

Definice metadatových formátů pro digitalizaci periodik

Vychází z obecné specifikace metadat v projektu NDK

verze 1.4
ze dne 4.4.2012

Autor: Jan Hutař
Úpravy: Pavla Švástová

2012

Historie verzí

jméno	datum	verze dokumentu	provedené změny
Pavla Švástová	4.4.2012	1.4	přidán element <structLink> do METS (zrušena druhá logická mapa), zpřesnění využití identifikátorů, sjednocena vnitřní ID u METS, přidán MODS pro ročník kvůli uuid, další drobné úpravy
Pavla Švástová	2.3.2012	1.3	Úpravy dle domluv s producenty dat a krajskými knihovnami, přidána druhá logická mapa
Jan Hutař	12.2.2012	1.2	doplněna možnost pro obrázek, který není součástí článku do logické strukturální mapy, TYPE <div> je PAGE
Jan Hutař	6.12.2011	1.2	úprava atributu USE v <fileGroup> u ALTO XML
Jan Hutař - NK	25.11.2011	1.2	úprava logické strukt. mapy – v ukázce oprava AUTHOR u obrazu na PICT_AUTHOR
Jan Hutař - NK	25.10.2011	1.2	úpravy překlepů, u elementu <genre> u vnitřní části přidána hodnota atributu „unspecified“; přidán element <url> k elementu <location>;
Jan Hutař - NK	19.10.2011	1.2	přidány informace o kořenovém elementu METS (LABEL a TYPE); opravy překlepů o omylů
Jan Hutař - NK	20.9- 12.10.2011	1.1	úpravy DC; povinností výskytu u některých elementů v MODS; forma zápisu data u vnitřních částí MODS; úpravy povinností plnění subelementů ALTO <preProcessingStep> <processingDateTime>; změna MD5 souborů, doplněna specifikace info.xml; nový sloupec pro povinnost elementů; <margin> v ALTO bude obsahovat text a textové bloky; upřesnění příloh; u všech tabulek přibyl sloupec povinnost; do logické mapy doplněn <div> pro přílohu; specifikace info.xml; + další drobnosti
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

Obsah dokumentu

HISTORIE VERZÍ	2
OBSAH DOKUMENTU	3
1 VÝCHODISKA	4
2 VÝSTUPY DIGITALIZACE	5
3 GRANULARITA METADATOVÉHO ZÁZNAMU	6
4 IDENTIFIKÁTORY	7
5 STRUKTURA PSP BALÍČKU	9
5.1 SOUBOR INFO.XML.....	9
5.2 SLOŽKA [MASTERCOPY].....	10
5.3 SLOŽKA [USERCOPY]	10
5.4 SLOŽKA [ALTO].....	10
5.5 SLOŽKA [TXT]	10
5.6 SLOŽKA [AMDSEC]	10
5.7 SOUBOR HLAVNI_METS.XML	11
5.8 SOUBOR MD5	11
6 NÁZVOVÁ KONVENCE SLOŽEK A SOUBORŮ	12
7 TRANSPORTNÍ BALÍK PRO JEDEN NEBO VÍCE PSP BALÍČKŮ	14
8 METADATA	15
8.1 KOŘENOVÝ ELEMENT HLAVNÍHO METS ZÁZNAMU	16
8.2 METS HLAVIČKA <METSHDR>.....	17
METS ČÁST <DMDSEC> - BIBLIOGRAFICKÁ METADATA.....	18
8.3.....	18
8.3.1 <i>Pole MODS a Dublin Core pro jednotlivé části periodika</i>	22
8.3.1.1 Pole MODS a DC pro titul periodika	23
8.3.1.2 Pole MODS a DC pro ročník periodika	28
8.3.1.3 Pole MODS a DC pro číslo periodika	30
8.3.1.4 Pole MODS a DC pro vnitřní část periodika (článek a obraz)	35
8.3.1.5 Pole MODS a DC pro přílohu.....	40
8.4 METS ČÁST <AMDSEC> - TECHNICKÁ A ADMINISTRATIVNÍ METADATA – FORMÁTY MIX A PREMIS..	45
8.4.1 <i>PREMIS Object</i>	47
8.4.2 <i>PREMIS Event</i>	53
8.4.3 <i>PREMIS Agent</i>	56
<i>Technická metadata MIX</i>	58
8.4.4.....	58
8.5 METS ČÁST <FILESEC>.....	67
8.5.1 <i><fileSec> hlavního záznamu METS</i>	67
8.5.2 <i><fileSec> vedlejšího záznamu METS (AMD_METS.xml)</i>	69
8.6 METS ČÁST <STRUCTMAP> - STRUKTURÁLNÍ METADATA.....	70
8.6.1 <i><structMap> hlavního záznamu METS</i>	70
8.6.1.1 Vyjádření fyzické strukturální mapy	71
8.6.1.2 Vyjádření logické strukturální mapy	72
8.6.2 <i><structMap> vedlejšího záznamu METS (AMD_METS.xml)</i>	79
8.7 METS ČÁST <STRUCTLINK> - VÝČET STRAN	80
8.7.1 <i>Výčet stran v případě popisu článků (s odkazy na bloky ALTA i bez nich)</i>	80
8.7.2 <i>Výčet stran bez popisu článků</i>	81
8.8 OCR (ALTO XML A TXT OCR).....	82

1 Východiska

- UC = user copy = uživatelské kopie
- MC = master copy = archivní kopie
- PS = původní sken – obrazový soubor vzniklý při digitalizaci, který se po zpracování (ořez, narovnání apod.) maže a dále se neukládá
- u všech metadatových formátů budou použity verze aktuální v době implementace projektu NDK nebo verze předchozí v případě, že nová verze je nová min. 3 měsíce
- základní intelektuální entita ve workflow digitalizace a následně i v LTP systému = číslo periodika
- PSP balíček – producer submission package
 - o balíček dat a metadat, který přichází od producenta dat (tedy např. z workflow digitalizace)
 - o PSP balíček bude obsahovat kompletní intelektuální entitu tj. číslo periodika
 - o z workflow digitalizace lze poslat více PSP balíčků v balíku např. [.tar] apod.
 - o pokud má dvousvazkové dílo v katalogu knihovny bibliografický záznam pro každý svazek, vznikne pro každý svazek PSP balíček a každý svazek bude brán jako jedna intelektuální entita; to samé platí i pro případ, že vícesvazkové dílo má pouze jeden záznam
- 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.

METS neobsahuje popisná ani technická metadata pro UC. Obrazové soubory UC jsou pouze součástí strukturální mapy <structMap> a <fileSec>. Součástí PSP balíčku jsou obrazy UC ve složce [userCopy].

3 Granularita metadatového záznamu

Periodika

- **základní intelektuální entitou periodik je 1 číslo**
- každé číslo periodika má svůj vlastní metadatový záznam (=METS)
- tento METS záznam 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 Identifikátory

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

- ISSN – pouze pro titul periodika, ISSN není přiděleno vždy (chybí např. u starých titulů z 19. století)
- ISBN – pouze pro titul monografie (jednosvazkové), nebo pro soubor monografií, které mají pouze jeden souborný záznam, ISBN není přiděleno vždy
-
- ČČNB – identifikátor entity tak jak odpovídá katalogizačnímu záznamu, tj. každá entita dle pravidel na <http://www.caslin.cz/spoluprace/sluzby/ccnb/> se záznamem v katalogu NK/MZK má tento identifikátor
- pokud není ani jeden z výše uvedených, lze použít čárový kód dokumentu, systémové číslo, signatura, nebo systémové číslo kombinované s polem 001 MARC záznamu apod.

Producent dat musí generovat do balíčku dle této specifikace:

- identifikátor UUID¹ ke každé úrovni bibliografických metadat ve formátu MODS (prvek <identifier> s atributem type=„uuid“ a DC (prvek <dc:identifier>)
 - o pro úroveň celého titulu (nenachází se v bibl. záznamu v knihovním katalogu)
 - o pro úroveň ročníku
 - o pro úroveň čísla a přílohy
 - o pro úroveň článků a obrázků
- URN:NBN, které bude možné nechat si přidělit nástrojem „Resolver URN:NBN“ pro dané úrovně²

Identifikátory pro úroveň titulu:

1. UUID (vygeneruje dodavatel) - povinné
2. URN:NBN (bude přiděleno aplikací „Resolver URN:NBN“) – povinné, pokud lze přidělit
3. Číslo České národní bibliografie – ČČNB (povinné, pokud je možné je záznamu přidělit)
4. ISSN (pokud je možné je záznamu přidělit)
5. ISBN (pokud je možné je záznamu přidělit)
6. Jiný identifikátor – lze využít různé identifikátory, které nějakým způsobem propojí digitální záznam s fyzickou jednotkou – např. čárový kód, systémové číslo nebo pole 001 z knihovního katalogu atd. Tato čísla lze kombinovat např. s pořadím čísla v ročníku, číslem čísla apod.

Identifikátory pro úroveň ročníku:

1. UUID (vygeneruje dodavatel) - povinné
2. URN:NBN (bude přiděleno aplikací „Resolver URN:NBN“) – povinné, pokud lze přidělit

Identifikátory pro úroveň čísla:

1. UUID (vygeneruje dodavatel) - povinné
2. URN:NBN (bude přiděleno aplikací „Resolver URN:NBN“) – povinné, pokud lze přidělit

¹ UUID se tvoří dle specifikace RFC - <http://www.ietf.org/rfc/rfc4122.txt>

² Pro bližší informace o Resolveru URN:NBN kontaktujte Odbor digitálních fondů v NK: http://www.nkp.cz/pages/page.php3?page=digi_ochrana.htm

3. ISBN (pokud bylo záznamu v minulosti přiděleno)
4. doi, handle apod. (pokud byly záznamu v minulosti přiděleny)
5. Jiné identifikátory (čárový kód, systémové číslo, pole 001 apod.)

Identifikátory pro úroveň článku nebo obrázku:

1. UUID (vygeneruje dodavatel) - povinné
2. doi, handle apod.
3. Jiný identifikátor

5 Struktura PSP balíčku

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

SLOŽKA >	OBSAHUJE >>	OBSAHUJE >>>
Číslo periodika	info.xml	
	masterCopy (složka)	obrazy JPEG 2000 lossless
	usecCopy (složka)	obrazy JPEG 2000 lossy
	ALTO (složka)	soubory ALTO.xml pro každou stranu
	TXT (složka)	soubory OCR.TXT pro každou stranu
	amdSec (složka)	AMD_METS.xml pro každou stranu
	hlavní_METS.xml	
	soubor.md5	

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:

5.1 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 mohou být údaje o obsahu PSP balíčku – počet a názvy souborů apod. Soubor obsahuje odkaz na soubor s MD5. Soubor info.xml by také mohl být vedle hlavního PSP balíčku.

Údaje a struktura info.xml souboru:

1. vznik balíčku – datum dle ISO8601 na úroveň vteřin
2. ID balíčku – použit identifikátor čísla periodika (URN:NBN) – viz názvová konvence v kap. 6
3. ID titulu - čČNB, ISBN nebo ISSN (opakovatelné)
4. údaje o větším celku (projektu), do kterého balíček patří - např. digitalizace pro ANL
5. název instituce, která je zadavatelem digitalizace
6. tvůrce balíčku – kód instituce (firmy), která balíček vytvořila
7. velikost balíčku – v kB
8. obsah balíčku
9. názvy souborů včetně directory path a koncovky (mime type) (nepovinné)
10. počet souborů v balíčku celkem
11. odkaz na soubor s MD5 a jeho MD5
12. poznámka – např. o tom, že balíček neobsahuje OCR apod.

```

<?xml version="1.0" encoding="UTF-8"?>
<info xmlns="http://www.ee.cz/schemas/NDK/info.xsd">
  <created>2012-01-26T13:53:11</created>
  <packageid>anl_123456</packageid>
  <titleid TYPE="issn">ISSN0009-2770</titleid>
  <titleid TYPE="ccnb">cnb123456789</titleid>
  <collection>ANL</collection>
  <institution>NKP</institution>
  <creator>Název firmy</creator>
  <size>123456789</size>
  <itemlist ITEMTOTAL="35"/>
    <item>
      ...
    </item>
  <checksum TYPE="md5" CHECKSUM="fe8e0172fb031cdc79a9f6002cb64f55"
>/MD5_anl001-
  000003.md5</checksum>
  <note />
</info>

```

5.2 složka [masterCopy]

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

5.3 složka [userCopy]

Složka s uživatelskými kopiemi, pro každou naskenovanou stránku čísla periodika obsahuje jeden JPEG2000 soubor se ztrátovou kompresí.

5.4 složka [ALTO]

Obsahuje ke každé stránce 1 ALTO XML soubor, tj. tolik ALTO XML souborů kolik je stránek čísla periodika.

5.5 složka [TXT]

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

5.6 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 části:

- technická metadata **techMD**, která:
 - ve formátu **PREMISobject** popisují vlastnosti:
 - archivní kopie
 - ALTO XML
 - původního TIFF souboru, ze kterého vznikly archivní kopie
 - ve formátu **MIX** popisují vlastnosti:
 - archivní kopie
 - původního TIFF souboru
- metadat o provenienci digitálních objektů **digiprovMD** – obsahuje části
 - **PREMISevent**
 - **PREMISagent**.
- **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.
- **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.

5.7 soubor Hlavní_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. článek, obrázek). Základ bude převzat z katalogu, případný další popis částí bude vyroben v digitalizaci případně dogenerován. Hlavním formátem bude MODS, nutná je i přítomnost zkráceného záznamu v Dublin Core.
- **fileSec** – hlavní část s linky na všechny digitální objekty (archivní kopie, uživatelské 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:
 - **fyzickou část** obsahující informace o všech reprezentacích konkrétní stránky (archivní kopie, uživatelská kopie, ALTO XML, OCR TXT a AMD_METS.xml)
 - **logickou část** – vyjadřuje logickou strukturu čísla periodika s odkazy na ALTO XML
 - mapování na ALTO XML areas
- **structLink** – výčet stran jednotlivých úrovní periodika na základě přidání vazeb mezi logickou a fyzickou strukturou

5.8 soubor MD5

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

6 Názvová konvence složek a souborů

Pojmenování PSP balíčku:

- každý PSP balíček přicházející z digitalizace musí obsahovat pouze jedinou intelektuální entitu (číslo periodika). **Pak musí název balíčku vycházet z identifikátoru této entity, v nejlepším případě URN:NBN, dále je možnost využít např. číslo čárového kódu použitého na fyzické jednotce v kombinaci s pořadovým číslem čísla periodika apod.**
- **každé číslo periodika musí mít svůj jednoznačný identifikátor, tím pádem pak každý PSP balíček a každý soubor v něm má vlastní jednoznačný identifikátor**
- názvy nesmí obsahovat mezery a diakritiku, doporučené oddělovače jsou podtržítka a pomlčka

Pojmenování složek:

- viz návrh struktur PSP balíčku (kap. 5)
- názvy nesmí obsahovat mezery a diakritiku, doporučené oddělovače jsou podtržítka a pomlčka

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 URN:NBN, 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.).
- názvy nesmí obsahovat mezery a diakritiku, doporučené oddělovače jsou podtržítka a pomlčka

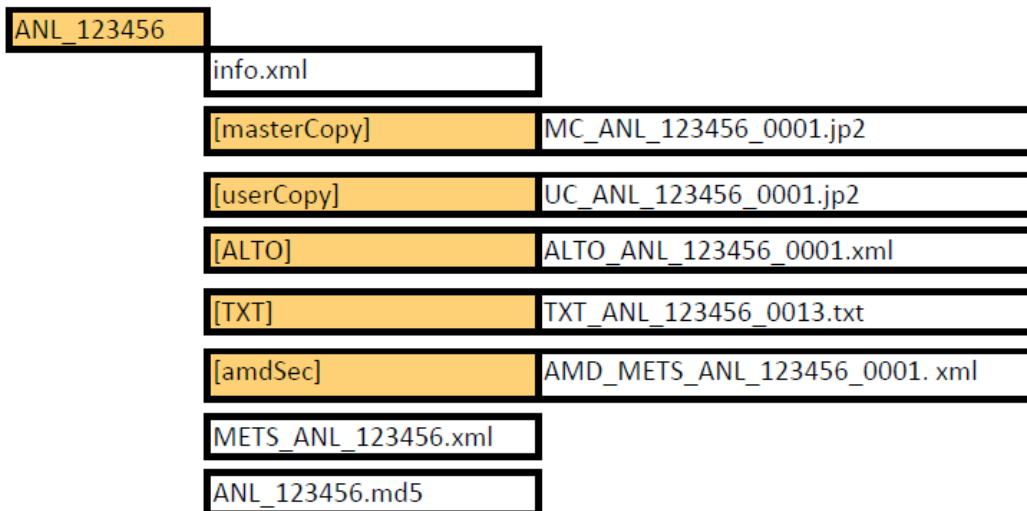
S využitím URN:NBN mohlo by to vypadat následovně (*POZOR: je použit příklad pojmenování pro projekt ANL+ - digitalizace periodik*):

typ souboru	název souboru	vysvětlení
PSP balíček (číslo, svazek)	ANL_123456	název celé složky PSP balíčku, u základních int. entit bude v názvu využito vždy URN:NBN
archivní kopie	MC_ANL_123456_0013.jp2	archivní JPEG2000 stránky 13 čísla periodika s urn:nbn:cz:anl-123456
uživatelská kopie	UC_ANL_123456_0013.jp2	uživatelská kopie ve formátu JPEG2000 stránky 13 čísla periodika s urn:nbn:cz:anl-123456
ALTO XML	ALTO_ANL_123456_0013.xml	ALTO soubor náležející ke 13té stránce z čísla periodika s urn:nbn:cz:anl-123456
OCR TXT	TXT_ANL_123456_0013.txt	TXT soubor s OCR náležející ke 13té stránce z čísla periodika s

		urn:nbn:cz:anl-123456
info.xml	INFO_ANL_123456.xml	info xml k celému PSP balíčku čísla periodika
MD5	ANL_123456.md5	soubor s kontrolními součty k celému PSP balíčku čísla periodika
Hlavni_METS.xml	METS_ANL_123456.xml	hlavní METS záznam k celému číslu periodika s urn:nbn:cz:anl-123456
AMD_METS.xml	AMD_METS_ANL_123456_0013.xml	METS záznam s technickými metadaty pro stránku 13 z čísla periodika s urn:nbn:cz:anl-123456

popis pojmenování složek viz struktura balíčku výše v kapitole 5

Složka jednoho balíčku PSP, který obsahuje jen jeden obrazový soubor k první stránce čísla periodika, pak může vypadat následovně (*příklad balíčku z digitalizace v projektu ANL+*):



7 Transportní balík pro jeden nebo více PSP balíčků

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

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

- v případě, že balík obsahuje 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_1234-5678.tar nebo CCNB_12345678910.tar apod.
- lze počítat s tím, že bude docházet k tomu, že sdružený balík nebude obsahovat např. všechna čísla určitého titulu periodika – tato skutečnost musí být patrná z názvu balíku (např. ISSN_1234-5678_YYYY kde YYYY může být pořadové číslo, datum, doba vzniku jednoho z více balíčků obsahujících čísla určitého titulu s identifikátorem ISSN 1234-5678).

Transportní balík by měl obsahovat následující části:

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

Do úvahy mohou přijít balící metody jako BagIt⁴, tar apod.

⁴ <https://confluence.ucop.edu/display/Curation/BagIt>

8 Metadata

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

Význam pole „Povinnost“

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

8.1 Kořenový element hlavního METS záznamu

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

element	atributy	popis	Povinnost
<mets>		kořenový element METS záznamu	M
	LABEL	LABEL – název titulu periodika, včetně čísla a data vydání čísla, např. Mladá fronta no. 5 29.06.1979	
	TYPE	TYPE – hodnota vždy „Periodical“	

Příklad:

```
<mets:mets
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:mods="http://www.loc.gov/mods/v3"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xsi:schemaLocation="http://www.w3.org/2001/XMLSchema-instance
http://www.w3.org/2001/XMLSchema.xsd http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/mods.xsd
http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
LABEL="Chemické listy no.6 1951"
TYPE="Periodical"
xmlns:mets="http://www.loc.gov/METS/">
```


8.2 METS hlavička <metsHdr>

- dokumentuje vznik a úpravy METS záznamu

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

Příklad:

```
<mets:metsHdr CREATEDATE="2012-01-26T13:49:12" LASTMODDATE="2012-01-26T13:49:12">
  <mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
    <mets:name>Elsyst EGINEERING</mets:name>
  </mets:agent>
  <mets:agent ROLE="ARCHIVIST" TYPE="ORGANIZATION">
    <mets:name>ABA001</mets:name>
  </mets:agent>
</mets:metsHdr>
```

8.3 METS část <dmdSec> - Bibliografická metadata

- na samotný bibliografický popis bude použit formát MODS, aktuální verze v době implementace, nebo verze předchozí (prosinec 2010 verze 3.4 viz <http://www.loc.gov/standards/mods/>) a formát Dublin Core (dále DC) kvalifikovaný (<http://dublincore.org/documents/dcmi-terms/>)
- DC je primárně určeno na poskytnutí dat přes OAI-PMH, bude odpovídat OAI XSD (viz http://www.openarchives.org/OAI/2.0/oai_dc.xsd) a bude se jednat o nekvalifikovaný Dublin Core
- DC bude použito, uloženo v METS apod. stejným způsobem jako formát MODS – viz struktura PSP balíčku výše
- pro vytvoření DC z MODS formátu může být použito (a dle potřeb knihovny upraveno) 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 – pro převod metadat k titulu periodika z MARC 21 resp. MARCXML do MODS lze využít a dle potřeb knihovny upravit oficiální mapování a konverzní šablonu ze stránek <http://www.loc.gov/standards/mods/mods-conversions.html>
- 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:
 1. **titul (Title)**
 2. **ročník (Volume)**
 3. **číslo (Issue)**
 4. **vnitřní část (InternalPart) – typy články (Article) a obraz (Picture)**
 5. **příloha (Supplement)**
1. **titul (Title)** – MODS záznam pro titul periodika
 2. **ročník (Volume)**
 3. **číslo (Issue)** – typy čísla jsou v elementu <genre> za použití atributu type
 4. **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>
 - a) 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
 5. **příloha (Supplement)** - přílohou se rozumí volně vložená entita do jednotlivého čísla, např. mapa, obsah celého ročníku, CD/DVD apod.
Rozlišujeme 3 druhy příloh periodik:
 - a) příloha, která **se neskenuje**, ale chceme o ní vytvořit bibliografický záznam, dát najevo

čtenáři, že existuje – to může být např. CD/DVD, pohlednice, plakát apod.

- digitální podoba přílohy (pokud existuje) není součástí balíčku PSP čísla (Issue)
 - popis lze udělat v rámci popisu přílohy (Supplement) v MODS – viz specifikace níže
 - taková příloha není součástí logické strukturální mapy formátu METS
- b) příloha podobného typu, tvaru a velikosti jako je popisované číslo periodika, která se spolu s číslem **skenuje**
- digitální podoba přílohy je, spolu s číslem (Issue), součástí PSP balíčku čísla (Issue) a je součástí hlavního METS záznamu
 - popis lze udělat v rámci popisu přílohy (Supplement) v MODS – viz specifikace níže
 - taková příloha může mít vnitřní části (InternalPart) stejně jako číslo (Issue) a jejich text je součástí ALTO XML, které je společné pro číslo (Issue) i přílohu (Supplement)
 - **taková příloha je součástí logické strukturální mapy formátu METS**
 - **taková příloha je součástí fyzické strukturální mapy formátu METS (linky mezi jednotlivými soubory reprezentujícími stránky a popisnými metadaty)**
- c) příloha odlišného typu, tvaru a velikosti než je popisované číslo periodika, která **se skenuje zvlášť** na čísle nezávisle
- taková příloha se zpracovává z pohledu katalogizace jako „nezávislé“ periodikum, z pohledu digitalizace pak jako „nezávislý“ časopis
 - může se jednat o přílohy časopiseckého typu vycházející u různých deníků (Pátek u Lidových Novin, čtvrtěční příloha MF Dnes apod.)
 - k těmto přílohám vznikají metadata podobně jako pro jednotlivá čísla deníků nebo klasické časopisy, ovšem na původním čísle, ke kterému příloha patřila, nezávisle - tj. pro „původní“ číslo, u kterého byla příloha, vznikne 1 popis (PSP balíček s jedním hlavním METS záznamem a ALTO XML souborem) a pro přílohu je vytvořen další 1 popis (a PSP balíček s METS záznamem), jako by šlo o běžný samostatný časopis
 - příloha se pak popisuje jako číslo (Issue)
- jednotlivé MODS záznamy pro části (titul, ročník, 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. **Identifikátory budou začínat prefixy:** MODSMD_TITLE, MODSMD_ISSUE, MODSMD_SUPPL, MODSMD_ART, MODSMD_PICT pro MODS, obdobně pro DC. Za ty se dále přidá podtržítka a číslo, identifikující pořadí identifikátoru, zarovnané a doplněné o nuly na 4 místa. Čtyřmístná pořadová čísla u ID jsou uvedena proto, aby byla v celém dokumentu jednotná. ID tedy vypadá následovně:
- titul (vždy jeden)
 - MODSMD_TITLE_0001
 - DCMD_TITLE_0001
 - ročník (vždy jeden)
 - MODSMD_VOLUME_0001

- DCMD_VOLUME_0001
- číslo (vždy jedno)
 - MODSMD_ISSUE_0001
 - DCMD_ISSUE_0001
- příloha (může být více)
 - MODSMD_SUPPL_XXXX, kde XXXX je pořadové číslo přílohy, např. MODSMD_SUPPL_0001 je první příloha atd.
 - DCMD_SUPPL_XXXX, kde XXXX je pořadové číslo přílohy, např. DCMD_SUPPL_0001 je první příloha atd.
- článek (může být více)
 - MODSMD_ART_XXXX, kde XXXX je pořadové číslo článku, např. MODSMD_ART_0001 je první článek atd.
 - DCMD_ART_XXXX, kde XXXX je pořadové číslo článku, např. DCMD_ART_0001 je první článek atd.
- obrázek (může být více)
 - MODSMD_PICT_XXXX, kde XXXX je pořadové číslo obrázku, např. MODSMD_PICT_0001 je první obrázek atd.
 - DCMD_PICT_XXXX, kde XXXX je pořadové číslo obrázku, např. DCMD_PICT_0001 je první obrázek atd.
- 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
 - a) tj. v METS záznamu vznikne 1 část <dmdSec> pro bibliografický záznam titulu periodika, 1 část <dmdSec> pro bibliografický záznam ročníku, 1 část <dmdSec> pro bibliografický záznam čísla periodika, několik <dmdSec> části pro vnitřní části (pro všechny články i obrázky) a odpovídající počet <dmdSec> částí pro přílohy, dle počtu příloh
 - i. 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)
 - a) typ stránky bude odpovídat novým pravidlům popisu periodik (dočasně na: https://docs.google.com/a/mzk.cz/document/d/1zSriHPdnUY5d_tKvOM8a6nEym56ODKh2H6XZ24tGAew/edit)

Jedná se o typy:

advertisement, backCover, backEndSheet, blank, cover, flyLeaf, frontCover, frontEndSheet, index, listOfIllustrations, listOfMaps, listOfTables, normalPage, spine, table, tableOfContents, titlePage
- **všechny top elementy MODS formátu jsou opakovatelné, kromě <recordInfo>**
- **všechny elementy Dublin Core jsou opakovatelné**
- každá část <dmdSec> musí mít ID a vnořený element <mdWrap> s atributy MDTYPE, MIMETYPE

element	atributy	popis	povinnost
<dmdSec>		identifikátor <dmdSec> části METS záznamu	M
	ID	pro <dmdSec> s popisem titulu periodika hodnota „MODSMD_TITLE_0001“ pro záznam v MODS nebo „DCMD_TITLE_0001“ pro záznam v Dublin Core	
		pro <dmdSec> s popisem ročníku periodika hodnota „MODSMD_VOLUME_0001“ pro záznam v MODS nebo „DCMD_VOLUME_0001“ pro záznam v Dublin Core	
		pro <dmdSec> s popisem čísla periodika hodnota „MODSMD_ISSUE_0001“ a „DCMD_ISSUE_0001“	
		pro <dmdSec> s popisem vnitřní části periodika hodnota dle typů vnitřní části (článek, obraz) - hodnoty „MODSMD_ART_XXXX“ a „DCMD_ART_XXXX“ pro článek, kde XXXX je pořadové číslo článku, např. 0001 pro první článek atd. a hodnoty „MODSMD_PICT_XXXX“ a „DCMD_PICT_XXXX“ pro obraz, kde XXXX je pořadové číslo obrázku např. 0001 pro první obrázek atd.	
	pro <dmdSec> s popisem přílohy periodika hodnota „MODSMD_SUPPL_XXXX“ a „DCMD_SUPPL_XXXX“, kde XXXX je pořadové číslo přílohy, např. 0001 pro první přílohu apod.		
<mdWrap>		element obsahující vložené záznamy MODS	M
	MDTYPE	MDTYPE – hodnota „MODS“ pro záznamy v MODS, hodnota „DC“ pro záznam v Dublin Core	
	MIMETYPE	MIMETYPE – hodnota „text/xml“	

Příklad:

```
<mets:dmdSec ID="MODSMD_TITLE_0001">
  <mets:mdWrap MDTYPE="MODS" MIMETYPE="text/xml">
  ...
  </mets:mdWrap>
</mets:dmdSec>
```

8.3.1 Pole MODS a Dublin Core pro jednotlivé části periodika

- ID u elementu <mods>: **Identifikátory budou začínat prefixy:** MODS_TITLE, MODS_ISSUE, MODS_SUPPL, MODS_ART, MODS_PICT pro MODS, obdobně pro DC. Za ty se dále přidá podtržítka a číslo, identifikující pořadí identifikátoru, zarovnané a doplněné o nuly na 4 místa. Čtyřmístná pořadová čísla u ID jsou uvedena proto, aby byla v celém dokumentu jednotná. ID tedy vypadá následovně:
 - o titul (vždy jeden)
 - MODS_TITLE_0001
 - DC_TITLE_0001
 - o ročník (vždy jeden)
 - MODS_VOLUME_0001
 - DC_VOLUME_0001
 - o číslo (vždy jedno)
 - MODS_ISSUE_0001
 - DC_ISSUE_0001
 - o příloha (může být více)
 - MODS_SUPPL_XXXX, kde XXXX je pořadové číslo přílohy, např. MODS_SUPPL_0001 je první příloha atd.
 - DC_SUPPL_XXXX, kde XXXX je pořadové číslo přílohy, např. DC_SUPPL_0001 je první příloha atd.
 - o článek (může být více)
 - MODS_ART_XXXX, kde XXXX je pořadové číslo článku, např. MODS_ART_0001 je první článek atd.
 - DC_ART_XXXX, kde XXXX je pořadové číslo článku, např. DC_ART_0001 je první článek atd.
 - o obrázek (může být více)
 - MODS_PICT_XXXX, kde XXXX je pořadové číslo obrázku, např. MODS_PICT_0001 je první obrázek atd.
 - DC_PICT_XXXX, kde XXXX je pořadové číslo obrázku, např. DC_PICT_0001 je první obrázek atd.

Obsah pole „Popis“:

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

Význam pole „Povinnost“

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

8.3.1.1 Pole MODS a DC pro titul periodika

Element MODS	Atributy	Popis	pov inn ost	Element DC
<mods>	ID	ID musí vyjadřovat název úrovně, tj. „MODS_TITLE_0001“ Kořenový element <mods> je ve skutečnosti o úroveň výš, než titleInfo, ale kvůli vzhledu tabulky je uveden takto, platí i dále)	M	
<titleInfo>		název titulu periodika pro plnění použít katalogizační záznam	M	
<title>		názvová informace – název periodika hodnoty převzít z katalogu	M	<dc.title>
<subTitle>		podnázev periodika	MA	<dc.title>
<partNumber>		číslo části, např. určité řady/edice (část 1, řada B), k použití u ročenek apod.	R	<dc:descrip tion>
<partName>		jméno edice nebo speciální ediční řady, např. Hygiena. k použití u ročenek a specializovaných periodik	R	<dc:descrip tion>
<typeOfResource>		popis charakteristiky typu nebo obsahu zdroje; pro periodika a monografie hodnota text; mělo by se vyčítat z MARC21 katalogizačního záznamu z pozice 06 návěští jedna z hodnot: - text - cartographic - notated music - sound recording-musical - sound recording-nonmusical - sound recording - still image - moving image - three dimensional object - software, multimedia - mixed material	R	<dc:type>
<genre>		bližší údaje o typu dokumentu hodnota „title“	M	<dc:type>
<originInfo>		informace o původu předlohy Poznámka: Jeden nebo více výskytů elementů se předpokládá pro vydavatele, další výskyt v případě nutnosti popsat tiskaře. Pokud je nutno vyjádřit tiskaře (pole 260 podpole „f“ a „e“ a „g“ v MARC21), je nutno element <originInfo> opakovat s atributem transliteration=“printer“ a elementy <place>,	M	

		<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	MA	<dc:coverage>
<placeTerm>		konkrétní určení místa, např. Praha odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „a“	MA	<dc:coverage>
	type	type – bude vždy „text“		
<publisher>		jméno entity, která dokument vydala, vytiskla nebo jinak vyprodukovala odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21; v případě, že existovalo více vydavatelů, jsou uvedeni v poznámce v poli 500 a měli by se objevit v elementu top elementu <note>	MA	<dc:publisher>
<dateIssued>		datum vydání předlohy, nutno zaznamenat v případě titulu roky v nichž časopis vycházel (např. 1900-1939), přebírat ve formě, jak je zapsáno v hodnotě pole v katalogu odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „c“	M	<dc:date>
<dateCreated>		datum vytvoření předlohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo> odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“	R	
<issuance>		údaje o vydávání hodnota „continuing“ odpovídá hodnotě uvedené návěští MARC21 na pozici 07	M	
<frequency>		údaje o pravidelnosti vydávání odpovídá údajům MARC21 v poli 310 nebo pozici 18 v poli 008	R	
<language>		údaje o jazyce dokumentu	M	
<languageTerm>		přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php	M	<dc:language>
	type	type: použít hodnotu code		
	authority	authority: použít hodnotu „iso639-2b“		
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy	M	

8.3.1.1 Pole MODS a DC pro titul periodika

<form>		údaje o fyzické podobě dokumentu, např. print, electronic apod. pro periodika hodnota „print“ odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21	M	<dc:format>
	authority	authority: hodnota „marcform“		
<extent>		údaje o rozsahu (stran, svazků nebo rozměrů); použití spíše u ročenek apod. odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat	RA	<dc:format>
<note>		poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element zde se zapíše defekty zjištěné při digitalizaci pro úroveň periodika (např. chybějící ročníky apod.)	RA	
<abstract>		shrnutí obsahu periodika jako celku odpovídá poli 520 MARC21	R	<dc:description>
<note>		obecná poznámka k periodiku jako celku odpovídá poli 500 v MARC21	RA	<dc:description>
<subject>		údaje o věcném třídění předpokládá se přebírání z katalogizačního záznamu	R	
	authority	authority: vyplnit hodnotu „czenas“		
<topic>		libovolný výraz specifikující nebo charakterizující obsah periodika; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma) nebo obsah pole 650 záznamu MARC21	M	<dc:subject>
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín) nebo obsah pole 651 záznamu MARC21	R	<dc:subject>
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj) nebo obsah pole 648 záznamu MARC21	R	<dc:subject>
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní) nebo obsah pole 600 záznamu MARC21	R	
<namePart>		celé jméno se zapíše do tohoto elementu	R	<dc:subject>
<classification>		klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění odpovídá poli 080 MARC21	M	<dc:subject>
	authority	authority: vyplnit hodnotu „udc“		

8.3.1.1 Pole MODS a DC pro titul periodika

<code><relatedItem></code>		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 <code><title></code> s jejím názvem element <code><relatedItem></code> může obsahovat jakýkoliv jiný element MODS – jejich použití se řídí pravidly popsanými pro tyto elementy;	RA	
	type	type: hodnota „series“		
<code><identifier></code>		údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které titul periodika má – viz přehled typů atributů níže	M	<code><dc:identifier></code>
	type	type: budou se povinně vyplňovat následující hodnoty, pokud existují: - doi - hdl - handle - issn - převzít z katalogizačního záznamu NK ČR - isbn - převzít z katalogizačního záznamu NK ČR - ccnb – čČNB - převzít z katalogizačního záznamu NK ČR - url – permalink záznamu z katalogu NK ČR, např. http://aleph.nkp.cz/F/?func=direct&doc_number=002186258&local_base=NKC - uuid – nutno vygenerovat, POVINNÉ! - jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu		
<code><location></code>		údaje o uložení popisovaného dokumentu, např. signatura, místo uložení apod.	MA	
<code><url></code>		pro uvedení lokace elektronického dokumentu	O	<code><dc:source></code>
<code><physicalLocation></code>		údaje o instituci, kde je fyzicky uložen popisovaný dokument, např. NK ČR nutno použít kontrolovaný slovník – sigly knihoven (ABA001 atd.) odpovídá poli 040 v MARC21 pozn. u dokumentů v digitální podobě není možné vyplnit	M	<code><dc:source></code>
	authority	authority: hodnota „siglaADR“		
<code><shelfLocator></code>		signatura nebo lokační údaje o dokumentu	M	<code><dc:source></code>
<code><recordInfo></code>		údaje o metadatovém záznamu – jeho vzniku,	M	

		změnách apod.		
<recordContentSource >		kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník	R	
<recordCreationDate>		datum prvního vytvoření záznamu, na úroveň minut	M	
	encoding	encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601		
<recordChangeDate>		datum změny záznamu	R	
	encoding	encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601		
<recordOrigin>		údaje o vzniku záznamu hodnoty: machine generated nebo human prepared	R	

Příklad:

```

<mets:dmdSec ID="MODSMD_TITLE_0001">
  <mets:mdWrap MDTYPE="MODS" MIMETYPE="text/xml">
    <mets:xmlData xmlns:mods="http://www.loc.gov/mods/v3">
      <mods:mods ID="MODS_TITLE_0001">
        <mods:titleInfo>
          <mods:title>Chemické listy</mods:title>
        </mods:titleInfo>
        <mods:genre>title</mods:genre>
        <mods:originInfo>
          <mods:place>
            <mods:placeTerm type="text">Praha</mods:placeTerm>
          </mods:place>
          <mods:publisher>Průmyslové vydavatelství</mods:publisher>
          <mods:dateIssued>1951</mods:dateIssued>
          <mods:issuance>continuing</mods:issuance>
        </mods:originInfo>
        <mods:language>
          <mods:languageTerm type="code" authority="iso639-2b">cze
          </mods:languageTerm>
        </mods:language>
        <mods:physicalDescription>
          <mods:form authority="marcform">print</mods:form>
        </mods:physicalDescription>
        <mods:classification authority="udc">54</mods:classification>
        <mods:classification authority="udc">050</mods:classification>
        <mods:identifier type="ccnb">cnb000356352</mods:identifier>
        <mods:identifier type="uuid">b6edddb3-51b6-4703-9585-fb9b9bfa858a
        </mods:identifier>
        <mods:location>
          <mods:physicalLocation authority="siglaADR">ABA001
          </mods:physicalLocation>
          <mods:shelfLocator>54 C 000412</mods:shelfLocator>
          <mods:shelfLocator>Nd 000329</mods:shelfLocator>
        </mods:location>
        <mods:recordInfo>
          <mods:recordCreationDate encoding="iso8601">2011-12-12T00:00:00
          </mods:recordCreationDate>
        </mods:recordInfo>
      </mods:mods>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>

```

8.3.1.2 Pole MODS a DC pro ročník periodika

Element MODS	Atributy	Popis	povinnost	Element DC
<mods>	ID	ID musí vyjadřovat název úrovně, tj. „MODS_VOLUME_0001“	M	
<titleInfo>		Informace o čísle ročníku	M	
<partNumber>		pořadové číslo vydání ročníku, např. 40	MA	<dc:description>
<genre>		Hodnota „volume“	M	<dc:type>
<originInfo>		Informace o původu předlohy	M	
<dateIssued>		datum vydání předlohy, v případě ročníku rok, případně rozsah let, kdy vyšel	M	<dc:date>
		- RRRR – pokud víme rok		
		- RRRR-RRRR – rozsah let		
qualifier	qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj			
<identifier>		údaje o identifikátorech čísla, obsahuje unikátní identifikátory mezinárodní nebo lokální	M	<dc:identifier>
	type	type: budou se povinně vyplňovat následující hodnoty, pokud existují:		
		- urnnbn - pro URN:NBN, např. zápis ve tvaru urn:nbn:cz:anl-123456 pro projekt ANL+; pozor, musí odpovídat URN:NBN, podle kterého je pojmenovaný PSP balíček a jeho jednotlivé soubory		
		- uuid – musí se vygenerovat - POVINNÉ		
- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu				
< physicalDescription >		obsahuje údaje o fyzickém popisu předlohy	O	
<note>		poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element zde se zapíše defekty zjištěné při digitalizaci pro úroveň ročníku (např. chybějící čísla apod.)	O	

Příklad:

```
<mets:dmdSec ID="MODSMD_VOLUME_0001">
  <mets:mdWrap MDTYPE="MODS" MIMETYPE="text/xml">
    <mets:xmlData xmlns:mods="http://www.loc.gov/mods/v3">
      <mods:mods ID="MODS_VOLUME_0001">
        <mods:titleInfo>
          <mods:partNumber>45</mods:partNumber>
        </mods:titleInfo>
        <mods:genre>volume</mods:genre>
        <mods:originInfo>
          <mods:dateIssued>1951</mods:dateIssued>
        </mods:originInfo>
        <mods:identifier type="uuid">b6edddb3-51b6-4703-9585-fb9b9bfa858a</mods:identifier>
      </mods:mods>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>
```

8.3.1.3 Pole MODS a DC pro číslo periodika

Element MODS	Atributy	Popis	povinnost	Element DC
<mods>	ID	ID musí vyjadřovat název úrovně, tj. např. „MODS_ISSUE_0001“	M	
<titleInfo>		název titulu periodika, kterého je číslo součástí, převzít z katalogizačního záznamu titulu periodika použít názvové autority nebo katalogizační záznam	M	
<title>		názvová informace – titul periodika převzít z katalogu	M	<dc:title>
<subTitle>		podnázev periodika	RA	<dc:title>
<partNumber>		pořadové číslo vydání (číslo), např. 40; nebo u ročenek číslo určité řady/edice (část 1, řada B)	MA	<dc:description>
<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	R	<dc:description>
<name>		údaje o odpovědnosti za číslo periodika; nepočítá se s vyplněním u deníků, ale např. u ročenek, zvláštních vydání čísel periodika apod., které mají vlastního autora/editora	MA	
	type	type: použít jeden z typů - personal - corporate - conference - family		
<namePart>		údaje o křestním jméně a příjmení apod. nutno vyjádřit pro křestní jméno i příjmení pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu <namePart>	MA	<dc:creator>
	type	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		nutno do jednoho pole DC spojit jméno i příjmení
<role>		specifikace role osoby nebo organizace uvedené v elementu <name>	MA	

<roleTerm>		popis role nutno použít kontrol. slovník např. z MARC21	MA				
	type	type: code – kód role z kontrolovaného slovníku rolí http://www.loc.gov/marc/relators/relaterm.html					
	authority	authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"					
<genre>		bližší údaje o typu dokumentu hodnota „issue“	M	<dc:type>			
	type	type: pro upřesnění typu čísla a jednotlivých vydání povinné hodnota může být:					
	- normal - běžné vydání	- morning – ranní vydání					
	- afternoon- odpolední vydání	- evening – večerní vydání					
	- sequence_X – pořadí vydání (sequence_1 = první vydání toho dne; sequence_2 = druhé vydání atd.)	- corrected – opravené vydání					
	- special – zvláštní vydání (např. k nějaké události)	- supplement – v případě, že se příloha časopiseckého typu popisuje jako číslo					
	<originInfo>				informace o původu předlohy doporučené kde lze vyplnit (např. u ročenek, kde se vydavatel měnil) nepovinné pro deníky a běžná čísla periodik 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.	MA	
		<place>			údaje o místě spojeném s vydáním, výrobou nebo původem popisovaného dokumentu		
		<placeTerm>			konkrétní určení místa, např. Praha odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „a“		

	type	type – bude vždy text		
<publisher>		jméno entity, která dokument vydala, vytiskla nebo jinak vyprodukovala odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21	MA	<dc:publisher>
<dateIssued>		datum vydání předlohy, v případě čísla datum dne, kdy vyšlo; musí vyjádřit den, měsíc a rok, dle toho jaké údaje jsou k dispozici; nutno zapsat v následujících podobách:	MA	<dc:date>
		- DD.MM.RRRR – pokud víme den, měsíc i rok vydání		
		- MM.RRRR – pokud víme jen měsíc a rok vydání		
		- RRRR – pokud víme pouze rok		
		- DD.-DD.MM.RRRR – vydání pro více dní		
		- MM.-MM.RRRR – vydání pro více měsíců		
	qualifier	qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj		
<dateCreated>		datum vytvoření předlohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo>; odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“	R	
	qualifier	qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj		
<language>		údaje o jazyce dokumentu	M	
<languageTerm>		přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php	M	<dc:language>
	type	type: použít hodnotu code		
	authority	authority: použít hodnotu „iso639-2b“		
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy	R	
<extent>		údaje o rozsahu (stran, svazků nebo rozměrů); použití spíše u ročenek apod. odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat; počet stránek bude vyjádřen ve fyzické strukturální mapě a bude tak vidět v aplikaci zpřístupnění i bez vyplnění tohoto pole	RA	<dc:format>
<note>		poznámka o fyzickém stavu dokumentu;	MA	

		pro každou poznámku je nutno vytvořit nový <note> element zde se zapíše poznámky o defektech zjištěné při digitalizaci pro úroveň čísla (chybějící a poškozené stránky apod.)		
<abstract>		shrnutí obsahu dokumentu, zvláště pro ročenky, zvláštní vydání a tematická čísla plnit pouze v případech, že se liší od abstraktu na úrovni titulu odpovídá poli 520 MARC21	RA	<dc:descripti on>
<note>		obecná poznámka k dokumentu odpovídá poli 500 v MARC21	RA	
<subject>	authority	údaje o věcném třídění plnit pouze pro tematická čísla, zvláštní vydání a ročenky – pouze pokud se liší od údajů v elementu <subject> na úrovni titulu	RA	
	authority	authority: vyplnit hodnotu „czenas“		
<topic>		libovolný výraz specifikující nebo charakterizující obsah čísla; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma)	M	<dc:subject>
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín)	R	<dc:subject>
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj)	R	<dc:subject>
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní)	R	
<namePart>		celé jméno se zapíše do elementu <namePart>, viz top element <name>		<dc:subject>
<identifier>		údaje o identifikátorech čísla, obsahuje unikátní identifikátory mezinárodní nebo lokální	M	<dc:identifier >
	type	type: budou se povinně vyplňovat následující hodnoty, pokud existují:		
		- doi		
		- hdl - handle		
		- isbn - převzít z katalogizačního záznam NK ČR (ročenky apod.)		
	- urnnbn - pro URN:NBN, např. zápis ve tvaru urn:nbn:cz:anl-123456 pro projekt ANL+; pozor, musí odpovídat URN:NBN, podle kterého je pojmenovaný PSP balíček a jeho jednotlivé soubory			
	- uuid – vygenerovat – POVINNÉ			

		- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu		
<location>		údaje o uložení popisovaného dokumentu, např. signatura, místo uložení apod. doporučené např. pro ročenky apod., kde se signatury jednotlivých čísel liší	R	
<url>		pro uvedení lokace elektronického dokumentu	O	<dc:source>
	note	note: pro poznámku o typu URL (na plný text, abstrakt apod.)		
<physicalLocation>		údaje o instituci, kde je fyzicky uložen popisovaný dokument, např. NK ČR nutno použít kontrolovaný slovník – sigly knihoven (ABA001 atd.) odpovídá poli 040 v MARC21	MA	<dc:source>
	authority	authority: hodnota „siglaADR“		
<shelfLocator>		signatura nebo lokační údaje o dokumentu	MA	<dc:source>
<part>		popis částí dokumentu, bude využit jen na zaznamenání <caption>	O	
	type	type: hodnota bude vždy „issue“		
<detail>				
<caption>		text před označením čísla, např. „č.“, „číslo“, „No.“ apod.	RA	

Příklad:

```

<mets:dmdSec ID="MODSMD_ISSUE_0001">
  <mets:mdWrap MDTYPE="MODS" MIMETYPE="text/xml">
    <mets:xmlData xmlns:mods="http://www.loc.gov/mods/v3">
      <mods:mods ID="MODS_ISSUE_0001">
        <mods:titleInfo>
          <mods:title>Chemické listy</mods:title>
          <mods:partNumber>6</mods:partNumber>
        </mods:titleInfo>
        <mods:genre type="normal">issue</mods:genre>
        <mods:originInfo>
          <mods:dateIssued>1.1.1956</mods:dateIssued>
        </mods:originInfo>
        <mods:language>
          <mods:languageTerm type="code" authority="iso639-2b">cze
          </mods:languageTerm>
        </mods:language>
        <mods:identifier type="uuid">fdee51ab-1547-4b9c-bf0e-cb0fdad7e40e
        </mods:identifier>
        <mods:identifier type="urnnbn">urn:nbn:boa001123456</mods:identifier>
      </mods:mods>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>

```

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

Element MODS	Atributy	Popis	Povinnost	Element DC
<mods>	ID	ID musí vyjadřovat název úrovně, tj. např. „MODS_PICT_0001“ pro obrázek v textu, „MODS_ART_0001“ pro článek apod.	M	
<titleInfo>		názvová informace vnitřní části	M	
<title>		vlastní název vnitřní části (článku, obrazu); u obrazu brát případně z popisku obrazu; pokud není titul, nutno vyplnit hodnotu „untitled“	M	<dc:title>
<subTitle>		podnázev vnitřní části (článku); za podnázev lze považovat i krátký text, který se před článkem objevuje tučným písmem (shrnutí obsahu článku)	MA	<dc:title>
<partNumber>		číslo vnitřní části např. článek na pokračování	RA	<dc:title>
<partName>		název pokračování vnitřní části (článku)	RA	<dc:title>
<name>		údaje o odpovědnosti za vnitřní část (článek i obraz)	MA	
	type	type: použít jeden z typů		
		- personal		
		- corporate		
		- conference		
		- family		
<namePart>		údaje o křestním jméně a příjmení apod. nutno vyjádřit pro křestní jméno i příjmení; pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu <namePart>	MA	<dc:creator> nutno do jednoho pole DC spojit jméno i příjmení
	type	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		
<role>		specifikace role osoby nebo organizace uvedené v elementu <name>	RA	
<roleTerm>		popis role nutno použít kontrol. slovník např. z MARC21	MA	

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

	type	type: code – kód role z kontrolovaného slovníku rolí http://www.loc.gov/marc/relators/relaterm.html		
	authority	authority – údaje o kontrolovaném slovníku využitém k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"		
<genre>		bližší údaje o typu vnitřní části povinné hodnota: article nebo picture type: doporučené	M	<dc:type>
	type	<p>hodnota pro article – možnost vyplnit bližší určení typu článku (možnost použít DTD periodika, Article Types)</p> <ul style="list-style-type: none"> - news - table of content - advertisement - abstract - introduction - review - dedication - bibliography - editorsNote - preface - main article - index (použije se pro všechny typy seznamů mimo hlavní obsah; např. seznam obrazů, tabulek apod.) - unspecified – pokud nepatří ani do jedné z výše uvedených kategorií - aj. <p>hodnota pro picture – možnost vyplnit další určení typu obrazu</p> <ul style="list-style-type: none"> - table - illustration <ul style="list-style-type: none"> - chart - photograph - graphic - map - advertisement - cover - unspecified – pokud nepatří ani do jedné z výše uvedených kategorií - aj. 		
<language>		údaje o jazyce vnitřní části	MA	

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

		nelze plnit u obrazu		
<languageTerm>		přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php nelze plnit u obrazu	M	<dc:language>
	type	type: použít hodnotu code		
	authority	authority: použít hodnotu „iso639-2b“		
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy; určeno spíše pro články než pro obrazy	R	
<form>		údaje o fyzické podobě vnitřní části, např. print, electronic apod. odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21	R	<dc:format>
	type	type: použít hodnotu code		
<abstract>		shrnutí obsahu vnitřní části	R	<dc:descripti on>
<note>		obecná poznámka k vnitřní části do poznámky by se měla dávat šifra autora vnitřní části, která se vyskytuje pod vnitřní částí odpovídá poli 500 v MARC21	RA	<dc:descripti on>
<subject>		údaje o věcném třídění	R	
<topic>		libovolný výraz specifikující nebo charakterizující obsah vnitřní části; lze (není ovšem nutno) použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma)	M	<dc:subject>
	authority (volitelné)	při použití autoritních záznamů použít AUT NK ČR a atribut authority: vyplnit hodnotu „czenas“; při použití volných klíčových slov atribut authority nepoužívat		
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín)	R	<dc:subject>
	authority	authority: vyplnit hodnotu „czenas“		
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj)	R	<dc:subject>
	authority	authority: vyplnit hodnotu „czenas“		
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní)	R	
	authority	authority: vyplnit hodnotu „czenas“		
<namePart>		vyplní se obdobně jako top element <name> -	R	<dc:subject>

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

		celé jméno autora do tohoto elementu			
<classification>		klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění plnit pouze pro článek odpovídá poli 080 MARC21	RA	<dc:subject>	
	authority	authority: vyplnit hodnotu „udc“			
<identifier>		údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které vnitřní část má – viz přehled typů atributů níže	M	<dc:identifier > povinné	
	type	type: budou se povinně vyplňovat následující hodnoty, pokud existují pro článek nebo obraz:			
		- doi			
		- hdl - handle			
		- urnnbn - pro URN:NBN			
		- uuid – generuje se – POVINNÉ			
- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu					
<part>		popis částí vnitřní části, bude využito na záznam rozsahu nelze u obrazu	RA		
<extent>		upřesnění popisu části – rozsah na stránkách	MA	<dc:format>	
<start>		první stránka, na které vnitřní část začíná	MA	<dc:coverage >	
<end>		poslední stránka, na které vnitřní část končí	MA	<dc:coverage >	
<recordInfo>		údaje o metadatovém záznamu vnitřní části – jeho vzniku, změnách apod.	M		
<recordContentSource >		kód nebo jméno instituce, která záznam vytvořila nebo změnila; nutno vytvořit kontrolovaný slovník	R		
<recordCreationDate>		datum prvního vytvoření záznamu vnitřní části	M		
	encoding	encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601			
<recordChangeDate>		datum změny záznamu vnitřní části	R		
	encoding	encoding: záznam bude podle normy ISO 8601 na úroveň minut, hodnota atributu tedy iso8601			
<recordOrigin>		údaje o vzniku záznamu vnitřní části hodnoty: machine generated nebo human prepared	R		

Příklad:

```

<mods:mods ID="MODS_ARTICLE_0001">
  <mods:titleInfo>
    <mods:title>Autoři a redaktoři (slovo do vlastních řad)</mods:title>
  </mods:titleInfo>
  <mods:name type="personal">
    <mods:namePart type="given">Bohumil</mods:namePart>
    <mods:namePart type="family">Kratochvíl</mods:namePart>
    <mods:role>
      <mods:roleTerm type="code" authority="marcrelator">aut</mods:roleTerm>
    </mods:role>
  </mods:name>
  <mods:genre>article</mods:genre>
  <mods:language>
    <mods:languageTerm type="code" authority="iso639-2b">cze</mods:languageTerm>
  </mods:language>
  <mods:identifier type="uuid">53779de4-1c10-40e5-a6ee-581796cb9a8c
</mods:identifier>
  <mods:part>
    <mods:extent>
      <mods:start>437</mods:start>
      <mods:end>437</mods:end>
    </mods:extent>
  </mods:part>
  <mods:recordInfo>
    <mods:recordContentSource>Elsyst Engineering</mods:recordContentSource>
    <mods:recordCreationDate encoding="iso8601">2012-01-13T06:28
</mods:recordCreationDate>
    <mods:recordOrigin>human prepared</mods:recordOrigin>
  </mods:recordInfo>
</mods:mods>

```

8.3.1.5 Pole MODS a DC pro přílohu

Element MODS	Atributy	Popis	Povinnost	Element DC
<mods>	ID	ID musí vyjadřovat název úrovně, tj. „MODS_SUPPL_0001“	M	
<titleInfo>		názvová informace přílohy použít názvové autority nebo katalogizační záznam	M	
<title>		názvová informace – název periodika, jehož součástí příloha je převzít z katalogu	M	<dc:title>
<partNumber>		číslo přílohy, pokud nějaké má doporučené pokud lze vyplnit	MA	<dc:description>
<partName>		název přílohy	MA	<dc:title>
<name>		údaje o odpovědnosti za přílohu	MA	
	type	type: použít jeden z typů: - personal - corporate - conference - family		
<namePart>		údaje o křestním jméně a příjmení apod. nutno vyjádřit pro křestní jméno i příjmení pokud nelze rozlišit křestní jméno a příjmení, nepoužije se type a jméno se zaznamená v podobě jaké je do jednoho elementu <namePart>	MA	<dc:creator> nutno do jednoho pole DC spojit jméno i příjmení
	type	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		
<role>		specifikace role osoby nebo organizace uvedené v elementu <name>	MA	
<roleTerm>		popis role nutno použít kontrol. slovník např. z MARC21	MA	
	type	type: code – kód role z kontrolovaného slovníku rolí http://www.loc.gov/marc/relators/relaterm.html		
	authority	authority – údaje o kontrolovaném slovníku		

		využitím k popisu role, k popisu výše uvedeného MARC seznamu nutno uvést authority="marcrelator"		
<typeOfResource>		popis charakteristiky typu nebo obsahu přílohy jedna z hodnot:	R	<dc:type>
		- text – např. pro přílohu typu časopis, kniha, brožura apod.		
		- cartographic – pro mapy		
		- notated music		
		- sound recording-musical - pro hudební CD/DVD		
		- sound recording-nonmusical		
		- sound recording		
		- still image – fotografie, plakáty apod.		
		- moving image – pro filmová DVD		
		- three dimensional object		
		- software, multimedia – pro CD/DVD se SW		
- mixed material				
<genre>		bližší údaje o typu dokumentu hodnota:	M	<dc:type>
		- volume_supplement (příloha k ročníku, např. obsah celého ročníku)		
		- issue_supplement (příloha k číslu)		
<originInfo>		informace o původu přílohy plnit pokud se liší od údajů v popisu čísla (platí i pro jednotlivé sub-elementy) 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.	MA	
<place>		údaje o místě spojeném s vydáním, výrobou nebo původem přílohy	MA	<dc:coverage>
<placeTerm>		konkrétní určení místa, např. Praha odpovídá hodnotě katalogizačního záznamu, pole 260, podpole „a“	MA	<dc:coverage>
	type	type – bude vždy text		
<publisher>		jméno entity, která přílohu vydala, vytiskla nebo	MA	<dc:publisher>

		jinak vyprodukovala odpovídá poli 260 podpoli „b“ katalogizačního záznamu v MARC21		
<dateIssued>		datum vydání přílohy, musí vyjádřit den, měsíc a rok, dle toho jaké údaje jsou k dispozici nutno zapsat v následujících podobách: - DD.MM.RRRR – pokud víme den, měsíc i rok vydání - MM.RRRR – pokud víme jen měsíc a rok vydání - RRRR – pokud víme pouze rok - DD.-DD.MM.RRRR – vydání pro více dní - MM.-MM.RRRR – vydání pro více měsíců možno použít hodnotu z katalogizačního záznamu, pole 260, podpole „c“	M	<dc:date>
	qualifier	qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj		
<dateCreated>		datum vytvoření přílohy bude použito pouze při popisu tiskaře, viz poznámka u elementu <originInfo> nebo např. u popisu CD/DVD apod. odpovídá hodnotě z katalogizačního záznamu, pole 260, podpole „g“	R	
	qualifier	qualifier – možnost dalšího upřesnění, hodnota „approximate“ pro data, kde nevíme přesný údaj		
<frequency>		údaje o pravidelnosti vydávání odpovídá údajům MARC21 v poli 310 nebo pozici 18 v poli 008	RA	
<language>		údaje o jazyce dokumentu	M	
<languageTerm>		přesné určení jazyka – kódem nutno použít kontrolovaný slovník ISO 639-2, http://www.loc.gov/standards/iso639-2/php/code_list.php	M	<dc:language>
	type	type: použít hodnotu code		
	authority	authority: použít hodnotu „iso639-2b“		
<physicalDescription>		obsahuje údaje o fyzickém popisu zdroje/předlohy	M	
<form>		údaje o fyzické podobě dokumentu, např. print, electronic apod. povinné	M	<dc:format>

		pro tištěné předlohy hodnota „print“, pro elektronické přílohy „electronic“ odpovídá hodnotám pozice 23 a 29 v poli 008 MARC21		
	authority	authority: hodnota „marcform“		
<extent>		údaje o rozsahu (stran, svazků nebo rozměrů) odpovídá hodnotám v poli 300 podpolích „a“ a „c“ MARC21, pokud jsou vyplněna obě pole, bude se element <extent> opakovat	RA	<dc:format>
<note>		poznámka o fyzickém stavu dokumentu; pro každou poznámku je nutno vytvořit nový <note> element zde se zapíše poznámky o defektech zjištěné při digitalizaci pro úroveň přílohy jako samostatného čísla (chybějící a poškozené stránky apod.)	RA	
<abstract>		shrnutí obsahu dokumentu odpovídá poli 520 MARC21	RA	<dc:descriptio n>
<note>		obecná poznámka k dokumentu odpovídá poli 500 v MARC21	RA	<dc:descriptio n>
<subject>		údaje o věcném třídění	R	
	authority	authority: vyplnit hodnotu „czenas“		
<topic>		libovolný výraz specifikující nebo charakterizující obsah přílohy; použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (věcné téma)	M	<dc:subject>
<geographic>		geografické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (geografický termín)	R	<dc:subject>
<temporal>		chronologické věcné třídění použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (chronologický údaj)	R	<dc:subject>
<name>		jméno použité jako věcné záhlaví použít kontrolovaný slovník - např. z báze autorit AUT NK ČR (jméno osobní)	R	<dc:subject>
<namePart>		celé jméno se zapíše do elementu <namePart>, viz top element <name>		
<classification>		klasifikační údaje věcného třídění podle Mezinárodního desetinného třídění odpovídá poli 080 MARC21	M	<dc:subject>
	authority	authority: vyplnit hodnotu „udc“		
<identifier>		údaje o identifikátorech, obsahuje unikátní identifikátory mezinárodní nebo lokální, které příloha má – viz přehled typů atributů níže	MA	<dc:identifier>
	type	type: budou se povinně vyplňovat následující		

		hodnoty, pokud existují:		
		- doi		
		- hdl - handle		
		- issn - převzít z katalogizačního záznam NK ČR		
		- isbn - převzít z katalogizačního záznam NK ČR		
		- ccnb – čČNB - převzít z katalogizačního záznam NK ČR		
		- permalink záznamu z katalogu NK ČR, např. http://aleph.nkp.cz/F/?func=direct&doc_number=002186258&local_base=NKC		
		- urnnbn - pro URN:NBN		
		- uuid		
		- jiný interní identifikátor, hodnota atributu „local“, lze použít např. k vyjádření čárového kódu		

8.4 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	Povinnost
<amdSec>		element obsahující technická metadata ve formátu PREMIS nebo MIX	M
	ID	ID – identifikátor konkrétní části <amdSec>, např. pro stránku 1 by hodnota mohla být „PAGE0001“	
<techMD> nebo <digiprovMD>		element rozlišující typy jednotlivých administrativních metadat	M
	ID	pro část <techMD>:	
		- pro části obsahující PREMIS-object hodnota „OBJ_001“ – objekt 1 (PREMIS object pro smazaný TIFF, OBJ_002 by bylo pro MC, OBJ_003 pro ALTO XML	
		- pro části obsahující MIX hodnota „MIX_001“ = MIX metadata pro původní TIFF, „MIX_002“ pro MC	
		pro část <digiprovMD>:	
	- pro části obsahující PREMIS-event hodnota „EVT_001“ apod.		
	- pro části obsahující PREMIS-agent hodnota „AGENT_001“ apod.		
<mdWrap>		element obsahující vložené záznamy PREMIS, MIX	
	MDTYPE	MDTYPE - pro záznamy PREMIS object, event i agent vždy hodnota „PREMIS“ - pro záznamy MIX hodnota „NISOIMG“	M

Příklad:

```
<mets:amdSec ID="PAGE0001">
  <mets:techMD ID="OBJ_001">
```

```

<mets:mdWrap MDTYPE="PREMIS" MIMETYPE="text/xml">
  <mets:xmlData>
    ...
  </mets:xmlData>
</mets:mdWrap>
</mets:techMD>
<mets:techMD ID="MIX_001">
  <mets:mdWrap MDTYPE="NISOIMG" MIMETYPE="text/xml">
    <mets:xmlData>
      ...
    </mets:xmlData>
  </mets:mdWrap>
</mets:techMD>
<mets:digiprovMD ID="EVT_001">
  <mets:mdWrap MDTYPE="PREMIS" MIMETYPE="text/xml">
    <mets:xmlData>
      ...
    </mets:xmlData>
  </mets:mdWrap>
</mets:digiprovMD>
<mets:digiprovMD ID="AGENT_001">
  <mets:mdWrap MDTYPE="PREMIS" MIMETYPE="text/xml">
    <mets:xmlData>
      ...
    </mets:xmlData>
  </mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>

```

- 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.)**
-

8.4.1 PREMIS Object

- bude odpovídat poslední aktuální verzi v době implementace (březen 2012 - 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 ani pro UC
- pro každý záznam PREMIS object bude existovat vlastní podčást <techMD>
- záznam PREMIS Object pro jeden soubor bude obsahovat linky na události, které jsou popsány v PREMIS Events ve stejném METS metadatovém záznamu konkrétního dokumentu (číslo, svazek) v části <digiprovdMD>; přes <premis:relatedEventIdentification>, to samé platí pro objekty, které budou nalinkovány v případě vztahu (např. UC vznikla z MC) s popisovaným objektem přes <premis:relatedObjectIdentification>.
 - o tj. např. PREMIS object popisující archivní soubor JPEG2000 je tímto způsobem nalinkován na původní sken ve formátu TIFF (resp. na jeho PREMIS object záznam) – pomocí tagu <relatedObjectIdentification>, který obsahuje ID původního objektu (např. TIFF)
 - o zároveň pomocí tagu <relatedEventIdentification> je záznam PREMIS object archivního souboru JPEG2000 nalinkován na událost, během které vznikl
- **POZOR – Premis Object bude vznikat a uchovávat se i pro neexistující data (původní a posléze smazaný TIFF)**

Pole záznamu PREMIS Object

Obsah pole „Popis“:

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

význam pole „povinnost“

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

Obsah pole „Použití pro“

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

Element	Popis	Povinnost	Použití pro
<objectIdentifier>	identifikátor k jednoznačnému odlišení objektu v určitém kontextu; 1-n	M	MC, XML, PS
<objectIdentifierType>	popis kontextu, ve kterém je identifikátor unikátní, např. NDK, ANL nebo název repozitáře; nutno použít kontrolovaný slovník; 1-1	M	MC, XML, PS
<objectIdentifierValue>	vlastní hodnota identifikátoru, např. img0001-master, urn.nbn.cz-123465 apod.; 1-1	M	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	M	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	M	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	M	MC, XML, PS
<preservationLevelDateAssigned>	datum, kdy byla přiřazena hodnota úrovně ochrany, zápis v ISO 8601, na úroveň dne (DD-MM-RRRR) 0-1	R	MC, XML, PS
<objectCharacteristics>	technické údaje o souboru 1-n	M	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	M	MC, XML, PS
<fixity>	údaje o kontrolním součtu 0-n	M	MC, XML, PS
<messageDigestAlgorithm>	použitý algoritmus kontrolního součtu, např. MD5 aj.	M	MC, XML,

	1-1		PS
<messageDigest>	hodnota kontrolního součtu 1-1	M	MC, XML, PS
<messageDigestOriginator>	agent (osoba, instituce, stroj, SW), který kontrolní součet vytvořil (např. JHOVE apod.) 0-1	M	MC, XML, PS
<size>	údaje o velikosti souboru v bytech 0-1	M	MC, XML, PS
<format>	údaje o formátu souboru 1-n	M	MC, XML, PS
<formatDesignation>	identifikace formátu souboru, výstup z JHOVE, PRONOM služeb apod. 0-1	M	MC, XML, PS
<formatName>	jméno formátu, např. image/tiff nebo Adobe PDF 1-1	M	MC, XML, PS
<formatVersion>	verze formátu, např. 6.0 0-1	M	MC, XML, PS
<formatRegistry>	identifikace formátu – dodatečná informace o záznamu formátů v registrech formátů (např. PRONOM aj.) 0-1	M	MC, XML, PS
<formatRegistryName>	jméno použitého registru formátů, např. UDFR, PRONOM aj. 1-1	M	MC, XML, PS
<formatRegistryKey>	unikátní identifikátor (označení) formátu v registru, např. fmt/155 z PRONOM 1-1	M	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	M	MC, XML, PS
<creatingApplicationName>	název aplikace, např. ImageGear, Kakadu apod.; 0-1	M	MC, XML, PS
<creatingApplicationVersion>	verze aplikace, např. 15.03.000 0-1	M	MC, XML, PS
<dateCreatedByApplication>	datum a čas vytvoření, např. 2008-11-10T12:37:46; musí být ve tvaru ISO 8601 (na úroveň vteřin);	M	MC, XML, PS

	0-1		
<originalName>	původní jméno souboru , např. digibok_2007081301091_0011.jp2 0-1	M	MC, XML, PS
<relationship>	vyjádření vztahu popisovaného souboru k jiným souborům a událostem (eventům) 0-n	M	MC, XML
<relationshipType>	typ vztahu, doporučené hodnoty: derivation= vztah kde objekt je výsledkem změny jiného objektu; structural= vztah mezi částmi objektu; tj. např. ALTO vytvořené z TIFFU bude mít vztah derivation, podobně jako JPEG2000 z TIFFu vytvořený; 1-1	M	MC, XML;
<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	M	MC, XML;
<relatedObjectIdentification>	identifikace souvisejícího souboru 1-n pro MC, XML pro vyjádření vztahu k původnímu objektu (skenu)	M	MC, XML
<relatedObjectIdentifierType>	specifikace kontextu, ve kterém je identifikátor souboru jedinečný, např. URN; temporary filepath; objectID 1-1	M	MC, XML
<relatedObjectIdentifierValue>	vlastní řetězec identifikátoru, např. URN:NBN:cz-1301091_011#0001 nebo název souboru, cesta k souboru apod. 1-1	M	MC, XML
<relatedEventIdentification>	identifikace s popisovaným souborem související události (eventu); seznam událostí viz PREMIS event 0-n	M	MC, XML
<relatedEventIdentifierType>	typ události, např. interní číslovací systém událostí jako no.nb.evt; NK repository event ID, UUID apod. 1-1	M	MC, XML
<relatedEventIdentifierValue>	hodnota identifikátoru události, např. NK_EVT_005 nebo hodnota UUID aj. 1-1	M	MC, XML
<relatedEventSequence>	pořadí události, např. 003; k určení pořadí lze	R	MC,

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

Příklad (PREMIS pro smazaný původní sken):

```
<premis:object xsi:type="premis:file" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:premis="info:lc/xmlns/premis-v2">
  <premis:objectIdentifier>

<premis:objectIdentifierType>EE_Sirius_NDK_FileID</premis:objectIdentifierType>
  <premis:objectIdentifierValue>614</premis:objectIdentifierValue>
</premis:objectIdentifier>
<premis:objectCategory>file</premis:objectCategory>
<premis:preservationLevel>
  <premis:preservationLevelValue>deleted</premis:preservationLevelValue>
  <premis:preservationLevelDateAssigned>2012-01-
26T13:49:25</premis:preservationLevelDateAssigned>
</premis:preservationLevel>
<premis:objectCharacteristics>
  <premis:compositionLevel>0</premis:compositionLevel>
  <premis:fixity>
    <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>

<premis:messageDigest>437d2c0462dfe2fb276bf0e4f86eeeal</premis:messageDigest>
    <premis:messageDigestOriginator>Elsyst
Engineering</premis:messageDigestOriginator>
  </premis:fixity>
  <premis:size>19535558</premis:size>
  <premis:format>
    <premis:formatDesignation>
      <premis:formatName>image/tiff</premis:formatName>
      <premis:formatVersion>42</premis:formatVersion>
    </premis:formatDesignation>
    <premis:formatRegistry>
      <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
      <premis:formatRegistryKey>x-cmp/12</premis:formatRegistryKey>
    </premis:formatRegistry>
  </premis:format>
  <premis:creatingApplication>
    <premis:creatingApplicationName>INSERTDOCUMENTSTOSIRIUS
    </premis:creatingApplicationName>
    <premis:creatingApplicationVersion>1.0.0.0
    </premis:creatingApplicationVersion>
    <premis:dateCreatedByApplication>2011-11-30T07:22:00
    </premis:dateCreatedByApplication>
```

```
</premis:creatingApplication>
</premis:objectCharacteristics>
<premis:originalName>D:\Test-Virtual\Vystup\ivo_import\nahledy\B300C__161289.tif
</premis:originalName>
<premis:linkingEventIdentifier>
  <premis:linkingEventIdentifierType>EE_Sirius_NDK_ANLs_LoggingID
  </premis:linkingEventIdentifierType>
  <premis:linkingEventIdentifierValue>1700</premis:linkingEventIdentifierValue>
</premis:linkingEventIdentifier>
</premis:object>
```

8.4.2 PREMIS Event

- bude odpovídat poslední aktuální verzi v době implementace (březen 2012 - 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

význam pole „povinnost“

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

Pole záznamu PREMIS Event

Element	Popis	Povinnost
<eventIdentifier>	údaje o identifikátoru události v kontextu digitalizace nebo repozitáře 1-1	M

<eventIdentifierType>	typ identifikátoru, např. no.nb.evt; NK_eventID, UUID apod. 1-1	M
<eventIdentifierValue>	hodnota identifikátoru, např. EVT_001; event_019 apod. 1-1	M
<eventType>	kategorizace události, nutno použít kontrolovaný slovník; typy událostí, které musí být zaznamenány: capture, migration, derivation, deletion 1-1	M
<eventDateTime>	datum a čas kdy byla událost provedena; nutno zapsat v ISO 8601 na úroveň vteřin 1-1	M
<eventDetail>	další údaje o události, doporučené hodnoty pro výše uvedené <eventType> následují za /: <ul style="list-style-type: none"> - capture/digitization – vznik prvního skenu - capture/XML_creation - capture/TXT_creation - migration/MC_creation - derivation/UC_creation - deletion/PS_deletion 0-1	M
<eventOutcomeInformation>	informace o výsledku události 0-n	R
<eventOutcome>	kategorizace výsledku události, např. slovy jako successful nebo failure, možno použít kódy – nutno používat kontrolovaný slovník nebo seznam kódů 0-1	M
<linkingAgentIdentifier>	identifikace jednoho nebo více agentů spojených s událostí 0-n	M
<linkingAgentIdentifierType>	označení typu identifikátoru, např. NK_AgentID, UUID apod. 1-1	M
<linkingAgentIdentifierValue>	hodnota identifikátoru, např. agent_softwareName_5.2; agent_novakJ apod. 1-1	M
<linkingAgentRole>	role agenta ve vztahu k události, např. software; SW component; operator; nutno používat kontrolovaný slovník 0-n	R
<linkingObjectIdentifier>	informace o objektu/souboru spojeného s událostí, link na něj 0-n	M
<linkingObjectIdentifierType>	označení typu identifikátoru, např. PhysUnitID; URN, NK_OBJ, OBJ_001 apod.; hodnoty by se měly brát	M

	z kontrolovaného slovníku 1-1	
<linkingObjectIdentifierValue>	hodnota identifikátoru, např. URN:NBN:cz- _0011#0001 aj. 1-1	M

Příklad (Událost – pořízení původního skenu):

```
<premis:event xmlns:premis="info:lc/xmlns/premis-v2">
  <premis:eventIdentifier>
    <premis:eventIdentifierType>EE_Sirius_NDK_ LoggingID
    </premis:eventIdentifierType>
    <premis:eventIdentifierValue>1700</premis:eventIdentifierValue>
  </premis:eventIdentifier>
  <premis:eventType>creation</premis:eventType>
  <premis:eventDateTime>2011-11-30T08:24:34</premis:eventDateTime>
  <premis:eventDetail>capture/digitization</premis:eventDetail>
  <premis:eventOutcomeInformation>
    <premis:eventOutcome>successful</premis:eventOutcome>
  </premis:eventOutcomeInformation>
  <premis:linkingAgentIdentifier>
    <premis:linkingAgentIdentifierType>EE_Sirius_NDK_UserID
    </premis:linkingAgentIdentifierType>
    <premis:linkingAgentIdentifierValue>1</premis:linkingAgentIdentifierValue>
    <premis:linkingAgentRole>operator</premis:linkingAgentRole>
  </premis:linkingAgentIdentifier>
  <premis:linkingObjectIdentifier>
    <premis:linkingObjectIdentifierType>EE_Sirius_NDK_ FileID
    </premis:linkingObjectIdentifierType>
    <premis:linkingObjectIdentifierValue>614</premis:linkingObjectIdentifierValue>
  </premis:linkingObjectIdentifier>
</premis:event>
```

8.4.3 PREMIS Agent

- bude odpovídat poslední aktuální verzi v době implementace (březen 2012 - PREMIS data dictionary v. 2.1), nebo verzi předchozí
- **využití PREMIS agent je spíše myšleno pro tzv. ochranné aktivity, které probíhají na archivních datech (AIP balíček) a je nutné pro každou událost na těchto datech mít přesnější informace o tom, kdo ji provedl (osoba administrátora nebo oprávněné osoby)**
 - o informace v PREMIS event a PREMIS object přicházející z procesu digitalizace v PSP balíčku jsou dostačující a dají nám dostatečné informace o události, kdy byla provedena, na jakém SW byla provedena (PREMIS object „creatingApplication“ + PREMIS event „eventDetail“ – tj. další upřesnění v PREMIS agent není nutné
- záznam PREMIS agent obsahuje charakteristiku tzv. agenta, který je spojen s provedenou a zaznamenanou událostí (PREMIS event)
 - o agent může být osoba, organizace nebo software
- z PREMIS Event je linkováno na agenta, který určitou akci provedl, typ ID agenta a jeho hodnota jsou uvedené v Premis Events (<premis:linkingAgentIdentifier>), plný popis agenta je pak v PREMIS Agent
- záznamy PREMIS agent budou uloženy v METS záznamu určeném pro administrativní a technická metadata (AMD_METS.xml) v jeho části <amdSec>, podčást <digiprovMD>
 - o AMD_METS.xml je linkován z hlavního METS záznamu dokumentu
- pro každého agenta, tj. jeden PREMIS agent záznam, bude vytvořena jedna <digiprovMD> část

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

význam pole „Povinnost“

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

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

Příklad:

```
<premis:agent xmlns:premis="info:lc/xmlns/premis-v2">
  <premis:agentIdentifier>
    <premis:agentIdentifierType>EE_App_Name</premis:agentIdentifierType>
    <premis:agentIdentifierValue>agent_Sirius</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentName>BATCHPROCESSOR</premis:agentName>
  <premis:agentType>software</premis:agentType>
  <premis:agentNote>C:\Program Files\Elsyst
Engineering\BatchProcessor\Utilities\image_to_j2k.exe -i
"C:\SiriusBatchProcessor\skenserver_Sirius_NDK_ANL\B300C__161289.tif" -o
"C:\SiriusBatchProcessor\skenserver_Sirius_NDK_ANL\B300C__161289.jp2" -c
[256,256],[256,256],[128,128],[128,128],[128,128],[128,128] -p RPCL -n 6 -t
4096,4096 -b 64,64 -SOP -EPH</premis:agentNote>
</premis:agent>
```

8.4.4 Technická metadata MIX

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

Pole formátu MIX pro popis archivní kopie a původního skenu

Obsah pole „Popis“:

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

Význam pole „povinnost“

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

Obsah pole „Použití pro“

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

Element	Popis	Povinnost	Použití pro
<BasicDigitalObjectInformation>			
<ObjectIdentifier>	údaje o identifikátoru obrazového dokumentu, který je formátem MIX popsán; 0-n	R	MC, PS
<objectIdentifierType>	např. jméno souboru, nebo jiný identifikátor; 0-1	M	MC, PS
<objectIdentifierValue>	hodnota identifikátoru, např. 20110306_001.jp2 nebo urn:nbn:123456; 0-1	M	MC, PS
<fileSize>	velikost souboru 0-1	R	MC + PS
<FormatDesignation>	údaje o formátu obrazového souboru 0-1	M	MC, PS
<formatName>	název formátu, např. lze využít MIME types ⁵ (Image/jp2 apod.) 0-1	M	MC, PS
<formatVersion>	verze formátu, např. 1.0 0-1	M	MC, PS
<byteOrder>	endianita, možnosti jsou little endian, middle (mix) endian a big endian 0-1	M	MC + PS
<Compression>	údaje o kompresi obrazového souboru (pokud 0-n	M	MC, PS
<compressionScheme>	informace o kompresním schématu, vyjádřeno číslem (např. 34712 je komprese JPEG2000) nebo slovy (např. JP2 Lossless) 0-1	M	MC, PS
<BasicImageInformation>	základní technické údaje o obrazovém dokumentu 0-1	M	MC, PS
<BasicImageCharacteristics>	0-1	M	MC, PS
<imageWidth>	šířka obrazu v pixelech, např. 3987 0-1	M	MC, PS
<imageHeight>	výška obrazu v pixelech, např. 2345 0-1	M	MC, PS
<PhotometricInterpretation>	photometrická interpretace 0-1	M	MC, PS
<colorSpace>	barevný prostor, např. RGB 0-1	M	MC, PS
<ColorProfile>	údaje o barevném profilu 0-1	MA	MC + PS

⁵ <http://www.iana.org/assignments/media-types/index.html>

	povinné pro dokumenty, kde je nutno uchovat přesnou reprezentaci barvy původního dokumentu a používá se ICC profil)		
<IccProfile>	ICC profil 0-1	M	MC + PS
<iccProfileName>	jméno profilu, např. sRGB, Adobe RGB aj. 0-1	M	MC + PS
<iccProfileVersion>	verze profilu, např. sRGB IEC61966-2.1 0-1	M	MC + PS
<iccProfileURI>	odkaz na profil, např. www.profil.cz/sRGB_v4_ICC_pref.icc ; 0-1	R	MC + PS
<SpecialFormatCharacteristics>	speciální technické údaje o obrazovém dokumentu, použití pro formát JPEG2000 0-1 povinný pro JPEG2000	MA	MC
<JPEG2000>	0-1	M	MC
<CodecCompliance>	údaje o kodeku 0-1	M	MC
<codec>	název kodeku, např. Kakadu, LuraWave aj. 0-1	M	MC
<codecVersion>	verze kodeku, např. 3.1 0-1	M	MC
< codestreamProfile >	popis codestream profilu JPEG2000, např. P0 a P1 (viz ISO/IEC 15444-4); 0-1	M	MC
< complianceClass >	specifikace největší výšky, šířky a počtu komponentů, které dekodér dokáže dekodovat, lze použít hodnoty C0, C1 a C2; 0-1	M	MC
<EncodingOptions >	obsahuje informace o kodování JPEG2000 0-1	M	MC
<Tiles >	popis pixelové velikosti dlaždic formátu JPEG2000 0-1	M	MC
< tileWidth>	šířka dlaždice, např. 128 0-1	M	MC
< tileHeight>	výška dlaždice, např. 128 0-1	M	MC
< qualityLayers>	číselná hodnota počtu vrstev, do kterých byl JPEG2000 rozdělen, např. 12 0-1	M	MC
< resolutionLevels>	popis počtu nižších rozlišení, které lze z obrazu získat, např. 6 0-1	M	MC
< ImageCaptureMetadata>	popis procesu skenování, je důležité vyplnit, protože tyto údaje nelze zjistit z finálního	M	PS

	master/archivního souboru 0-1		
<SourceInformation>	informace o předloze 0-1	R	PS
<sourceType>	Book, Newspaper aj.; nutno používat kontrolovaný slovník 0-1	M	PS
<SourceID>	identifikátor předlohy 0-n	R	PS
<sourceIDType>	typ identifikátoru, např. čČNB, URN:NBN 0-1	M	PS
<sourceIDValue>	vlastní hodnota identifikátoru 0-1 povinné	M	PS
<GeneralCaptureInformation>	základní údaje o skenování 0-1	M	PS
<dateTimeCreated>	údaj o datu a čase skenování, např. 2009-01-03T08:25:28; zapsat v ISO 8601 na úroveň vteřin 0-1	M	PS
<imageProducer>	entita provádějící skenování, např. The National Library of the Czech Republic, osoba apod. 0-1	M	PS
<captureDevice>	typ skenovacího zařízení, např. reflection print scanner; doporučené využívání hodnot z kontrolovaného slovníku 0-1	M	PS
<ScannerCapture>	údaje o skeneru 0-1	M	PS
<scannerManufacturer>	výrobce skeneru, např. 4DigitalBooks, Treventus, Zeutschel 0-1	M	PS
<ScannerModel>	údaje o konkrétním typu skeneru 0-1	M	PS
<scannerModelName>	jméno modelové řady skeneru, např. DL 0-1	M	PS
<scannerModelNumber>	číslo/označení modelu, např. 3000 0-1	M	PS
<scannerModelSerialNo>	výrobní číslo skeneru, např. E4R0003649 0-1	M	PS
<MaximumOpticalResolution>	údaje o maximálním optickém rozlišení skeneru 0-1	M	PS
<xOpticalResolution>	optické rozlišení na ose x, např. 300 0-1	M	PS
<yOpticalResolution>	optické rozlišení na ose y, např. 300 0-1	M	PS

< opticalResolutionUnit>	jednotka optického rozlišení, např. inch (in.) 0-1	M	PS
<scannerSensor>	popis typu snímacího senzoru skenovacího zařízení, např. matrix, linear, undefined aj. 0-1	M	PS
<ScanningSystemSoftware>	údaje o softwaru skenovacího zařízení 0-1	M	PS
<scanningSoftwareName>	název softwaru, např. Copinet 0-1	M	PS
<scanningSoftwareVersionNo>	číslo verze softwaru, např. 3.7 0-1	M	PS
<DigitalCameraCapture>	údaje o snímacím zařízení (fotoaparát) 0-1 povinné, pokud je používán fotoaparát a není používán skener	MA	PS
<digitalCameraManufacturer>	výrobce fotoaparátu, např. Canon 0-1	M	PS
<DigitalCameraModel>	popis modelu fotoaparátu 0-1	M	PS
<digitalCameraModelName>	název modelové řady, např. EOS 0-1	M	PS
< digitalCameraModelNumber>	označení modelu fotoaparátu, např. 1000D 0-1	M	PS
< digitalCameraModelSerialNo>	výrobní číslo přístroje, např. E12345 0-1	M	PS
<camerarSensor>	typ senzoru fotoaparátu, např. matrix aj. 0-1	M	PS
<CameraCaptureSettings>	údaje o nastavení fotoaparátu použitého ke snímání předloh 0-1	M	PS
<ImageData>	v rámci tohoto kontejnerového elementu budou použity následující sub-elementy: fNumber exposureTime isoSpeedRatings shutterSpeedValue apertureValue brightnessValue exposureBiasValue maxApertureValue subjectDistance meteringMode lightSource flash focalLength backLight	M	PS

	exposureIndex sensingMethod cfaPattern autoFocus PrintAspectRatio všechny hodnoty budou přebrány v případě použití fotoaparátu z údajů Exif		
<orientation>	popis orientace obrazu tak, jak je uložen vzhledem k jeho řádkům a sloupcům, např. normal*; normal, image flipper; normal, rotated 180°; unknown apod. 0-1	M	PS
<ImageAssessmentMetadata>	informace o digitálním obrazu pro jeho hodnocení a využití z hlediska dlouhodobé ochrany apod. 0-1	M	MC, PS
<SpatialMetrics>	rozměry obrázku, 2 rozměrná projekce objektů tak jak ji „vidí“ snímací zařízení 0-1	M	MC, PS
<samplingFrequencyPlane>	popis základní roviny, např. object plane (pro přímo ze předlohy digitalizované dokumenty), source object plane (pro digitalizaci mikrofilmů), camera/scanner focal plane (indikace sampl. frekvence fyzického senzoru); 0-1	R	MC + PS
<samplingFrequencyUnit>	jednotka měření sampl. frekvence, např. hodnoty 1= žádná pevná jednotka ; 2= inch, 3=centimetr; 0-1	M	MC, PS
<xSamplingFrequency>	údaje o počtu pixelů na jednotku samplovací frekvence pro šířku obrázku 0-1 povinné, pokud hodnota samplingFrequencyUnit je 2 nebo 3	MA	MC, PS
<numerator>	čítatel, číselné vyjádření, např. 300 0-1	M	MC, PS
<denominator>	jmenovatel, číselné vyjádření např. 1 0-1	M	MC, PS
<ySamplingFrequency>	údaje o počtu pixelů na jednotku samplovací frekvence pro výšku obrázku 0-1 povinné, pokud hodnota samplingFrequencyUnit je 2 nebo 3	MA	MC, PS
<numerator>	čítatel, číselné vyjádření, např. 300 0-1	M	MC, PS
<denominator>	jmenovatel, číselné vyjádření např. 1 0-1	M	MC, PS

<ImageColorEncoding>	doplňující údaje o barvě obrazu 0-1	M	MC, PS
<BitsPerSample>	počet bitů na kanál 0-1	M	MC, PS
<bitsPerSampleValue>	hodnota počtu bitů, např. 8, 1, 4 nebo 8,8,8 apod. 0-n POZOR – pro každou hodnotu je nutno element opakovat, tj. např. 3x element <bitsPerSampleValue> s hodnotou 8 <mix:BitsPerSample> <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue> <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue> <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue> </mix:BitsPerSample>	M	MC, PS
<bitsPerSampleUnit>	specifikace jednotky, např. integer nebo floating point 0-1	R	MC, PS
<samplesPerPixel>	počet barevných komponentů na pixel, např. 1, 3, 4 0-1	M	MC, PS
<TargetData>	informace o kalibračních tabulkách 0-1 povinné pro obrazy, kde se dělá kontrola oproti kalibrační tabulce	MA	MC
<targetType>	typ kalibrační tabulky; 0= external (kalibrační tabulka se neobjeví na dig. obraze, je to oddělený dig. soubor); 1= internal (tabulka je naskenována spolu s přelohou a objeví se na dig. obraze); 0-n	M	MC
<targetID>	údaje o původu kalibrační tabulky 0-n	M	MC
<targetManufacturer>	výrobce/původce kalibrační tabulky, např. Eastman Kodak nebo NK ČR, oddělení kontroly kvality apod. 0-1	M	MC
<targetName>	název kalibrační tabulky, např. ColorChecker, MicrofilmScanTarget aj. 0-1	M	MC
<targetNo>	číslo nebo verze kalibrační tabulky 0-1	M	MC
<targetMedia>	údaj o tom, na jakém médiu je kalibrační tabulka, např. film, paper aj. 0-1	R	MC
<externalTarget>	údaje o externí kalibrační tabulce; např. link na http://skenservis.cz/target-00000001 nebo název a cesta ke konkrétnímu souboru 0-n povinné v případě, že byla použita externí kalibrační	MA	MC

	tabulka (targetType = 0)		
<performaceData>	odkaz na soubor obsahující charakteristiku výkonu systému vzhledem k nastaveným hodnotám rozlišení atd.; možné hodnoty plnění – link URN nebo URL, nebo název souboru 0-n	R	MC
<ChangeHistory>	dokumentace procesů provedených na obrazovém souboru v jeho životním cyklu 0-1	M	MC
<ImageProcessing>	údaje o zpracování obrazového souboru 0-n	M	MC
<dateTimeProcessed>	2009-01-04T15:12:06; zapsat v ISO 8601 na úroveň vteřin 0-1	M	MC
<sourceData>	odkaz na původní zdrojová data, ze kterých byl vytvořen finální obrazový soubor; může to být např. URL nebo cesta do složky s původním skenem včetně názvu souboru; 0-1	M	MC
<processingAgency>	The National Library of the Czech Republic 0-n	R	MC

Příklad:

```
<mix:mix xmlns:mix="http://www.loc.gov/mix/v20">
  <mix:BasicDigitalObjectInformation>
    <mix:FormatDesignation>
      <mix:formatName>image/tif</mix:formatName>
      <mix:formatVersion>42</mix:formatVersion>
    </mix:FormatDesignation>
    <mix:byteOrder>little endian</mix:byteOrder>
    <mix:Compression>
      <mix:compressionScheme>LZW</mix:compressionScheme>
    </mix:Compression>
  </mix:BasicDigitalObjectInformation>
  <mix:BasicImageInformation>
    <mix:BasicImageCharacteristics>
      <mix:imageWidth>2504</mix:imageWidth>
      <mix:imageHeight>3301</mix:imageHeight>
      <mix:PhotometricInterpretation>
        <mix:colorSpace>Bgr</mix:colorSpace>
      </mix:PhotometricInterpretation>
    </mix:BasicImageCharacteristics>
  </mix:BasicImageInformation>
  <mix:ImageCaptureMetadata>
    <mix:SourceInformation>
      <mix:sourceType>Newspaper</mix:sourceType>
    </mix:SourceInformation>
    <mix:GeneralCaptureInformation>
      <mix:dateTimeCreated>2011-11-24T08:36:21</mix:dateTimeCreated>
      <mix:imageProducer>Elsyst Engineering</mix:imageProducer>
      <mix:captureDevice>reflection print scanner</mix:captureDevice>
    </mix:GeneralCaptureInformation>
  </mix:ImageCaptureMetadata>
</mix:mix>
```

```

<mix:ScannerCapture>
  <mix:scannerManufacturer>Contex</mix:scannerManufacturer>
  <mix:ScannerModel>
    <mix:scannerModelName>HD</mix:scannerModelName>
    <mix:scannerModelNumber>4250</mix:scannerModelNumber>
    <mix:scannerModelSerialNo>KE67E07001D</mix:scannerModelSerialNo>
  </mix:ScannerModel>
  <mix:MaximumOpticalResolution>
    <mix:xOpticalResolution>300</mix:xOpticalResolution>
    <mix:yOpticalResolution>300</mix:yOpticalResolution>
    <mix:opticalResolutionUnit>in.</mix:opticalResolutionUnit>
  </mix:MaximumOpticalResolution>
  <mix:ScanningSystemSoftware>
    <mix:scanningSoftwareName>NextImage</mix:scanningSoftwareName>
  </mix:ScanningSystemSoftware>
</mix:ScannerCapture>
<mix:scanningSoftwareVersionNo>1.6.9</mix:scanningSoftwareVersionNo>
</mix:ScanningSystemSoftware>
</mix:ImageCaptureMetadata>
<mix:ImageAssessmentMetadata>
  <mix:SpatialMetrics>
    <mix:samplingFrequencyUnit>in.</mix:samplingFrequencyUnit>
    <mix:xSamplingFrequency>
      <mix:numerator>300</mix:numerator>
      <mix:denominator>1</mix:denominator>
    </mix:xSamplingFrequency>
    <mix:ySamplingFrequency>
      <mix:numerator>300</mix:numerator>
      <mix:denominator>1</mix:denominator>
    </mix:ySamplingFrequency>
  </mix:SpatialMetrics>
  <mix:ImageColorEncoding>
    <mix:BitsPerSample>
      <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue>
      <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue>
      <mix:bitsPerSampleValue>8</mix:bitsPerSampleValue>
    </mix:BitsPerSample>
    <mix:samplesPerPixel>3</mix:samplesPerPixel>
  </mix:ImageColorEncoding>
</mix:ImageAssessmentMetadata>
</mix:mix>

```

8.5 METS část <fileSec>

8.5.1 <fileSec> hlavního záznamu METS

file group

- pro obrazy i texty (ALTO XML/OCR.TXT) budou v hlavním METS záznamu použity elementy <fileGrp>
 - jeden element <fileGrp> bude existovat pro obrazy archivních kopií, další pro user kopie, 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 (ideálně v podobě nějakého identifikátoru) na obrazový soubor (xlink:href) a atribut LOCTYPE
 2. **<fileGrp> pro obrazy uživatelských kopií**, bude mít tyto atributy: ID="UC_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 (ideálně v podobě nějakého identifikátoru) na obrazový soubor (xlink:href) a atribut LOCTYPE
 3. **<fileGrp> pro ALTO XML** bude mít následující atributy: ID="ALTOGRP" USE="Layout"
 - každý ALTO XML soubor bude mít vlastní element <file> s následujícími atributy:
 - ID – 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 (ideálně v podobě nějakého identifikátoru) na xml soubor obsahující ALTO (xlink:href) a atribut LOCTYPE
4. **<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 (ideálně v podobě nějakého identifikátoru) na xml soubor AMD_METS.xml (xlink:href) a atribut LOCTYPE
5. **<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 (ideálně v podobě nějakého identifikátoru) na txt soubor (xlink:href) a atribut LOCTYPE

Příklad níže – společný s následující kapitolou.

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

- <fileSec> ve vedlejších 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

Příklad:

```
<mets:fileSec>
  <mets:fileGrp ID="MC_IMGGRP" USE="Images">
    <mets:file ID="MC_anl001-000003_0001" MIMETYPE="image/jp2" SIZE="9801586"
      CHECKSUMTYPE="MD5" CHECKSUM="e5fba17c73bf736e5da06de847f2b9af" SEQ="0"
      CREATED="2012-01-26T09:32:40">
      <mets:FLocat xlink:href="./masterCopy/MC_anl001-000003_0001.jp2"
        LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp ID="UC_IMGGRP" USE="Images">
    <mets:file ID="UC_anl001-000003_0001" MIMETYPE="image/jp2" SIZE="5014649"
      CHECKSUMTYPE="MD5" CHECKSUM="fe8e0172fb031cdc79a9f6002cb64f55" SEQ="0"
      CREATED="2012-01-26T11:35:03">
      <mets:FLocat xlink:href="./userCopy/UC_anl001-000003_0001.jp2"
        LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp ID="ALTOGRP" USE="Layout">
    <mets:file ID="ALTO_anl001-000003_0001" MIMETYPE="text/xml" SIZE="26328"
      CHECKSUMTYPE="MD5" CHECKSUM="19e5a62e5283c274a2809a3c00ec13c4" SEQ="0"
      CREATED="2012-01-26T09:32:42">
      <mets:FLocat xlink:href="./ALTO/ALTO_anl001-000003_0001.xml" LOCTYPE="URL"
    />
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp ID="TXTGRP" USE="Text">
    <mets:file ID="TXT_anl001-000003_0001" MIMETYPE="text/plain" SIZE="360"
      CHECKSUMTYPE="MD5" CHECKSUM="02ab00b4bf830fd8e6980d828998ec2a" SEQ="0"
      CREATED="2012-01-26T09:32:42">
      <mets:FLocat xlink:href="./txt/TXT_anl001-000003_0001.txt" LOCTYPE="URL"
    />
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp ID="TECHMDGRP" USE="Technical Metadata">
    <mets:file ID="AMD_METS_anl001-000003_0001" MIMETYPE="text/xml" SIZE="50460"
      CHECKSUMTYPE="MD5" CHECKSUM="d6b86ff8547c3ef63cf7c33dcebf86bf" SEQ="0"
      CREATED="2012-01-26T14:07:04">
      <mets:FLocat xlink:href="./amdSec/AMD_METS_anl001-000003_0001.xml"
        LOCTYPE="URL" />
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>
```

8.6 METS část <structMap> - Strukturální metadata

8.6.1 <structMap> hlavního záznamu METS

- strukturální mapy v METS záznamu existují dvojího typu, **fyzická a logická**
 - **fyzická mapa** zaznamenává hierarchické informace o dokumentu včetně vazeb na fyzické soubory, ze kterých se skládají jednotlivé úrovně dokumentu
 - **logická mapa** zaznamenává logickou strukturu dokumentu, provazuje fyzické strany a jejich popis z dmdSec
- strukturální mapa logická i fyzická včetně odkazů na ALTO XML bude v hlavním METS záznamu hlavní_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.

8.6.1.1 Vyjádření fyzické strukturální mapy

- bude mít následující atributy <structMap LABEL="Physical_Structure" TYPE="PHYSICAL">
- fyzická strukturální mapa obsahuje rodičovský <div>, který obsahuje tyto atributy:
 - o LABEL- obsahuje název titulu periodika včetně čísla a roku vydání
 - o TYPE – např. newspaper
 - o ID – identifikátor div
 - o DMDID – identifikátor části popisných metadat na úroveň čísla
- jednotlivé stránky jsou zanořeny do rodičovského elementu <div> jako dceřiné <div> elementy
 - o <div> pro soubory stránky bude mít tyto atributy:
 - TYPE – bude se plnit typem stránky (viz typy stránek v pravidlech popisu periodik⁶)
 - ID – identifikátor div
 - ORDERLABEL – pořadové číslo stránky, jak je na ní vytištěno
 - ORDER – pořadí stránky v čísle periodika
 - o <div> pro soubory stránky vždy obsahují link <fptr> na soubor obrazu archivní kopie, na soubor obrazu uživatelské kopie, na ALTO XML, na OCR.TXT a na AMD_METS.xml
 - link na obrazový soubor archivní kopie má v elementu <fptr> následující atributy: FILEID, který obsahuje ID souboru archivní kopie
 - link na obrazový soubor uživatelské kopie má v elementu <fptr> následující atributy: FILEID, který obsahuje ID souboru uživatelské kopie
 - link na ALTO XML má v elementu <fptr> následující atributy: FILEID, který obsahuje ID ALTO XML souboru
 - link na OCR.TXT soubor má v elementu <fptr> následující atributy: FILEID, který obsahuje ID souboru OCR.TXT
 - link na AMD_METS.xml soubor má v elementu <fptr> následující atributy: FILEID, který obsahuje ID souboru AMD_METS.xml

Příklad:

```
<mets:structMap LABEL="Physical_Structure" TYPE="PHYSICAL">
  <mets:div LABEL="Chemické listy no.6 1951" TYPE="newspaper" ID="DIV_P_0000"
    DMDID="MODSMD_ISSUE_0001">
    <mets:div ID="DIV_P_PAGE_0000" ORDER="0" ORDERLABEL="[437a]"
      TYPE="titlePage">
      <mets:fptr FILEID="MC_an1001-000003_0001"/>
      <mets:fptr FILEID="UC_an1001-000003_0001"/>
      <mets:fptr FILEID="ALTO_an1001-000003_0001"/>
      <mets:fptr FILEID="TXT_an1001-000003_0001"/>
      <mets:fptr FILEID="AMD_METS_an1001-000003_0001"/>
    </mets:div>
    ... další stránky ...
  </mets:div>
</mets:structMap>
```

⁶

dočasně

umístěno

zde:

https://docs.google.com/a/mzk.cz/document/d/1zSriHPdnUY5d_tKv0M8a6nEym560DKh2H6XZ24tGAew/edit

8.6.1.2 Vyjádření logické strukturální mapy

8.6.1.2.1 Vyjádření logické struktury pro články s vazbou na ALTO bloky

- bude mít následující atributy <structMap LABEL="Logical_Structure " TYPE="LOGICAL">
- logická struktura na úroveň článků nebo např. ilustrací se popisuje pomocí do sebe zanořených elementů <div>
- pokud stránka obsahuje jen obraz a žádný text, pak je popsána jedním elementem <div> s atributem TYPE="PAGE" a link do souboru ALTO XML vede přímo na element <ComposedBlock>
 - o <div TYPE="PAGE"> lze využít jako kontejner na obrazy a další části stránky, které nejsou součástí článku
 - o pro obraz je možno využít atributy a typy podřízených elementů <div> jako je specifikováno v tabulce níže pro PICTURE, který je součástí článku
- 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"
 - o 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
 - o pro každou část článku existuje vlastní <div> element, podřízený hlavnímu <div> elementu článku
 - o 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
- do logické struktury PSP balíčku může být v případě její existence zakomponována i příloha (Supplement), která má vlastní <div> element s atributem TYPE="SUPPLEMENT"
 - o vnořené <div> elementy pro obraz a články i jejich použití je shodné se způsobem popisu logické struktury u elementu <div> s atributem TYPE="ISSUE"
- výčet stránek k jednotlivým článkům a obrázkům (popisují se pouze ty, které nejsou součástí článku) je obsažen v elementu <structLink> - popsáno v kapitole 8.7

Příklad:

```

<mets:structMap LABEL="Logical_Structure" TYPE="LOGICAL">
  <mets:div LABEL="Mezinárodní vztahy no. 1 1987" DMDID="MODSMD_TITLE_0001"
    TYPE="PERIODICAL_TITLE" ID="TITLE_0001">
    <mets:div DMDID="MODSMD_VOLUME_0001" TYPE="PERIODICAL_VOLUME"
      ID="VOLUME_0001">
      <mets:div LABEL="Mezinárodní vztahy no. 1 1987" TYPE="ISSUE" ID="ISSUE_01"
        DMDID="MODSMD_ISSUE_0001">
        <mets:div ID="ART_0001" TYPE="ARTICLE" ORDER="0" LABEL="PAPRSKY SMRTI
          prezidenta Beneše" DMDID="MODSMD_ART_0001">
          <mets:div ID="ART_0001_0001" TYPE="TITLE" ORDER="0">
            <mets:fptr>
              <mets:area FILEID="ALTO_an1001-000001" BETYPE="IDREF"
                BEGIN="TextBlock1" />
            </mets:fptr>
          </mets:div>
          <mets:div ID="ART_0001_0002" TYPE="NORMAL_TEXT" ORDER="1">
            <mets:fptr>
              <mets:area FILEID="ALTO_an1001-000007_0005" BETYPE="IDREF"
                BEGIN="TextBlock4" />
            </mets:fptr>
          </mets:div>
          <mets:div LABEL="Paprasek" ID="ART_0001_0003" TYPE="PICTURE"
            ORDER="2" DMDID="MODSMD_PIC_0001">
            <mets:div TYPE="IMAGE" ID="ART_0001_0008">
              <mets:fptr>
                <mets:area FILEID="ALTO_an1001-000007_0006" BETYPE="IDREF"
                  BEGIN="GraphicalElement0" />
              </mets:fptr>
            </mets:div>
            <mets:div TYPE="CAPTION" ID="ART_0001_0035">
              <mets:fptr>
                <mets:area FILEID="ALTO_an1001-000007_0007" BETYPE="IDREF"
                  BEGIN="TextBlock20" />
              </mets:fptr>
            </mets:div>
          </mets:div>
        </mets:div>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>

```

Význam pole „Povinnost“

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

<div> type	Atributy	Popis	Povinnost
TITLE		<div> obsahuje údaje o titulu periodika	M
	LABEL	LABEL – název titulu periodika, včetně čísla a data vydání čísla, např. Mladá fronta no. 5 29.06.1979	
	TYPE	TYPE – hodnota „PERIODICAL_TITLE“	
	ID	ID – identifikátor <div>, např. hodnota „TITLE_1“	
	DMDID	DMDID – obsahuje identifikátor DMD popisné části MODS titulu	
VOLUME		<div> obsahuje údaje o ročníku	M
	TYPE	TYPE – hodnota „PERIODICAL_VOLUME“	
	ID	ID – identifikátor <div>, např. hodnota „VOLUME_1“	
	DMDID	DMDID – obsahuje identifikátor DMD popisné části MODS ročníku	
ISSUE nebo SUPPLEMENT		<div> obsahuje údaje o čísle/příloze čísla periodika	M
	LABEL	LABEL – název titulu periodika, ve stejné podobě jako u titulu, tedy např. „Mladá fronta no. 5 29.06.1979“	
	TYPE	TYPE- hodnota ISSUE nebo SUPPLEMENT	
	ID	ID – identifikátor <div>, např. hodnota „ISSUE_1“ nebo „SUPPL_1“	
	DMDID	DMDID – obsahuje identifikátor DMD popisné části MODS čísla/přílohy	
ARTICLE		<div> obsahující údaje o jednom článku a jeho částech	M
	LABEL	LABEL – název článku	
	TYPE	TYPE – hodnota ARTICLE s pořadovým číslem, např. ARTICLE_1	
	ID	ID – identifikátor <div> elementu	
	DMDID	DMDID – identifikátor popisných metadat	
	ORDER	ORDER – pořadí článku	
	<div>	<div> TYPE=“ARTICLE“ může obsahovat další vnořený <div> různých typů popisující různé části článku, rozlišujeme tyto části (typy):	
		- TITLE	
		- SUBTITLE	
		- AUTHOR	
	- TRANSLATOR		
	- NORMAL_TEXT – běžný text bez dalšího upřesnění		
	- PICTURE		

		<ul style="list-style-type: none"> - NOTE - ARTICLE_PART - u článků, které jsou rozděleny na více míst na jedné stránce nebo více stránkách - tento <div> pro jednu součást rozděleného článku pak může obsahovat stejné části jako <div> pro článek, tj. (TITLE, SUBTITLE, AUTHOR, TRANSLATOR, NORMAL_TEXT, PICTURE) 	
<div> type pro ARTICLE:			
TITLE		<div> obsahující link na textový blok s nadpisem	MA
	TYPE	TYPE – hodnota „TITLE“	
	ID	ID – identifikátor <div> elementu, který popisuje jednu část článku (nadpis), např. hodnota „ARTICLE_PART_1“	
	ORDER	ORDER – pořadí části článku	
<fptr> <area>	FILEID	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID textového bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
SUBTITLE	TYPE	<div> obsahující link na textový blok s podnadpisem	MA
	ID	-----	
	ORDER	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	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID textového bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
AUTHOR	TYPE	<div> obsahující link na textový blok se jménem autora	MA
	ID	-----	
	ORDER	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	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID textového bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
TRANSLATOR	TYPE	<div> obsahující link na textový blok se jménem překladatele	MA
	ID	-----	
	ORDER	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	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID textového bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
NORMAL_TEXT	TYPE	<div> obsahující link na textový blok s běžným textem	M
	ID	-----	

	ORDER	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	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID textového bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
NOTE	TYPE ID ORDER	<div> obsahující link na textový blok s např. poznámkou pod čarou ----- TYPE – hodnota „NOTE“ ID – identifikátor <div> elementu, který popisuje jednu část článku (poznámku pod čarou), např. hodnota „ARTICLE_PART_5“ ORDER – pořadí části článku	
ARTICLE_PART	TYPE ID ORDER	<div> obsahující další vnořené <div> odkazující na jednotlivé části konkrétní části rozděleného článku; povinné pro dělený článek Pozn: pod <div> TYPE=“ARTICLE_PART“ lze vnořit všechny typy <div> jako pod <div> TYPE=“ARTICLE“ ----- TYPE – hodnota „ARTICLE_PART“ ID – identifikátor <div> konkrétní části, pro první část děleného článku např. „ARTICLE_2-1“, tj. první část článku 2 ORDER – pořadí konkrétní části děleného článku	
PICTURE -	LABEL	<div> pro obraz náležející k článku	MA
	TYPE	plní se pokud se obraz vyskytuje	
	ID	-----	
	DMDID	LABEL – název obrazu pokud existuje	
	ORDER	TYPE - PICTURE ID – identifikátor <div> elementu, který popisuje jednu část článku (běžný text), např. hodnota „ARTICLE_PART_6“ DMDID – link na bibliogr. popis obrazu ORDER – pořadí obrazu	
<div>	<div> element s typem PICTURE může obsahovat další <div> elementy s typy CAPTION, PICT_AUTHOR, PICT_TITLE a IMAGE;		
		- CAPTION obsahuje text případného popisku	
		- PICT_AUTHOR obsahuje text se jménem případného autora obrazu	
		- PICT_TITLE obsahuje text názvu obrazu, pokud nějaký název existuje	
		- IMAGE – obsahuje link do souboru ALTO XML na blok popisující vlastní obraz	
<div> type pro PICTURE			
CAPTION	TYPE ID	<div> obsahující link na textový blok s popisem obrazu ----- TYPE – hodnota CAPTION	MA

		ID – identifikátor <div> elementu, např. „ARTICLE_PART_7“	
<fptr> <area>	FILEID	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID komponovaného bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
PICT_AUTHOR	TYPE ID	<div> obsahující link na textový blok s autorem obrazu ----- TYPE – hodnota PIT_AUTHOR ID – identifikátor <div> elementu, např. „ARTICLE_PART_7“	MA
<fptr> <area>	FILEID	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID komponovaného bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
PICT_TITLE	TYPE ID	<div> obsahující link na textový blok s názvem obrazu ----- TYPE – hodnota PICT_TITLE ID – identifikátor <div> elementu, např. „ARTICLE_PART_7“	MA
<fptr> <area>	FILEID	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID komponovaného bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
IMAGE	TYPE ID	<div> obsahující link na komponovaný blok ALTO XML obsahující souřadnice vlastního obrazu ----- TYPE – hodnota IMAGE ID – identifikátor <div> elementu, např. „ARTICLE_PART_8“	MA
<fptr> <area>	FILEID	FILEID – ID ALTO XML souboru, např. „ALTO_PAGE_1“	
	BEGIN	BEGIN – ID komponovaného bloku v ALTO XML souboru	
	BETYPE	BETYPE – hodnota IDREF	
NOTE	ID	<div> obsahující link na textový blok s poznámkami k článku ----- ID – identifikátor <div> elementu, např. „ARTICLE_PART_9“	
ARTICLE_PART	TYPE ID ORDER	<div> obsahující další vnořené <div> odkazující na jednotlivé části konkrétní části rozděleného článku; povinné pro dělený článek Pozn: pod <div> TYPE=„ARTICLE_PART“ lze vnořit všechny typy <div> jako pod <div> TYPE=„ARTICLE“ ----- TYPE – hodnota „ARTICLE_PART“ ID – identifikátor <div> konkrétní části, pro první část děleného článku např. „ARTICLE_2-1“, tj. první část článku 2 ORDER – pořadí konkrétní části děleného článku	MA

8.6.1.2.2 Vyjádření logické struktury pro články bez vazby na ALTO bloky (pouze s vazbami na strany)

- bude mít následující atributy <structMap LABEL="Logical_Structure " TYPE="LOGICAL">
- logická struktura na úroveň článků nebo např. ilustrací se popisuje pomocí do sebe zanořených elementů <div>, ale dále už neobsahuje elementy <fptr> a <area>
- existuje tedy jen výčet popsanych článků a obrázků jako v případě předchozí kapitoly, ale bez vazeb na ALTO bloky
- výčet stránek k jednotlivým článkům je obsažen v elementu <structLink> - popsáno v kapitole 8.7

Příklad:

```
<mets:structMap LABEL="Logical_Structure" TYPE="LOGICAL">
  <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="PERIODICAL_TITLE"
    ID="TITLE_1" DMDID="MODSMD_TITLE_0001">
    <mets:div LABEL="Mezinárodní vztahy no. 1" DMDID="MODSMD_VOLUME_0001"
      TYPE="PERIODICAL_VOLUME" ID="VOLUME_1">
      <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="ISSUE" ID="ISSUE_1"
        DMDID="MODSMD_ISSUE_0001">
        <mets:div ID="ART_0001" TYPE="ARTICLE" ORDER="0" LABEL="PAPRSKY SMRTI
          prezidenta Beneše" DMDID="MODSMD_ART_0001"/>
        <mets:div ID="ART_0002" TYPE="ARTICLE" ORDER="1" LABEL="PROKLETÝ"
          DMDID="MODSMD_ART_0002"/>
        </mets:div>
      <mets:div LABEL="Příloha 1" TYPE="SUPPLEMENT" ID="SUPPLEMENT_1"
        DMDID="MODSMD_SUPPL_0001">
        <mets:div TYPE="PAGE">
          <mets:div ID="PICT_0001" TYPE="PICTURE" ORDER="1" LABEL="LES"
            DMDID="MODSMD_PICT_0001"/>
          </mets:div>
        </mets:div>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
```

8.6.1.2.3 Vyjádření logické struktury bez článků

- bude mít následující atributy <structMap LABEL="Logical_Structure " TYPE="LOGICAL">
- logická struktura na úroveň čísla a případné přílohy periodika se vyjadřuje pomocí zanořených elementů <div>
- výčet stránek k číslu a příloze je obsažen v elementu <structLink> - popsáno v kapitole 8.7

Příklad:

```
<mets:structMap LABEL="Logical_Structure" TYPE="LOGICAL">
  <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="PERIODICAL_TITLE"
    ID="TITLE_1" DMDID="MODSMD_TITLE_0001">
    <mets:div LABEL="Mezinárodní vztahy no. 1" DMDID="MODSMD_VOLUME_0001"
      TYPE="PERIODICAL_VOLUME" ID="VOLUME_1">
      <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="ISSUE" ID="ISSUE_1"
        DMDID="MODSMD_ISSUE_0001" />
      <mets:div LABEL="Příloha 1" TYPE="SUPPLEMENT" ID="SUPPL_1"
        DMDID="MODSMD_SUPPL_0001" />
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
```

8.6.2 <structMap> 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

Příklad:

```
<mets:structMap TYPE="PHYSICAL">
  <mets:div TYPE="PERIODICAL_PAGE">
    <mets:fptr FILEID="MC_an1001-000003_0001" />
    <mets:fptr FILEID="UC_an1001-000003_0001" />
    <mets:fptr FILEID="ALTO_an1001-000003_0001" />
    <mets:fptr FILEID="TXT_an1001-000003_0001" />
  </mets:div>
</mets:structMap>
```

8.7 METS část <structLink> - Výčet stran

- element <structLink> obsahuje výčet stran jednotlivých úrovní periodika na základě přidání vazeb mezi logickou a fyzickou strukturální mapou
- element <structLink> obsahuje subelement <smLink>, který obsahuje atributy „xlink:from“ a „xlink:to“
 - o „xlink:from“ obsahuje ID divu z logické strukturální mapy
 - o „xlink:to“ obsahuje ID divu stránky z fyzické strukturální mapy

8.7.1 Výčet stran v případě popisu článků (s odkazy na bloky ALTA i bez nich)

- vztahuje se ke kapitolám 8.6.1.2.1 a 8.6.1.2.2
- element <structLink> obsahuje:
 - o výčet stran jednotlivých článků a případných samostatných obrázků (obrázky, které jsou součástí článku se nepopisují zvlášť) (ID divu z logické strukturální mapy pro úroveň článků a obrázků)
 - o výčet stran celého výtisku a případné přílohy (ID divu z logické strukturální mapy pro úroveň čísla periodika a přílohy)

Příklad (číslo obsahující přílohu, které má popsáno dva články a jeden samostatný obrázek (není součástí článku) – ukázka včetně logické strukturální mapy):

```
<mets:structMap LABEL="Logical_Structure" TYPE="LOGICAL">
  <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="PERIODICAL_TITLE"
    ID="TITLE_1" DMDID="MODSMD_TITLE_0001">
    <mets:div LABEL="Mezinárodní vztahy no. 1" DMDID="MODSMD_VOLUME_0001"
      TYPE="PERIODICAL_VOLUME" ID="VOLUME_1">
      <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="ISSUE" ID="ISSUE_1"
        DMDID="MODSMD_ISSUE_0001">
        <mets:div ID="ART_0001" TYPE="ARTICLE" ORDER="0" LABEL="PAPRSKY SMRTI
          prezidenta Beneše" DMDID="MODSMD_ART_0001"/>
        <mets:div ID="ART_0002" TYPE="ARTICLE" ORDER="1" LABEL="PROKLETÝ"
          DMDID="MODSMD_ART_0002"/>
        </mets:div>
      <mets:div LABEL="Příloha 1" TYPE="SUPPLEMENT" ID="SUPPLEMENT_1"
        DMDID="MODSMD_SUPPL_0001">
        <mets:div TYPE="PAGE">
          <mets:div ID="PICT_0001" TYPE="PICTURE" ORDER="2" LABEL="LES"
            DMDID="MODSMD_PICT_0001"/>
          </mets:div>
        </mets:div>
      </mets:div>
    </mets:div>
  </mets:div>
</mets:structMap>
<mets:structLink>
  <mets:smLink xlink:from="ART_0001" xlink:to="DIV_P_PAGE_0000"/>
  <mets:smLink xlink:from="ART_0001" xlink:to="DIV_P_PAGE_0001"/>
  <mets:smLink xlink:from="ART_0002" xlink:to="DIV_P_PAGE_0002"/>
  <mets:smLink xlink:from="ART_0002" xlink:to="DIV_P_PAGE_0003"/>
  <mets:smLink xlink:from="PICT_0001" xlink:to="DIV_P_PAGE_0004"/>
  <mets:smLink xlink:from="PICT_0001" xlink:to="DIV_P_PAGE_0005"/>
  <mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0000"/>
  <mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0001"/>
```



```

<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0002"/>
<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0003"/>
<mets:smLink xlink:from="SUPPLEMENT_1" xlink:to="DIV_P_PAGE_0004"/>
<mets:smLink xlink:from="SUPPLEMENT_1" xlink:to="DIV_P_PAGE_0005"/>
</mets:structLink>

```

8.7.2 Výčet stran bez popisu článků

- vztahuje se ke kapitole 8.6.1.2.3
- element <structLink> obsahuje:
 - o výčet stran celého výtisku a případné přílohy (ID divu z logické strukturální mapy pro úroveň čísla periodika a přílohy)

Příklad (číslo obsahující přílohu – ukázka včetně logické strukturální mapy):

```

<mets:structMap LABEL="Logical_Structure" TYPE="LOGICAL">
  <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="PERIODICAL_TITLE"
    ID="TITLE_1" DMDID="MODSMD_TITLE_0001">
    <mets:div LABEL="Mezinárodní vztahy no. 1" DMDID="MODSMD_VOLUME_0001"
      TYPE="PERIODICAL_VOLUME" ID="VOLUME_1">
      <mets:div LABEL="Mezinárodní vztahy no.1 1966-" TYPE="ISSUE" ID="ISSUE_1"
        DMDID="MODSMD_ISSUE_1" />
      <mets:div LABEL="Příloha 1" TYPE="SUPPLEMENT" ID="SUPPLEMENT_1"
        DMDID="MODSMD_SUPPL_0001" />
    </mets:div>
  </mets:div>
</mets:structMap>
<mets:structLink>
<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0000"/>
<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0001"/>
<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0002"/>
<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0003"/>
<mets:smLink xlink:from="ISSUE_1" xlink:to="DIV_P_PAGE_0004"/>
<mets:smLink xlink:from="SUPPLEMENT_1" xlink:to="DIV_P_PAGE_0005"/>
</mets:structLink>

```

8.8 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, předšátky apod.
- ALTO XML i OCR TXT budou vznikat na úrovni stránky
- ALTO XML soubor pro zcela prázdné stránky bude obsahovat element `/alto/Layout/Page/PrintSpace`, ten ovšem nebude obsahovat podelementy `/alto/Layout/Page/PrintSpace/TextBlock`; `/alto/Layout/Page/PrintSpace/TextBlock/Illustration`; `/alto/Layout/Page/PrintSpace/TextBlock/GraphicalElement` ani `/alto/Layout/Page/PrintSpace/TextBlock/ComposedBlock`
- struktura ALTO XML bude generovaná na úrovni rozpoznání slova generovaná OCR
- kvalita rozpoznání znaků bude akceptována do určité hranice, výstupy nebudou ručně opravovány
- struktura ALTO umožní vyhledávání textu a jeho zvýraznění na úrovni slova, pokud bude použit odpovídající prohlížeč
- obrazy reprezentující stránku, které budou použity jako UC, musí odpovídat rozměry, orientaci a natočením obrazu, který byl použit pro vytvoření OCR
- OCR TXT bude vznikat z hotových ALTO XML během procesu digitalizace
- ALTO XML se bude vytvářet pouze pro novodobé dokumenty, nebo dokumenty s určitou hranicí kvality OCR
- jméno OCR souboru musí odpovídat jménu obrazového souboru, ke kterému náleží; např. `pr_0007.jp2` a `al_0007.xml` nebo např. `123456_006_alto.xml` a `123456_006_archiv.jp2`
- kódování ALTO XML i TXT OCR musí být v UTF-8
- souřadnice pozic (HPOS, VPOS, WIDTH, HEIGHT) musí být vyjádřeny v pixelech
- v této specifikaci ALTO XML se počítá s OCR i pro text mimo tzv. textové „zrcadlo“, tj. mimo hlavní text, jako jsou např. čísla stránek, běžící nadpisy ani jiné části vyskytující se na okrajích stránky (top, left, top a bottom margin)
 - o elementy `topMargin`, `leftMargin`, `rightMargin`, `bottomMargin` budou obsahovat elementy `<TextBlock>`, pro které platí stejná pravidla, jako pro element `<textBlock>` pro hlavní text stránky
 - o pozor: údaje z OCR mimo hlavní text stránky by neměly být vyhledávatelné v aplikaci zpřístupnění, docházelo by ke zmatení uživatele a výsledků (např. při hledání titulu kapitoly by byly zobrazeny výsledky pro každou stránku, která obsahuje běžící nadpis apod.)
- pokud je na konci věty dělicí znaménko, ALTO XML i OCR TXT musí obsahovat oba fragmenty slova s dělítkem a současně také kompletní slovo – je vysvětleno dále v tabulce

- ilustrace, reklamy a jiné grafické části stránky **nebudou** vyjádřeny v tazích /alto/Layout/Page/PrintSpace/Illustration ani Layout/Page/PrintSpace/GraphicalElement, tyto nejsou v popisu/tabulce níže vůbec uvedeny
- ilustrace, reklamy a jiné grafické části stránky **budou** vyjádřeny v tagu /alto/Layout/Page/PrintSpace/ComposedBlock/ s vyjádřením atributu TYPE, který bude označovat typ bloku (illustration, advertisement aj.)
 - o např. ilustrace bude popsána v elementu /alto/Layout/Page/PrintSpace/ComposedBlock/GraphicalElement, kde ComposedBlock TYPE je Illustration
 - o reklama s textem v rámečku bude popsána v elementu Layout/Page/PrintSpace/ComposedBlock/TextBlock, kde ComposedBlock TYPE je Advertisement
 - o tabulky, grafy obdobně
- elementy /alto/Layout/Page/PrintSpace/ComposedBlock/Illustration a Layout/Page/PrintSpace/ComposedBlock/ComposedBlock také **nebudou** využity
- /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock a /alto/Layout/Page/PrintSpace/ComposedBlock/GraphicalElement **nebudou** obsahovat elementy <Shape>; tvar těchto bloků je vyjádřen v elementu <Shape> samotného elementu <ComposedBlock>; logicky pak souřadnice tvaru <TextBlock> nebo <GraphicalElement> obsaženého v /alto/Layout/Page/PrintSpace/ComposedBlock jsou většinou shodné, pokud není tvarů nebo bloků v rámci /alto/Layout/Page/PrintSpace/ComposedBlock více
- všechny vyplněné hodnoty jsou příklady plnění, plnění v konkrétní instituci je nutno specifikovat vlastními pravidly a kontrolovanými slovníky
- ALTO XML bude využíváno pro tzv. pořadí čtení, tj. články vyskytující se na více stránkách nebo na více různých místech jedné stránky bude možné zobrazit celý a ve správném pořadí. K tomu je nutno znát jeho strukturu. Struktura bude vyjádřena v korespondujícím METS záznamu v logické strukturální mapě. Ta bude obsahovat odkazy na jednotlivé textové bloky článku, pomocí ID textových bloků použitých v ALTO XML.

Obsah pole popis:

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

význam pole „povinnost“

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

povinnost uvádění apod.)

Element	Atribut	Popis	Povinnost
<Description>			
<MeasurementUnit>		měřicí jednotka pro souřadnice v ALTO XML; možné hodnoty – dpi, pixel, inch1200 a mm10); inch1200 = 1/1200 inche; doporučené plnění je „mm10“ nebo „pixel“; 0-1	M
<sourceImageInformation>		informace o obrazovém souboru, ze kterého vzniklo ALTO XML; 0-1	M
<fileName>		jméno obrazového souboru, ze kterého bylo ALTO XML vytvářeno; ideálně i s filesystem cestou jeho uložení; např. nImageSeq-33386- b.tif//produkce/OCR/digibok_XY/XY_011.tif 0-1	M
<fileIdentifier>		jedinečný identifikátor obrazového souboru; 0-n	R
<OCRProcessing>	ID	popis procesu vzniku OCR; 0-n ----- ID OCR procesu, např. <OCRProcessing ID="OCRPROCES_1">; povinné	M
<preProcessingStep>		procesy před vznikem OCR, které provádí SW pro OCR (např. natočení obrazu) 0-n	M
<processingDateTime>		určení času procesu, který předcházel samotnému OCR; např. 2008-03-29T19:42:23 dle ISO 8601 na úrovni vteřin; 0-1	O
<processingAgency>		jméno nebo kód instituce, např. NK CZ, název externí firmy apod.; doporučujeme použít kontrolovaný slovník hodnot; 0-1	R
<processingStepDescription>		popis procesu (např. zarovnání, ořez apod.); 0-n	O
<processingStepSettings>		nastavení kroku popsaného v <processingStepDescription>, např. CCS OCR Processing Filter 0-1	O
<processingSoftware>		popis SW, který upravoval obrázek před vznikem OCR; 0-1	M

<softwareCreator>		výrobce softwaru - např. CCS Content Conversion Specialists GmbH, Germany; 0-1	M
<softwareName>		jméno softwaru - např. CCS docWORKS; 0-1	M
<softwareVersion>		verze SW, např. 6.2-1.16; 0-1	M
<ocrProcessingStep>		popis procesu vzniku OCR 1-1 – povinné pole	M
<processingDateTime>		okamžik kdy bylo OCR vytvořeno; nutno zapsat v ISO 8601 na úroveň vteřin; 0-1	M
<processingAgency>		jméno nebo kód instituce, např. NK CZ doporučujeme použít kontrolovaný slovník hodnot; 0-1	M
<processingSoftware>		popis SW, který dělal vlastní OCR; 0-1	M
<softwareCreator>		výrobce softwaru - např. ABBYY, Russia; 0-1	M
<softwareName>		jméno softwaru - např. FineReader; 0-1	M
<softwareVersion>		např. 8.0; 0-1	M
<Styles>		styly definují vlastnosti jednotlivých grafických prvků stránky. styl definovaný v elementu vrchní úrovně je použit jako výchozí pro podřízené elementy; 0-1	M
<TextStyle>	ID FONTSTYLE FONTFAMILY FONTSIZE	definuje font textu; 0-n ----- ID pro každý text style použitý v OCR souboru – povinné FONTSTYLE – např. bold, italics apod.; doporučujeme používat kontrolovaný slovník; doporučené FONTFAMILY – např. arial, calibri apod.; doporučujeme používat kontrolovaný slovník; povinné FONTSIZE – velikost fontu, např. 10, 12 apod.; povinné	M
<ParagraphStyle>	ID ALIGN	definuje formátování textových bloků; 0-n	M

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

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

	HEIGHT	ID: unikátní ID pro element RightMargin, např. P1_RM0001 (page 1, rightMargin0001); povinné HPOS: horizontální pozice; povinné VPOS: vertikální pozice; povinné WIDTH – šířka pravého okraje; povinné HEIGHT – výška pravého okraje; povinné	
<TextBlock>		stejně plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace>	MA
<BottomMargin>	ID HPOS VPOS WIDTH HEIGHT	pravý okraj – prostor mezi spodním okrajem stránky a textem; 0-1 ----- ID: unikátní ID pro element BottomMargin, např. P1_BM0001 (page 1, bottomMargin0001); povinné HPOS: horizontální pozice; povinné VPOS: vertikální pozice; povinné WIDTH – šířka spodního okraje; povinné HEIGHT – výška spodního okraje; povinné	M
<TextBlock>		stejně plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace>	MA
<PrintSpace>	ID HPOS VPOS WIDTH HEIGHT	popis tvaru pokrývajícího textové pole stránky; 0-1 ----- ID: unikátní ID pro element <printSpace>, např. P1_PS0001 (page 1, printSpace0001); - povinné HPOS: horizontální pozice; povinné	M

		<p>VPOS: vertikální pozice; povinné</p> <p>WIDTH – šířka textového pole; povinné</p> <p>HEIGHT – výška textového pole; povinné</p>	
<TextBlock>	<p>ID</p> <p>STYLEREFS</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p>	<p>popisy textových bloků na konkrétní stránce; 0-n</p> <p>pokud je stránka prázdná, TextBlock není potřeba uvádět; pokud je na stránce text tak ano</p> <p>-----</p> <p>ID obsahuje identifikátor textového bloku na stránce, např. "BLOCK1" nebo P1_TB0002 (stránka 1, textový blok 2); povinné</p> <p>STYLEREFS: reference na ID definice formátování textových bloků <ParagraphStyle>; povinné</p> <p>HPOS: horizontální pozice bloku; povinné</p> <p>VPOS: vertikální pozice bloku; povinné</p> <p>WIDTH – šířka textového bloku; povinné</p> <p>HEIGHT – výška textového bloku; povinné</p>	MA
<Shape>		<p> tvar textového bloku; 0-1 – pro jeden výskyt <TextBlock> jeden nebo žádný výskyt <Shape>; plnit v případě, že je tvar textového bloku nestandardní (víceúhelník)</p>	RA
<Polygon>	POINTS	<p>popis (souřadnice) tvaru víceúhelníku; 0-1</p> <p>-----</p> <p>POINTS – vyjádření jednotlivých bodů víceúhelníku; povinné</p>	M
<TextLine>	ID	popis jedné řádky textu v rámci textového bloku;	M

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

		<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 řetězce - slova (word confidence); hodnota mezi 0 (nejistá) a 1 (jistá); např. WC="0,99"; povinné</p> <p>SUBS_CONTENT – obsah chybějící části řetězce v případě, že je slovo na konci řádku rozdělené i do druhého řádku; obsahuje celý řetězec - aby byl vyhledatelný i v případě, že slovo se na stránce vyskytuje, ale je rozděleno; povinné</p> <p>SUBS_TYPE – označení typu substituce; možné hodnoty: HypPart1; HypPart2; Abbreviation; povinné - při výskytu SUBS_CONTENT</p> <p><i>HypPart1</i> se vyskytuje při rozdělení slova u jeho první OCR části (u první části tagu <CONTENT> ve větě (stringu) první; <i>HypPart2</i> se vyskytuje u následujícího tagu <CONTENT> v následující větě (stringu), který obsahuje druhou část rozděleného slova/řetězce; <i>Abbreviation</i> – typ substituce používaný při rozepisování zkratk v textu na jejich plný text; při dělení slov v textu HypPart1 a HypPart2 povinné, abbreviation nepovinné</p>	
<ALTERNATIVE>		<p>alternativní hodnota OCR řetězce pro jednotlivá slova; 0-n lze použít v případě nejistoty rozpoznání řetězce;</p>	O
<HYP>	<p>CONTENT WIDTH HPOS VPOS</p>	<p>zápis znaku rozdělovníku slov 0-1 pro jeden výskyt <TextLine>; vždy pro poslední <String>; může se vyskytnout pouze na konci řádku (1x) ----- CONTENT – obsahuje řetězec znaků, které jsou v textu použity na rozdělení slova, nejčastěji „-“; povinné</p>	MA

		<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</p> <p>WIDTH</p> <p>HPOS</p> <p>VPOS</p>	<p>prázdný prostor mezi řádky; 0-n v rámci jednoho <TextLine>; vždy mezi řádky, tj. mezi tagy <String>; -----</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>	M
<ComposedBlock>	<p>ID</p> <p>TYPE</p> <p>HPOS</p> <p>VPOS</p> <p>WIDTH</p> <p>HEIGHT</p> <p>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>ID: unikátní ID komponovaný blok, např. P6_CB0001 (stránka 6, komponovaný blok 0001); povinné</p> <p>TYPE – označení typu komponovaného bloku; nutné používat kontrolovaný slovník (illustration, Advertisement, apod.);</p>	MA

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

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