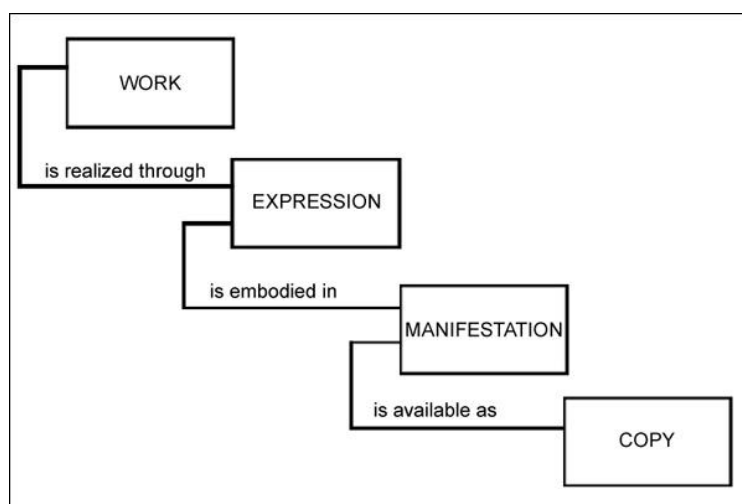


## Digitálna knižnica zvukových záznamov

Digitálne knižnice sa za posledné roky stali veľmi využívaným zdrojom rôznorodých informácií. Základným stavebným kameňom digitálnych knižníc je podľa Kahn-Wilenského architektúry digitálny objekt. Za digitálny objekt sa považuje samostatná informačná jednotka, ktorá je tvorená jednoznačným identifikátorom, metadátami a samotným obsahom. Táto práca je zameraná na digitálne objekty, ktorých obsah je reprezentovaný pomocou zvukových záznamov a na digitálne knižnice, ktoré takéto objekty tvoria [1].

Základným kameňom každého digitálneho systému je vymedzenie jeho architektúry. Architektúra digitálnej knižnice býva najčastejšie vyjadrená jej dátovým modelom. Charakter každej digitálnej knižnice je často nejakým spôsobom odlišný. Líšiť sa môže napríklad vo formáte uchovávaných digitálnych objektov, ich sprístupňovaní, uložení, atď. Kvôli interoperabilite by však architektúry jednotlivých digitálnych knižníc mali byť čo najviac podobné. Preto prišla organizácia IFLA v roku 1998 s modelom FRBR (obrázok č.1) [2]. FRBR (Functional Requirements for Bibliographic Records) má za úlohu simulovať hierarchiu intelektuálnych prác. Tento model sa skladá zo štyroch druhov abstrakcií. Každé dielo, vyjadrenie, prejav ale aj jednotka by mala mať svoj identifikátor a popisné, štrukturálne a administratívne metadáta. Tento model je však primárne tvorený pre tlačene monografie, preto nie je vždy úplne vhodný pre popis zvukového intelektuálneho diela.



*Obrázok č.1: Model FRBR [2]*

Na základe modelu FRBR sa dajú intelektuálne diela realizovať rôznymi spôsobmi. Napríklad album „Amnesiac“ od skupiny „Radiohead“. V tomto prípade sa za dielo môže považovať celé album, za vyjadrenie štúdiová nahrávka albumu, za prevedenie

jeho špeciálna verzia pre Veľkú Britániu a za jednotku CD so sériovým číslom „xyz“. Takúto interpretáciu používajú napríklad knižnice. Výhodou tejto interpretácie je hlavne ľahká čitateľnosť (človekom), avšak strojové spracovanie takto štruktúrovanej informácie nemusí byť príliš efektívne. Ďalšou nevýhodou je napríklad to, že táto interpretácia považuje jednotlivé skladby za pevné súčasti daného diela, pričom v praxi to tak často nebýva. Napríklad skladba „Pyramid song“ sa nachádza na albume „Amnesiac“ ale aj na albume „Best of“. V reálnom svete sa jedná o jednu a tú istú skladbu, vo svete digitálnych knižníc to tak môže, ale nemusí byť. Páve kvôli tomu existujú rôzne spôsoby interpretovania týchto metadát pomocou rôznych štandardov a metadátových formátov. V nasledujúcich oddieloch sa budem venovať tým, ktoré sú používané v digitálnych knižniciach zvukových záznamov.

## Rodina štandardov MARC

Na popis zvukových nahrávok, sa hlavne vo svete knižníc, používa metadátový štandard MARC 21 a jeho XML verzia MARXML [3]. Tieto formáty slúžia na vyjadrenie deskriptívnych (popisných) metadát. Často sú však pomocou formátov MARC vyjadrené aj štrukturálne a administratívne metadáta formou voliteľných podpolí určených pre alternatívny záznam.

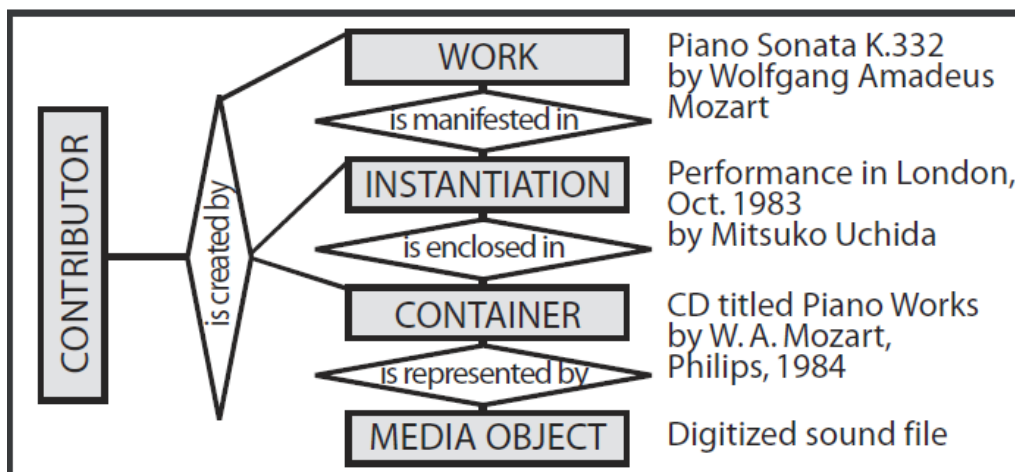
Takto utvorené metadáta sú bohaté na informácie a zároveň dobre čitateľné (hlavne vo formáte MARCXML). Kvôli týmto dôvodom sa považuje formát MARC za bezpečnú formu dlhodobého uchovávaní metadát. Naopak nevýhodou formátov MARC je neefektívne strojové spracovanie na moderných počítačoch, nie úplne priamočiara podpora diel na úrovni zvukových skladieb a častá duplicita dát. V prípade zapisovania štrukturálnych a administratívnych metadát môže dôjsť k častej nejednoznačnosti obsahu (môžu sa líšiť na úrovni národov, knižníc alebo dokonca aj samotných katalogizátorov). Náročné je aj tieto dáta následne upravovať, väčšinou sa chybné pole alebo podpole necháva v zázname s upozornením, že ide o chybné informácie (v lepšom prípade) [4].

Formáty MARC využíva napríklad databáza freeDB.org<sup>1</sup>, rozsiahla metadátová databáza zvukových nahrávok a napríklad aj digitálna knižnica Variations<sup>2</sup> Indianskej univerzity. Digitálna knižnica Variations sa snaží zjednotiť formu zvukových digitálnych knižníc. Variations je postavená na upravenom modeli FRBR a na formáte USMARC. Dátový model digitálnej knižnice Variations považuje za diela jednotlivé skladby zvukových nahrávok (obrázok č.2) [5].

---

<sup>1</sup> <http://www.freedb.org/en/>

<sup>2</sup> <http://dml.indiana.edu/>



Obrázok č.2: Dátový model digitálnej knižnice Variations [5]

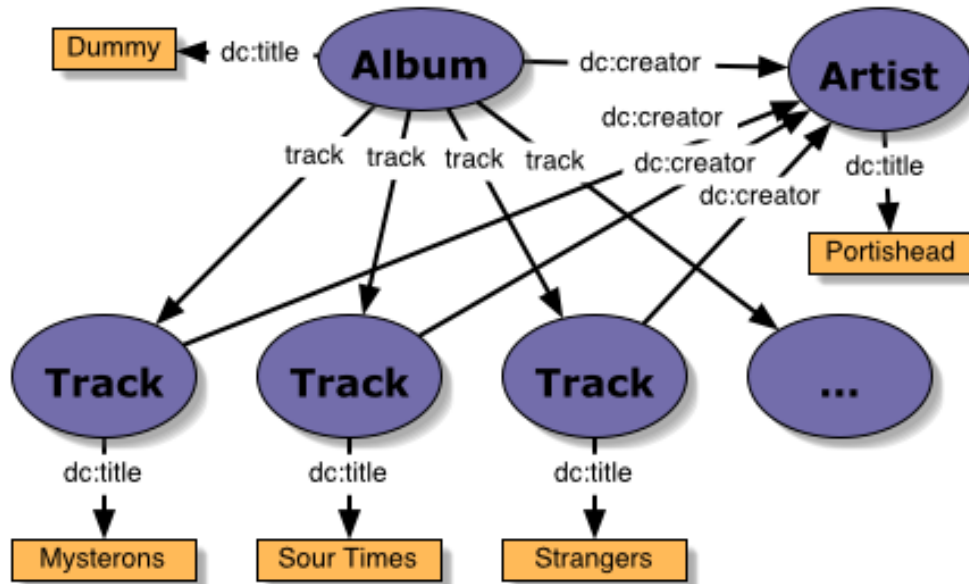
## Dublin Core a RDF

RDF (Resource Description Framework) sa považuje za základný štandard sémantického webu. V spojení s formátmi sieťových metadát môže RDF tvoriť veľmi komplexný spôsob popisu dát na webe. Vo svete zvukových digitálnych knižníc býva najčastejšie formát RDF nositeľom popisných metadát vo formáte kvalifikovaného alebo nekvalifikovaného Dublin Core (DC). DC je dobre definovaný, jednoduchý a štruktúrovaný štandard. Táto kombinácia je vhodná na popis zvukových záznamov na úrovni skladieb. Takisto je veľmi jednoduché dopĺňať, alebo opravovať štruktúru dát v RDF pomocou nových alebo upravených metadát v DC. V zvukových digitálnych knižniciach je toto riešenie veľmi populárne ale takisto obsahuje niekoľko väčších, či menších nedostatkov. Napríklad absencia administratívnych či štruktúrnych informácií v špecifikácii DC. Administratívne a štruktúrne metadáta sú u zvukových záznamov obzvlášť dôležité. V prípade použitia DC a RDF sa buďto metadáta dopĺňajú o iné formáty metadát, čo ešte viac komplikuje už tak komplexnú štruktúru záznamov v tomto riešení alebo sú zapisované ako nepodporované komponenty DC, čím narušujú jeho dobrú definovanosť. Ďalšou nevýhodou použitia RDF s DC je neistota ohľadom dlhodobého uchovávanía dát. Za štandardmi ako MARC stoja veľké a silné organizácie (prevažne knižnice), ktoré ich udržiavajú [6].

Na štandardoch RDF a DC je postavená napríklad otvorená encyklopédia hudobných informácií MusicBrainz<sup>3</sup>. Táto databáza vznikla v roku 2000, aby poskytovala voľne šíriteľné informácie o hudobných dielach. MusicBrainz je jeden s prvých projektov, ktorý reálne postavili svoj model na FRBR [7]. MusicBrainz sa snaží poskytovať informácie, ktoré sú určené predovšetkým pre obyčajných užívateľov, ktorý si chcú zistiť informácie o svojej obľúbenej skladbe alebo popísať svoju hudobnú kolekciu. Taktiež sa prispôbuje moderným trendom webu a snaží

<sup>3</sup> <http://musicbrainz.org/>

sa prepojiť svoje informácie so servermi ako je napríklad Wikipédia alebo freeDB. Často sa MusicBrainz uvádza ako príklad služby sémantického webu [8].



Obrázok č.3: Ukážka záznamu v Dátovom modeli MusicBrainz [8]

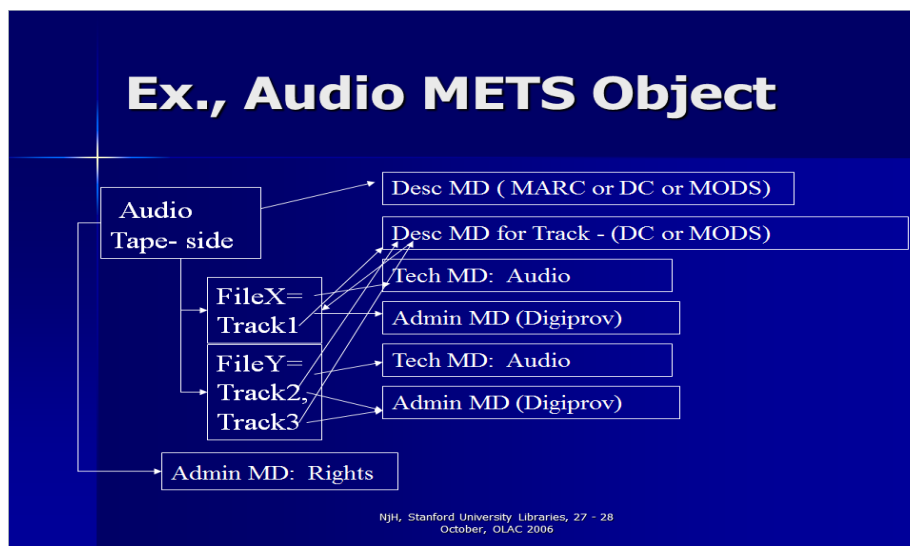
## METS a MODS/Dublin Core /MARC

METS (Metadata encoding & Transmission Standard) [9] je schéma pre deskriptívne, administratívne a štrukturálne metadáta. Využíva rôzne metadátové popisné štandardy ako napríklad MODS (Metadata Object Description Schema) [10], ktorý poskytuje kompromis medzi MARC a DC metadátami. Podporuje vnáranie metadátových popisov, čo je vhodné na popis jednotlivých skladieb hudobného diela. METS dokáže takisto dobre pracovať už so spomínanými metadátovými štandardmi MARC a DC. Popisné metadáta sa u zvukových diel dopĺňajú administratívnymi AdminMD a technickými metadátami pre hudbu AudioMD (Obrázok č.4) [11].

Najznámejší projekt, ktorý je postavený na METS je kolekcia metadát pre Monterey Jazz Festival<sup>4</sup>. Tento projekt mal ukázať funkčnosť schémy METS. Ďalšia služba postavená predovšetkým na štandarde MODS je MusicAustralia<sup>5</sup>.

<sup>4</sup> <http://www.montereyjazzfestival.org>

<sup>5</sup> <http://trove.nla.gov.au/general/australian-music-in-trove>



Obrázok č.4: Štruktúra zvukových metadát v METS [11]

## Identifikátory zvukových záznamov

Dôležitým prvkom identifikácie zvukových záznamov v digitálnych knižniciach sú ich identifikátory. Zvukové záznamy sú identifikované klasickými, digitálnymi ale aj rôznymi špecializovanými identifikátormi. Zatiaľ čo klasické a digitálne identifikátory slúžia prevažne pre jednoznačnú identifikáciu diela a digitálneho objektu, špeciálne identifikátory pomáhajú poskytovať komplexnejšie informácie o zvukových nahrávkach. Napríklad v už spomínanej databáze MusicBrainz sa pri dielach vyskytuje identifikátor MusicBrainz Identifier (MBID) [12], ale takisto aj identifikátory tretích strán ako napríklad discogs\_id<sup>6</sup>, allmusic id<sup>7</sup>, BBC Music ID, alebo perzistentné URL na WikiData. Metadáta autorov bývajú navyše obalené identifikátorom autority (viaf.org<sup>8</sup>). Týmto spôsobom sa užívateľ rýchlo a jednoducho dostane k užitočným a presným informáciám ohľadom hľadaného diela.

## Záver

Svet zvukových digitálnych knižníc je veľmi rozmanitý. Táto práca zďaleka nepopisuje všetky možné technológie a riešenia spojené z digitálnymi knižnicami zvukových záznamov. Ukazuje však, že v prípade tvorenia takejto digitálnej knižnice existuje mnoho spôsobov, ktoré majú svoje klady, ale takisto aj zápory.

<sup>6</sup> Databáza Discogs sa zaoberá popisom a predajom zvukových albumov. Odkaz:

<http://www.discogs.com/>

<sup>7</sup> Server poskytujúce hĺbkové informácie o zvukových dielach, ich vzorky, hodnotenia, ... Odkaz:

<http://www.allmusic.com/>

<sup>8</sup> Virtuálna medzinárodná databáza autorít. Odkaz: <http://viaf.org/>

## Literatúra

- [1] BARTOŠEK, Miroslav. Digitální knihovny-teorie a praxe. *Národní knihovna: knihovnická revue*, 2005, 15.4: 233-254.
- [2] TILLET, Barbara. What is FRBR? A Conceptual Model for the Bibliographic Universe. *Technicalities*. Los Angeles, 2003, roč. 25, č. 5, s. 8. Dostupné z: <http://www.loc.gov/cds/downloads/FRBR.PDF>
- [3] LIBRARY OF CONGRESS. *MARC STANDARDS: Library of Congress - Network Development and MARC Standards Office* [online]. 2013 [cit. 2013-12-01]. Dostupné z: <http://www.loc.gov/marc/>
- [4] LESTYAN, Tomáš. Návrh infrastruktury pro virtuální Národní fonotéku [online]. 2013 [cit. 2013-12-01]. Bakalářská práce. Masarykova univerzita, Fakulta informatiky. Vedoucí práce Petr Žabička. Dostupné z: <[http://is.muni.cz/th/359357/fi\\_b/](http://is.muni.cz/th/359357/fi_b/)>.
- [5] MINIBAYEVA, Natalia a Jon W. DUNN. A digital library data model for music. *Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries - JCDL '02* [online]. New York, New York, USA: ACM Press, 2002, s. 154- [cit. 2013-11-15]. DOI: 10.1145/544220.544249. Dostupné z: <http://portal.acm.org/citation.cfm?doid=544220.544249>
- [6] Open Metadata Handbook/Technical Overview. *Wikibooks* [online]. 2013 [cit. 2013-12-03]. Dostupné z: [http://en.wikibooks.org/wiki/Open\\_Metadata\\_Handbook/Technical\\_Overview](http://en.wikibooks.org/wiki/Open_Metadata_Handbook/Technical_Overview)
- [7] FRBR. MUSICBRAINZ. *MusicBrainz Wiki* [online]. 2012 [cit. 2013-12-03]. Dostupné z: <http://wiki.musicbrainz.org/FRBR>
- [8] MusicBrainz: A Semantic Web Service. *LogicError* [online]. LogicError [cit. 2013-12-03]. Dostupné z: <http://logicerror.com/musicbrainzArticle>
- [9] LIBRARY OF CONGRESS. *METS: Metadata encoding & Transmission Standard Official Web Site* [online]. 2013 [cit. 2013-12-03]. Dostupné z: <http://www.loc.gov/standards/mets/>
- [10] LIBRARY OF CONGRESS. *MODS: Metadata Object Description Schema Official Web Site* [online]. 2013 [cit. 2013-12-03]. Dostupné z: <http://www.loc.gov/standards/mods/>
- [11] HOEBELHEINRICH, Nancy J. STANFORD UNIVERSITY LIBRARIES. *Gathering Audio Metadata for the Monterey Jazz Festival Concerts*. Stanford, 2006. Dostupné z: <https://www.google.cz/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC0QFjAA&url=http%3A%2F%2Fwww.loc.gov%2Fstandards%2Fmets%2Fpresentations%2>

[Fnancysolac%2FOLAC\\_2006\\_MJF.ppt&ei=HUOeUqbYJoLPtAbN9ICoBg&usg=AFQjCNG-x13MMj3GqRLqpRYyM2UKvA16TA&bvm=bv.57155469,d.Yms](http://fnancysolac%2FOLAC_2006_MJF.ppt&ei=HUOeUqbYJoLPtAbN9ICoBg&usg=AFQjCNG-x13MMj3GqRLqpRYyM2UKvA16TA&bvm=bv.57155469,d.Yms)

[12] MusicBrainz Identifier. MUSICBRAINZ. *MusicBrainz* [online]. 2012 [cit. 2013-12-03]. Dostupné z: [http://musicbrainz.org/doc/MusicBrainz\\_Identifier](http://musicbrainz.org/doc/MusicBrainz_Identifier)

## Metadáta

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Digitálna knižnica zvukových záznamov" />
<meta name="DC.Creator" content="Tomáš Lestyan" />
<meta name="DC.Subject" content="Audio Digital Object" />
<meta name="DC.Subject" content="Metadata" />
<meta name="DC.Subject" content="Identifier" />
<meta name="DC.Subject" content="MARC" />
<meta name="DC.Subject" content="RDF" />
<meta name="DC.Subject" content="Dublin Core" />
<meta name="DC.Subject" content="METS" />
<meta name="DC.Subject" content="MODS" />
<meta name="DC.Subject" content="MusicBrainz" />
<meta name="DC.Format" scheme="IMT" content="application/pdf" />
<meta name="DC.Description.tableOfContents" content="Táto práca popisuje jednotlivé riešenia zvukových digitálnych knižníc, uvádza ich príklady a hodnotí klady a zápory jednotlivých riešení" />
<meta name="DC.Date" content="4.12.2013" />
<meta name="DC.Source" scheme="URL"
content="http://www.loc.gov/cds/downloads/FRBR.PDF" />
<meta name="DC.Source" scheme="URL" content="http://www.loc.gov/marc/" />
<meta name="DC.Source" scheme="URL"
content="http://portal.acm.org/citation.cfm?doid=544220.544249" />
<meta name="DC.Source" scheme="URL" content="
http://is.muni.cz/th/359357/fi_b"/>
<meta name="DC.Source" scheme="URL"
content="http://en.wikibooks.org/wiki/Open_Metadata_Handbook/Technical_Overview"/>
<meta name="DC.Source" scheme="URL"
content="http://wiki.musicbrainz.org/FRBR" />
<meta name="DC.Source" scheme="URL"
content="http://logicerror.com/musicbrainzArticle " />
<meta name="DC.Source" scheme="URL"
content="http://www.loc.gov/standards/mets/" />
```

```
<meta name="DC.Source" scheme="URL"
content="http://www.loc.gov/standards/mods/" />
<meta name="DC.Source" scheme="URL" content="
https://www.google.cz/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC0QFj
AA&url=http%3A%2F%2Fwww.loc.gov%2Fstandards%2Fmets%2Fpresentations%2
Fnancysolac%2FOLAC_2006_MJF.ppt&ei=HUOeUqbYJoLPtAbN9ICoBg&usg=AFQj
CNG-x13MMj3GqRLqpRYyM2UKvA16TA&bvm=bv.57155469,d.Yms/>
<meta name="DC.Source" scheme="URL"
content="http://musicbrainz.org/doc/MusicBrainz_Identifier" />
<meta name="DC.Language" scheme="RFC3066" content="sk" />
```