Národní knihovna

MORAVSKÁ ZEMSKÁ KNIHOVNA

DEFINICE METADATOVÝCH FORMÁTŮ

17. prosince

2024

Dokument verze 1.0

OCR (ALTO XML a TXT OCR)

PŘEDPIS PRO ZÁPIS ALTO XML PRO UŽITÍ V RÁMCI STANDARDU NDK

Autoři aktuální verze: Filip Pavčík, Boris Lehečka

Autoři předchozích verzí: Jan Hutař, Pavla Švástová, Jaroslav Kvasnica, Iveta Lodrová

OCR (ALTO XML a TXT OCR)

Předpis pro zápis ALTO XML pro užití v rámci Standardu NDK

Tento předpis pro zápis ALTO XML je závazný od **DMF pro monografie, verze 2.2** a **DMF pro periodika, verze 2.1**.

Předpis doporučuje poslední verzi formátu ALTO XML, aktuální v době vydání dokumentu; v současnosti (rok 2024) verze 4.4; předpis ale umožňuje použít i starší verze ALTO XML (avšak verze ALTO XML nesmí být starší než verze 2.0) – viz http://www.loc.gov/standards/alto/.

- níže uvedená specifikace **neobsahuje všechny elementy a atributy formátu ALTO XML, ale jenom ty, které jsou pro tuto konkrétní specifikaci relevantní** ; elementy a atributy, které v této specifikaci nejsou uvedeny, nepovažujeme pro účely specifikace za důležité
- každý uvedený element má vyjádřenou míru závaznosti výskytu ve sloupci **Povinnost**, a to pomocí zkratek: M (Mandatory), příp. MA (Mandatory if available) = **povinné**, R (Recommended) = **doporučené** a O (Optional) = **nepovinné**
- závaznost výskytu atributů, popř. jejich hodnot, je uvedena slovy ve sloupci **Popis**
- kódování ALTO XML i TXT OCR musí být v UTF-8
- ALTO XML i OCR TXT vzniknou pro všechny obrazové soubory náležející k jedné intelektuální entitě (číslu periodika nebo svazku monografie) včetně prázdných stran, fotografií hřbetu, předsádky apod.
- ALTO XML i OCR TXT budou vznikat pro úroveň stránky
- ALTO XML soubor pro zcela prázdné stránky bude obsahovat element /alto/Layout/Page/PrintSpace, ten ovšem nebude obsahovat tyto podelementy:

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

- struktura ALTO XML bude generovaná na úrovni rozpoznání slova generovaná OCR
- struktura ALTO umožní vyhledávání textu a jeho zvýraznění na úrovni slova, pokud bude použit odpovídající prohlížeč
- obrazy reprezentující stránku, které budou použity jako UC, musí odpovídat rozměry, orientací a natočením obrazu, který byl použit pro vytvoření OCR
- OCR TXT bude vznikat z hotových ALTO XML během procesu digitalizace
- jméno OCR souboru musí odpovídat jménu obrazového souboru, ke kterému náleží, liší se příponou a/nebo sufixem k základnímu pojmenování; např. pr_0007.jp2 a al_0007.xml nebo např. 123456_006_alto.xml a 123456_006_archiv.jp2
- souřadnice pozic (HPOS, VPOS, WIDTH, HEIGHT) musí být vyjádřeny v pixelech
- v této specifikaci ALTO XML se počítá s OCR i pro text mimo tzv. textové „zrcadlo“, tj. mimo hlavní text, jako jsou např. čísla stránek, běžící nadpisy a jiné části vyskytující se na okrajích stránky (top, left, top a bottom margin)

  - Elementy TopMargin, LeftMargin, RightMargin, BottomMargin budou obsahovat elementy, pro které platí stejná pravidla, jako pro element pro hlavní text stránky
  - POZOR: údaje z OCR mimo hlavní text stránky by neměly být vyhledavatelné 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 řádku dělící znaménko, ALTO XML i OCR TXT musí obsahovat oba fragmenty slova s dělítkem a současně také kompletní slovo – je vysvětleno dále v tabulce

- ilustrace, reklamy a jiné grafické části stránky nebudou vyjádřeny v tazích /alto/Layout/Page/PrintSpace/Illustration ani Layout/Page/PrintSpace/GraphicalElement, tyto nejsou v popisu/tabulce níže vůbec uvedeny
- ilustrace, reklamy a jiné grafické části stránky budou vyjádřeny v tagu /alto/Layout/Page/PrintSpace/ComposedBlock/ s vyjádřením atributu TYPE, který bude označovat typ bloku (Illustration, Advertisement aj.)

    - např. ilustrace bude popsána v elementu /alto/Layout/Page/PrintSpace/ComposedBlock/GraphicalElement, kde ComposedBlock TYPE je Illustration
    - reklama s textem v rámečku bude popsána v elementu Layout/Page/PrintSpace/ComposedBlock/TextBlock, kde ComposedBlock TYPE je Advertisement
    - tabulky, grafy obdobně

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

Legenda pro čtení specifikace ALTO XML:

- sloupec **Element** obsahuje název elementu. Počet znaků „x“ před názvem elementu značí stupeň zanoření elementu v zápisu.
- sloupec **Atribut** obsahuje názvy atributů, pokud se k danému elementu nějaké vážou.
- sloupec **Popis** obsahuje vysvětlení a příklad užití příslušného elementu a jeho atributů. Kde je to možné, je uvedeno doporučené plnění.
- uvedena je i číselná hodnota pro výskyt elementu, tak jak je definována formátem ALTO XML (dle XSD):

    - 0-1 element je nepovinný, neopakovatelný
    - 0-n element je nepovinný, opakovatelný
    - 1-1 element je povinný a neopakovatelný
    - 1-n element je povinný a opakovatelný

    Tato číselná povinnost má informativní charakter zejména s ohledem na opakovatelnost elementu. Z hlediska povinnosti použití elementu **je závazná písmenná povinnost uvedená ve sloupci Povinnost**, která může být stejná, nebo přísnější, než jakou definuje příslušný standard. Pokud je tedy například opakovatelnost dle mezinárodního standardu ALTO XML "0-n" a povinnost v rámci DMF je "M", pak je element povinný a opakovatelný.

- sloupec **Povinnost** určuje povinnost použití elementu. Pokud je rodičovský element např. doporučený a dceřiný element povinný, znamená to, že je dceřiný element povinný pouze tehdy, pokud je použit element rodičovský

Element	Atribut	Popis	Povinnost
<alto>		kořenový element schématu ALTO XML	M
	SCHEMAVERSION	použitá verze schématu ALTO XML	R
x<Description>		kontejnerový element sloužící pro zápis obecných nastavení souboru ALTO XML	M
xx<MeasurementUnit>		měřící jednotka pro souřadnice v ALTO XML možné hodnoty: - dpi, pixel, inch1200 a mm10 - inch1200 = 1/1200 palce - doporučené plnění je "pixel" **0-1**	M
xx<sourceImageInformation>		informace o obrazovém souboru, ze kterého vzniklo ALTO XML **0-1**	MA
xxx<fileName>		jméno obrazového souboru, ze kterého bylo ALTO XML vytvářeno; ideálně i s filesystem cestou jeho uložení - např. nlaImageSeq-33386-b.tif//produkce/OCR/digibok_XY/XY_011.tiff **0-1**	M
xxx<fileIdentifier>		jedinečný identifikátor obrazového souboru **0-n**	R
xx<Processing>		popis procesu před vznikem OCR **0-n** (v případě používání verze ALTO 2.0 je možné místo elementu <Processing> využít <OCRProcessing>)	R

Element	Atribut	Popis	Povinnost
	ID	ID procesu OCR např. <Processing ID="OCR_1"> **povinné**	M
xxx<processingCategory>		procesy před vznikem OCR, které provádí SW pro OCR (např. natočení obrazu) možné hodnoty: "preOperation", "other" **0-n**	R
xxx<processingDateTime>		určení času procesu, který předcházel samotnému OCR např. 2008-03-29T19:42:23 dle ISO 8601 alespoň na úroveň vteřin **0-1**	O
xxx<processingAgency>		jméno nebo kód instituce; např. NK CZ, název externí firmy apod. doporučujeme použít kontrolovaný slovník hodnot **0-1**	R
xxx<processingStepDescription>		popis procesu (např. zarovnání, ořez apod.) **0-n**	O
xxx<processingStepSettings>		nastavení kroku popsaného v <processingStepDescription> např. CCS OCR Processing Filter **0-1**	O
xxx<processingSoftware>		popis SW, který upravoval obrázek před vznikem OCR **0-1**	MA
xxxx<softwareCreator>		výrobce softwaru - např. CCS Content Conversion Specialists GmbH, Germany **0-1**	M

Element	Atribut	Popis	Povinnost
xxxx<softwareName>		jméno softwaru - např. CCS docWORKS **0-1**	M
xxxx<softwareVersion>		verze SW, např. 6.2-1.16 **0-1**	M
xxxx<applicationDescription>		popis komponent nebo nastavení SW, např. použité modely a jejich verze **0-n**	O
xx<Processing>		popis procesu vzniku OCR **0-n** (v případě používání verze ALTO 2.0 je možné místo elementu <Processing> využít <OCRProcessing>)	M
	ID	ID procesu OCR, např. <Processing ID="OCR_2"> **povinné**	M
xxx<processingCategory>		popis procesu vzniku OCR; možné hodnoty: - "contentGeneration", "contentModification", "postOperation", "other" **0-1 - povinné pole**	R
xxx<processingDateTime>		okamžik kdy bylo OCR vytvořeno - nutno zapsat v ISO 8601 alespoň na úroveň vteřin **0-1**	M
xxx<processingAgency>		jméno nebo kód instituce, např. NK CZ - doporučujeme použít kontrolovaný slovník hodnot **0-1**	M

Element	Atribut	Popis	Povinnost
xxx<processingSoftware>		popis SW, který dělal vlastní OCR **0-1**	M
xxxx<softwareCreator>		výrobce softwaru - např. ABBYY, Russia **0-1**	M
xxxx<softwareName>		jméno softwaru - např. FineReader **0-1**	M
xxxx<softwareVersion>		např. 8.0 **0-1**	M
xxxx<applicationDescription>		popis komponent nebo nastavení SW, např. použité modely a jejich verze **0-n**	O
x<Styles>		styly definují typografické vlastnosti jednotlivých textových prvků stránky; styl definovaný v elementu vrchní úrovně je použit jako výchozí pro podřízené elementy **0-1**	MA
xx<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 povinné - FONTFAMILY – např. arial, calibri apod.; doporučujeme používat kontrolovaný slovník povinné - FONTSIZE – velikost fontu v bodech (1/72 palce), např. 10, 12 apod. povinné	M

Element	Atribut	Popis	Povinnost
xx<ParagraphStyle>	ID ALIGN	definuje formátování textových bloků **0-n** ---------- - ID – pro každý odstavec + zarovnání; např. "PAR_01", "PAR_02" apod. povinné - ALIGN – zarovnání; povolené hodnoty: Left, Right, Center, Block aj. povinné	M
x<Tags>		definuje repertoár specifických popisných prvků (viz elementy níže), na které se odkazuje z rozpoznaného textu pomocí atributu TAGREFS	R
	ID TYPE LABEL DESCRIPTION URI	následující atributy se uvádějí u prvků podřízené úrovně: - ID – pro každý identifikovaný prvek povinné - TYPE – obecnější specifikace typu prvku, např.: Table, Formula, Typesetting apod..; doporučujeme používat kontrolovaný slovník doporučené - LABEL – specifický účel prvku na stránce, např. MathFormula, Map, MusicalScore, DropCap apod; doporučujeme používat kontrolovaný slovník doporučené - DESCRIPTION – doplňující popis označovaného prvku, popř. jeho textový obsah doporučené - URI – odkaz na autoritu nebo relevantní informace k popisu doporučené	
xx<LayoutTag>		popisuje prvek na stránce, který má vliv na uspořádání a grafický vzhled (např. tabulka, vzorec, reklama apod.)	R
xx<StructureTag>		popisuje prvek na stránce, který strukturuje intelektuální obsah (např. titulní stránka, obsah, nadpis, poznámka pod čarou apod.)	R

Element	Atribut	Popis	Povinnost
xx<RoleTag>		pro označení lidí a institucí, kteří jsou odpovědní za vznik publikace (např. autor, ilustrátor, nakladatel apod.)	R
xx<NamedEntityTag>		popisuje tzv. rozpoznané entity, tj. pojmy nebo údaje (např. osoba, organizace, geografické místo, číslo apod.)	R
xx<OtherTag>		popisuje jakýkoliv jiný prvek, který neodpovídá žádnému z výše uvedených	R
x<Layout>		layout – rozložení struktur (slov, odstavců apod.) na jedné stránce dokumentu **1-1 povinný výskyt** (element není opakovatelný)	M
xx<Page>	ID ACCURACY POSITION QUALITY PHYSICAL_IMG_NR HEIGHT WIDTH PC LANG OTHERLANGS ROTATION	element popisující jednu stránku dokumentu **1-n** ---------- - ID – vygenerovaný identifikátor stránky, např. "PAGE1", nebo "P1" apod. povinné - ACCURACY – procentuální odhad přesnosti OCR (0-100) doporučené - POSITION – pozice stránky; hodnoty k plnění: Left, Right, Foldout, Single, Cover nepovinné - QUALITY – krátký údaj o kvalitě předlohy stránky; hodnoty k plnění: OK, Missing, Missing in original, Damaged, Retained, Target, As in original nepovinné - PHYSICAL_IMG_NR – fyzické (pořadové) číslo stránky v dokumentu; vyjádřeno číslem, např. 1,2,3 apod. povinné - WIDTH – šířka stránky vyjádřená v pixelech povinné - HEIGHT – výška stránky vyjádřená v pixelech povinné	M

Element	Atribut	Popis	Povinnost
		- PC – Confidence level OCR souboru – hodnota mezi 0 (nejistá kvalita) a 1 (dobrá kvalita) nepovinné; pokud nevyplníte ACCURACY – tak je vyplnění doporučené - LANG – označení jazyka použitého na stránce doporučené - OTHERLANGS – označení dalších jazyků, které jsou použité na stránce doporučené - ROTATION – popisuje základní natočení na stránce. Výchozí hodnota je ve stupních proti směru hodinových ručiček doporučené	
xxx<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
xxxx<TextBlock>		stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace>	MA

Element	Atribut	Popis	Povinnost
xxx<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
xxxx<TextBlock>		stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace>	MA
xxx<RightMargin>	ID HPOS VPOS WIDTH HEIGHT	pravý okraj – prostor mezi pravým okrajem stránky a textem **0-1** ---------- - ID – unikátní ID pro element RightMargin, např. "P1_RM0001" (page 1, rightMargin0001) povinné - HPOS – horizontální pozice povinné - VPOS – vertikální pozice povinné - WIDTH – šířka pravého okraje povinné - HEIGHT – výška pravého okraje povinné	M

Element	Atribut	Popis	Povinnost
xxxx<TextBlock>		stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace>	MA
xxx<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
xxxx<TextBlock>		stejné plnění a pravidla jako pro element <TextBlock> vnořený do elementu <PrintSpace>	MA
xxx<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 - VPOS – vertikální pozice povinné	M

Element	Atribut	Popis	Povinnost
		- WIDTH – šířka textového pole povinné - HEIGHT – výška textového pole povinné	
xxxx<TextBlock>	ID STYLEREFS HPOS VPOS WIDTH HEIGHT LANG	popisy textových bloků na konkrétní stránce **0-n** - pokud je stránka prázdná, TextBlock není potřeba uvádět - pokud je na stránce text tak ano --- - ID – obsahuje identifikátor textového bloku na stránce, např. "BLOCK1" nebo "P1_TB0002" (stránka 1, textový blok 2); povinné - STYLEREFS – reference na ID definice formátování textových bloků <ParagraphStyle> povinné - HPOS – horizontální pozice bloku povinné - VPOS – vertikální pozice bloku povinné - WIDTH – šířka textového bloku povinné - HEIGHT – výška textového bloku povinné - LANG – označení jazyka použitého v rámci textového bloku povinné	MA
xxxxx<Shape>		tvar textového bloku **0-1** – pro jeden výskyt <TextBlock> jeden nebo žádný výskyt <Shape>; plnit v případě, že je tvar textového bloku nestandardní (víceúhelník)	RA

Element	Atribut	Popis	Povinnost
xxxxxx<Polygon>	POINTS	popis (souřadnice) tvaru víceúhelníku **0-1** ---------- - POINTS – vyjádření jednotlivých bodů víceúhelníku povinné	M
xxxxx<TextLine>	ID STYLEREFS HPOS VPOS WIDTH HEIGHT LANG	popis jedné řádky textu v rámci textového bloku **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 (povinné v případě vyplnění <TextStyle> nebo <ParagraphStyle>) - HPOS – horizontální pozice řádky povinné - VPOS – vertikální pozice řádky povinné - WIDTH – šířka řádky povinné - HEIGHT – výška řádky povinné - LANG – označení jazyka použitého v rámci řádku doporučené	M

Element	Atribut	Popis	Povinnost
xxxxxx<String>	ID CONTENT HEIGHT WIDTH HPOS VPOS CC WC LANG V případě dělení slov také: SUBS_TYPE SUBS-CONTENT	řetězec znaků – vlastní obsah OCR; znaky tvoří jednotlivá slova a více tagů <String> tvoří řádek textu <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 - 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" nepovinné - WC – úroveň důvěry v přesnost OCR výstupu celého řetězce-slova (word confidence); hodnota mezi 0 (nejistá) a 1 (jistá); např. WC="0,99" nepovinné - LANG – označení jazyka použitého v rámci řetězce znaků (tj. slova) doporučené - 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	M

Element	Atribut	Popis	Povinnost
		- SUBS_TYPE – označení typu substituce; možné hodnoty: HypPart1; HypPart2; Abbreviation povinné – při výskytu SUBS_CONTENT **HypPart1** se vyskytuje při rozdělení slova u jeho první OCR části (u první části tagu <CONTENT> ve větě (stringu)) první; **HypPart2** se vyskytuje u následujícího tagu <CONTENT> v následující větě (stringu), který obsahuje druhou část rozděleného slova/řetězce; **Abbreviation** – typ substituce používaný při rozepisování zkratek v textu na jejich plný text Při dělení slov v textu HypPart1 a HypPart2 povinné, abbreviation nepovinné	
xxxxxxx<ALTERNATIVE>		alternativní hodnota OCR řetězce pro jednotlivá slova **0-n** - lze použít v případě nejistoty rozpoznání řetězce	O
xxxxxx<HYP>	CONTENT WIDTH HPOS VPOS	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 - WIDTH – šířka dělícího znaku doporučené - HPOS – horizontální pozice dělícího znaku doporučené - VPOS – vertikální pozice dělícího znaku doporučené	MA
xxxxxx<SP>	ID WIDTH HPOS VPOS	prázdný prostor mezi řádky **0-n** v rámci jednoho <TextLine>; vždy mezi řádky, nikdy na začátku řádku, tj. mezera mezi tagy <String> ----------	M

Element	Atribut	Popis	Povinnost
		- ID – unikátní ID pro prázdný prostor mezi řádky, např. "P1_SP0001" (stránka 1, prázdný prostor 0001) povinné - HPOS – horizontální pozice povinné - VPOS – vertikální pozice povinné - WIDTH – šířka prázdného prostoru povinné	
xxxx<ComposedBlock>	ID TYPE HEIGHT WIDTH HPOS VPOS STYLEREFS TAGREFS	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 ---------- - ID – unikátní ID pro komponovaný blok, např. "P6_CB0001" (stránka 6, komponovaný blok 0001) povinné - TYPE – označení typu komponovaného bloku; nutné používat kontrolovaný slovník (illustration, Advertisement, apod.) povinné - HEIGHT – výška komponovaného bloku povinné - WIDTH – šířka komponovaného bloku povinné - HPOS – horizontální pozice bloku povinné	MA

Element	Atribut	Popis	Povinnost
		- VPOS – vertikální pozice bloku povinné - STYLEREFS – reference na ID definice formátování textových bloků <ParagraphStyle> povinné - TAGREFS – odkazuje na použité specifické prvky (tabulky, ilustrace apod.) doporučené	
xxxxx<Shape>		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)	RA
xxxxxx<Polygon>	POINTS	popis tvaru víceúhelníku **0-1** --- - POINTS – vyjádření jednotlivých bodů víceúhelníku povinné	M
xxxxx<TextBlock>	ID STYLEREFS HPOS VPOS WIDTH HEIGHT	v případě, že komponovaný blok (např. orámovaný tvar) obsahuje text; platí stejná pravidla jako pro normální element /alto/Layout/Page/PrintSpace/TextBlock **0-n** – pro jeden výskyt <ComposedBlock> 0 nebo více elementů /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock> - plnit, pokud je v komponovaném bloku text --- - ID – obsahuje identifikátor textového bloku v komponovaném bloku, např. "P1_CB0002_SUB" (stránka 1, textový blok 2, SUB značí komponovaný blok) povinné - STYLEREFS – reference na ID definice formátování textových bloků /alto/Styles/ParagraphStyle povinné	MA

Element	Atribut	Popis	Povinnost
		- HPOS – horizontální pozice bloku povinné - VPOS – vertikální pozice bloku povinné - WIDTH – šířka textového bloku povinné - HEIGHT – výška textového bloku povinné	
xxxxxx<TextLine>		/alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock/TextLine a ostatní elementy v rámci /alto/Layout/Page/PrintSpace/ComposedBlock/TextBlock mají stejná pravidla a výskyty jako jako ve vrchním elementu /alto/Layout/Page/PrintSpace/TextBlock	M
xxxxx<GraphicalElement>	ID HPOS VPOS WIDTH HEIGHT	popis grafického tvaru; v případě využití v rámci /alto/Layout/Page/PrintSpace/ComposedBlock označuje rozměry tvaru v rámci něhož je tabulka, reklama apod. **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 tabulka apod. ---------- - ID – unikátní identifikátor grafického tvaru povinné - HEIGHT – výška grafického tvaru povinné - WIDTH – šířka grafického tvaru povinné - HPOS – horizontální pozice grafického tvaru povinné - VPOS – vertikální pozice grafického tvaru povinné	MA

Historie verzí

Jméno	Datum	Verze	Provedené změny
Filip Pavčík Boris Lehečka	17. 12. 2024	1.0	- první oficiální draft samostatného dokumentu