

# KOMPATIBILITA, INTEROPERABILITA FORMÁTU MARC ← → TEI A KATALOGIZACE STARÝCH TISKŮ

*Jaroslava Kašparová*

Domnívám se, že **potřeba rychlé, informačně korektní a dostatečné katalogizace** bez ohledu na formát je dnes již samozřejmým požadavkem jak knihovníků, tak tvůrců elektronických databází a katalogů. Ta s sebou nese samozřejmě i potřebu spolupráce a kooperace, sdílení informací mezi různými systémy a katalogizačními prostředími.

**Existence, koexistence, kompatibilita, průnik čili interoperabilita, konverze formátů, univerzální přístupnost dat** apod. jsou pojmy, které vytanou na mysl, hovořme-li v knihovnickém prostředí o spolupráci při tvorbě popisných dat starých tisků, nebo o sdílené katalogizaci. Koexistence primárních popisných dat bez ohledu na formát, v němž vznikly, je v rámci jednoho společného vyhledávacího rozhraní nejvolnější formou kompatibility. Optimálním řešením se jeví koexistence různých metadatových formátů, tj. např. MARC a TEI, a to tak, aby byly kompatibilní a interoperabilní. Je zapotřebí definovat průnik obou formátů a určit to, co je jim společné, v čem existuje shoda, či naopak v čem se liší. Popisná metadata i primární data zůstávají v primárních formátech a mohou být součástí popisných metadat dalších schémat. Konverze jsou jistou variantou průniku formátů a realizovat by se měly jen v nejnnutnějších případech a měla by jim předcházet podrobná analýza. Ta by měla zjistit, jaké jsou podmínky této interoperability a jak široký tento průnik bude, aby konverze dopadla co nejlépe. Jednou z podmínek úspěšnosti je rovněž použití určitých společných metodických postupů při samotné katalogizaci.

**Konverze z MARC21 (UNIMARC) do TEI**, které se v minulosti realizovaly, nedopadly nejhůře. Také dnes, kdy se XML popis v oblasti starých tisků standardizuje a metodicky profiluje (nová verze TEI P5 provedla některé úpravy právě ve prospěch starých tisků), aby převod záznamů pořízených v MARC-formátu do formátu XML neměl činit z hlediska technického potíže a ztrátovost dat nehrozí. Otázkou však zůstává, do jaké míry bude možno v novém prostředí zachovat selekčnost některých informací.

**Konverze z TEI do MARC** v prvních pokusech nedopadla nejlépe. Důvodů bylo více, ovšem hlavní překážkou byla malá standardizace zápisu informací a nedostatečná propracovanost metadatové DTD struktury vzhledem ke starým tiskům. Nová verze TEI P5 otevírá možnost zapsat (převést) všechna důležitá data dostatečně standardně a korektně, a to jak v rámci minimálního, tak i podrobného

popisu, a posílením řízenosti metodických postupů směřuje k větší kompatibilitě s MARC-formáty tak, aby se jejich vzájemná interoperabilita zlepšila. Domnívám se, že výsledky tohoto typu konverze (tj. TEI → MARC), která je v domácích poměrech velmi potřebná např. z hlediska portálu Jednotné informační brány nebo Souborného katalogu CASLIN, mohou být mnohem uspokojivější, než tomu bylo v minulosti.

### **Kompatibilita, interoperabilita formátu MARC ↔ TEI?**

Dnes za hlavní katalogizační formáty v oblasti tisků považujeme MARC – formáty a strukturu XML podle formátu TEI. Spor týkající se toho, který formát je pro katalogizaci starých tisků lepší, zda MARC či TEI, je dnes už úsměvnou minulostí. I když formát UNIMARC z knihovnického prostředí nezmizel, má jeho nástupce – formát MARC21 – jasnou převahu a také otevřenější budoucnost. Z formátů XML, které se v současnosti nabízejí, je formát TEI pro knihovnickou katalogizaci nejvíce propracován a díky existenci modulu pro popis historických fondů zjevně nejvhodnější.

**Formát MARC21** bývá často označován za rigidní knihovnický formát předepisující katalogizátorovi způsob zápisu dat dle striktně sestavené struktury a podle striktních pravidel (norma IBSD(A), AACR2, citační norma apod.). Na první pohled takto „sešněrovaný“ popis může vyvolávat odpor jisté části knihovnické obce, netušící ovšem, jaké zálužnosti si na ně připravil konkurenční formát, tedy formát TEI. Pocit sešněrovanosti je umocněn faktem, že formát MARC je skutečně v mnoha ohledech méně flexibilní než prostředí XML, což ostatně vyplývá z podstaty samotného formátu. Tvzení o tom, že je třeba „strnulý knihovnický formát“ MARC zavrhnout, že už dosloužil a v knihovnickém světě nemá co dělat, je samozřejmě absurdní. Sama knihovnická praxe ukázala, že je to formát, který lze dokonce rozvíjet a „modernizovat“ a jehož existence je opodstatněná. Je to knihovnický formát s jasně danou strukturou a standardy, s fungujícími vazbami na soubory jmenných a věcných autorit usnadňujících tvorbu záznamů i uživatelské vyhledávání. Vedle svých předností má ovšem zcela logicky i svoje zápory.

Je to formát, který vznikl jako bibliografický formát, a tudíž je vhodnější pro tvorbu záznamů bibliografických než záznamů exemplářových. Ty se v něm dají pořizovat, a taky se tak děje, nicméně kombinace postupů bibliografického a exemplářového popisu vytváří z MARCu z hlediska potřeb katalogizace historických dokumentů poněkud těžkopádnější hybrid (rozpor se projevuje zejména při popisu konvolutů). MARC21 byl, jak je známo, ve srovnání s UNIMARC-formátem v některých oblastech určitým krokem zpět – zejména z hlediska katalogizátorského komfortu (ruční zápis interpunkce, nebo redukce některých polí a podpolí, pře-

devším v bloku 500 pro zápis diferencovaných poznámek, apod.). Formát MARC je málo vhodný pro analytický popis (ten je řešen možná až příliš komplikované přes vazební pole) a není schopen víceúrovňové katalogizace v rámci jednoho záznamu atd.

**Formát TEI** – formát s pružnou a variabilní víceúrovňovou strukturou dopřávající katalogizátorovi značnou volnost a kreativnost v zápise dat. Tato volnost může být ovšem zneužitelná, zejména tehdy, přidá-li s k němu slabá či nulová standardizace. Zjistila jsem však, že absence standardizace a knihovnických katalogizačních norem padá na vrub knihovníkům a metodikům formátu. Formát XML má totiž poměrně velmi dobře standardizováno použití jednotlivých tagů, atributů i jejich pořadí. To, že tyto možnosti nebyly využity, je chybou těch, kdo jej nedokážou používat, není to nedostatek samotného formátu. Na druhé straně i zjevná přednost formátu – jeho víceúrovňovitá struktura určená pro exemplářovou katalogizaci – může představovat mnohdy pro katalogizátora značné úskalí (např. popis konvolutu o 80 samostatných tiscích v rámci jednoho katalogizačního záznamu, který je i uživatelům na zobrazení málo přehledný). TEI verze 4 a z ní odvozený formát MASTER je dnes již do jisté míry překonán. Od letošního roku vstoupila v platnost verze TEI P5, v níž byl formát MASTER, původně určený pro popis středověkých rukopisů, zapracován do modulu msDesc a v detailech pro popis rukopisů upraven. Tento modul pro popis rukopisů (vycházející ze standardu MASTER TEI P4, který se používal i pro popis prvotisků a starých tisků) je i v nové verzi výchozí strukturou pro popis starých tisků – samostatný modul pro staré tisky, jak jsme očekávali, nevznikl. K této verzi TEI schválené na podzim 2007 vznesli v létě 2008 pracovníci AiP Beroun několik zásadních připomínek (v souvislosti s výzkumným úkolem nazvaným „*Interoperabilita MARC21 a TEI P5 a katalogizace historických dokumentů*“, který AiP Beroun řeší a který úzce souvisí s mezinárodním projektem ENRICH).<sup>1</sup> Požadavky na úpravu DTD struktury TEI P5 se týkaly především starých tisků a vyplývaly ze zmíněné analýzy vzájemné kompatibility a interoperability formátu MARC a TEI. Z výsledků analýzy vyplynulo, že úspěšná interoperabilita MARC ↔ TEI je podmíněna úpravami modulu MsDesc TEI P5, a to tak, aby byl schopen pojmout popis starých tisků a zohlednit specifika jejich katalogizace v daleko větší míře, než jak tomu bylo ve formátu MASTER.

Analýza interoperability ukázala, že každý formát má svoje přednosti i zápory. Odhalila slabá místa obou formátů, pokusila se je buď odstranit, nebo oslabit vzájemným sblížením, potlačením jejich záporů a vyzdvihnutím jejich kladů, bylo-li to možné. Dá se říci, že z valné většiny se podařilo nalézt vhodný postup, který zvýší

<sup>1</sup> K mezinárodnímu projektu ENRICH viz blíže ENRICH European Networking Resources and Information concerning Cultural Heritage, dostupné z: <<http://enrich.manuscriptorium.com/index.php?q=node/1>>.

kompatibilitu a interoperabilitu a umožní dokonalejší průnik obou formátů – a to především úpravou TEI. I když pokus iniciovat vznik samostatného modulu pro popis starých tisků v rámci TEI P5 nevyšel, Konsorcium TEI sídlící v Londýně se požadavky AiP Beroun vážně zaobíralo a většinu z nich akceptovalo.<sup>2</sup>

### **Obecně k průniku dat TEI a MARC21**

Z hlediska úspěšné interoperability TEI  $\leftrightarrow$  MARC21 vyplývá rovněž to, že je nutno v oblasti katalogizace historických dokumentů rozlišovat jednotlivé samostatné kolekce dokumentů. Toto rozlišení kolekcí je určující pro jejich zpracování, zobrazování i uživatelské vyhledávání. Ukázala se nutnost odlišného přístupu k jednotlivým druhům historických dokumentů, což by se mělo promítnout i do fungování digitálních knihoven. Platí to především pro digitální databázi MANUSCRIPTORIUM, které bude takto rozlišovat minimálně kolekci rukopisů, starých tisků a map. Mapy nebyly detailně předmětem analýzy, a to proto, že představují na jedné straně specifický kartografický dokument a na druhé straně dokument, který je charakterem popisu možno přiřadit jak k rukopisům, tak k tiskům. Kartografická povaha dokumentů je určující a jejich popis v prostředí TEI by měl být podroben samostatné analýze z pohledu průniku TEI  $\leftrightarrow$  MARC21 tak, aby mohl být eventuálně vypracován ve formátu TEI také samostatný modul pro popis map (v MARC-formátu existuje pro popis map katalogizační schéma, není však příliš používáno).

V oblasti tisků, prvotisky počínaje a tisky 19. století konče, je naopak MARCovská metodika velice dobře propracovaná, a to jak metadatovou strukturou, tak i standardy. Pružná struktura TEI je schopna metadata MARC21 pojmout (viz výše v souvislosti s konverzí). Tato pružnost spolu s některými nespornými přednostmi, které XML popis přináší (možnost několikaúrovňového členění dat, flexibilita zápisu dat a někdy i mnohem větší možnost zpřesňujícího zápisu dat, možnost snadného přiřazování digitálních metadat apod.), staví formát TEI do role favorizovaného prostředí, které nejenže může být velmi dobře kompatibilní s klasickými knihovnickými formáty, ale pokud bude optimálně katalogizátorsky použito, může je šířit a způsobem zápisu údajů dokonce překonat. Rozdíly přetrvávají především v popise konvolutů, které vyplývají z odlišné filozofie popisu: MARC chápe konvolut v souladu s tradičním pojmáním bibliografie jako dokument o dvou a více bibliografických jednotkách, jejichž těsné sepětí v rámci jedné knihy (svazku) je záležitost až sekundární a de facto majitelská, čtenářská. Tyto

<sup>2</sup> Viz P5: Guidelines for Electronic Text Encoding and Interchange, dostupné z: <http://www.tei-c.org/release/doc/tei-p5-doc/html/MS.html> a též Extending msDescription to Describe Printed Books (and to describe manuscripts better), dostupné z: [http://www.cch.kcl.ac.uk/cocoon/tei2008/programme/workingpapers/wp\\_004.htm](http://www.cch.kcl.ac.uk/cocoon/tei2008/programme/workingpapers/wp_004.htm).

dvě jednotky mohou být zcela odlišné, nemusí na sobě z hlediska vydání nikterak záviset. TEI naopak vychází z pojetí knihy jako muzejního předmětu. Celistvost knihy je pro XML popis primární a toto pravidlo se při popisu dokumentu dodržuje bez ohledu na obsah knihy. Různorodost obsahová, různorodý výběr tisků jsou až druhotné. Bude nutno patrně tyto dvě odlišné katalogizační perspektivy akceptovat a přizpůsobit jim interoperabilitu.

### **Co katalogizace starých tisků?**

#### **v oblasti MARC:**

Limity MARC-formátů, tedy i MARC21, jsou dány (zejména malá strukturovanost a omezený počet polí bloku 500). Uvidíme, co ukáže další vývoj, bude-li nová metodika Prozatím lze doporučit dodržování stávajících standardů, a to v míře maximální, a co nejkorektnější zápis dat. V poznámkových polích využívat konkrétní diferencovaná pole a vyhýbat se opakování polí 500 pro všeobecnou poznámku.

#### **v oblasti TEI:**

Hlavním úkolem je vypracování optimální DTD struktury pro popis starých tisků a vytvoření funkčního on-line editoru M-TOOL pro přímou katalogizaci v XML, zvlášť pro rukopisy a pro staré tisky. Vytvoření nové generace XML editoru v současné době probíhá a editor by měl být k dispozici během první poloviny příštího roku. Tento on-line editor bude mít význam nejen pro spolupráci domácích a zahraničních institucí s MANUSCRIPTORIEM, ale určen je také všem katalogizátorům starých tisků popisujícím své dokumenty v XML. Měl by usnadnit tvorbu záznamů, nabídnout standardní zápis informací a stát se významným nástrojem v procesu digitalizace dokumentů.<sup>3</sup>

### **Průnik metodiky řízené katalogizace (MARC21) a „volné“ katalogizace TEI P5**

Analýza ukázala, že kromě úprav datové XML struktury pro staré tisky bude třeba přehodnotit i přístup katalogizátorů pracujících v XML. Prostředí XML, stejně jako MARC21, potřebuje mít definovány jisté katalogizační standardy, postupy, a to nejen vzhledem k interoperabilitě k MARC ← → TEI, ale obecně, z hlediska svého optimálního fungování. Předností struktury TEI je to, že skýtá velice flexibilní a kreativní nástroj pro katalogizaci různých druhů dokumentů, který je schopen interoperability s jinými formáty. Metodika popisu starých tisků však ve formátu TEI P4 DTD MASTER neexistovala, nebo existovala jen velice nedostatečně. Na překážku interoperability bylo (a ještě je) především nedodržování základních ka-

---

<sup>3</sup> Rysuje se změna postupu v pořizování dat bibliografických a technických tak, že bibliografický záznam se bude pořizovat až po zdigitalizování dokumentu a bude se včleňovat do struktury dat automaticky vzniklé při digitalizaci.

talogizačních pravidel běžných v knihovnickém světě (nejen ISBD(A) a AACR2) ze strany katalogátorů.

Pokud má dojít k průniku informací obou prostředí, musí se přiblížit metodiky a standardy obou struktur a možno říci i knihovnických kultur. Přijetí některých katalogizačních pravidel, v míře rozumné, se ukazuje nutností. Jejich absence ve formátu způsobuje to, že formát není dobře použitelný, což v minulosti vedlo k tomu, že jeho konverze byly příliš ztrátové. Pracuje se na tom, jak tyto katalogizační předpisy zahrnout do XML struktury, jak katalogizátora navigovat a jak do formátu XML implementovat co nejoptimálněji základní standardy knihovnického popisu.

Jestliže cílem analýzy bylo zjistit, je-li informační a metodický průnik obou prostředí možný a za jakých podmínek, pak na knihovnické veřejnosti je, aby se dokázala sjednotit na základních knihovnických katalogizačních postupech, platných napříč formáty, a aby je ve své praxi také dodržovala. Aplikovat rozumné a propracované standardy prověřené klasickou „předelektronickou“ katalogizací i elektronickou katalogizací v MARCovských formátech nemusí být nikterak v rozporu s flexibilním a kreativním katalogizačním XML TEI prostředím. Sebelepší nástroj ovšem nemůže apriorně zaručit standardní zápis z hlediska obsahu a formy. Korekce a doplnění popisných dat je záležitostí katalogizátora. Jaká bude úroveň poskytovaných informací a jak bude uživatelská veřejnost spokojena s jejich vyhledáváním, to záleží nejen na optimálním nastavení metadatové struktury a na použitých technických nástrojích, ale především na těch, kteří pořizují popisná data.

Relativní volnost katalogizace se ukazuje v rutinní katalogizaci téměř jako přítěž a je vítáno, je-li v některých směrech přece jen řízena. Nezbytná je metodika, která bude oběma formátům společná, velice blízká a v zásadních věcech jednotná. Katalogizace historických dokumentů je v podstatě jen jedna, má obecné zákonitosti vyplývající z povahy dokumentu, jež jsou nadřazeny prostředím, v němž popis dokumentu probíhá, a nástrojům, které se při tom používají. Formáty a prostředí nejsou rozhodující, rozhodující je vlastní způsob katalogizace a vznikající informační obsah a jeho struktura. Nestojí proti sobě na jedné straně stručná a na druhé straně podrobná katalogizace. Je jen katalogizace informačně akceptovatelná a využitelná pro všechny zainteresované a katalogizace nějakým způsobem informačně nevyhovující.

Metodickou inspiraci může TEI struktura nalézt v těchto oblastech:

- standardizace zápisu jmen osobních, korporací, rodů,
- aplikace ISBDa v základním popise – stejná forma zápisu rozsahu, datace dokumentů,
- stejná citační pravidla,

- aplikace metodiky věcného popisu alespoň tak, jak byla vypracována pro minimální katalogizaci historických fondů v MARC21,
- stejné mezinárodní zkratky pro jména zemí, jazyků,
- využití propracované anglické terminologie a jí odpovídajících zkratk pro jednotlivé role odpovědnosti,
- tvorba a používání autoritních souborů, včetně autorit „měst starých tisků“, apod.

Lze konstatovat, že neexistují zásadní formální překážky pro sblížení mezi standardy a katalogizační praxí v prostředí typu MARC21 a v prostředí TEI. Mám na mysli takové sblížení, kdy budou i v oblasti historických dokumentů, zejména starých tisků, vznikat **data univerzálně přístupná**. Rozhodující pro interoperabilitu formátů bude však katalogizátorská praxe – schopnost katalogizátorů vytvářet **informačně nasycený a formálně korektní katalogizační záznam**.

