolními součty pro celý balíček, pro každý objekt a pro každou složku. Z tohoto důvodu nejsou kontrolní součty součástí složek s objekty. Kontrolní součty jsou také samozřejmě v technických metadatach.

## 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**

### 8.1 METS hlavička <metsHdr>

- dokumentuje vznik a úpravy METS záznamu

element	atributy	popis
<metsHdr>	LASTMODDATE CREATEDATE	hlavička METS záznamu povinné ----- LASTMODDATE – datum poslední úpravy záznamu, musí být ve tvaru ISO 8601 (na úrovni vteřin)  CREATEDATE – datum vytvoření záznamu, musí být ve tvaru ISO 8601 (na úrovni vteřin)
<agent>	ROLE TYPE	údaje o tvůrci záznamu METS povinné ----- ROLE – hodnota „CREATOR“



		TYPE – hodnota „ORGANIZATION“
<name>		jméno jednotlivce nebo organizace; povinné ----- tvůrce záznamu, buď dodavatel (firma XY) nebo v případě tvorby záznamu v knihovně bude využita sigla knihoven, tj. pro NK ČR hodnota „ABA001“

## 8.2 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
<dmdSec>	ID	identifikátor <dmdSec> části METS záznamu povinné ----- ID: pro <dmdSec> s popisem titulu periodika hodnota „MODSMD_TITLE“ pro záznam v MODS nebo „DCMD_TITLE“ pro záznam v Dublin Core  pro <dmdSec> s popisem čísla periodika hodnota „MODSMD_ISSUE“ a „DCMD_ISSUE“  pro <dmdSec> s popisem vnitřní části periodika hodnota dle typů vnitřní části (článek, obraz) - hodnoty „MODSMD_ART“ a „DCMD_ART“ pro článek a hodnoty „MODSMD_PICT“ a „DCMD_PICT“ pro obraz  pro <dmdSec> s popisem přílohy periodika hodnota „MODSMD_SUPPL“ a „DCMD_SUPPL“
<mdWrap>	MDTYPE MIMETYPE	element obsahující vložené záznamy MODS povinné ----- MDTYPE – hodnota „MODS“ pro záznamy v MODS, hodnota „DC“ pro záznam v Dublin Core  MIMETYPE – hodnota „text/xml“

- 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](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 struktura 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 struktura PSP balíčku v kap. 7
- základním zdrojem pro popisná metadata je katalog NK a MZK
- u digitalizovaných dokumentů je bibliografický popis vytvářen primárně z pohledu popisu fyzické předlohy, nejde o popis elektronického dokumentu

### Periodika

- základní intelektuální entitou pro popis je číslo periodika, tj. v jednom METS záznamu, který bude obsahovat metadata a strukturu jednoho čísla periodika, budou MODS záznamy k tomuto číslu
- metadata budou popisovat následující entity:
  - o titul (Title)
  - o číslo (Issue)
  - o vnitřní část (InternalPart) – typy články (Article) a obraz (Picture)
  - o příloha (Supplement)
- 
- ad titul (Title) – MODS záznam bude obsahovat i číslo ročníku
- ad číslo (Issue) – typy čísla jsou v elementu <genre> za použití atributu type
- ad vnitřní část (InternalPart) - typy vnitřní části články a obraz by měly pokrýt veškerou variabilitu možností, které mohou texty a obrázky na tištěné stránce mít; bližší určení typů článku (novinky, zprávy, reklama apod.) 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 článků bude vznikat manuálně a je žádoucí o vzniku záznamu článku držet kontrolní údaje; element je volitelný
- ad příloha (Supplement) - přílohou se rozumí volně vložená entita do jednotlivého čísla, např. mapa, obsah celého ročníku, CD/DVD apod.
- jednotlivé MODS záznamy pro části (titul, číslo, vnitřní část a příloha) nejsou samopopisné, tj. neobsahují vždy údaje o vrchních entitách (článek neobsahuje informace o titulu apod.)
- pro každou entitu vznikne jeden MODS záznam s vlastním ID, které bude označovat i typ části (např. článek, ilustrace apod.) v případě opakování částí se bude opakovat odpovídající počet MODS záznamů

- každý MODS záznam bude uložen ve vlastní METS části <dmdSec> pomocí mdWrap
- u úrovní kde je to potřeba (vnitřní část, příloha apod.) se budou opakovat <dmdSec> části tolikrát, kolik je konkrétních částí
  - o tj. v METS záznamu vznikne 1 část <dmdSec> pro bibliografický záznam titulu periodika, 1 <dmdSec> část pro bibliografický záznam čísla periodika, několik <dmdSec> částí pro vnitřní části (pro všechny články i obrázky) a odpovídající počet <dmdSec> částí pro přílohy, dle počtu příloh
    - bibliografický popis obrazů bude velmi minimalistický
- záznam periodika v katalogu – v katalozích NK a MZK existuje záznam pouze pro titul periodika, neexistují samostatné záznamy pro čísla, ročníky apod. – tj. vnitřní členění a popis musí vzniknout v digitalizaci, popis titulu periodika musí být stažen z katalogu do workflow digitalizace
- stránka se nebude popisovat, její logické i fyzické číslování i typ stránky je obsaženo ve struktuře METS dokumentu (část structMap)
  - o typ stránky (Advertisement, Blank, Index aj.) budou odpovídat přesně seznamu typů z DTD periodika – viz <http://digit.nkp.cz/DigitizedPeriodicals/DTD/2.10/Periodical.xsd>
- **všechny top elementy MODS formátu jsou opakovatelné, kromě <recordInfo>**
- **všechny elementy Dublin Core jsou opakovatelné**

### 8.2.1 Pole MODS a Dublin Core pro jednotlivé části periodika

#### Obsah pole „Popis“:

- vysvětlení a příklad
- doporučené plnění tam, kde je to možné uvést
- povinnost plnění dle NK ČR (slovní vyjádření: povinné, doporučené, nepovinné)

#### Pole MODS a DC pro titul periodika

Element MODS	Atributy	Popis	Element Dublin Core
<titleInfo>	ID	název titulu periodika povinné – použít katalogizační záznam ----- ID musí vyjadřovat název úrovně, tj. např. „MODS_TITLE“	
<title>		názvová informace – název periodika povinné – převzít z katalogu	<dc.title> povinné
<subTitle>		podnázev periodika povinné pokud lze uvést	<dc.title> povinné pokud lze uvést
<partNumber>		číslo části, např. určité řady/edice (část 1, řada B), k použití u ročenek apod.	<dc:description> doporučené

		doporučené	
<partName>		jméno edice nebo speciální ediční řady, např. Hygiena. k použití u ročenek a specializovaných periodik doporučené	<dc:description> doporučené
<typeOfResource>		popis charakteristiky typu nebo obsahu zdroje doporučené jedna z hodnot: <ul style="list-style-type: none"> <li>- text</li> <li>- cartographic</li> <li>- notated music</li> <li>- sound recording-musical</li> <li>- sound recording-nonmusical</li> <li>- sound recording</li> <li>- still image</li> <li>- moving image</li> <li>- three dimensional object</li> <li>- software, multimedia</li> <li>- mixed material</li> </ul> pro periodika a monografie hodnota text; mělo by se vyčítat z MARC21 katalogizačního záznamu z pozice 06 návěští	<dc:type> doporučené
<genre>		bližší údaje o typu dokumentu povinné hodnota: title	<dc:type> povinné
<originInfo>		informace o původu předlohy povinné  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.	

		Pokud bylo za dobu vydávání více vydavatelů, nutno vzít z katalogizačního záznamu pole 260 indikátor 02 a údaje o vydavatelích opakovat.	
<place>		údaje o místě spojeném s vydáním, výrobou nebo původem popisovaného dokumentu povinné pokud lze uvést	<dc:coverage> povinné pokud lze uvést
<placeTerm>	type	konkrétní určení místa, např. Praha povinné pokud lze uvést odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „a“ ----- type – bude vždy text	<dc:coverage> povinné pokud lze uvést
<publisher>		jméno entity, která dokument vydala, vytiskla nebo jinak vyprodukovala povinné pokud lze uvést odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21	<dc:publisher> povinné pokud lze uvést
<dateIssued>		datum vydání předlohy, nutno zaznamenat v případě titulu roky v nichž časopis vycházel (např. 1900-1939) povinné odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „c“	<dc:date> povinné
<dateCreated>		datum vytvoření předlohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo> odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“ doporučené	
<issuance>		údaje o vydávání povinné hodnota continuing odpovídá hodnotě uvedené návěští MARC21 na pozici 07	
<frequency>		údaje o pravidelnosti vydávání odpovídá údaji MARC21 v poli 310 nebo pozici 18 v poli 008 doporučené	
<language>		údaje o jazyce dokumentu povinné	

<languageTerm>	type authority	přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, <a href="http://www.loc.gov/standards/iso639-2/php/code_list.php">http://www.loc.gov/standards/iso639-2/php/code_list.php</a> povinné ----- type: použít hodnotu code  authority: použít hodnotu „iso639-2b“	<dc:language> povinné
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy povinné	
<form>	authority	údaje o fyzické podobě dokumentu, např. print, electronic apod. povinné- pro periodika hodnota print odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21 ----- authority: hodnota „marcform“	<dc:format> povinné
<extent>		údaje o rozsahu (stran, svazků nebo rozměrů); použití spíše u ročenek apod. doporučené pokud lze uvést odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat	
<note>		poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element doporučeno pokud lze vyplnit	
<abstract>		shrnutí obsahu periodika jako celku doporučené odpovídá poli 520 MARC21	<dc:description> doporučené
<note>		obecná poznámka k periodiku jako celku doporučeno pokud lze vyplnit odpovídá poli 500 v MARC21	<dc:description> doporučené pokud lze vyplnit
<subject>	authority	údaje o věcném třídění předpokládá se přebírání z katalogizačního záznamu doporučené ----- authority: vyplnit hodnotu „czenas“	

<topic>		libovolný výraz specifikující nebo charakterizující obsah vnitřní části; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) nebo obsah pole 650 záznamu MARC21 povinné	<dc:subject> doporučené
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) nebo obsah pole 651 záznamu MARC21 doporučené	<dc:subject> doporučené
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) nebo obsah pole 648 záznamu MARC21 doporučené	<dc:subject> doporučené
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) nebo obsah pole 600 záznamu MARC21 doporučené	<dc:subject> doporučené
<classification>	authority	klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění povinné odpovídá poli 080 MARC21  ----- authority: vyplnit hodnotu „udc“	<dc:subject> povinné
<relatedItem>	type	informace o dalších dokumentech/částech/zdrojích, které jsou ve vztahu k popisovanému dokumentu;  použití pro vyjádření edice, ve které je dokument vydán, údaj o edici musí obsahovat minimálně element <title> s jejím názvem doporučené pokud lze uvést  Poznámka: element <relatedItem> může obsahovat jakýkoliv jiný element MODS – jejich použití se řídí pravidly popsány pro tyto	

		<p>elementy;</p> <p>-----</p> <p>type: hodnota „series“</p>	
<identifier>	type	<p>údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které titul periodika má – viz přehled typů atributů níže</p> <p>povinné</p> <p>-----</p> <p>type: budou se povinně vyplňovat následující hodnoty, pokud existují:</p> <ul style="list-style-type: none"> <li>- doi</li> <li>- hdl - handle</li> <li>- issn - převzít z katalogizačního záznam NK ČR</li> <li>- isbn - převzít z katalogizačního záznam NK ČR</li> <li>- ccnb – čČNB - převzít z katalogizačního záznam NK ČR</li> <li>- permalink záznamu z katalogu NK ČR, např. <a href="http://aleph.nkp.cz/F/?func=direct&amp;doc_number=002186258&amp;local_base=NKC">http://aleph.nkp.cz/F/?func=direct&amp;doc_number=002186258&amp;local_base=NKC</a></li> <li>- urn - pro URN:NBN</li> <li>- uuid</li> <li>- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu</li> </ul>	<dc:identifier> povinné
<location>		<p>údaje o uložení popisovaného dokumentu, např. signatura, místo uložení apod.</p> <p>povinné</p>	
<physicalLocation>	authority	<p>údaje o instituci, kde je fyzicky uložen popisovaný dokument, např. NK ČR</p> <p>povinné</p> <p>nutno použít kontrolovaný slovník – sigly knihoven (ABA001 atd.)</p> <p>odpovídá poli 040 v MARC21</p> <p>-----</p> <p>authority: hodnota „siglaADR“</p>	<dc:source> povinné
<shelfLocator>		sigla nebo lokační údaje o dokumentu	<dc:source>



		povinné	povinné
<part>	type	popis částí dokumentu, bude využit jen na popis ročníku (volume) periodika povinné ----- type: hodnota bude vždy „volume“	
<detail>	type	upřesnění popisu části povinné ----- type: hodnota bude vždy „volume“	
<number>		číslo části (ročníku) povinné pokud lze uvést	<dc:description> povinné pokud lze uvést; nutno doplnit slovo „volume number“, viz <dc:description>volume number: 25 </dc:description>
<date>		datum vztahující se k části povinné v případě, že se ročník vycházel během více let (přelom roku), nutno uvést oba roky, např. 1920-1921	
<recordInfo>		údaje o metadatovém záznamu – jeho vzniku, změnách apod. povinné	
<recordContentSource>		kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník doporučené	
<recordCreationDate>	encoding	datum prvního vytvoření záznamu, na povinné ----- encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601	
<recordChangeDate>	encoding	datum změny záznamu doporučené ----- encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu	

		tedy iso8601	
<recordOrigin>		údaje o vzniku záznamu doporučené hodnoty: machine generated nebo human prepared	

### Pole MODS a DC pro číslo periodika

Element MODS	Atributy	Popis	Element Dublin Core
<titleInfo>	ID	název titulu periodika, kterého je číslo součástí povinné – použít názvové authority nebo katalogizační záznam ----- ID musí vyjadřovat název úrovně, tj. např. „MODS_ISSUE“	
<title>		názvová informace – titul periodika povinné – převzít z katalogu	<dc:title> povinné
<subTitle>		podnázev periodika doporučené pokud lze uvést	<dc:title> povinné pokud lze vyplnit
<partNumber>		pořadové číslo vydání (čísla), např. 40; nebo u ročenek číslo určité řady/edice (část 1, řada B) povinné pokud lze vyplnit	<dc:description> povinné pokud lze vyplnit
<partName>		jméno edice nebo speciální ediční řady, např. Hygiena; lze uvést i název tématického čísla nebo zvláštního vydání; k použití u ročenek a specializovaných periodik nebo u tematických čísel nebo zvláštních vydání doporučené	<dc:description> doporučené
<name>	type	údaje o odpovědnosti za číslo periodika povinné pokud lze uvést; nepočítá se s vyplněním u deníků, ale např. u ročenek, zvláštních vydání apod. ----- type: použít jeden z typů - personal - corporate	

		<ul style="list-style-type: none"> <li>- conference</li> <li>- family</li> </ul>	
<namePart>	type	<p>údaje o křestním jméně a příjmení apod. povinné kde lze uvést nutno vyjádřit pro křestní jméno i příjmení</p> <p>-----</p> <p>type: použít jednu z hodnot:</p> <ul style="list-style-type: none"> <li>- date – doporučené pokud lze uvést</li> <li>- family – povinné pokud lze uvést</li> <li>- given – povinné pokud lze uvést</li> <li>- termsOfAddress – doporučené pokud lze uvést</li> </ul> <p>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu &lt;namePart&gt;</p>	<dc:creator> povinné pokud lze uvést nutno do jednoho pole DC spojit jméno i příjmení
<role>		specifikace role osoby nebo organizace uvedené v elementu <name> povinné kde lze uvést	
<roleTerm>	type authority	<p>popis role nutno použít kontrol. slovník např. z MARC21 povinné kde lze uvést</p> <p>-----</p> <p>type: code – kód role z kontrolovaného slovníku rolí (<a href="http://www.loc.gov/marc/relators/relaterm.html">http://www.loc.gov/marc/relators/relaterm.html</a>)</p> <p>authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"</p>	
<genre>	type	<p>bližší údaje o typu dokumentu povinné hodnota: issue</p> <p>-----</p> <p>type: pro upřesnění typu čísla a jednotlivých vydání povinné</p>	<dc:type> povinné

		<p>hodnota může být:</p> <ul style="list-style-type: none"> <li>- normal - běžné vydání</li> <li>- morning – ranní vydání</li> <li>- afternoon- odpolední vydání</li> <li>- evening – večerní vydání</li> <li>- sequence_X – pořadí vydání (sequence_1 = první vydání toho dne; sequence_2 = druhé vydání atd.)</li> <li>- corrected – opravené vydání</li> <li>- special – zvláštní vydání (např. k nějaké události)</li> </ul>	
<originInfo>		<p>informace o původu předlohy doporučené kde lze vyplnit (např. u ročenek, kde se vydavatel měnil) nepovinné pro deníky</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 &lt;originInfo&gt; opakovat s atributem transliteration=“printer“ a elementy &lt;place&gt;, &lt;publisher&gt;, &lt;dateCreated&gt;, které budou obsahovat údaje o tiskaři.</p>	
<place>		<p>údaje o místě spojeném s vydáním, výrobou nebo původem popisovaného dokumentu povinné pokud lze uvést</p>	<dc:coverage> povinné pokud lze uvést
<placeTerm>	type	<p>konkrétní určení místa, např. Praha povinné pokud lze uvést odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „a“</p> <p>----- type – bude vždy text</p>	<dc:coverage> povinné pokud lze uvést
<publisher>		<p>jméno entity, která dokument vydala, vytiskla nebo jinak vyprodukovala povinné pokud lze uvést odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21</p>	<dc:publisher> povinné pokud lze uvést

<dateIssued>		datum vydání předlohy, v případě čísla datum dne, kdy vyšlo; musí vyjádřit den, měsíc a rok povinné nutno zapsat v podobě DD.MM.RRRR	<dc:date> povinné
<dateCreated>		datum vytvoření předlohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo> odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“ doporučené	
<language>		údaje o jazyce dokumentu povinné	
<languageTerm>	type authority	přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, <a href="http://www.loc.gov/standards/iso639-2/php/code_list.php">http://www.loc.gov/standards/iso639-2/php/code_list.php</a> povinné ----- type: použít hodnotu code  authority: použít hodnotu „iso639-2b“	<dc:language> povinné
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy povinné	
<extent>		údaje o rozsahu (stran, svazků nebo rozměrů); použití spíše u ročenek apod. doporučené pokud lze uvést odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat	
<note>		poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element doporučeno pokud lze vyplnit	
<abstract>		shrnutí obsahu dokumentu, zvláště pro ročenky, zvláštní vydání a tematická čísla doporučené pokud lze vyplnit, pouze v případech, že se liší od abstraktu na úrovni titulu odpovídá poli 520 MARC21	<dc:description> doporučené

<note>		obecná poznámka k dokumentu doporučeno pokud lze vyplnit odpovídá poli 500 v MARC21	
<subject>	authority	údaje o věcném třídění doporučené pro tématická čísla, zvláštní vydání a ročenky – pouze pokud se liší od údajů v elementu <subject> na úrovni titulu ----- authority: vyplnit hodnotu „czenas“	
<topic>		libovolný výraz specifikující nebo charakterizující obsah čísla; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) povinné	<dc:subject> doporučené
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) doporučené	<dc:subject> doporučené
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) doporučené	<dc:subject> doporučené
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) doporučené	<dc:subject> doporučené
<identifier>	type	údaje o identifikátorech čísla, obsahuje unikátní identifikátory mezinárodní nebo lokální povinné ----- type: budou se povinně vyplňovat následující hodnoty, pokud existují: - doi - hdl - handle - isbn - převzít z katalogizačního záznam NK ČR (ročenky apod.) - urn - pro URN:NBN - uuid - jiný interní identifikátor, hodnota atributu „local“, lze použít např.	<dc:identifier> povinné

		k vyjádření čárového kódu	
<location>		údaje o uložení popisovaného dokumentu, např. signatura, místo uložení apod. doporučené - pro ročenky apod., kde se signatury jednotlivých čísel liší	
<physicalLocation>	authority	údaje o instituci, kde je fyzicky uložen popisovaný dokument, např. NK ČR povinné nutno použít kontrolovaný slovník – sigly knihoven (ABA001 atd.) odpovídá poli 040 v MARC21 ----- authority: hodnota „siglaADR“	<dc:source> doporučeno pokud lze vyplnit
<shelfLocator>		sigla nebo lokační údaje o dokumentu povinné pokud lze uvést	<dc:source> doporučeno pokud lze vyplnit

#### Pole MODS a DC pro vnitřní část periodika (článek a obraz)

Element MODS	Atributy	Popis	Dublin Core Element
<titleInfo>	ID	názvová informace vnitřní části povinné ----- ID musí vyjadřovat název úrovně, tj. např. „MODS_PICTURE“ pro obrázek v textu, „MODS_ARTICLE“ pro článek apod.	
<title>		vlastní název vnitřní části (článku) u obrazu brát případně z popisku obrazu povinné, pokud není vyplnit hodnotu „untitled“	<dc:title> povinné
<subTitle>		podnázev vnitřní části (článku); za podnázev lze považovat i krátký text, který se před článkem objevuje tučným písmem (shrnutí obsahu článku) povinné pokud lze vyplnit	<dc:title> povinné
<name>	type	údaje o odpovědnosti za vnitřní část povinné pokud lze uvést; ----- type: použít jeden z typů - personal - corporate	

		<ul style="list-style-type: none"> <li>- conference</li> <li>- family</li> </ul>	
<namePart>	type	<p>údaje o křestním jméně a příjmení apod. povinné pokud lze uvést nutno vyjádřit pro křestní jméno i příjmení</p> <p>-----</p> <p>type: použít jednu z hodnot:</p> <ul style="list-style-type: none"> <li>- date – doporučené pokud lze uvést</li> <li>- family – povinné pokud lze uvést</li> <li>- given – povinné pokud lze uvést</li> <li>- termsOfAddress – doporučené pokud lze uvést</li> </ul> <p>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu &lt;namePart&gt;</p>	<dc:creator> povinné pokud lze uvést nutno do jednoho pole DC spojit jméno i příjmení
<role>		specifikace role osoby nebo organizace uvedené v elementu <name> povinné kde lze uvést	
<roleTerm>	type authority	<p>popis role nutno použít kontrol. slovník např. z MARC21 povinné kde lze uvést</p> <p>-----</p> <p>type: code – kód role z kontrolovaného slovníku rolí (<a href="http://www.loc.gov/marc/relators/relaterm.html">http://www.loc.gov/marc/relators/relaterm.html</a>)</p> <p>authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"</p>	
<genre>	type	<p>bližší údaje o typu vnitřní části povinné hodnota: article nebo picture</p> <p>-----</p> <p>type: doporučené</p>	<dc:type> povinné



		<p>hodnota pro article – možnost vyplnit bližší určení typu článku (možnost použít DTD periodika, Article Types)</p> <ul style="list-style-type: none"> <li>- news</li> <li>- advertisement</li> <li>- abstract</li> <li>- introduction</li> <li>- review</li> <li>- dedication</li> <li>- remark</li> <li>- bibliography</li> <li>- editorsNote</li> <li>- preface</li> <li>- research article</li> <li>- main article</li> <li>- aj.</li> </ul> <p>hodnota pro picture – možnost vyplnit další určení typu obrazu</p> <ul style="list-style-type: none"> <li>- table</li> <li>- illustration</li> <li>- chart</li> <li>- photograph</li> <li>- graphic</li> <li>- map</li> <li>- advertisement</li> <li>- aj.</li> </ul>	
<language>		údaje o jazyce vnitřní části povinné pokud lze uvést (nelze u obrazu)	
<languageTerm>	type authority	<p>přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, <a href="http://www.loc.gov/standards/iso639-2/php/code_list.php">http://www.loc.gov/standards/iso639-2/php/code_list.php</a> povinné</p> <p>-----</p> <p>type: použít hodnotu code</p> <p>authority: použít hodnotu „iso639-2b“</p>	<dc:language> povinné
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy doporučené	

<form>	authority	údaje o fyzické podobě vnitřní části, např. print, electronic apod. doporučené odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21 ----- authority: hodnota „marcform“	<dc:format> doporučené
<abstract>		shrnutí obsahu vnitřní části doporučené	<dc:description> doporučené
<note>		obecná poznámka k vnitřní části doporučeno pokud lze vyplnit odpovídá poli 500 v MARC21	<dc:description> doporučené
<subject>	authority	údaje o věcném třídění doporučené ----- authority: vyplnit hodnotu „czenas“	
<topic>		libovolný výraz specifikující nebo charakterizující obsah vnitřní části; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) povinné	<dc:subject> doporučené
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) doporučené	<dc:subject> doporučené
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) doporučené	<dc:subject> doporučené
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) doporučené	<dc:subject> doporučené
<classification>	authority	klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění doporučené odpovídá poli 080 MARC21 ----- authority: vyplnit hodnotu „udc“	<dc:subject> doporučené
<identifier>	type	údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální,	<dc:identifier> povinné

		<p>které vnitřní část má – viz přehled typů atributů níže</p> <p>povinné</p> <p>-----</p> <p>type: budou se povinně vyplňovat následující hodnoty, pokud existují:</p> <ul style="list-style-type: none"> <li>- doi</li> <li>- hdl - handle</li> <li>- urn - pro URN:NBN</li> <li>- uuid</li> <li>- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu</li> </ul>	
<part>		<p>popis částí vnitřní části, bude využito na záznam rozsahu</p> <p>povinné pokud lze uvést (nelze u obrazu)</p>	
<extent>		<p>upřesnění popisu části – rozsah na stránkách</p> <p>povinné</p>	
<start>		<p>první stránka, na které vnitřní část začíná</p> <p>povinné pokud lze uvést</p>	<dc:coverage> povinné
<end>		<p>poslední stránka, na které vnitřní část končí</p> <p>povinné pokud lze uvést</p>	<dc:coverage> povinné
<recordInfo>		<p>údaje o metadatovém záznamu vnitřní části – jeho vzniku, změnách apod.</p> <p>povinné</p>	
<recordContentSource>		<p>kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník</p> <p>doporučené</p>	
<recordCreationDate>	encoding	<p>datum prvního vytvoření záznamu vnitřní části</p> <p>povinné</p> <p>-----</p> <p>encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601</p>	
<recordChangeDate>	encoding	<p>datum změny záznamu vnitřní části</p> <p>doporučené</p> <p>-----</p> <p>encoding: záznam bude podle normy ISO</p>	

		8601 na úroveň minut, hodnota atributu tedy iso8601	
<recordOrigin>		údaje o vzniku záznamu vnitřní části doporučené hodnoty: machine generated nebo human prepared	

### Pole MODS a DC pro přílohu

Element MODS	Atributy	Popis	Dublin Core Element
<titleInfo>	ID	názvová informace přílohy povinné – použít názvové autority nebo katalogizační záznam ----- ID musí vyjadřovat název úrovně, tj. „MODS_SUPPLEMENT“	
<title>		názvová informace – název periodika, jehož součástí příloha je povinné – převzít z katalogu	<dc:title> povinné
<partNumber>		číslo přílohy, pokud nějaké má doporučené pokud lze vyplnit	<dc:description> povinné
<partName>		název přílohy povinné	<dc:title> povinné
<name>	type	údaje o odpovědnosti za přílohu povinné pokud lze uvést; ----- type: použít jeden z typů - personal - corporate - conference - family	
<namePart>	type	údaje o křestním jméně a příjmení apod. povinné kde lze uvést 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 uvést - termsOfAddress – doporučené pokud lze uvést	<dc:creator> povinné pokud lze uvést nutno do jednoho pole DC spojit jméno i příjmení

		<p>pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu &lt;namePart&gt;</p>	
<role>		<p>specifikace role osoby nebo organizace uvedené v elementu &lt;name&gt; povinné kde lze uvést</p>	
<roleTerm>	<p>type authority</p>	<p>popis role nutno použít kontrol. slovník např. z MARC21 povinné kde lze uvést</p> <p>-----</p> <p>type: code – kód role z kontrolovaného slovníku rolí <a href="http://www.loc.gov/marc/relators/relaterm.html">http://www.loc.gov/marc/relators/relaterm.html</a>)</p> <p>authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"</p>	
<typeOfResource>		<p>popis charakteristiky typu nebo obsahu přílohy doporučené jedna z hodnot:</p> <ul style="list-style-type: none"> <li>- text – např. pro přílohu typu časopis, kniha, brožura apod.</li> <li>- cartographic – pro mapy</li> <li>- notated music</li> <li>- sound recording-musical - pro hudební CD/DVD</li> <li>- sound recording-nonmusical</li> <li>- sound recording</li> <li>- still image – fotografie, plakáty apod.</li> <li>- moving image – pro filmová DVD</li> <li>- three dimensional object</li> <li>- software, multimedia – pro CD/DVD se SW</li> <li>- mixed material</li> </ul>	<p>&lt;dc:type&gt; doporučené</p>
<genre>		<p>blíže údaje o typu dokumentu</p>	<p>&lt;dc:type&gt;</p>

		<p>povinné hodnota:</p> <ul style="list-style-type: none"> <li>- volume_supplement (příloha k ročníku, např. obsah celého ročníku)</li> <li>- issue_supplement (příloha k číslu)</li> </ul>	povinné
<originInfo>		<p>informace o původu přílohy povinné - pokud lze vyplnit a <i>pokud se liší od údajů v popisu čísla (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 &lt;originInfo&gt; opakovat s atributem transliteration="printer" a elementy &lt;place&gt;, &lt;publisher&gt;, &lt;dateCreated&gt;, které budou obsahovat údaje o tiskaři.</p>	
<place>		<p>údaje o místě spojeném s vydáním, výrobou nebo původem přílohy povinné pokud lze uvést</p>	<dc:coverage> povinné pokud lze uvést
<placeTerm>	type	<p>konkrétní určení místa, např. Praha povinné pokud lze uvést odpovídá hodnotě katalogizačního záznamu, pole 260, podpole „a“</p> <p>----- type – bude vždy text</p>	<dc:coverage> povinné pokud lze uvést
<publisher>		<p>jméno entity, která přílohu vydala, vytiskla nebo jinak vyprodukovala povinné pokud lze uvést odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21</p>	<dc:publisher> povinné pokud lze uvést
<dateIssued>		<p>datum vydání přílohy, musí vyjádřit den, měsíc a rok povinné nutno zapsat v podobě DD.MM.RRRR možno použít hodnotu z katalogizačního záznamu, pole 260, podpole „c“</p>	<dc:date> povinné

<dateCreated>		datum vytvoření přílohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo> nebo např. u popisu CD/DVD apod. doporučené odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“	
<frequency>		údaje o pravidelnosti vydávání doporučené pokud lze vyplnit odpovídá údaji MARC21 v poli 310 nebo pozici 18 v poli 008	
<language>		údaje o jazyce dokumentu povinné	
<languageTerm>	type authority	přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, <a href="http://www.loc.gov/standards/iso639-2/php/code_list.php">http://www.loc.gov/standards/iso639-2/php/code_list.php</a> povinné ----- type: použít hodnotu code  authority: použít hodnotu „iso639-2b“	<dc:language> povinné
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy povinné	
<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“	<dc:format> povinné
<extent>		údaje o rozsahu (stran, svazků nebo rozměrů) doporučeno pokud lze uvést odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat	
<note>		poznámka o fyzickém stavu dokumentu;	

		pro každou poznámku je nutno vytvořit nový <note> element doporučeno pokud lze vyplnit	
<abstract>		shrnutí obsahu dokumentu doporučené pokud lze vyplnit odpovídá poli 520 MARC21	<dc:description> doporučené
<note>		obecná poznámka k dokumentu doporučeno pokud lze vyplnit odpovídá poli 500 v MARC21	<dc:description> doporučené pokud lze vyplnit
<subject>	authority	údaje o věcném třídění doporučené ----- authority: vyplnit hodnotu „czenas“	
<topic>		libovolný výraz specifikující nebo charakterizující obsah přílohy; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) povinné	<dc:subject> doporučené
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) doporučené	<dc:subject> doporučené
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) doporučené	<dc:subject> doporučené
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) doporučené	<dc:subject> doporučené
<classification>	authority	klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění povinné odpovídá poli 080 MARC21 ----- authority: vyplnit hodnotu „udc“	<dc:subject> povinné
<identifier>	type	údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které příloha má – viz přehled typů atributů níže povinné pokud lze uvést	<dc:identifier> povinné



		<p>-----</p> <p>type: budou se povinně vyplňovat následující hodnoty, pokud existují:</p> <ul style="list-style-type: none"> <li>- doi</li> <li>- hdl - handle</li> <li>- issn - převzít z katalogizačního záznam NK ČR</li> <li>- isbn - převzít z katalogizačního záznam NK ČR</li> <li>- ccnb – čČNB - převzít z katalogizačního záznam NK ČR</li> <li>- permalink záznamu z katalogu NK ČR, např. <a href="http://aleph.nkp.cz/F/?func=direct&amp;doc_number=002186258&amp;local_base=NKC">http://aleph.nkp.cz/F/?func=direct&amp;doc_number=002186258&amp;local_base=NKC</a></li> <li>- urn - pro URN:NBN</li> <li>- uuid</li> <li>- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu</li> </ul>	
--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

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

- technická a administrativní metadata pro různé reprezentace jedné strany čísla periodika (původní TIFF, MC, ALTOXML a OCR.TXT) budou zabalena v části <amdSec> vedlejšího METS záznamu (AMD\_METS.xml) ve vlastních formátech (MIX, PREMIS – části object; events; agent)
- **pro všechny reprezentace jedné strany čísla periodika bude v AMD\_METS.xml existovat jedna část <amdSec>, která bude obsahovat metadata v <techMD> a <digiprovMD> podčástech pro jednotlivé soubory**
- část <amdSec> musí mít ID a vnořený element <techMD> nebo <digiprovMD>, oba s ID a vnořeným elementem <mdWrap> s atributem MDTYPE

element	atributy	popis
<amdSec>	ID	<p>element obsahující technická metadata ve formátu PREMIS nebo MIX</p> <p>povinné</p> <p>-----</p> <p>ID – identifikátor konkrétní části &lt;amdSec&gt;, např. pro stránku 1 by hodnota mohla být</p>

		„PAGE0001“
<techMD> nebo <digiprovMD>	ID	<p>element rozlišující typy jednotlivých administrativních metadat</p> <p>povinné</p> <p>-----</p> <p>ID</p> <p>pro část &lt;techMD&gt;:</p> <ul style="list-style-type: none"> <li>- 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</li> <li>- pro části obsahující MIX hodnota „MIX_001“ = MIX metadata pro původní TIFF, „MIX_002“ pro MC</li> </ul> <p>pro část &lt;digiprovMD&gt;:</p> <ul style="list-style-type: none"> <li>- pro části obsahující PREMIS-event hodnota „EVT_001“ apod.</li> <li>- pro části obsahující PREMIS-agent hodnota „AGENT_001“ apod.</li> </ul>
<mdWrap>	MDTYPE	<p>element obsahující vložené záznamy PREMIS, MIX</p> <p>povinné</p> <p>-----</p> <p>MDTYPE</p> <ul style="list-style-type: none"> <li>- pro záznamy PREMIS object, event i agent vždy hodnota „PREMIS“</li> <li>- pro záznamy MIX hodnota „NISOIMG“</li> </ul>

- pro všechna digitalizovaná data se bude využívat formát PREMIS (jeho části object, event a agent), pro obrazová data dále i formát MIX
- technická a administrativní metadata budou vznikat i pro prvotní sken (většinou TIFF), který se po nutných úpravách maže a dále neuchovává – viz specifikace
- technická metadata jsou určena primárně pro zachycení technických informací o formátech souborů, o výsledcích validací a kontrol
- administrativní metadata zachycují veškeré změny, procesy apod., které byly na datech i metadatach provedeny
- všechny PREMIS a MIX záznamy budou obsaženy v tzv. vedlejším METS záznamu (AMD\_METS.xml), který je určen pro administrativní a technická metadata
  - 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.)

### 8.3.1 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
- 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 eventy, které jsou popsány v PREMIS Events ve stejném METS metadatovém záznamu konkrétního dokumentu (číslo, svazek) 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
- povinnost plnění dle NK ČR (slovní vyjádření: povinné, doporučené, nepovinné)

##### Obsah pole „Použití pro“

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

Element	Popis	Použití pro
<objectIdentifier>	identifikátor k jednoznačnému odlišení objektu v určitém kontextu; 1-n povinné	MC, XML, PS
<objectIdentifierType>	popis kontextu, ve kterém je identifikátor unikátní, např. NDK, ANL nebo název repozitáře; nutno použít kontrolovaný slovník; 1-1 povinné	MC, XML, PS
<objectIdentifierValue>	vlastní hodnota identifikátoru, např. img0001-master, urn.nbn.cz-123465 apod.; 1-1 povinné	MC, XML, PS
<objectCategory>	typ objektu, ke kterým se metadata (PREMIS object) vztahuje, např. file pro soubor, representation pro dig. reprezentaci, bitstream pro bitstream; 1-1 povinné	MC, XML, PS
<preservationLevel>	údaje o úrovni ochrany souboru, která se na něj vztahuje; některé soubory nejsou tak důležité jako jiné, mají menší úroveň ochrany; 0-n povinné	MC, XML, PS
<preservationLevelValue>	hodnota úrovně ochrany, která je pro soubor relevantní, pro původní sken PS hodnota deleted, pro MC a XML hodnota preservation; 1-1 povinné	MC, XML, PS
<preservationLevelDateAssigned>	datum, kdy byla přiřazena hodnota úrovně ochrany, zápis v ISO 8601, na úroveň dne (DD-MM-RRRR) 0-1 doporučené	MC, XML, PS
<objectCharacteristics>	technické údaje o souboru 1-n povinné	MC, XML, PS

<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 povinné	MC, XML, PS
<fixity>	údaje o kontrolním součtu 0-n povinné	MC, XML, PS
<messageDigestAlgorithm>	použitý algoritmus kontrolního součtu, např. MD5 aj. 1-1 povinné	MC, XML, PS
<messageDigest>	hodnota kontrolního součtu 1-1 povinné	MC, XML, PS
<messageDigestOriginator>	agent (osoba, instituce, stroj, SW), který kontrolní součet vytvořil (např. JHOVE apod.) 0-1 povinné	MC, XML, PS
<size>	údaje o velikosti souboru v bytech 0-1 povinné	MC, XML, PS
<format>	údaje o formátu souboru 1-n povinné	MC, XML, PS
<formatDesignation>	identifikace formátu souboru, výstup z JHOVE, PRONOM služeb apod. 0-1 povinné	MC, XML, PS
<formatName>	jméno formátu, např. image/tiff nebo Adobe PDF 1-1 povinné	MC, XML, PS
<formatVersion>	verze formátu, např. 6.0 0-1 povinné	MC, XML, PS
<formatRegistry>	identifikace formátu – dodatečná informace o záznamu formátů v registrech	MC, XML, PS

	formátů (např. PRONOM aj.) 0-1 povinné	
<formatRegistryName>	jméno použitého registru formátů, např. UDFR, PRONOM aj. 1-1 povinné	MC, XML, PS
<formatRegistryKey>	unikátní identifikátor (označení) formátu v registru, např. fmt/155 z PRONOM 1-1 povinné	MC, XML, PS
<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 povinné	MC, XML, PS
<creatingApplicationName>	název aplikace, např. ImageGear, Kakadu apod.; 0-1 povinné	MC, XML, PS
<creatingApplicationVersion>	verze aplikace, např. 15.03.000 0-1 povinné	MC, XML, PS
<dateCreatedByApplication>	datum a čas vytvoření, např. 2008-11-10T12:37:46; musí být ve tvaru ISO 8601 (na úroveň vteřin); 0-1 povinné	MC, XML, PS
<originalName>	původní jméno souboru , např. digibok_2007081301091_0011.jp2 0-1 povinné	MC, XML, PS
<relationship>	vyjádření vztahu popisovaného souboru k jiným souborům a událostem (eventům) 0-n povinné	MC, XML
<relationshipType>	typ vztahu, doporučené hodnoty: derivation= vztah kde objekt je výsledkem změny jiného objektu; structural= vztah mezi částmi objektu;	MC, XML;

	tj. např. ALTO vytvořené z TIFFU bude mít vztah derivation, podobně jako JPEG2000 z TIFFu vytvořený; 1-1 povinné	
<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 povinné	MC, XML;
<relatedObjectIdentification>	identifikace souvisejícího souboru 1-n povinné pro MC, XML pro vyjádření vztahu k původnímu objektu (skenu)	MC, XML
<relatedObjectIdentifierType>	specifikace kontextu, ve kterém je identifikátor souboru jedinečný, např. URN; temporary filepath; objectID 1-1 povinné	MC, XML
<relatedObjectIdentifierValue>	vlastní řetězec identifikátoru, např. URN:NBN:cz-1301091_011#0001 nebo název souboru, cesta k souboru apod. 1-1 povinné	MC, XML
<relatedEventIdentification>	identifikace s popisovaným souborem související události (eventu); seznam událostí viz PREMIS event 0-n povinné	MC, XML
<relatedEventIdentifierType>	typ události, např. interní číslovací systém událostí jako no.nb.evt; NK repository event ID, UUID apod. 1-1 povinné	MC, XML
<relatedEventIdentifierValue>	hodnota identifikátoru události, např. NK_EVT_005 nebo hodnota UUID aj. 1-1	MC, XML

	povinné	
<relatedEventSequence>	pořadí události, např. 003; k určení pořadí lze určit datum události 0-1 doporučené	MC, XML
<linkingEventIdentifier>	identifikátor události týkající původního skenu PS; typy událostí mohou být např. vytvoření, smazání 0-n povinné – pro PS nutný link na události vytvoření (digitalizace) a jeho vymazání	PS
<linkingEventIdentifierType>	typ identifikátoru události, např. UUID, NK_eventID, vlastní číslovací systém apod. 1-1 povinné	PS
<linkingEventIdentifierValue>	hodnota identifikátoru, např. event_01; img0001-master-event001 apod. 1-1 povinné	PS

### 8.3.2 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
  - o vygenerování UC
  - o 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>
  - o 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)
  - 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
- povinnost plnění dle NK ČR (slovní vyjádření: povinné, doporučené, nepovinné)

**Pole záznamu PREMIS Event**

Element	Popis
<eventIdentifier>	údaje o identifikátoru události v kontextu digitalizace nebo repozitáře 1-1 povinné
<eventIdentifierType>	typ identifikátoru, např. no.nb.evt; NK_eventID, UUID apod. 1-1 povinné
<eventIdentifierValue>	hodnota identifikátoru, např. EVT_001; event_019 apod. 1-1 povinné
<eventType>	kategorizace události, nutno použít kontrolovaný slovník; typy událostí, které musí být zaznamenány: capture, migration, derivation, deletion 1-1 povinné
<eventDateTime>	datum a čas kdy byla událost provedena; nutno zapsat v ISO 8601 na úroveň vteřin 1-1 povinné
<eventDetail>	další údaje o události, doporučené hodnoty pro výše uvedené <eventType> následují za /: <ul style="list-style-type: none"><li>- capture/digitization – vznik prvního skenu</li><li>- capture/XML_creation</li><li>- capture/TXT_creation</li><li>- migration/MC_creation</li><li>- derivation/UC_creation</li></ul>

	- deletion/PS_deletion 0-1 povinné
<eventOutcomeInformation>	informace o výsledku události 0-n doporučené
<eventOutcome>	kategorizace výsledku události, např. slovy jako successful nebo failure, možno použít kódy – nutno používat kontrolovaný slovník nebo seznam kódů 0-1 povinné
<linkingAgentIdentifier>	identifikace jednoho nebo více agentů spojených s událostí 0-n povinné
<linkingAgentIdentifierType>	označení typu identifikátoru, např. NK_AgentID, UUID apod. 1-1 povinné
<linkingAgentIdentifierValue>	hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod. 1-1 povinné
<linkingAgentRole>	role agenta ve vztahu k události, např. software; SW component; operator; nutno používat kontrolovaný slovník 0-n doporučené
<linkingObjectIdentifier>	informace o objektu/souboru spojeného s událostí, link na něj 0-n povinné
<linkingObjectIdentifierType>	označení typu identifikátoru, např. PhysUnitID; URN, NK_OBJ, OBJ_001 apod.; hodnoty by se měly brát z kontrolovaného slovníku 1-1 povinné
<linkingObjectIdentifierValue>	hodnota identifikátoru, např. URN:NBN:cz-_0011#0001 aj. 1-1 povinné

### 8.3.3 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)
- 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

#### Navrhovaná 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
- povinnost plnění dle NK ČR (slovní vyjádření: povinné, doporučené, nepovinné)

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

<agentIdentifierType>	označení typu identifikátoru, např. NK_AgentID, UUID apod. 1-1 povinné
<agentIdentifierValue>	hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod. 1-1 povinné
<agentName>	textové upřesnění agenta, např. přesný název SW, plné jméno osoby apod. - FixImage1.3; Jan Novák; CCS docWorks 6.2.1; 0-n doporučené
<agentType>	obecné označení agenta – pro osoby např. osoba, pro SW např. software apod. hodnoty: organization; person; software 0-1 povinné
<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 povinné pokud lze vyplnit

#### 8.3.4 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 bude vznikat 1) pro archivní kopii, 2) pro původní soubor vzniklý prvotním skenováním (nejčastěji TIFF)** a to i přesto, že tento TIFF se v průběhu výroby maže a není archivován
- 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 čísla periodika
- **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>
- 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

### Navrhovaná 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
- povinnost plnění dle NK ČR (slovní vyjádření: povinné, doporučené, nepovinné)

#### Obsah pole „Použití pro“

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

Element	Popis	Použití pro
<BasicDigitalObjectInformation>		
<ObjectIdentifier>	údaje o identifikátoru obrazového dokumentu, který je formátem MIX popsán; 0-n doporučené	MC, PS
<objectIdentifierType>	např. jméno souboru, nebo jiný identifikátor; 0-1 povinné	MC, PS
<objectIdentifierValue>	hodnota identifikátoru, např. 20110306_001.jp2 nebo urn:nbn:123456; 0-1 povinné	MC, PS
<fileSize>	velikost souboru 0-1 doporučené	MC + PS
<FormatDesignation>	údaje o formátu obrazového souboru 0-1 povinné	MC, PS

<formatName>	název formátu, např. lze využít MIME types <sup>2</sup> (Image/jp2 apod.) 0-1 povinné	MC, PS
<formatVersion>	verze formátu, např. 1.0 0-1 povinné	MC, PS
<byteOrder>	endianita, možnosti jsou little endian, middle (mix) endian a big endian 0-1 povinné	MC + PS
<Compression>	<b>údaje o kompresi obrazového souboru</b> <b>(pokud</b> <b>0-n</b> <b>povinné</b>	<b>MC, PS</b>
<compressionScheme>	informace o kompresním schématu, vyjádřeno číslem (např. 34712 je komprese JPEG2000) nebo slovy (např. JP2 Lossless) 0-1 povinné	MC, PS
<BasicImageInformation>	<b>základní technické údaje o obrazovém</b> <b>dokumentu</b> <b>0-1</b> <b>povinný</b>	<b>MC, PS</b>
<BasicImageCharacteristics>	<b>0-1</b>	<b>MC, PS</b>
<imageWidth>	šířka obrazu v pixelech, např. 3987 0-1 povinné	MC, PS
<imageHeight>	výška obrazu v pixelech, např. 2345 0-1 povinné	MC, PS
<PhotometricInterpretation>	<b>photometrická interpretace</b> <b>0-1</b> <b>povinné</b>	<b>MC, PS</b>
<colorSpace>	barevný prostor, např. RGB 0-1 povinné	MC, PS
<ColorProfile>	<b>údaje o barevném profilu</b> <b>0-1</b>	<b>MC + PS</b>

<sup>2</sup> <http://www.iana.org/assignments/media-types/index.html>

	<b>povinné pro dokumenty, kde je nutno uchovat přesnou reprezentaci barvy původního dokumentu a používá se ICC profil)</b>	
<b>&lt;IccProfile&gt;</b>	<b>ICC profil 0-1 povinné</b>	<b>MC + PS</b>
<iccProfileName>	jméno profilu, např. sRGB, Adobe RGB aj. 0-1 povinné	MC + PS
<iccProfileVersion>	verze profilu, např. sRGB IEC61966-2.1 0-1 povinné	MC + PS
<iccProfileURI>	odkaz na profil, např. <a href="http://www.profil.cz/sRGB_v4_ICC_pref.icc">www.profil.cz/sRGB_v4_ICC_pref.icc</a> ; 0-1 doporučené	MC + PS
<b>&lt;SpecialFormatCharacteristics&gt;</b>	<b>speciální technické údaje o obrazovém dokumentu, použití pro formát JPEG2000 0-1 povinný pro JPEG2000</b>	<b>MC</b>
<b>&lt;JPEG2000&gt;</b>	<b>0-1 povinné</b>	<b>MC</b>
<b>&lt;CodecCompliance&gt;</b>	<b>údaje o kodeku 0-1 povinné</b>	<b>MC</b>
<codec>	název kodeku, např. Kakadu, LuraWave aj. 0-1 povinné	MC
<codecVersion>	verze kodeku, např. 3.1 0-1 povinné	MC
< codestreamProfile >	popis codestream profilu JPEG2000, např. P0 a P1 (viz ISO/IEC 15444-4); 0-1 povinné	MC
< complianceClass >	specifikace největší výšky, šířky a počtu komponentů, které dekodér dokáže dekodovat, lze použít hodnoty C0, C1 a C2; 0-1 povinné	MC

<b>&lt;EncodingOptions &gt;</b>	<b>obsahuje informace o kodování JPEG2000</b> <b>0-1</b> <b>povinné</b>	MC
<b>&lt;Tiles &gt;</b>	<b>popis pixelové velikosti dlaždic formátu JPEG2000</b> <b>0-1</b> <b>povinné</b>	MC
< tileWidth>	šířka dlaždice, např. 128 0-1 povinné	MC
< tileHeight>	výška dlaždice, např. 128 0-1 povinné	MC
< qualityLayers>	číselná hodnota počtu vrstev, do kterých byl JPEG2000 rozdělen, např. 12 0-1 povinné	MC
< resolutionLevels>	popis počtu nižších rozlišení, které lze z obrazu získat, např. 6 0-1 povinné	MC
<b>&lt; ImageCaptureMetadata&gt;</b>	<b>popis procesu skenování, je důležité vyplnit, protože tyto údaje nelze zjistit z finálního master/archivního souboru</b> <b>0-1</b> <b>povinné</b>	PS
<b>&lt;SourceInformation&gt;</b>	<b>informace o předloze</b> <b>0-1</b> <b>doporučené</b>	PS
<sourceType>	Book, Newspaper aj.; nutno používat kontrolovaný slovník 0-1 povinné	PS
<b>&lt;SourceID&gt;</b>	<b>identifikátor předlohy</b> <b>0-n</b> <b>doporučené</b>	PS
<sourceIDType>	typ identifikátoru, např. ČČNB, URN:NBN 0-1 povinné	PS
<sourceIDValue>	vlastní hodnota identifikátoru 0-1	PS



	povinné	
<b>&lt;GeneralCaptureInformation&gt;</b>	<b>základní údaje o skenování</b> <b>0-1</b> <b>povinné</b>	<b>PS</b>
<dateTimeCreated>	údaj o datu a čase skenování, např. 2009-01-03T08:25:28; zapsat v ISO 8601 na úroveň vteřin 0-1 povinné	PS
<imageProducer>	entita provádějící skenování, např. The National Library of the Czech Republic, osoba apod. 0-1 povinné	PS
<captureDevice>	typ skenovacího zařízení, např. reflection print scanner; doporučené využívání hodnot z kontrolovaného slovníku 0-1 povinné	PS
<b>&lt;ScannerCapture&gt;</b>	<b>údaje o skeneru</b> <b>0-1</b> <b>povinné</b>	<b>PS</b>
<scannerManufacturer>	výrobce skeneru, např. 4DigitalBooks, Treventus, Zeuschel 0-1 povinné	PS
<b>&lt;ScannerModel&gt;</b>	<b>údaje o konkrétním typu skeneru</b> <b>0-1</b> <b>povinné</b>	<b>PS</b>
<scannerModelName>	jméno modelové řady skeneru, např. DL 0-1 povinné	PS
<scannerModelNumber>	číslo/označení modelu, např. 3000 0-1 povinné	PS
<scannerModelSerialNo>	výrobní číslo skeneru, např. E4R0003649 0-1 povinné	PS
<b>&lt;MaximumOpticalResolution&gt;</b>	<b>údaje o maximálním optickém rozlišení skeneru</b> <b>0-1</b>	<b>PS</b>

	<b>povinné</b>	
< xOpticalResolution>	optické rozlišení na ose x, např. 300 0-1 povinné	PS
< yOpticalResolution>	optické rozlišení na ose y, např. 300 0-1 povinné	PS
< opticalResolutionUnit>	jednotka optického rozlišení, např. inch (in.) 0-1 povinné	PS
<scannerSensor>	popis typu snímacího senzoru skenovacího zařízení, např. matrix, linear, undefined aj. 0-1 povinné	PS
<b>&lt;ScanningSystemSoftware&gt;</b>	<b>údaje o softwaru skenovacího zařízení</b> <b>0-1</b> <b>povinné</b>	PS
<scanningSoftwareName>	název softwaru, např. Copinet 0-1 povinné	PS
<scanningSoftwareVersionNo>	číslo verze softwaru, např. 3.7 0-1 povinné	PS
<b>&lt;DigitalCameraCapture&gt;</b>	<b>údaje o snímacím zařízení (fotoaparát)</b> <b>0-1</b> <b>povinné, pokud je používán fotoaparát a není používán skener</b>	PS
<digitalCameraManufacturer>	výrobce fotoaparátu, např. Canon 0-1 povinné	PS
<b>&lt;DigitalCameraModel&gt;</b>	<b>popis modelu fotoaparátu</b> <b>0-1</b> <b>povinné</b>	PS
<digitalCameraModelName>	název modelové řady, např. EOS 0-1 povinné	PS
< digitalCameraModelNumber>	označení modelu fotoaparátu, např. 1000D 0-1 povinné	PS
< digitalCameraModelSerialNo>	výrobní číslo přístroje, např. E12345 0-1	PS

	povinné	
<camerarSensor>	typ senzoru fotoaparátu, např. matrix aj. 0-1 povinné	PS
<CameraCaptureSettings>	údaje o nastavení fotoaparátu použitého ke snímání předloh 0-1 povinné	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	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. 0-1 povinné	PS
<ImageAssessmentMetadata>	informace o digitálním obrazu pro jeho hodnocení a využití z hlediska dlouhodobé ochrany apod.	MC, PS

	<b>0-1 povinné</b>	
<b>&lt;SpatialMetrics&gt;</b>	<b>rozměry obrázku, 2 rozměrná projekce objektů tak jak ji „vidí“ snímací zařízení 0-1 povinné</b>	<b>MC, PS</b>
<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 doporučené	MC + PS
<samplingFrequencyUnit>	jednotka měření sampl. frekvence, např. hodnoty 1= žádná pevná jednotka ; 2= inch, 3=centimetr; 0-1 povinné	MC, PS
<b>&lt;xSamplingFrequency&gt;</b>	<b>údaje o počtu pixelů na jednotku smplovací frekvence pro šířku obrázku 0-1 povinné, pokud hodnota samplingFrequencyUnit je 2 nebo 3</b>	<b>MC, PS</b>
<numerator>	čítatel, číselné vyjádření, např. 300 0-1 povinné	MC, PS
<denominator>	jmenovatel, číselné vyjádření např. 1 0-1 povinné	MC, PS
<b>&lt;ySamplingFrequency&gt;</b>	<b>údaje o počtu pixelů na jednotku smplovací frekvence pro výšku obrázku 0-1 povinné, pokud hodnota samplingFrequencyUnit je 2 nebo 3</b>	<b>MC, PS</b>
<numerator>	čítatel, číselné vyjádření, např. 300 0-1 povinné	MC, PS
<denominator>	jmenovatel, číselné vyjádření např. 1 0-1 povinné	MC, PS
<b>&lt;ImageColorEncoding&gt;</b>	<b>doplňující údaje o barvě obrazu</b>	<b>MC, PS</b>

	<b>0-1 povinné</b>	
<b>&lt;BitsPerSample&gt;</b>	<b>počet bitů na kanál 0-1 povinné</b>	<b>MC, PS</b>
<bitsPerSampleValue>	hodnota počtu bitů, např. 8, 1, 4 nebo 8,8,8 apod. 0-n povinné	MC, PS
<bitsPerSampleUnit>	specifikace jednotky, např. integer nebo floating point 0-1 doporučené	MC, PS
<samplesPerPixel>	počet barevných komponentů na pixel, např. 1, 3, 4 0-1 povinné	MC, PS
<b>&lt;TargetData&gt;</b>	<b>informace o kalibračních tabulkách 0-1 povinné pro obrazy, kde se dělá kontrola oproti kalibrační tabulce</b>	<b>MC</b>
<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 povinné	MC
<b>&lt;targetID&gt;</b>	<b>údaje o původu kalibrační tabulky 0-n povinné</b>	<b>MC</b>
<targetManufacturer>	výrobce/původce kalibrační tabulky, např. Eastman Kodak nebo NK ČR, oddělení kontroly kvality apod. 0-1 povinné	MC
<targetName>	název kalibrační tabulky, např. ColorChecker, MicrofilmScanTarget aj. 0-1 povinné	MC
<targetNo>	číslo nebo verze kalibrační tabulky	MC

	0-1 povinné	
<targetMedia>	údaj o tom, na jakém médiu je kalibrační tabulka, např. film, paper aj. 0-1 doporučené	MC
<externalTarget>	údaje o externí kalibrační tabulce; např. link na <a href="http://skenservis.cz/target-00000001">http://skenservis.cz/target-00000001</a> nebo název a cesta ke konkrétnímu souboru 0-n povinné v případě, že byla použita externí kalibrační tabulka (targetType = 0)	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 doporučené	MC
<ChangeHistory>	<b>dokumentace procesů provedených na obrazovém souboru v jeho životním cyklu</b> 0-1 povinné	MC
<ImageProcessing>	<b>údaje o zpracování obrazového souboru</b> 0-n povinné	MC
<dateTimeProcessed>	2009-01-04T15:12:06; zapsat v ISO 8601 na úrovni vteřin 0-1 povinné	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 povinné	MC
<processingAgency>	The National Library of the Czech Republic 0-n doporučené	MC

## 8.4 METS část <fileSec>

### 8.4.1 <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

#### 8.4.2 <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

### 8.5 METS část <structMap> - Strukturální metadata a ALTO XML

#### 8.5.1 <structMap> hlavního záznamu METS

- zaznamenávají hierarchické informace o dokumentu, včetně vazeb na fyzické soubory, ze kterých se skládají jednotlivé úrovně dokumentu
- 1 strukturální mapa v hlavním METS záznamu popisuje 1 číslo periodika a musí popisovat strukturu až na úroveň všech článků čísla
- strukturální mapa logická i fyzická včetně linků na ALTO XML bude v hlavním METS 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:
  - LABEL- může obsahovat titul periodika
  - TYPE – např. newspaper



- ID – identifikátor div
- 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/DigitizedPeriodicals/DTD/2.10/DocumentationPeriodical/Periodical.html#element\\_PeriodicalPage\\_Link031EEEEAO](http://digit.nkp.cz/DigitizedPeriodicals/DTD/2.10/DocumentationPeriodical/Periodical.html#element_PeriodicalPage_Link031EEEEAO))
    - ID – identifikátor div
    - ORDERLABEL – pořadové číslo stránky, jak je na ní vytištěno
    - ORDER – pořadí stránky v čísle periodika
  - <div> pro soubory stránky vždy obsahují link <fptr> na soubor obrazu archivní kopie, na ALTO XML, na OCR.TXT a na AMD\_METS.xml pomocí elementu <par>
    - link na obrazový soubor archivní kopie má v elementu <area> následující atributy: FILEID, který obsahuje ID souboru archivní kopie
    - link 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 úrovni článků nebo např. ilustrací se popisuje pomocí do sebe zanořených elementů <div>
- stránky sestávající pouze z jedné logické oblasti (např. stránka monografie bez nadpisů apod.), jsou popsány jedním div elementem (TYPE="page") a link do souboru ALTO XML vede přímo na element <Page>, ne na jednotlivé textové bloky <TextBlock>
  - pokud stránka obsahuje jen obraz a žádný text, pak je popsána jedním elementem <div> (TYPE="page") 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. článku, ilustraci.
  - a. pokud se jedná o jednoduchý, celistvý článek na jedné straně, tak je popsán jen jedním <div> elementem s atributem TYPE="article"
    - 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 <fptr>

- u bloku tvořeného obrazem je odkaz do ALTO XML na příslušný komponovaný blok <ComposedBlock>; je realizováno pomocí struktury <area> v elementu <fptr>
  - v případě použití atributu ORDER umožňuje tento princip u článků vyjádřit i tzv. pořadí čtení jeho částí, jako jsou např. nadpis, autor, obrázek apod.
- b. pokud článek není celistvý a je rozdělen na více částí, které se vyskytují na jedné nebo více stránkách, je nutné určit pořadí čtení těchto částí, opět pomocí atributu ORDER
- pro každou část článku existuje vlastní <div> element, podřízený hlavnímu <div> elementu článku
  - element <div> každé části má atribut TYPE hodnotu „article-part“ a atribut ID musí vyjadřovat o jakou z částí se jedná, tj. např. ID=“article5-1“ odpovídá první části článku číslo pět

#### Příklad

Logická mapa pro číslo periodika obsahující 2 články. První článek je jen na straně jedna a má 3 součásti (titul a odstavec s normálním textem) a obrázek s popiskem i uvedeným autorem. Druhý článek začíná na straně první a jeho další odstavec je na straně druhé. První část děleného článku má titulní část a běžný text a druhá část na straně druhé obsahuje jen odstavec s běžným textem.

```

<structMap LABEL="Logical Structure" TYPE="LOGICAL">
  <div LABEL="Mladá fronta no. 5 29.06.1979" TYPE="PERIODICAL_TITLE" ID="TITLE_1" DMDID="XY">
    <div LABEL="Mladá fronta no.5 29.06.1979" TYPE="ISSUE" ID="ISSUE_1" DMDID="XY">
      <div LABEL="Boj o zrno" TYPE="ARTICLE" ID="ARTICLE_1" DMDID="XY" ORDER="0">
        <div TYPE="TITLE" ID="ARTICLE_PART_1" ORDER="1">
          <fptr>
            <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK1"/>
          </fptr>
        </div>
        <div TYPE="NORMAL_TEXT" ID="ARTICLE_PART_2" ORDER="2">
          <fptr>
            <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK2"/>
          </fptr>
        </div>
        <div LABEL="Obilí na poli" TYPE="PICTURE" ID="ARTICLE_PART_3" DMDID="XY" ORDER="3">
          <div TYPE="CAPTION" ID="ARTICLE_PART_4">
            <fptr>
              <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK3"/>
            </fptr>
          </div>
          <div TYPE="AUTHOR" ID="ARTICLE_PART_5">
            <fptr>
              <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK4"/>
            </fptr>
          </div>
          <div TYPE="IMAGE" ID="ARTICLE_PART_6">
            <fptr>
              <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="COMPOSED_BLOCK1"/>
            </fptr>
          </div>
        </div>
      </div>
    </div>
    <div LABEL="XVI. sjezd strany" TYPE="ARTICLE" ID="ARTICLE_2" DMDID="XY" ORDER="1">
      <div TYPE="ARTICLE_PART" ID="ARTICLE_2-1" ORDER="1">
        <div TYPE="TITLE" ID="ARTICLE_PART_1" ORDER="1">
          <fptr>
            <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK5"/>
          </fptr>
        </div>
        <div TYPE="NORMAL_TEXT" ID="ARTICLE_PART_2" ORDER="2">
          <fptr>
            <area FILEID="ALTO_PAGE_1" BETYPE="IDREF" BEGIN="BLOCK6"/>
          </fptr>
        </div>
      </div>
      <div TYPE="ARTICLE_PART" ID="ARTICLE_2-2" ORDER="2">
        <div TYPE="NORMAL_TEXT" ID="ARTICLE_PART_1" ORDER="1">
          <fptr>
            <area FILEID="ALTO_PAGE_2" BETYPE="IDREF" BEGIN="BLOCK1"/>
          </fptr>
        </div>
      </div>
    </div>
  </div>
</div>
</div>
</div>
</structMap>

```

kde jednotlivé části obsahují a popisují...

<div> type	Atributy	Popis
TITLE	LABEL TYPE ID DMDID	<div> obsahuje údaje o titulu periodika povinné ----- LABEL – název titulu periodika, včetně čísla a data vydání čísla, např. Mladá fronta no. 5 29.06.1979 TYPE – hodnota „PERIODICAL_TITLE“ ID – identifikátor <div>, např. hodnota „TITLE_1“ DMDID – obsahuje identifikátor DMD popisné části MODS titulu
ISSUE	LABEL TYPE ID DMDID	<div> obsahuje údaje o čísle periodika povinné ----- LABEL – název titulu periodika, ve stejné podobě jako u titulu, tedy např. „Mladá fronta no. 5 29.06.1979“ TYPE- hodnota ISSUE ID – identifikátor <div>, např. hodnota „ISSUE_1“ DMDID – obsahuje identifikátor DMD popisné části MODS čísla
ARTICLE	LABEL TYPE ID DMDID ORDER	<div> obsahující údaje o jednom článku a jeho částech povinné ----- LABEL – název článku TYPE – hodnota ARTICLE s pořadovým číslem, např. ARTICLE_1 ID – identifikátor <div> elementu DMDID – identifikátor popisných metadat ORDER – pořadí článku
<p>&lt;div&gt; TYPE="ARTICLE" může obsahovat další vnořený &lt;div&gt; různých typů popisující různé části článku, rozlišujeme tyto části (typy):</p> <ul style="list-style-type: none"> <li>- TITLE</li> <li>- SUBTITLE</li> <li>- AUTHOR</li> <li>- TRANSLATOR</li> <li>- NORMAL_TEXT – běžný text bez dalšího upřesnění</li> <li>- PICTURE</li> <li>- ARTICLE_PART - u článků, které jsou rozděleny na více míst na jedné stránce nebo více</li> </ul>		

stránkách		
<ul style="list-style-type: none"> <li>tento &lt;div&gt; pro jednu součást rozděleného článku pak může obsahovat stejné části jako &lt;div&gt; pro článek, tj. (TITLE, SUBTITLE, AUTHOR, TRANSLATOR, NORMAL_TEXT, PICTURE)</li> </ul>		
TITLE	TYPE ID ORDER	<div> obsahující link na textový blok s nadpisem povinné - pokud se plní článek ----- ----- TYPE – hodnota „TITLE“ ID – identifikátor <div> elementu, který popisuje jednu část článku (nadpis), např. hodnota „ARTICLE_PART_1“ ORDER – pořadí části článku
<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 povinné pokud lze vyplnit ----- TYPE – hodnota „SUBTITLE“ ID – identifikátor <div> elementu, který popisuje jednu část článku (podnadpis), např. hodnota „ARTICLE_PART_2“ ORDER – pořadí části článku
<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 povinné pokud lze vyplnit ----- TYPE – hodnota „AUTHOR“ ID – identifikátor <div> elementu, který popisuje jednu část článku (autor), např. hodnota „ARTICLE_PART_3“ ORDER – pořadí části článku
<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 povinné pokud lze vyplnit ----- TYPE – hodnota „TRANSLATOR“ ID – identifikátor <div> elementu, který popisuje jednu část článku (překladatel), např. hodnota „ARTICLE_PART_3“ ORDER – pořadí části článku
<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 s běžným textem povinné ----- TYPE – hodnota „NORMAL_TEXT“ ID – identifikátor <div> elementu, který popisuje jednu část článku (běžný text), např. hodnota „ARTICLE_PART_4“ ORDER – pořadí části článku
<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 článku povinné pokud se obraz vyskytuje ----- LABEL – název obrazu pokud existuje TYPE - PICTURE ID – identifikátor <div> elementu, který popisuje jednu část článku (běžný text), např. hodnota „ARTICLE_PART_5“ DMDID – link na bibliogr. popis obrazu ORDER – pořadí obrazu
<p>&lt;div&gt; element s typem PICTURE může obsahovat další &lt;div&gt; elementy s typy CAPTION, PICT_AUTHOR, PICT_TITLE a IMAGE;</p> <ul style="list-style-type: none"> <li>- CAPTION obsahuje text případného popisku k obrazu</li> <li>- AUTHOR obsahuje text se jménem případného autora obrazu</li> <li>- TITLE obsahuje text názvu obrazu, pokud nějaký název existuje</li> </ul>		

- IMAGE – obsahuje link do souboru ALTO XML na blok popisující vlastní obraz		
CAPTION	TYPE ID	<div> obsahující link na textový blok s popisem obrazu povinné pokud lze vyplnit ----- TYPE – hodnota CAPTION ID – identifikátor <div> elementu, např. „ARTICLE_PART_6“
<fptr> <area>	FILEID BEGIN BETYPE	FILEID – ID ALTO XML souboru BEGIN – ID textového bloku v ALTO XML souboru BETYPE – hodnota IDREF
PICT_AUTHOR	TYPE ID	<div> obsahující link na textový blok s autorem obrazu povinné pokud lze vyplnit ----- TYPE – hodnota PICT_AUTHOR ID – identifikátor <div> elementu, např. „ARTICLE_PART_7“
<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 povinné pokud lze vyplnit ----- TYPE – hodnota PICT_TITLE ID – identifikátor <div> elementu, např. „ARTICLE_PART_7“
<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 povinné pokud lze vyplnit ----- TYPE – hodnota IMAGE ID – identifikátor <div> elementu, např. „ARTICLE_PART_8“
<fptr> <area>	FILEID BEGIN BETYPE	FILEID – ID ALTO XML souboru BEGIN – ID komponovaného bloku v ALTO XML souboru

		BETYPE – hodnota IDREF
ARTICLE_PART	TYPE ID ORDER	<p>&lt;div&gt; obsahující další vnořené &lt;div&gt; odkazující na jednotlivé části konkrétní části rozděleného článku; povinné pro dělený článek</p> <p>Pozn: pod &lt;div&gt; TYPE="ARTICLE_PART" lze vnořit všechny typy &lt;div&gt; jako pod &lt;div&gt; TYPE="ARTICLE"</p> <p>-----</p> <p>TYPE – hodnota „ARTICLE_PART“</p> <p>ID – identifikátor &lt;div&gt; konkrétní části, pro první část děleného článku např. „ARTICLE_2-1“, tj. první část článku 2</p> <p>ORDER – pořadí konkrétní části děleného článku</p>

**Jednotlivé <div> elementy lze kombinovat a vytvářet nové struktury.**

### 8.5.2 <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="PERIODICAL\_PAGE"
- do <div> budou vnořeny odkazy na jednotlivé reprezentace stránky periodika (MC, ALTO XML a OCR.TXT) pomocí elementu <fptr> s atributem FILEID

```

<structMap TYPE="PHYSICAL">
  <div TYPE="PERIODICAL_PAGE">
    <fptr FILEID="JP2_0001"/>
    <fptr FILEID="ALTOXML_0001"/>
    <fptr FILEID="OCRTXT_0001"/>
  </div>
</structMap>

```

### 8.6 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 nebo číslu periodika) včetně prázdných stran, fotografií hřbetu, předsádky 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 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 jen pro hlavní text stránky, tj. nebudou popsány např. čísla stránek, běžící nadpisy ani jiné části vyskytující se na okrajích stránky mimo hlavní text (top, left, top a bottom margin)
  - o elementy `topMargin`, `leftMargin`, `rightMargin`, `bottomMargin` tedy nebudou obsahovat elementy `<TextBlock>` ani jiné, jen atributy (viz níže)
- 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. 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
- povinnost plnění dle NK ČR (slovní vyjádření: povinné, doporučené, nepovinné)
- u elementů, které obsahují atributy, jsou atributy rozepsány pod čarou (vysvětlení, povinnost uvádění apod.)

Element	Atribut	Popis
<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 povinné
<sourceImageInformation>		informace o obrazovém souboru, ze kterého vzniklo ALTO XML; 0-1 povinné
<fileName>		jméno obrazového souboru, ze kterého bylo ALTO XML vytvářeno; ideálně i s filesystem cestou jeho uložení; např. n1aImageSeq-33386- b.tif//produkce/OCR/digibok_XY/XY_011.tiff 0-1

		povinné
<fileIdentifier>		jedinečný identifikátor obrazového souboru; 0-n doporučené
<OCRProcessing>	ID	popis procesu vzniku OCR; 0-n povinné ----- ID OCR procesu, např. <OCRProcessing ID="OCRPROCES_1">; povinné
<preProcessingStep>		procesy před vznikem OCR 0-n povinné
<processingDateTime>		určení času procesu, který předcházal samotnému OCR; např. 2008-03-29T19:42:23 dle ISO 8601 na úroveň vteřin; 0-1 povinné
<processingAgency>		jméno nebo kód instituce, např. NK CZ; doporučujeme použít kontrolovaný slovník hodnot; 0-1 doporučené
<processingStepDescription>		popis procesu (např. zarovnání, ořez apod.); 0-n doporučené
<processingStepSettings>		např. CCS OCR Processing Filter 0-1 doporučené
<processingSoftware>		popis SW, který upravoval obrázek před vznikem OCR; 0-1 povinné
<softwareCreator>		výrobce softwaru - např. CCS Content Conversion Specialists GmbH, Germany; 0-1 povinné
<softwareName>		jméno softwaru - např. CCS docWORKS; 0-1 povinné
<softwareVersion>		verze SW, např. 6.2-1.16; 0-1

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

		<p>používat kontrolovaný slovník; povinné</p> <p>FONTSIZE – velikost fontu, např. 10, 12 apod.; povinné</p>
<ParagraphStyle>	<p>ID ALIGN</p>	<p>definuje formátování textových bloků; 0-n povinné</p> <p>-----</p> <p>ID pro každý odstavec + zarovnání; např. PAR_LEFT, PAR_RIGHT, PAR_CENTER, PAR_BLOCK; povinné</p> <p>ALIGN – zarovnání; povolené hodnoty: Left, Right, Center, Block aj.; povinné</p>
<Layout>		<p>layout - rozložení struktur (slov, odstavců apod.) na jedné stránce dokumentu; 1-1 povinný výskyt element není opakovací povinné</p>
<Page>	<p>ID ACCURACY POSITION QUALITY PHYSICAL_IMG_NR HEIGHT WIDTH PC</p>	<p>element popisující jednu stránku dokumentu; 1-n povinné</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>

		<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</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p>	<p>horní okraj – prostor mezi vrchní hranou listu a vrchní linkou textu;</p> <p>0-1</p> <p>povinné</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>
<LeftMargin>	<p>ID</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p>	<p>levý okraj – prostor mezi levým okrajem stránky a textem;</p> <p>0-1</p> <p>povinné</p> <p>-----</p>

	HEIGHT	<p>ID: unikátní ID pro element LeftMargin, např. P1_LM0001 (page 1, leftMargin0001); povinné</p> <p>HPOS: horizontální pozice; povinné</p> <p>VPOS: vertikální pozice; povinné</p> <p>WIDTH – šířka levého okraje; povinné</p> <p>HEIGHT – výška levého okraje; povinné</p>
<RightMargin>	ID HPOS VPOS WIDTH HEIGHT	<p>pravý okraj – prostor mezi pravým okrajem stránky a textem; 0-1 povinné</p> <p>-----</p> <p>ID: unikátní ID pro element RightMargin, např. P1_RM0001 (page 1, rightMargin0001); povinné</p> <p>HPOS: horizontální pozice; povinné</p> <p>VPOS: vertikální pozice; povinné</p> <p>WIDTH – šířka pravého okraje; povinné</p> <p>HEIGHT – výška pravého okraje; povinné</p>
<BottomMargin>	ID HPOS VPOS WIDTH HEIGHT	<p>pravý okraj – prostor mezi spodním okrajem stránky a textem; 0-1 povinné</p> <p>-----</p> <p>ID: unikátní ID pro element BottomMargin, např.</p>

		<p>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>
<PrintSpace>	<p>ID HPOS VPOS WIDTH HEIGHT</p>	<p>popis tvaru pokrývajícího textové pole stránky; 0-1 povinné</p> <p>-----</p> <p>ID: unikátní ID pro element &lt;printSpace&gt;, 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>
<TextBlock>	<p>ID STYLEREFS HPOS VPOS WIDTH HEIGHT</p>	<p>popisy textových bloků na konkrétní stránce; 0-n pokud je stránka prázdná, TextBlock není potřeba uvádět; pokud je na stránce text &gt; povinné</p> <p>-----</p> <p>ID obsahuje identifikátor textového bloku na stránce, např. "BLOCK1" nebo P1_TB0002 (stránka 1, textový blok 2); povinné</p>



		<p>STYLEREFS: reference na ID definice formátování textových bloků &lt;ParagraphStyle&gt;; 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 &lt;TextBlock&gt; jeden nebo žádný výskyt &lt;Shape&gt;; doporučeno – v případě, že je tvar textového bloku nestandardní (víceúhelník)</p>
<Polygon>	POINTS	<p>popis (souřadnice) tvaru víceúhelníku; 0-1 povinné</p> <p>-----</p> <p>POINTS – vyjádření jednotlivých bodů víceúhelníku; povinné</p>
<TextLine>	ID STYLEREFS HPOS VPOS WIDTH HEIGHT	<p>popis jedné řádky textu v rámci textového bloku; 1-n povinný - 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 &lt;ParagraphStyle&gt;; nepovinné</p> <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</p> <p>CONTENT</p> <p>HEIGHT</p> <p>WIDTH</p> <p>HPOS</p> <p>VPOS</p> <p>CC</p> <p>WC</p> <p>V případě dělení slov také: SUBS_TYPE SUBS-CONTENT</p>	<p>řetězec znaků – vlastní obsah OCR;</p> <p>znaky tvoří jednotlivá slova a více tagů &lt;String&gt; větu &lt;TextLine&gt;;</p> <p>1-n v rámci &lt;TextLine&gt;</p> <p>povinné</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 CONTENT="TEXT"; povinné</p> <p>WC – úroveň důvěry v přesnost OCR výstupu celého</p>

		<p>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 substitute; 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 &lt;CONTENT&gt; ve větě (stringu) první; <i>HypPart2</i> se vyskytuje u následujícího tagu &lt;CONTENT&gt; v následující větě (stringu), který obsahuje druhou část rozděleného slova/řetězce; <i>Abbreviation</i> – typ substitute 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 lze použít v případě nejistoty rozpoznání řetězce; nepovinné</p>
<HYP>	<p>CONTENT WIDTH HPOS VPOS</p>	<p>zápis znaku rozdělovníku slov 0-1 pro jeden výskyt &lt;TextLine&gt;; vždy pro poslední &lt;String&gt;; 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>
<SP>	<p>ID WIDTH HPOS VPOS</p>	<p>prázdný prostor mezi řádky; 0-n v rámci jednoho &lt;TextLine&gt;; vždy mezi řádky, tj. mezi tagy &lt;String&gt;; povinné</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; povinné</p>
<ComposedBlock>	<p>ID TYPE HPOS VPOS WIDTH HEIGHT STYLEREFS</p>	<p>blok sestávající z jiných bloků; může obsahovat PrintSpace/ComposedBlock/TextBlock, PrintSpace/ComposedBlock/Illustration, PrintSpace/ComposedBlock/GraphicalElement, /PrintSpace/ComposedBlock/ComposedBlock, 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>-----</p> <p>ID: unikátní ID komponovaný blok, např. P6_CB0001 (stránka 6, komponovaný blok 0001); povinné</p> <p>TYPE – označení typu komponovaného bloku; nutné</p>

		<p>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; povinné</p>
<Shape>		<p> tvar 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 tvar komponovaného bloku nestandardní (víceúhelník)</p>
<Polygon>	POINTS	<p>popis tvaru víceúhelníku; 0-1 povinné</p> <p>-----</p> <p>POINTS – vyjádření jednotlivých bodů víceúhelníku povinné</p>
<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 &lt;ComposedBlock&gt; 0 nebo více elementů /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock&gt;; povinné pokud je v komponovaném bloku text</p> <p>-----</p> <p>ID obsahuje identifikátor textového bloku v komponovaném bloku, např. P1_CB0002_SUB (stránka 1,</p>

		<p>textový blok 2, SUB značí komponovaný blok); povinné</p> <p>STYLEREFS: reference na ID definice formátování textových bloků /alto/Styles/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>
<TextLine>	/alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock/TextLine a ostatní elementy v rámci /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock mají stejná pravidla a výskyty jako jako ve vrchním elementu /alto/Layout/Page/PrintSpace/TextBlock	
<GraphicalElement>	<p>ID</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p>	<p>popis grafického tvaru; v případě využití v rámci /alto/Layout/Page/PrintSpace/ComposedBlock označuje rozměry tvaru v rámci něhož je tabulka, ilustrace, reklama apod.;</p> <p>0-1 - pro jeden výskyt /alto/Layout/Page/PrintSpace/ComposedBlock 0 nebo max. 1 výskyt &lt;GraphicalElement&gt;; povinné - 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>

		HPOS – horizontální pozice grafického tvaru; povinné  VPOS – vertikální pozice grafického tvaru; povinné
--	--	----------------------------------------------------------------------------------------------------------------------