

PŘEDPOKLADY MOŽNÉHO TECHNICKÉHO ŘEŠENÍ VIRTUÁLNÍHO ZPŘÍSTUPNĚNÍ KULTURNÍHO DĚDICTVÍ

Adolf Knoll
Národní knihovna ČR

Obsah článku byl autorem přednesen na konferenci Archivy, knihovny a muzea v digitálním světě 2001, kterou pořádal SKIP dne 6. 12. 2001 v Městské knihovně v Praze.

Článek byl již uveřejněn ve slovenském odborném periodiku Knižnica, 2002, roč. 3, č. 2, s. 57-64.

Encyklopedie udávají, že kultura je specifický způsob organizace a rozvoje lidské životní činnosti, který je zpředmětněn ve výsledcích fyzické a duševní práce. Dalo by se říci, že kultura je vše, co člověk nejen vytvořil, ale také co reflektoval a poznal v okolním světě. Z tohoto pohledu nelze kulturu jednoduše zaměňovat pouze s uměním nebo s činnostmi, které k němu mají blízko.

A protože jsme hovořili o zpředmětněných výsledcích lidské činnosti, bude velmi vhodné nazývat tyto výsledky kulturními objekty. Tyto objekty se nenacházejí pouze v často citovaných institucích, jako jsou knihovny, muzea a archivy, ale mohou se nacházet kdekoli, a to i v přírodě a ve městech, a mohou mít i výrazně pomíjivý charakter.

Objekty mohou být nejen dokumenty, tak jak je známe z knihoven a archivů, ale i různé exponáty z muzeí a galerií, jako jsou obrazy, umělecké předměty, technické předměty nebo zástupci živé a neživé přírody. Dále to mohou být objekty v krajině, a to jak ty, které stvořil člověk, tak ty, které stvořila sama příroda. Hovoříme tak o architektonických památkách, archeologických nalezištích nebo přírodních památkách. Ale mohou to být i události a různá představení a umělecké i jiné výkony (events and performances), které proběhly v čase a mohly být určitým způsobem zaznamenány.

Pro všechny tyto objekty platí, že mohou existovat samy o sobě, nebo že o nich mohou existovat nepřímé informace: záznamy, a to ať textové, obrazové statické či pohyblivé nebo zvukové, případně smíšené. Tyto nepřímé informace reprezentují dané objekty reálného světa, přičemž vždy platí, že úroveň reprezentace stanovuje její tvůrce. Tak jako konkrétní kniha reprezentuje svérázným způsobem určité dílo nebo myšlenku, tak i jiné typy informací mohou reprezentovat tuto knihu. Může to být bibliografický záznam, anotace nebo i její xerokopie nebo digitální kopie.

Obdobným způsobem ve světě trojrozměrných předmětů může dochovaný exemplář francouzské stíhačky z 1. světové války Nieuport 17C reprezentovat tento stroj jako určitý typ stíhacího letadla, zatímco fotografie tohoto exempláře již je další nepřímou informací o tomto exempláři. Na druhé straně dobová fotografie mechaniků táhnoucích tento stroj na holandském letišti Schiphol v době války představuje jiný exemplář téhož typu, exemplář, jenž se jistě liší od exempláře muzeálního nebo od exempláře, na kterém se prosla-

vil stíhací eso Charles Nungesser a který je dobře zdokumentován v popisech, obrazech a na fotografiích zrovna tak, jako některé další slavné a nedochované kusy.

Z těchto jednoduchých náznaků vychází, že o témže objektu reálného světa může existovat celá řada dalších informací, které jej přibližují. Archeologické naleziště keltského opevnění na Závisti nad Zbraslaví má určitě svou podrobnou dokumentaci a určitě vydalo řadu svých svědectví v podobě různých nalezených předmětů, které jsou uloženy v různých sbírkách. Kromě toho však o tomto nalezišti existuje řada knih nebo časopiseckých článků, ale také elektronických dokumentů na webu, nesčetné množství fotografií a zcela jistě i řada dokumentárních filmových záběrů.

Všechny tyto další informační objekty, které vypovídají nějakým způsobem o objektu našeho zájmu, jímž je archeologické naleziště, mohou mít významnou poznávací hodnotu. Jen například rozdíl mezi dvěma fotografiemi téhož naleziště pořízenými v roce 1990 a 2001 jasně dokumentuje jeho devastaci a z pohledu současného návštěvníka starší z nich i stav, který je badatelsky cennější.

O řadě kulturních objektů se shánějí informace velmi těžce, neboť jsou rozestry na velkém množství míst. Často ani nevíme, kde je hledat. A když ano, i přesto bývá cesta k nim leckdy složitá a ti, kteří se o ně zajímají, ztratí se sháněním informací spoustu času, který by lépe zhodnotili v tvůrčí práci. Kromě docházení do míst a pracného pořizování vlastních poznámek a dokumentace se však s úspěchem využívá především obrazové anebo jiné smyslové reprezentace příslušných objektů.

1 Základní struktura reprezentace kulturního objektu

Ten, kdo se zajímá o určité kulturní objekty, musí především vědět o tom, že vůbec existují, zda se dochovaly v originále nebo zda o nich existují nějaké další informace. Aby bylo možné tomuto požadavku vyhovět, musí být tyto objekty identifikovány. V příslušných informačních systémech, tj. kartotékách, katalozích nebo databázích, k tomu slouží identifikační záznam. Jeho role může být různá: někde plní základní inventární funkci ve smyslu majetkové evidence, jinde je díky své bohatosti a organizovanosti do větších celků součástí rozsáhlých a všestrannějších informačních systémů; proto se mu většinou říká katalogizační záznam, a to i v oblasti muzeí (catalogue record).

Tento záznam popisuje obecné rysy daného druhu a typu objektu a specifické rysy konkrétního exempláře. Obecné rysy se obvykle promítají do příslušných katalogizačních pravidel neboli pravidel popisu daného typu objektu (dokument, muzeální předmět, archiválie) a dále se projevují v konkrétní specifikaci pro daný typ objektu (periodikum, rukopis, mapa, obraz, socha, technický výrobek, archeologické naleziště atp.). Specifické rysy pak označují vlastnosti, kterými se daný objekt specifikuje (liší) v dané kategorii, jako například jiná kamufláž toho kterého exempláře téhož typu historického letadla.

Popsaný objekt sám může však také být reprezentován jiným typem objektu, který může mít textovou, zvukovou nebo obrazovou podobu, případně podobu smíšenou. Lze říci, že v tomto reprezentujícím objektu je původní objekt obsažen.

Tyto reprezentující objekty samy o sobě mohou a měly by také vytvářet sbírky, to znamená, že i ony mohou být identifikovány v nějakém systému, popisovány nebo katalogizovány. Je však třeba si uvědomit, že předmětem příslušného záznamu bude reprezentující objekt a nikoli objekt reprezentovaný, který pak v rámci záznamu reprezentujícího objektu může být popsán pouze částečně, a tudíž někdy i nedostatečně. Identifikace popisovaného objektu v rámci popisu objektu, který ho zahrnuje, může tedy – a často tomu tak i bývá – být pouze generická, tedy z věcného hlediska omezující se jen na hrubé zařazení.

Toto je například zřetelné z katalogizačního záznamu muzejního exponátu, ve kterém najdeme řadu technických dat o předmětu reprezentovaném exponátem, včetně specifických dat o exponátu samém, kdežto v katalogizačním záznamu publikace (knihy nebo článku) se odpovídající data (fyzický popis) zaměřují na tuto publikaci, kdežto původní objekt je zmíněn v předmětovém popisu a většinou i v názvu (může být součástí názvu). Původní technická data jsou pak obsažena v textu a obrazové výbavě dané publikace.

Lze tedy říci, že tentýž objekt je v různých identifikačních záznamech popsán na různých úrovních a v různých popisných prvcích, které spolu formálně nemusejí vůbec souviset. Na druhé straně lze však také říci, že je každopádně obsažen v popisných textech, ale s tím rozdílem, že jeden relevantní popisný text se může nazývat katalogizační záznam, kdežto text druhý je například prostě textem monografického díla.

Pokud si však představíme, že oba texty mohou být obsahově značkovány – tagovány, což je typické pro katalogizační záznam a je možné i pro souvislý text, pak lze tvrdit, že analogické a často navzájem se obohacující informace je možné nalézt v obou textech.

Tato úvaha je zjednodušenou ilustrací komplexnosti popisu a klasifikace kulturních předmětů a různých přístupů k nim.

Pokud si jakýkoli objekt zvolíme jako objekt popisovaný, je možno pro něj zhotovit katalogizační záznam. Pokud vyrobíme další objekt, který tento původní objekt reprezentuje – fotografie, ilustrace, text, videozáznam – pak můžeme vyrobit katalogizační nebo identifikační záznam tohoto nového objektu: popis fotografie, ilustrace, textu, videozáznamu. Velmi často v tomto popisu pak uvádíme specifická technická popisná data odlišná od technických dat popisujících původní předmět.

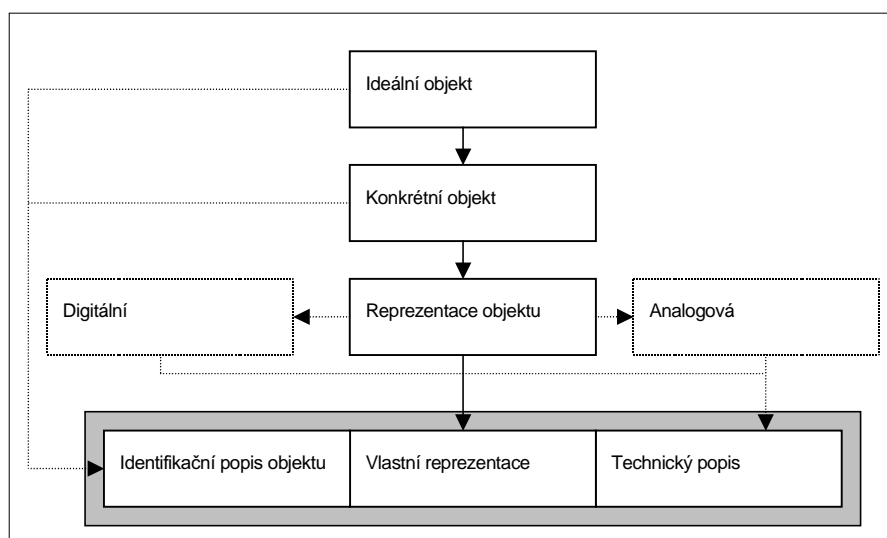
Technická data popisující muzejní exemplář letadla jsou například rozpětí, délka, hmotnost, druh motoru, stoupavost atp., kdežto technická data popisující jeho digitální obraz jsou například rozlišení, barevná hloubka, formát, komprese dat atp.

Z uvedeného vyplývá, že kromě popisů originálních předmětů musíme zvládat i popisy jejich vizuálních a dalších reprezentací, přičemž určité jádro popisu by mělo být totožné v obou.

Klasický knihovnický popis – katalogizace – tento problém nezná, protože popisovaný předmět byl vždy dokumentem, zatímco jeho obsah byl vyjádřen předmětovým popisem, například na základě desetinného třídění (MDT, DDC) nebo předmětového popisného jazyka, jako jsou třeba Library of Congress Subject Headings. Při popisu vizuální reprezentace muzejního exponátu se však v mnoha systémech¹⁾ předmětem identifikace opět stává originální předmět, tzn., že jak originál, tak i jeho vizuální reprezentace se shodují v základních identifikačních datech.

Z tohoto lze usoudit, že přímá smyslová reprezentace původního objektu (většinou vizuální) má shodná identifikační data s tímto objektem, kdežto komplexní dokument (knihy, články, filmy), jehož předmětem je (též) původní objekt, má popis, který se o původním objektu vyjadřuje nepřímě. Je to vždy záležitost pohledu a klasické informační systémy většinou umožňují pouze pohled popisovaného předmětu. Pouze tezaury nebo komplexní soubory některých autorit mohou umožnit i odlišné pohledy, často však se značnými omezeními, danými rozdíly v pojetí jednotlivých disciplín lidské činnosti a poznání.

Vizuální reprezentací knihy, která by byla obdobou vizuální reprezentace trojrozměrného předmětu, je pak například její digitální kopie. A právě s příchodem digitalizace do knihoven dochází i k zásadnímu zlomu v pojmání popisu dokumentu. Již nevystačíme pouze s jeho základní identifikací včetně jeho věcného zařazení, ale nutně potřebujeme další data popisující vlastnosti digitálních kopií částí dokumentu, data umožňující virtuální výstavbu dokumentu z digitálních komponent a data umožňující více než kdy jindy i popis různých objektů nacházejících se na úrovni částí dokumentů: objekty v textu i objekty ikonografické.



Obr. 1: Schéma vzniku reprezentace kulturního objektu a jejího popisu

Výše uvedené schéma znázorňuje jednoduchý případ vzniku komplexní analogové nebo digitální reprezentace konkrétního objektu. Hlavními částmi této reprezentace jsou:

- její vlastní vyjádření analogové (fotografie, analogová zvuková nahrávka, mikrofilm, ...) nebo digitální (počítačový soubor nebo počítačové soubory grafické, zvukové, videosoubory)
- identifikační popis reprezentovaného objektu, tj. jeho katalogizační záznam, případně popisy součástí tohoto objektu (jednotlivých stran, komponent, dílčích staveb, částí, ...)
- technický (fyzický) popis reprezentace objektu, který směřuje především k lepšímu a věrnějšímu užití a obnovitelnosti vlastní reprezentace, což má velký význam hlavně u digitálních reprezentací.

1.1 Identifikace kulturního objektu

Pro identifikační (katalogový, inventární) popis objektu se vyvinula řada pravidel. V knihovnictví jsou to například Angloamerická katalogizační pravidla AACR2, pravidla pro popis rukopisů MASTER²⁾, pravidla pro jakousi katalogizaci v elektronické publikaci Dublin Core a řada dalších. Ve světě muzeí pak existuje celá řada pravidel pro popis různých předmětů³⁾.

Mezi neznámější patří především AMICO⁴⁾, CDWA⁵⁾, Object ID⁶⁾ (používaný v prevenci proti krádežím uměleckých předmětů), Systèmes descriptifs⁷⁾, CIDOC⁸⁾, Van Eyck⁹⁾, AFRICOM (muzejní předměty mající vztah k Africe), MIDAS (architektura, archeologie), SPECTRUM¹⁰⁾ a další.

Tyto systémy řeší, co se má popisovat, tj. které vlastnosti daných objektů se stanou popisnými termíny a co se jimi má rozumět. Nejvíce propracované, ale zároveň jednostranné, jsou systémy pro popis uměleckých předmětů jako AMICO, CDWA nebo Object ID. Velmi přesvědčivým dojmem a kvalitní dokumentací působí zejména francouzské Systèmes descriptifs, neboť umožňují i kvalitní popis technických objektů (i v místě jejich původního fungování) a architektury. Samotné systémy se mezi sebou liší, ale společným jmenovatelem je v řadě případů snaha používat různé tezaury pro dodržení konzistence popisu.

Velmi výmluvně a systematicky se staví k užití kontrolovaných slovníků například CDWA, které doporučuje jejich použití pro identifikaci umělce, identifikaci místa, identifikaci generického zařazení a předmětu díla. Tezaurů a slovníků je k dispozici v oblasti muzeí celá řada, například Thesaurus of Geographic Names, Union List of Artists Names, The Art and Architecture Thesaurus, Multilingual Egyptological Thesaurus a další.

Pokud srovnáme tuto snahu o vnitřní konzistenci s používáním tzv. souborů autorit v knihovnictví, vidíme, že přinejmenším oblast fyzických jmen osob je společná a určitým způsobem i oblast generického zařazení a předmětu díla. Zde je však podstatný rozdíl v tom, že v knihovnách se více méně pod předmětovým popisem používá často vlastně pouze generické zařazení, což může být nedostatečné. Například publikace o konkrétním typu letadla z období 1. světové války je věcně popsána v databázi Národní knihovny ČR pojmy jako letadla vojenská, letadla historická a letadla

vrtulová, což je z pohledu odborníka velmi hrubé určení velmi obsáhlé skupiny předmětů. Navíc není indikováno místo, kterého se tyto předměty týkají nebo kde vznikly (zde Francie), ani bližší datace jejich vzniku či užití. Datace se vztahuje v popisu dokumentu na dobu jeho vzniku. K těmto rozdílům nedochází vždy, ale přece jen velmi často a je třeba se s nimi vypořádat, pokud bychom chtěli vytvářet z různých systémů jedno homogennější informační prostředí.

1.1.1 Strukturace standardizovaného popisu

Při bližším studiu dostupných předpisů pro popis muzeálních předmětů jsme se vždy setkali s určitou doporučenou strukturou, ale již ne vždy s normativním předpisem, jak doporučovaná pole a jejich skupiny označovat. Jestliže to například AMICO nebo Systèmes descriptifs uvádějí celkem jednoznačně, pak na druhé straně například Object ID nebo CDWA se omezují na definování popisných kategorií.

To neznamená, že uvedené systémy nejsou používány pro potřeby různých databází, ale znamená to, že označení jednotlivých popisných prvků není jednotné, tedy například takové, jaké knihovníci znají z různých druhů formátu MARC (UNIMARC, INTERMARC, UKMARK, MARC21 – dříve USMARC a další), Dublin Core nebo speciálnějších předpisů, jako jsou MASTER nebo DIEPER¹¹⁾.

Samozřejmě, že primárním požadavkem je vnitřní konzistence a přesné deklarování datové struktury, protože ostatní už může být jen záležitostí jednoduché konverze. Tuto věc je třeba rozumně vážit ve všech heterogenních prostředích, neboť někdy lze velmi dobrých výsledků dosáhnout i bez přílišného tlaku na unifikaci tagování nebo případně syntaxe, pokud jsou dodávaná data konzistentní a jejich struktura se řídí dobře definovanými pravidly. Tohoto si je například vědoma americká skupina knihoven RLG (Research Libraries Group), která ve svém projektu Cultural Materials Alliance apeluje především na spolupráci různých subjektů při dodržení určitých základních pravidel¹²⁾, tj. na konzistentní chování, při kterém velmi často zdánlivá jednoduchost nebo nedokonalost může znamenat velice moc.

1.1.2 Syntaktické formální vyjádření strukturovaného popisu

Syntaktickým formálním vyjádřením standardizovaného strukturovaného popisu je vlastně již předpis (forma, formát), který umožňuje jeho praktické využití. Tatáž strukturace velmi často snese různé formáty nebo různé způsoby (jazyky) její praktické implementace.

Nemusíme hledat daleko od knihoven, vždyť třeba UNIMARC lze vyjádřit vlastně jednoduchým řádkovým formátem s návěstími (značkami polí a podpolí) nebo i pomocí XML, pokud je sepsáno příslušné DTD jako například francouzská aplikace BiblioML¹³⁾. Rovněž tak i Dublin Core může být převedeno do metatagů v HTML nebo i do vlastního XML DTD.

Totéž lze učinit i s ostatními popisnými standardy, přičemž jako nejvíce perspektivní se jeví platforma XML, která je schopna zajistit snadnou interoperabilitu mezi různými přístupy a systémy.

Lze tedy říci, že tatož katalogizační pravidla mohou být v různých systémech různě tagována, tj., že stejné popisné prvky budou různě značeny, a že mohou též být i různě formálně vyjádřena, přičemž se názvy tagů mohou (ale nemusí) rovněž změnit.

Tato skutečnost bude jistě dále probíhat, protože se objevují nové a nové možnosti přemostění různých informačních zdrojů. Důležité však je, aby byla vždy zachována vnitřní obsahová konzistence popisovaných vlastností kulturních objektů, která je důležitým předpokladem pro jejich informační slučitelnost v heterogenním prostředí.

V oblasti popisu muzeálních objektů existuje v řadě případů řádkový formát, i když někdy je a někdy není jednoznačně předepsaný. Existují též pokusy formalizovat popis v XML¹⁴⁾ nebo alespoň použít XML pro výměnu dat¹⁵⁾.

2 Vlastní reprezentace kulturního objektu a její popis

2.1 Existující požadavky a řešení

Jak již bylo řečeno, kulturní objekt může být reprezentován (dokumentován) textovými popisy a především vizuálními a zvukovými objekty. Tato reprezentace se vztahuje k původnímu objektu, proto část jejího popisu musí obsahovat identifikační popis objektu. Dále jde však o něco jiného a specifického. Jednak jsou to data vztahující se například k vizuální reprezentaci jako k něčemu, co je dílem samo o sobě, jednak to jsou data související s použitými technikami při pořizování této reprezentace.

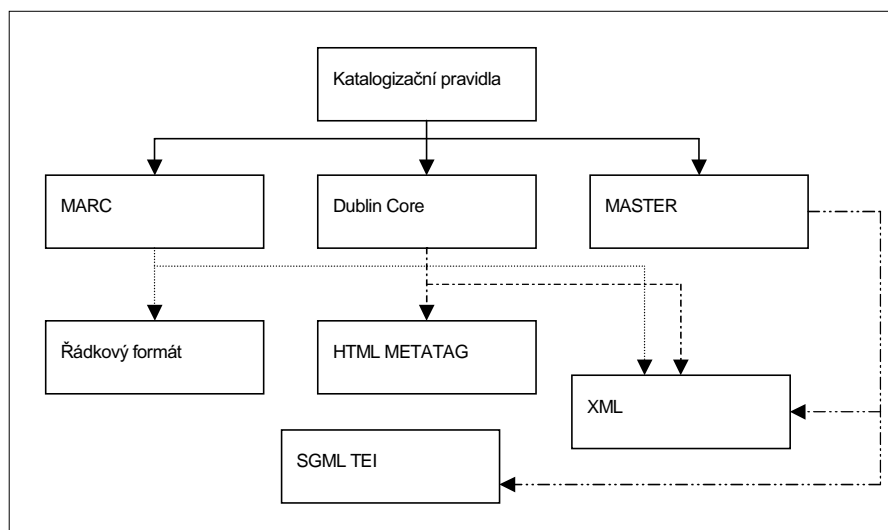
AMICO například vychází při stanovení metadat pro popis médií (media metadata) důsledně z Dublin Core, které doplňuje dalšími svými specifickými poli, která vyjadřují technické náležitosti reprezentace: především jde o formát dat (komprese, rozměry, kódování, velikost souboru, barevná paleta).

Van Eyck hovoří o záznamu obrazu (image record) a omezuje se na devět popisných prvků: identifikační číslo, umístění obrazu, zdroj obrazu, datum pořízení, datum zařazení do sbírek, fotograf, číslo negativu a copyright. Ani základní kategorie VRA¹⁶⁾ (Visual Resources Association) nepřekračují v podstatě tento rozsah.

2.2 Možné perspektivní řešení

Při posuzování tohoto v podstatě identifikačně technického popisu je třeba vycházet z požadavků kladených na roli popisných metadat v příslušných systémech. Ve všech třech výše uvedených případech jde spíše o velmi obecnou identifikaci média, tzn. informaci, která má sloužit veřejnosti pro lepší představu o tom, co lze využít.

Nicméně digitální svět s sebou přináší jeden velký problém, a tím je požadavek dlouhodobého uchování vizuální nebo jiné smyslové reprezentace původního objektu. Pokud



Obr. 2: Možnosti formalizace katalogizačních pravidel

se omezíme jen na grafické soubory, víme, že vlastnosti snímacích zařízení se mezi sebou různí a zároveň se odlišují i od vlastností výstupních zařízení, jejichž pomocí pořízené soubory zobrazujeme (tiskárny, monitory). Z praxe víme, že výstupní zařízení stárnou, mění se (zlepšuje se) jejich kvalita, a tím se mění i jejich charakteristiky, takže tentýž grafický soubor, například barevný obraz iluminace, je na každém zařízení jiný, nemluvě už o běžném uživatelském tisku.

Dalším problémem je způsob uložení dat, tj. grafický formát a jeho vlastnosti. I když tento formát bývá dobře popsán, řada informací o tom, jak jsou data uspořádána – tedy de facto, co od nich lze a nelze očekávat – by měla být zapsána i jinde ve srozumitelnější podobě, aby bylo možné po delší době případně soubory rekonstruovat.

Vlastnosti snímacích a výstupních zařízení lze popsat tzv. ICC¹⁷⁾ profily, které umožňují dosáhnout věrného zobrazení bez ohledu na typ technického zařízení, pokud jsou použité profily snímacího a výstupního zařízení známé. Je proto vhodné ICC profil snímacího zařízení v technickém popisu vizuální reprezentace uložit.

Rovněž tak je třeba uložit v technických metadatach i řadu dalších údajů charakterizujících dané grafické soubory, jako jsou užitý formát, příslušný počítačový soubor, rozměry obrazu a rozměry snímaného originálu atp. Tato záležitost je řešena dostatečně v NISO návrhu na standard technických metadat¹⁸⁾ pro statické digitální obrazy a ještě podrobněji v tzv. DIG35 specifikaci¹⁹⁾ o tomtéž. Zaslouhou těchto doporučení je především to, že dostávají obrazová metadata vně vlastního obrazového souboru, i když třeba řadu popisných prvků přebírají z definice TIFF 6.0²⁰⁾.

Tento fakt je významný i proto, že v digitálních knihovnách se uživatelé nedostanou zpravidla k tzv. archivním obrazům, ale k obrazům uživatelské kvality v horším rozlišení, někdy i menší barevné hloubce, zpracovaným větší ztrátovou kompresí, případně ve zcela jiném formátu. Na druhé straně pak jde o to – a to především – umět věrně reprodukovat originál nezávisle na čase, neboť digitální obraz poskytuje nadčasovou dokumentaci jeho stavu. K tomu se ještě přidružuje požadavek kvalitní kalibrace zařízení a uchování informace o ní.

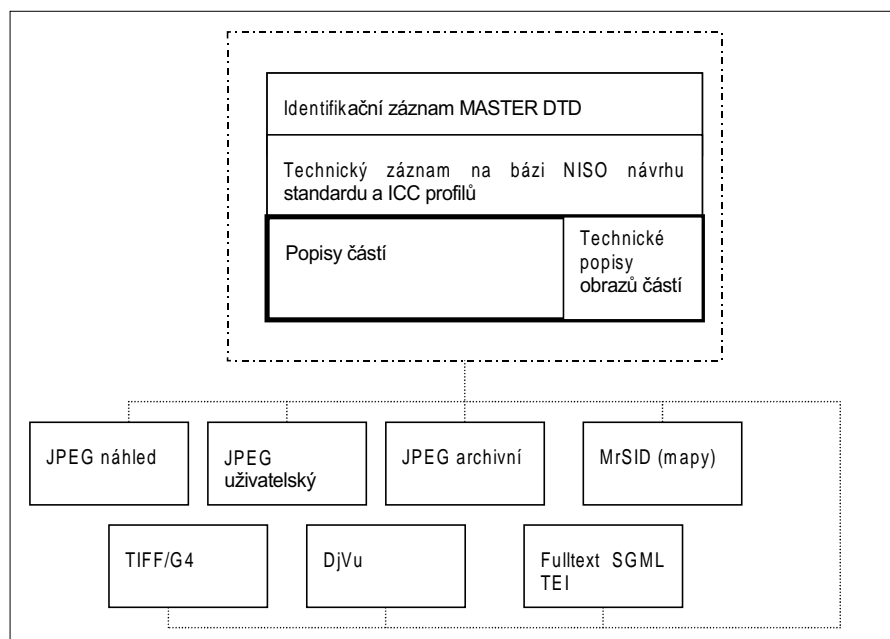
V programu Memoriae Mundi Series Bohemica dochází v současné době k radikálnímu posunu v oblasti tzv. technického popisu, a to právě zavedením využití ICC profilů, využitím aplikace uvedeného NISO standardu a optimalizované kalibrace, mj. jde i o nasnímání kvalitních a stabilních kalibračních tabulek spolu s originálem a jejich fyzické uložení v digitálním archivu spolu s příslušným médiem, na kterém jsou obrazová data zapsána.

Jeden objekt může být složen z více objektů a každý může mít svou vizuální reprezentaci – například kniha má vizuální reprezentaci pro každou stranu zvlášť a zvlášť pro desky a hřbet. Kromě toho tentýž snímání objekt může být reprezentován paralelně různými způsoby: různě kvalitními obrazy (náhled, obraz pro internet, archivní obraz, dílčí obraz detailu) nebo jinými typy reprezentace, jako jsou videosoubor nebo text či zvukový soubor.

3 Možné řešení záznamu digitální reprezentace kulturního objektu

V České republice jsou v plném rutinním provozu dva programy digitalizace knihovních materiálů v rámci programu Veřejné informační služby knihoven, a to Memoriae Mundi Series Bohemica²¹⁾ (ke konci roku 2001 cca 300 000 digitalizovaných stran rukopisů a desítky tisíc stran dalších dokumentů) a Kramerius²²⁾ (ke konci roku 2001 cca 600 000 digitalizovaných stran převážně starších periodik vytištěných na kyselém papíře a dalších dokumentů). V plánu je digitalizace zvukových dokumentů (mluvené slovo) a historických map. Lze konstatovat, že pokud jde o různé typy dokumentů, existuje dnes v ČR obecně přijatý standard jejich popisu a struktury, jinými slovy jejich elektronický formát.

V současné době probíhá adaptace formátu rukopisů na nedávno přijaté nebo přijímané standardy evropského významu, totéž je plánováno pro oblast periodik. Dojde tak k zajištění kompatibility se souvisejícími projekty v těch segmentech struktury digitálního dokumentu, které jsou pro komunikaci významné. Tím dojde k optimalizaci formátu popisu a přístupu DOBM SGML, který je doporučeným postupem a standardem programu UNESCO Paměť světa pro digitalizované knihovní materiály²³⁾. Datové



Obr. 3: Schéma digitalizovaného rukopisu

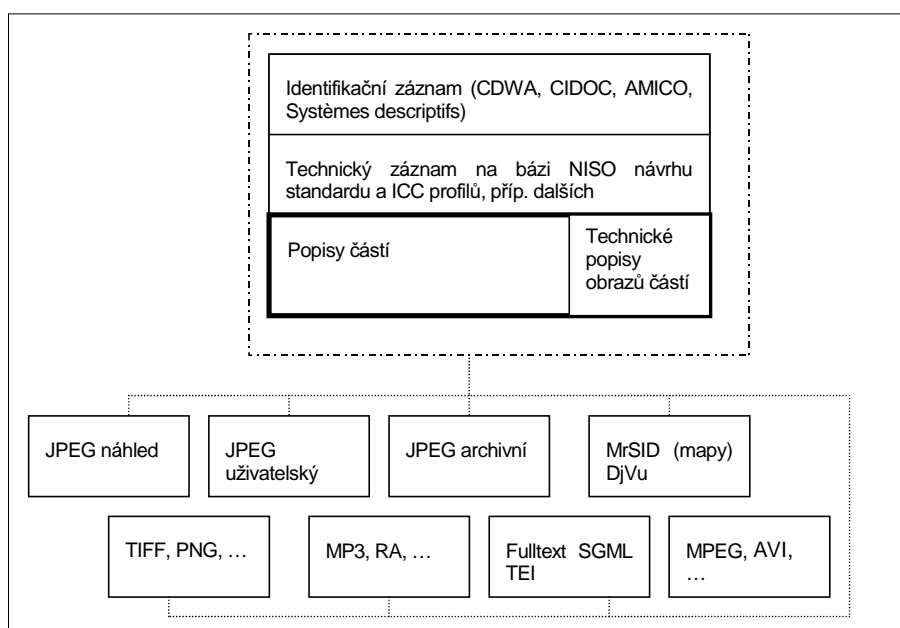
struktury budou migrovat na platformu XML, prakticky k tomu dochází již dnes například ve zpracování nových definic struktury rukopisu. První verze DTD (Document Type Definition) je již k dispozici.

Obr. 3 schematicky znázorňuje popis digitalizovaného rukopisu a uvádí jen některé typy obrazových formátů a vždy jen jeden obraz. Samozřejmě, že mohou být odkázány ještě další typy souborů a že počet popisů částí a jejich reprezentací je přímo úměrný počtu stran jednoho rukopisu.

Tento model zabezpečuje dobrou evropskou komunikaci především díky implementaci MASTER DTD pro katalogizační záznam rukopisu. Obdobně pak lze řešit i formát záznamu digitální reprezentace muzeálního objektu.

Obr. 4 pak znázorňuje obecný model reprezentace muzeálního objektu.

Obr. 4: Schéma digitalizovaného muzeálního objektu



Pokud si všimneme blíže rozdílů, pak strukturálně neexistují žádné, pouze se musí oblast muzeí sjednotit na identifikačním záznamu muzeálního objektu nejlépe aplikací již některých existujících modelů, kterých může být víc podle typu popisovaných a digitalizovaných objektů.

V oblasti technického záznamu budou při aplikaci dalších typů digitálních formátů pro zvuk a video případně nutné ještě nějaké úpravy, nejlépe však opět na bázi existujících mezinárodních doporučení. Záleží však na roli, kterou budou příslušné soubory ve zpřístupnění objektů hrát. Pokud bude tato role méně významná, nebude třeba ukládat příliš mnoho metadat, nebo jen ta zásadní.

V oblasti užitých datových formátů je obecně možné takřka cokoli, měla by však existovat rozumná doporučení i pro další typy digitálních formátů pro zvuk a video.

Realizace uvedeného schématu je konkrétní a pragmatickou záležitostí a je k ní možno přistoupit na základě většiny shody představitelů příslušných institucí a financování prostřednictvím Ministerstva kultury ČR. V maximální možné míře by mělo být využito zkušeností výše uvedených knihovnických digitalizačních programů a měla by být dodržena vzájemná maximální možná kompatibilita.

To potom umožní využít pro oblast muzeí většiny již vyvinutých nástrojů pro tvorbu a zpřístupnění dat (AIP Safe, ImGear, výrobní procesy, atp.), čímž se celý program nesmírně zlevní. Takto bude možné velmi rychle přistoupit k vlastní tvorbě dat jako základnímu předpokladu digitálního zpřístupnění muzejního i dalšího kulturního dědictví. Vhodným řešením by byl obdobný program, jako je například Memoriae Mundi Series Bohemica pro vzácné knihovní materiály.

4 Vytvoření virtuálního prostředí pro přístup ke kulturnímu dědictví

Výše uvedeným způsobem bychom mohli dosáhnout slušné strukturální kompatibility systémů zpřístupnění obou světů – knihovního a muzeálního – po internetu. Rozhodně by se daly používat tytéž nástroje a vzhledem k tomu, že by společnou platformou byl XML (formát pro muzeální objekty by měl být od počátku napsán v tomto jazyce), byly by vytvořeny i velmi dobré předpoklady pro další zacházení s daty.

Řada popisných prvků je podobná a mohou se v nich vyskytovat podobné nebo tytéž hodnoty, jde například o jména fyzických osob, věcně popisné prvky, dataci i názvy míst. V první fázi přibližování obou světů by mohlo dojít k poměrně slušné digitální reprezentaci kulturního dědictví z více oblastí. Bohužel, strukturální identita nebo podobnost může však být – a velmi často bývá – jen povrchní. Nedovoluje totiž – nebo jen velmi ztěžka – změny pohledů při vyhledávání, což je ztěženo především tím, že i přes formální příbuznost řady popisných prvků jsou příslušné systémy řízeny různými katalogizačními pravidly.

Datace u dokumentu znamená rok jeho vzniku nebo vydání. Totéž znamená i u muzeálního objektu, ale pokud dokument pojednává o tomto objektu, jen zřídka se vyskytne ve věcném popisu časové určení. Datace se ale může týkat i osob, a to jak autorů, tak osob, kterých se dané objekty týka-

jí; může se týkat i různých míst nebo provozních operací knihoven a muzeí, jako jsou akvizice nebo vyřazování.

I když se zdánlivě dva objekty jeden druhého netýkají, například olejomalba a fotografie domu nebo kniha, mohou se týkat nebo být shodná místa, osoby, časová rozmezí nebo různé datace, které se k oběma těmto objektům vztahují. Pokud se vůbec tyto překrývající údaje v popisu objeví, je to velmi často v úplně jiných popisných prvcích a v jiných vztazích k objektu popisu. Autor obrazu může být totožný s osobou, která v domě na fotografii žila nebo k němu má určitý vztah, jenž může být vyjádřen v úplně jiném díle (například knize), nebo může být prostě jen součástí autorovy biografie, případně může být v řadě domů na historické pohlednici jeho rodný dům v čase a v podobě doby kolem jeho narození.

Ukazuje se, že více než homogenní zdroje dat potřebujeme pravděpodobně jejich sémantickou kompatibilitu, která je možná jen z pohledu vyšší a neutrální abstrakce, neboť nutit pohled jedné disciplíny disciplíně jiné může být velmi zavádějící.

Z těchto důvodů vznikl na půdě ICOM tzv. objektově orientovaný Konceptuální referenční model (Conceptual Reference Model CRM)²⁴. Tento model se snaží abstrahovat od konkrétních způsobů popisu a snaží se stanovit právě takové entity, jejichž pomocí by bylo možné výše uvedené a další příklady kompatibilně řešit. Model hovoří o tzv. CRM entitě, k níž se vztahují další významné entity; na nejvyšší úrovni to jsou časová entita, existence, časový interval, místo a rozměr. Jde o poměrně složitý koncept²⁵, nyní již ISO standard, který se v podstatě snaží vyjádřit komplexitu vztahů mezi popisnými prvky na vyšší úrovni, než jsou příslušná popisná pravidla. Model je schopen vyjádřit například právě vztah času k různým popisným prvkům nebo různý vztah osob k různým popisným prvkům.

Tento model po aplikaci na konkrétní popisná pravidla dovoluje formálně vyjádřit znalost těchto pravidel (sad dat) pro vyšší úroveň aplikací. Tato znalost je uložena ve znalostní databázi a říká se jí ontologie. Pokud je v určitém mediačním softwaru obsažena tato znalost o různých heterogenních zdrojích dat, mediační software ví, jak s každým zdrojem zacházet, aby dokázal dát uživateli sémanticky kompatibilní odpověď na jeho dotaz směřující simultánně do různých prostředí. Mediátor v rámci dotazování stěhuje data mezi poli a entitami, přejmenovává pole, překládá jejich hodnoty atp. Velice důležité je propojení hierarchií CRM entit na příslušné oborové tezaury.

Nicméně, tyto otázky je třeba řešit nad bohatými zdroji dat, které je třeba teprve vytvořit. K tomu je třeba založit pracovní skupinu, která by zpracovala datový model (formát) digitalizovaných muzeálních objektů v těsné vazbě na již běžící programy z oblasti knihoven. Teprve větší množství dobře strukturovaných a po síti přístupných dat pak otevírá plodnou cestu od obecných informačních portálů k badatelsky dobře využitelným integrovaným informačním prostředím.

Poznámky pod čarou:

- 1) VRA Core Categories, Version 3.0, a project of the Visual Resources Association Data Standards Commit-

- tee. URL: <http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm> nebo záznam média v: AMICO Data Specification. Art Museum Image Consortium, 1999. (URL: <http://amico.org>)
- 2) Reference Manual for the MASTER Document Type Definition. URL: <http://www.hcu.ox.ac.uk/TEI/Master/Reference/>
 - 3) Viz například Museum Information Standards. URL: <http://www.diffuse.org/museums.html> nebo Museum Information Standards / Les normes et standards documentaires pour les musées. URL: <http://www.cidoc.icom.org/stand0.htm>
 - 4) AMICO Data Specification. Art Museum Image Consortium, 1999. (URL: <http://www.amico.org>)
 - 5) Categories for the Description of Works of Art. Definitions. J. Paul Getty Trust – College Art Association. 34 s.
 - 6) Introduction to Object ID. Guidelines for Making Records that Describe Art, Antiques, and Antiquities / Robin Thornes with Peter Dorrell and Henry Lie. The J. Paul Getty Trust, 1999.
 - 7) Système descriptif des objets mobiliers / Sous la direction de Hélène Verdier par Aline Magnien, ... Paris, Inventaire general, E.L.P., Editions du patrimoine, 1999. 372 s.
 - 8) CIDOC Guidelines for Museum Object Information: The Information Groups and Categories. URL: <http://www.cidoc.icom.org/guide/guidecat.htm>
 - 9) Van Eyck Core Record Structure: Proposal for a Standard. URL: <http://www.rkd.nl/pblctns/stndr-e.htm>
 - 10) SPECTRUM: The UK Museum Documentation Standard. URL: <http://www.mda.org.uk/spectrum.htm>
 - 11) Digitised European Periodicals. URL: <http://gdz.sub.uni-goettingen.de/dieper/>
 - 12) RLG Cultural Materials Alliance Description Guidelines, Version 1.0.1. 2 November 2001. URL: <http://www.rlg.org/culturalres/descguide.html>
 - 13) BiblioML. An XML application for bibliographic records, based on the Unimarc Bibliographic Format, and for authority records, based on UNIMARC / Authorities // Une application XML pour des références bibliographiques, basée sur le Format Bibliographique Unimarc, et pour des notices d'autorités, basée sur Unimarc / Autorités. Creation: December 25th, 1999; last modification date : May 5th, 2001 // Création : 25 décembre 1999 ; dernière mise à jour : 5 mai 2001. URL: <http://www.culture.fr/BiblioML/>
 - 14) SPECTRUM XML-DTD Testbed Project. URL: http://www.cimi.org/wg/xml_spectrum/xml_testbed_desc.html
 - 15) RLG Cultural Materials Alliance Description Guidelines, Version 1.0.1. 2 November 2001. URL: <http://www.rlg.org/culturalres/descguide.html>
 - 16) VRA Core Categories, Version 3.0, a project of the Visual Resources Association Data Standards Committee. URL: <http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm>
 - 17) Spec ICC.1:1998-09, File Format for Color Profiles. International Color Consortium, 1998. 134 s.
 - 18) NISO Draft Standard. Data Dictionary. Technical Metadata for Digital Still Images. Working Draft, 1.0, July 5, 2000. 45 s. URL: <http://www.niso.org/pdfs/DataDict.pdf>
 - 19) DIG35 Specification. Metadata for Digital Images. Version 1.1, June 18, 2001. Digital Imaging Group, 2001. 219 s.
 - 20) TIFF™ – Revision 6.0, Final – June 3, 1992. Mountain View, Adobe Developers Association, 1992. 121 s.
 - 21) URL: http://www.nkp.cz/o_knihovnach/konsorcia/VISK/VISK6.htm
 - 22) URL: http://www.nkp.cz/o_knihovnach/konsorcia/VISK/VISK7.htm
 - 23) Digitization of Rare Library Materials: Storage and Access to Data / Project Management by Adolf Knoll and Stanislav Psohlavec. Authors: Adolf Knoll, Stanislav Psohlavec, Jan Mottl, Jan Vomlel, Tomáš Mayer, ... Prague, National Library – Albertina income Praha, 1999. CD-ROM (Memoriae Mundi Series Bohemica). URL: <http://digit.nkp.cz/rare99/enter.htm>
 - 24) Definition of the CIDOC object-oriented Conceptual Reference Model. Version 3.2 / Editors: Nick Crofts, Ifigenia Dionissiadou, Martin Doerr, Matthew Stiff. B. m., ICOM/CIDOC Documentation Standards Group – CRM Special Interest Group, 2001. 13, 123 s.
 - 25) Electronic Communication on Diverse Data – The Role of the oo CIDOC Reference Model / Martin Doerr, Nicholas Croft. URL: <http://cidoc.ics.forth.gr/docs/crmrole.htm>