

## Projekt DC07P02OUK002

### Národní autority v prostředí muzeí a galerií - interoperabilita s NK ČR

# Výměnný formát XML pro muzejní autority

Jana Šubová, Cosmotron Bohemia s.r.o.  
Jazyková úprava Z.Lenhart a J.Podolníková, MZM - CITeM  
29. 1. 2012

## 1. Úvod

Tento text je komentářem k XML schématu v souboru XML-definice.pdf.

Výměnný formát XML je založen na principech dokumentu CIDOC CRM (publikovaného také jako mezinárodní norma ISO 21127:2006) a principech formátu MARCXML. Využívá dva jmenné prostory: muzejní (*muz*) a marcový (*marc*). Kombinace těchto dvou jmenných prostorů přináší výhody zápisu různých informací pomocí obecných tříd a vlastností tak, jak je definuje CIDOC CRM, a současně respektuje nejběžnější praxi při zpracování informací o autoritách – tj. využívání formátu MARC21/Autority v knihovnách. Schéma XML tak respektuje metodiku Tvorba autoritních záznamů pro potřeby muzeí a galerií, která kombinuje data potřebná pro zachování systému autorit v knihovnách a data potřebná pro muzea a galerie. Současně využití dvou jmenných prostorů odstraňuje problémy CIDOC CRM právě v jeho nejslabších místech. CIDOC CRM je vytvořen na zpracování informací o předmětech kulturního dědictví. Zpracování dat o osobách, korporacích a pod. je v CIDOC CRM rozpracováno pouze do takového rozsahu, který je nezbytný při evidenci předmětů. Zápis některých specifík vyplývajících z metodických postupů při zpracování jmen osob, korporací, míst a pod. v takové podrobnosti, jakou vyžadují především knihovnická pravidla, by v reprezentaci podle CIDOC CRM byl příliš nepřehledný a nejednoznačný. Jmenný prostor *marc* je tedy využíván tam, kde data vycházejí především z knihovnické praxe a reprezentace podle CIDOC CRM není jednoznačná.

Jmenný prostor „*muz*“ obsahuje prvky schématu vytvořeného podle dostupných informací o transformaci CIDOC CRM do XML. V XML schématu zápis takového prvku začíná textem *muz*: (např. `<muz:E1.CRM_Entity>`). Prvky jsou založeny na vztazích mezi třídami, vztahy jsou vyjádřeny tzv. vlastnostmi. Třídy se značí písmenem E (od entity), vztahy se značí písmenem P (od property). Pro přiřazení dat do tříd je využíván prvek *in\_class* se standardizovaným zápisem „E“, číslo třídy, tečka, název třídy (např. `<muz:in_class>E21.Person</muz:in_class>`). Používá se standardizované označení vlastnosti: „P“, číslo vlastnosti, písmeno označující směrování vlastnosti, tečka, název vlastnosti (např. `<muz:P48F.has_preferred_identifier>`). Pro indikaci směrování vlastnosti se používají písmena F (od forward) a B (backward). Pro vlastnost odkazující z třídy definičního oboru na třídu oboru hodnot se použije F, pro opačný směr B.

Například:

Vlastnost P48 má označení *has preferred identifier* nebo *is preferred identifier of*

Definiční obor je: E1 CRM Entity

Obor hodnot je: E42 Identifier

V případě, že vlastnost spojuje třídu E1 se třídou E42 použije se v XML označení vlastnosti: *P48F.has\_preferred\_identifier*.

V případě, že vlastnost spojuje třídu E42 se třídou E1 použije se v XML označení vlastnosti: *P48B.is\_preferred\_identifier\_of*

Pro označení tříd a vlastností jsme zvolili kombinaci číselného a slovního označení z toho důvodu, aby výsledné XML bylo volně čitelné. Slovní označení vlastností a tříd vypovídá o obsahu. Nevýhodou takového kombinovaného označení je to, že v případě změny slovního označení v CIDOC CRM bude nutné podle potřeby aktualizovat označení prvků v XML schématu.

Označení pro tuto verzi XML schématu vychází z verze 5.0.4 dokumentu CIDOC CRM (z listopadu 2011).

Jmenný prostor „`marc`“ obsahuje prvky schématu MARCXML. Pro potřeby tohoto XML jsou využívány pouze dva prvky schématu MARCXML: `datafield` a `subfield`. V jmenném prostoru „`marc`“ se také využívají kvalifikátory prvků. Vztahují se k nejbližšímu prvku uvedeném ve stejném sloupci nad nimi. V dokumentaci jsou kvalifikátory uváděny tečkou (např. `.code=""`) a hodnota se uvádí do zpětných uvozovek (např. `.code="a"`). V XML formátu se na místo tečky uvádí mezera (např. `<marc:subfield code="a">`).

V dokumentaci schématu jsou jmenné prostory rozlišeny barvou. Jmenný prostor „`muz`“ je uveden černou a jmenný prostor „`marc`“ modrou barvou.

## 2. Hodnoty prvků

Hodnoty prvků se uvádějí podle metodiky Tvorba autoritních záznamů pro potřeby muzeí a galerií. V případě, že se v prvku uvádí konkrétní hodnota, je v dokumentaci zapsána v uvozovkách. Některé opakující se části jsou v dokumentaci vedeny samostatně. Odkaz na samostatný popis hodnot prvků je uveden kurzívou v pomlčkách (*-- ... --*).

## 3. Opakování prvků

Některé prvky musí být uvedeny právě jednou, jiné mohou být i vícekrát, jiné zase nemusí být uvedeny vůbec. Vyjádřeno je to ve sloupci Opakování prvku takto:

- 0 nepovinný, může být uveden nejvýše jednou,
- 1 povinný, musí být uveden právě jednou,
- ∞ nepovinný, může být uveden bez omezení vícekrát.

Slovně jsou popsány ty případy, kdy platí složitější pravidla pro (opakovaný) výskyt.