

Standardy v prezentaci digitalizovaných sbírek

Veronika Mikešová

zprávy

Presented article discusses current trends in the presentation of digitized museum collections. The author focuses on the use of metadata and explains why this issue is so important for museum work. Addition of metadata to digitized collections is nowadays a unique way to preserve the long-term usability of information about collections object. To ensure the effectivity of using metadata it is necessary to keep a unified standard both of the format of the metadata and of the form of the digitized objects themselves.

Keywords: Digitisation, Standards, Metadata, Multimedia

V současné době se v souvislosti s prezentací digitalizovaných sbírek stále více hovoří o metadatach. Metadata jsou strukturovaná data o datech, která umožňují snadné a rychlé vyhledávání ve velkém množství informací. Může se tedy jednat například o popisné informace, které obsahují jednotlivé objekty v určité databázi, nebo také informace, které jsou uvedeny ve zdrojovém kódu a poskytují základní údaje o dané webové stránce. V době digitalizace sbírek je třeba metadata využívat i v muzeu, neboť každý sbírkový předmět či dokument nese konkrétní datum vzniku, autora, formát a také autorská práva. *A právě všechny tyto informace se dají jasně a přehledně shrnout ve formě metadat, která by měla mít maximální a co nejširší srozumitelnost, a to nejen v současnosti, ale také v dlouhodobém časovém měřítku. Z tohoto důvodu jsou vytvářeny standardy pro jejich tvorbu. Za posledních dvacet let takových standardů vzniklo hned několik.*

Celosvětově nejrozšířenější je soubor metadatových prvků zvaný Dublin Core,¹ který je v současnosti využíván v několika desítkách zemí Evropy, Ameriky, Asie i Austrálie. Svůj název dostal tento způsob označování metadat podle města Dublin ve státě Ohio, ve kterém se v roce 1995 konala konferen-

ce, na které byl Dublin Core navržen. Jeho záměrem je usnadnit vyhledávání elektronických zdrojů. Původně byl vytvořen pro použití pouze na webu, postupně se ale díky své univerzálnosti a srozumitelnosti rozšířil na knihovny a posléze i na popisy zdrojů, které využívají státní i soukromé instituce včetně muzeí.

Dalším, v muzeích také hojně rozšířeným, standardem je Spectrum,² které bylo navrženo v roce 1994 ve Velké Británii jako metadatový standard pro dokumentaci muzejních sbírek. Umožňuje přiřazovat sbírkovým předmětům navíc popisné informace o způsobu jejich nabytí, zápůjčkách a reprodukčních právech. Podobným standardem byl také museumdat, jenž byl vytvořen německou pracovní skupinou.³

V současné době se však v muzejní prezentaci stále více ujímá standard LIDO (Lightweight Information Describing Objects).⁴ Ten vznikl v roce 2008 za mezinárodní spolupráce mnoha pracovníků muzeí za účelem jednotného označení shromážděných digitalizovaných muzejních sbírek. Standard LIDO byl vytvořen tak, aby propojil starší standardy jako je Spectrum či museumdat, a sjednotil tak již existující metadatové prvky. Tento standard však není vhodný pro

1 Více o standardu Dublin Core na http://www.ics.muni.cz/dublin_core/index.html, http://www.webarchiv.cz/generator/dc_locale/cz/napovedalobecnapravidla.html.

<http://jib-info.cuni.cz/tematbrany/doporuceni.html>.

2 Více na <http://www.collecti-onslink.org.uk/programmes/spectrum>.

3 Více na <http://www.museum-dat.org/>.

4 Více na <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/>.

Mgr. Veronika Mikešová
Národní muzeum
Historické muzeum
veronika_mikesova@nm.cz

katalogizaci sbírek. Je navržen pro harvestování dat ze sbírek evropských muzeí a jejich online prezentaci.⁵

Metadata můžeme v základu rozdělit do tří kategorií: informace o obsahu, intelektuálním vlastnictví a identifikačních údajích zdroje. Jednotlivé složky daných kategorií (dle standardu Dublin Core) jsou uvedeny v tabulce 1.

Jedná se tedy o logické informace, které lze uvést k většině databázových objektů. Budeme-li všichni používat stejný standard popisných informací, usnadníme tak nejen práci s digitalizovanými objekty našim kolegům v jiných českých i evropských muzeích, ale také vzájemné předávání informací mezi jednotlivými portály. Na tomto metadatovém principu pracuje český portál digi-

talizovaného kulturního dědictví eSbírkky, kde mohou jednotlivá muzea publikovat své sbírky, ale také celoevropský portál Europeana⁶, kam je možno data z eSbírek zdarma převádět.

Používání metadat také podporuje současný trend otevřených dat (tzv. open data). To jsou všechna data, která zveřejňují s použitím informačních technologií orgány veřejné správy ve formátu umožňujícím jejich další strojové zpracování. Šíření těchto dat není omezeno licenční ani jinou právní podmínkou, jsou tedy volně dostupné i k opakovanému užití. A právě otevřenými daty by se díky digitalizaci a užívání metadat měly stát i muzejní sbírky, aby tak široká veřejnost získala přístup ke svému kulturnímu dědictví.

Tab. 1: Metadatové prvky podle standardu Dublin Core

Obsah	Intelektuální vlastnictví	Identifikační údaje zdroje
Název (vyjádřen konkrétním názvem dokumentu či souboru)	Tvůrce (autor dokumentu, fotografie, obrazu,...)	Datum (datum zpřístupnění objektu ve stávající podobě)
Předmět (vyjádřen klíčovými slovy, charakterizujícími daný objekt, jednotlivá slova se oddělují čárkou)	Vydavatel (osoba či instituce zodpovědná za obsah zdroje)	Formát (velikost a formát objektu – např. JPEG, 2 MB)
Popis (podrobnější informace o objektu)	Příspěvatel (osoba či instituce, která se podílela na vytvoření objektu)	Identifikátor zdroje (unifikovaný lokátor zdroje – URL či např. u knih ISBN)
Typ zdroje (uvádí, o jaký objekt se jedná – fotografie, obraz, archiválie, webová stránka, sbírkový předmět,...)	Správa autorských práv (komu náleží současná autorská práva)	Jazyk (dvouznakový kód jazyky dle normy ISO 639)
Zdroj (odkaz na zdroj, kde je digitalizovaný objekt umístěn)		
Vztah (odkazy na příbuzné objekty a informace)		
Pokrytí (rozsah obsahu zdroje – geografický, časový, prostorový)		

⁵ Standard LIDO byl využit pro evropský projekt Athena (Access to cultural heritage networks across Europe), který podporoval muzea v poskytování digitálního obsahu Europeana.

⁶ Sbírky i Europeana využívají metadatové prvky Europeana Semantic Elements (ESE), které jsou založeny na standardu Dublin Core – <http://www.europeana.eu/schemas/ese/>.

Je-li objekt či webová stránka opatřena metadaty, jsou napevno zakotveny v HTML kódu (ten si můžeme prohlédnout, zvolíme-li v internetovém prohlížeči možnost zobrazit zdrojový kód) hned v úvodu v sekci začínající tagem <head>. Díky takto zakódovaným metadatům, která obsahují i klíčová slova, se dají konkrétní objekty snadno vyhledat v nepřehledném množství informací na celosvětovém internetu. Zpřístupníme-li tedy objekty s metadaty, zabráníme tak tomu, aby se časem „utopily“ v gigantickém informačním prostoru.

Jako názorný příklad si můžeme uvést, jak vypadají metadata ve zdrojovém kódu dokumentu o standardu Dublin Core. Prostudujeme-li si kód důkladně, nalezneme v něm všechny prvky, popsané v tabulce č. 1:

```
<meta name="DC.Title" content="Dublin Core Czech Homepage" />
```

```
<meta name="DC.Creator" content="ÚVT MU" />
```

```
<meta name="DC.Subject" content="Dublin Core" />
```

```
<meta name="DC.Subject" content="metadata" />
```

```
<meta name="DC.Subject" scheme="PHNK" content="metadata" />
```

```
<meta name="DC.Subject" scheme="PHNK" content="metadatové formáty" />
```

```
<meta name="DC.Subject" scheme="PHNK" content="Dublin Core" />
```

```
<meta name="DC.Description.abstract" content="České stránky iniciativy Dublin Core. Obsahují přeložené dokumenty týkající se metadatového formátu Dublin Core a další informace pro české uživatele Dublin Core." />
```

```
<meta name="DC.Date.available" scheme="W3C-DTF" content="2006-08-23" />
```

```
<meta name="DC.Type" scheme="DCMIType" content="Text" />
```

```
<meta name="DC.Format" scheme="IMT" content="text/html" />
```

```
<meta name="DC.Format.medium" content="computerFile" />
```

```
<meta name="DC.Language" scheme="RFC3066" content="cze" />7
```

Další příklad uveďme již pro oblast muzejnictví. Jako vzor vybereme měděnou hřivnu z archeologické sbírky Národního muzea. Metadata, vztahující se k tomuto předmětu by podle standardu Dublin Core vypadala následovně:

```
<meta name="DC.Title" content="Hřivna" />
```

```
<meta name="DC.Subject" content="doba bronzová" />
```

```
<meta name="DC.Subject" content="měď" />
```

```
<meta name="DC.Subject" content="hřivna" />
```

```
<meta name="DC.Description.abstract" content="Hřivna nákrčníkovitého typu" />
```

```
<meta name="DC.Publisher" content="Národní muzeum" />
```

```
<meta name="DC.Type" scheme="DCMIType" content="PhysicalObject" />
```

```
<meta name="DC.Format" scheme="IMT" content="image/jpeg" />
```

```
<meta name="DC.Source" scheme="URL" content="http://www.esbirky.cz/detail/26675/" />
```

```
<meta name="DC.Rights" content="Národní muzeum" />
```

⁷ Zdrojový kód z http://www.ics.muni.cz/dublin_core/charakteristika.html.

Pozor je ale třeba dávat na pojmy, které použijeme u jednotlivých popisných prvků. Měli bychom se snažit dodržovat základní slovníkové názvy, obzvláště u klíčových slov. V češtině lze využít podrobné hesláře, které se již používají v knihovnictví.⁸ Standardizovaný heslář, který by obsahoval popisné pojmy pro oblast muzejnictví, však dosud chybí. Do budoucna by tedy jistě bylo užitečné takový heslář vytvořit, a zajistit tak co největší interoperabilitu sbírkových informací.

Hovoříme-li o standardech v prezentaci digitalizovaných sbírek, musíme kromě popisných metadat zmínit také samotné formáty digitalizovaného obsahu. Veškerá obrazová dokumentace, která je určena pro prezentaci, by měla být ve formátu JPEG (Joint Photographic Experts Group). Ten je nejjednodušším a nejpoužívanějším formátem obrazových informací a práci s ním podporují všechny softwary. Naopak formát TIFF (Tag Image File Format) je vhodný pro archivaci a tisk obrázků, neboť zachovává jejich původní kvalitu a soubory, na rozdíl od formátu JPEG, nekomprimuje.

Co se týče audio souborů, pro prezentaci je určený formát MP3 (založený na algoritmu skupiny MPEG – Moving Picture Experts Group). Ten umožňuje výrazně snížit velikost zvukových souborů, aniž by výrazně zhoršil jejich kvalitu. Tento formát lze tedy velice snadno sdílet a prezentovat. U video souborů je pak nejvhodnějším formátem AVI (Audio Video Interleave) či MPEG (Moving Pictures Experts Group), které jsou nejrozšířenější a dokáže s nimi pracovat téměř každý software. Pro online prezentaci na webu a streamování videí se nyní stále hojněji používá formát FLV (Flash Video). Ten je podporovaný větši-

nou operačních systémů i prohlížečů, hodí se tedy pro videa, která chceme volně šířit.

Na rozdíl od metadatových standardů, které jsou tvořeny tak, aby byly použitelné po dlouhé roky, multimediální standardy se neustále mění. Výše uvedené formáty tak jsou aktuální nyní, za několik let však nejspíše budeme používat úplně jiné. To je však nezvratná situace v oboru prezentace digitalizovaných sbírek a nám nezbyvá, než se přizpůsobit.

Literatura:

ATHENA WP3: Digitisation: Standards landscape for european museums, archives, libraries. Řím 2009.

ATHENA WP3. Lightweight information describing objects (LIDO): The international harvesting standard for museums. Řím 2011.

CIGÁNEK, D. Digitální dokumentace objektů kulturní, historické a vědecké hodnoty. Brno: Moravské zemské muzeum, 2005 [cit. 16. 11. 2012]. [PDF dokument]. Dostupný z: <http://www.citem.cz/wp-content/plugins/downloads-manager/upload/digitalizace.pdf>.

CHLAPEK, D. – KUČERA, J. Otevřená data veřejné správy. Praha: Vysoká škola ekonomická v Praze, 2011 [cit. 21. 11. 2012]. [PDF dokument] dostupný z: <http://www.datakon.cz/media/2011/prezentace/studie/chlappek.pdf>.

Data harvesting and interchange. network.icom.museum [www portál]. Dostupné z: <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange> [cit. 15. 11. 2012].

⁸ Seznam českých tezaurů např. na http://www.nkp.cz/pages/page.php3?page=fond_kodovnik.htm, světových na <http://www.citem.cz/wp-content/plugins/downloads-manager/upload/14-TezaurySlovnikyMezinarodni.pdf>.

- Dublin Core. ics.muni.cz [www portál].
Dostupné z: http://www.ics.muni.cz/dublin_core/ [cit. 14. 11. 2012].
- Europeana Semantic Elements. Europeana.eu [www portál]. Dostupné z: <http://www.europeana.eu/schemas/ese> [cit. 15. 11. 2012].
- Metadata standards and interoperability. JISC Digital Media [www portál]. Dostupné z: <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/metadata-standards-and-interoperability> [cit. 14. 11. 2012].
- Metodické centrum pro informační technologie v muzejnictví [www stránky]. Dostupné z: <http://www.citem.cz> [cit. 16. 11. 2012].
- museumdat. museumdat.org [www stránky]. Dostupné z: <http://www.museumdat.org/> [cit. 14. 11. 2012].
- Návrh požadavků na metadata používaná pro provoz oborových informačních bran v České republice. jib-info.cuni.cz [www portál]. Dostupné z: <http://jib-info.cuni.cz/tematbrany/doporuceni.html> [cit. 1. 11. 2012].
- Spectrum standard. Collections Link [www portál]. Dostupné z: www.collectionslink.org.uk/programmes/spectrum.
- What is LIDO? network.icom.museum [www portál]. Dostupné z: <http://lido-schema.org> [cit. 15. 11. 2012].