

Univerzita Karlova v Praze

Filozofická fakulta

Ústav informačních studií a knihovnictví

Studijní program: informační studia a knihovnictví

Studijní obor: informační studia a knihovnictví

Bakalářská práce

Martina Čulíková

**Ontologie využívané při organizaci webového obsahu se zaměřením na
oblast výtvarného umění**

Ontology used in the organization of web content focused on visual arts

Praha 2010

Vedoucí práce: PhDr. Helena Kučerová

Oponent bakalářské práce:

Datum obhajoby:

Hodnocení:

Zadání – pdf soubor

Děkuji PhDr. Heleně Kučerové za její cenné komentáře, rady a připomínky, své rodině za motivaci a Andrejovi Štefáněkovi za inspiraci.

Prohlášení:

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Praze, dne 22. července 2010

.....
podpis studenta

Identifikační záznam:

ČULÍKOVÁ, Martina. *Ontologie využívané při organizaci webového obsahu se zaměřením na oblast výtvarného umění = Ontology Used in the Organization of Web Content Focused on Visual Arts*. Praha, 2010-07-28. 79s., 12 s. příl. Bakalářská práce (Bc.). Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí bakalářské práce Helena Kučerová.

Abstrakt (česky)

Práce v prvních kapitolách uvádí do širšího kontextu ontologií, jakožto technologií sdílených formálních konceptuálních modelů pro systémy pořádání informací a znalostí. Definiuje nejdůležitější pojmy a termíny, představuje problematiku a vymezuje oblast zájmu, která je v rámci práce zahrnuta – především se zaměřuje na znalostní doménu výtvarného umění. V dalších kapitolách práce nastiňuje hlavní činnosti a procesy ontologického inženýrství, včetně obecně použitelné metodiky pro tvorbu ontologií. Dále přináší typologii ontologií členěnou podle několika různých kritérií. Nejpodstatnější částí práce je pak popis a charakteristika ontologií používaných při organizaci informací o výtvarném umění dostupných na webu. Stěžejními jsou pak sofistikované systémy – ontologie – ULAN, AAT, Iconclass a konceptuální model CIDOC CRM. V závěru se práce zaměřuje na otázky možné integrace těchto doménových ontologií do jednoho multifunkčního systému.

Abstract (English)

The paper in first chapters provides a broader context of ontology technologies as shared conceptual models for formal systems of organizing information and knowledge. It also defines key terms and concepts, introduces the issue of ontologies and defines areas of interest which are included in the paper - mainly focuses on the knowledge domain of fine art. In subsequent chapters it describes the main activities and processes of ontological engineering, including a generic methodology for building ontologies. Furthermore provides a typology of ontologies created by several different criteria. The most important part of the thesis is a description and characterization of ontologies used for organization of information about fine art which are available on the web. The cores of the paper are sophisticated systems - ontologies - ULAN, AAT, Iconclass and conceptual reference model CIDOC. In conclusion, the paper focuses on questions of possible integration of these domain ontologies into one multifunctional system.

Klíčová slova (česky):

ontologie, aplikace ontologií, systémy pořádání znalostí, Iconclass, AAT, ULAN, CIDOC CRM, ontologické inženýrství, metadata, výtvarné umění

Key words (English):

ontology/ontologies, application of ontologies, knowledge organization systems, Iconclass, AAT, ULAN, CIDOC CRM, ontological engineering, metadata, visual fine arts

OBSAH

PŘEDMLUVA: uvedení do otázky ontologií v prostředí WWW	10
1. ÚVOD	12
1.1. charakteristika, obsah a metodika práce.....	14
1.2. definice domény výtvarného umění.....	15
2. ZÁKLADNÍ KONCEPCE ONTOLOGIÍ	16
2.1. definice a termíny	16
2.1.1. ontologie	16
2.1.2. RDF a XML	19
2.1.3. deskripční logika.....	21
2.1.4. další pojmy.....	22
2.2. proč jsou ontologie vyvíjeny a používány v prostředí WWW	23
3. RŮZNÁ POJETÍ ONTOLOGIÍ V OBLASTI INFORMATIKY	25
3.1. formální typologie ontologií.....	25
3.2. co je a co není ontologie.....	26
3.2.1. řízené slovníky, glosáře, katalogy.....	27
3.2.2. taxonomie a tezaury	27
4. ONTOLOGICKÉ INŽENÝRSTVÍ	30
4.1. jazyky pro reprezentaci znalostí.....	31
4.2. obecná metodika pro tvorbu ontologií.....	33
5. APLIKACE ONTOLOGIÍ V OBLASTI VÝTVARNÉHO UMĚNÍ	35
5.1. ontologie pro popis uměleckých děl	35
5.1.1. Iconclass.....	35
5.1.1.1. vývoj a charakteristika	35
5.1.1.2. obsah	36
5.1.1.3. struktura a princip	37
5.1.1.4. použití Iconclassu v praxi.....	40
5.1.2. ULAN	41
5.1.2.1. vývoj a charakteristika	41
5.1.2.2. obsah	42
5.1.2.3. struktura a princip	45
5.1.2.4. použití ULANu v praxi	46
5.1.3. AAT	48
5.1.3.1. vývoj a charakteristika	49
5.1.3.2. obsah	50
5.1.3.3. struktura a princip	50
5.1.3.4. použití AAT v praxi	52
5.1.4. CIDOC CRM	54
5.1.4.1. vývoj a charakteristika CRM – informace vs. data	56
5.1.4.2. pole působnosti a použití CRM.....	57
5.1.4.3. použitá forma, struktura a principy modelování.....	58
5.1.4.4. CIDOC CRM a FRBR.....	60
5.1.4.5. použití CIDUCu CRM v praxi	61
5.2. ontologie pro vizuální popis uměleckých děl.....	64

5.2.1.	prostorový popis.....	64
5.2.2.	ontologický popis vizuálních rysů	65
5.3.	metadatová schémata pro popis uměleckých děl.....	65
5.3.1.	Dublin Core.....	66
5.3.2.	CDWA The Categories for the Description of Works of Art.....	66
	Obrázek 20 Tabulka s kategoriemi CDWA; [Breitman, 2006, s. 254].....	67
5.3.3.	ISO 21127	67
5.3.4.	VRA – The Visual Resources Association Core Categories.....	67
6 .	<i>I N T E R O P E R A B I L I T A O N T O L O G I Í</i>	69
6.1.	teoretický návrh integrace existujících ontologií do jednoho systému	69
6.2.	možný budoucí vývoj ontologií.....	71
7 .	<i>Z Á V Ě R</i>	73
	<i>Seznam použité literatury</i>	<i>74</i>

PŘEDMLUVA: uvedení do otázky ontologií v prostředí WWW

"Bylo by hezké, kdyby byl vyhledávač schopen chápání významu hledané fráze, a ne slova, která jsou ve frázi obsažena."

Eric Schmidt, generální ředitel Google Inc., leden 2009 [Brůcha, 2009]

Otázka, zda by výše citované obohacení neznámějšího webového vyhledávače o schopnost sémantiky bylo či nebylo hezké, je sice subjektivní, nicméně objektivně lze konstatovat, že technologie, jež by podobný efekt přinesla, existuje již pár desetiletí. Její implementace do stávajícího webového prostředí je, a pravděpodobně i nadále bude, náročná ve všech svých aspektech. I přes to jsou však možnosti, které nabízí natolik zajímavé, že otázky vzniku nového, lepšího webu – webu nové generace, webu sémantického – představují jednu z největších výzev informačního věku.

Je nasnadě, že rozšíření internetu od počátku 90. zaznamenalo neuvěřitelný boom. Nicholas Negroponte ve své knize *Being Digital* [Negroponte, 1995, s. 8/9], jež byla publikována poprvé v roce 1995, uvádí, že toho roku bylo k Internetu připojeno kolem 30 milionů lidí. Roku 2005, kdy vyšel první český překlad knihy, bylo připojeno něco mezi 300 a 350 miliony uživatelů. A podle odhadů společnosti Cisco, uveřejněných 15. 7. 2009 v Lidových novinách [Pešek, 2009], by měl během příštích čtyř let vrůst celkový objem internetových přenosů 3,5krát (to by znamenalo, že na konci roku 2013 projde za měsíc internetem 40,5 exabytů). Jak je tedy možné, že je Internet (a potažmo i WWW) tak úspěšný?

Faktorů zapříčiňujících jeho úspěch je několik: na scéně se objevil v příhodné době, začal se rychle rozšiřovat – to právě z toho důvodu, že nepředstavitelným způsobem zefektivnil, zjednodušil a zrychlil komunikaci informací v čase i prostoru. Díky geniální technologii přenosu bitů z jednoho konce světa na druhý v reálném čase světelnou rychlostí se svět změnil. Vývoj samotné technologie přenosu pokročil k optickým vláknům a bezdrátovým technologiím, které dokážou přenášet nezměrné množství bitů každou sekundou. Čím objemnější a rychlejší přenos dat, tím větší je jejich spotřeba.

A právě zde se nám začíná rýsovat spojitost mezi technologií přenosu, objemem informací a uživatelem, který je nucen vynakládat větší a větší úsilí na zpracování a třídění informací, které mu je na jeho dotaz vyhledávač během okamžiku schopný poskytnout. Web, jako největší databáze lidských znalostí, jako nepředstavitelné množství syntakticky strukturovaných dat, obsahuje většinu odpovědí na otázky plynoucích z našich individuálních

informačních potřeb. Data z jednotlivých webových zdrojů jsou navzájem propojena a dnes tvoří největší otevřený celosvětový informační systém. Jeho architektura musí tedy zohledňovat typy reprezentací zdrojů v něm obsažených - může se jednat o stránky, feedy, obrázky a mnoho dalších zdrojů (i přesto, že počet reprezentací je omezen – viz, tzv. MIME typy – příloha č. 1).

Pokud si uvědomíme, kolik různých dat a zdrojů je na webu k dispozici, přirozeně bychom si měli položit otázku, jak je možné ze změní informací dostupných na webu vydolovat právě tu jednu, kterou hledáte? Odpověď zní: zatím stále těžce.

I v kyberprostoru (který na rozdíl od reálného světa není prostorově ohraničen) platí, že pokud někam uloží data a někdo jiný bude tyto data chtít nalézt a použít pro svoje potřeby, jediná jeho šance na úspěch je postupovat podle logického systému. Především musí vědět (ve smyslu chápat - což je právě zásadní rozdíl od vyhledávacích robotů), co hledá a musí znát směr, kterým se za hledanou informací vydat. Řeč je tedy o názvu hledané entity (nebo klíčových slov) a navigaci k jejímu umístění. Dotazy tohoto typu nejsou dnes pro počítače větší problém (stále je ale nutné počítat s tím, že - samozřejmě v přeneseném smyslu slova – Vám bude dodán obsah i s často zbytečně obsáhlým kontextem, který pro vás bude pravděpodobně nezajímavý).

Tedy každá z entit musí mít jedinečný identifikační prvek, kterým je její globální unikátní název, zkráceně URI (Unified Resource Identifier), jehož aplikací je obecně známější URL (Unified Resource Locator). Díky němu je možné entitu zpětně vyhledat. Kromě svého jedinečného jména jsou zdroje reprezentovány určitým typem serializace (kanonického zápisu), díky níž mohou být dále transferovány přenosovou soustavou. Tak mohou být rozesílány v reálném čase například z Austrálie do Jižní Afriky během jediného okamžiku. To je možné díky značkovacím jazykům (Markup Language), které mimo vlastní text, který je pro člověka z hlediska obsahu podstatný, disponují instrukcemi pro zpracování textu počítačem. Základem pro definování značkovacích jazyků se stal univerzální značkovací metajazyk SGML.

SGML (Standard Generalized Markup Language) byl původně navrhnutý pro sdílení strojem čitelných dokumentů ve velkých vládních, právních i průmyslových projektech. Udává strukturu, ale ne sémantiku.

Z něho, jako aplikace SGML, vychází **HTML** (Hypertext Markup Language) - jazyk pro komunikaci dokumentů pro lidskou potřebu. HTML svou jednoduchostí umožnil rapidní rozvoj celého webu. Na druhou stranu právě jeho jednoduchost omezovala pokročilé webové aplikace v mnoha oblastech a úlohách.

To bylo příčinou vzniku dalšího značkovacího jazyka **XML**, který dovoluje nadefinovat široké rozšíření, včetně vytváření vlastních tagů (metadat) samotným uživatelem.

Jeho primárním účelem bylo usnadnit sdílení strukturovaných dat napříč odlišnými informačními systémy. Mohou do něj být implementovány některé aplikační jazyky, včetně XHTML, RSS, GraphML a mnoho jiných. Jazyk XML je zjednodušením obecného jazyka SGML a pro tvorbu ontologií je jedním ze základních kamenů, proto se jím bude text zabývat i níže.

Aplikace XHTML (viz obrázek 1) je úpravou HTML do jazyka, který je aplikací obecného XML. Zároveň je možné definovat jej jako schéma, které určitým způsobem omezuje možnosti jazyka XML, a to za účelem zjednodušení, aby například nedocházelo k možnosti pojmenování jednoho objektu více možnými jmény. Proto je pouze logickým důsledkem vymezení technické pozadí vylepšeného sémantického webu na aplikaci XML [Kashyap, 2008, s. 293/295].

The image shows a screenshot of the ClickforArt website interface on the left and a corresponding XHTML code snippet on the right. The website features a search bar, navigation links, and a grid of art prints. The code snippet is an XHTML document with the following structure:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html>
<head>
<title>ClickforArt | Art Styles - Art Prints & Canvas</title>
<meta name="keywords" content="ClickforArt.com, Style Page, Meet
the StyleArt Prints & Canvas">
<meta name="description" content="ClickforArt, Style Page">
<meta http-equiv="Content-Type" content="text/html; charset=iso-
8859-1">
<meta http-equiv="expires" content="0">
<link href="/css/style.css" rel="stylesheet" type="text/css">
<link rel="shortcut icon" href="/favicon.ico" type="image/x-icon"
/>
<script type="text/javascript" src="/js/script.js"></script>
<script language="javascript">AC_FL_RunContent = 0;</script>
<script src="AC_RunActiveContent.js"
language="javascript"></script>
<script>
var gblPage = "style";
var gblDisable = "";
</script>
<script src="http://www.google-analytics.com/urchin.js"
type="text/javascript">
</script>
<script type="text/javascript">
_uacct = "UA-1484843-1";
urchinTracker();
</script>
</body>
</html>
</head>
<body>
<div id="container">
...
```

Obrázek 1 V hlavičce XHTML kódu jsou uvedeny metadata popisující obsah dokumentu na základě klíčových slov – ovšem bez popisu vztahů mezi těmito klíčovými slovy a pojmy; [ClickforArt, 2010]

Ve značkovacích jazycích existují nástroje, kterými lze popsat strukturu dokumentu a odkázat na jiné webové zdroje – řeší tedy syntax webových dat, jejichž ukládání a přístup k nim se poté realizuje pomocí technologie REST (REpresentational State Transfer) definovanou řadou protokolů jako HTTP(S) (HyperText Transfer Protocol (Secure)), SOAP (Simple Object Access Protocol) nebo třeba Java. Dnešní web by mohl díky své nejtýpčtější vlastnosti být označen jako web syntaktický.

Díky značkovacím jazykům můžeme na webu najít mnohé aplikace reprezentující

rozličné informací zdroje. Ještě před několika málo lety aplikace nebyly schopné spolu spolupracovat.

Jádrem problému spolupráce je nedostatek dat, jež by popisovala data zdrojová, respektive metadata (data o datech). Právě metadata jsou klíčem k nové generaci internetu, na kterých stojí veškeré ontologické inženýrství, jehož praktický i teoretický význam bude v základech přiblížen v dalších kapitolách práce.

1.1. charakteristika, obsah a metodika práce

Celá práce je systematicky řazena od kapitol nejobecnějších po nejkonkrétnější, což čtenáři pomůže vytvořit si co nejkomplexnější představu o daném tématu.

Již v úvodu jsme se obeznámili se základním kontextem nutným pro nasazení této technologie do praxe. V druhé kapitole nazvané Základní koncepce ontologií se seznámíme s nejobecnějšími pojmy, definicemi a termíny pojícími se s ontologiemi a nejdůležitějšími systémy pro popis metadat – s RDF(S) a XML.

Ve třetí kapitole se zaměříme různá pojetí a typologii ontologií a vymezíme koncept ontologie pro účely této práce.

Ve čtvrté kapitole, pojednávající o ontologickém inženýrství, budou představeny významné oblasti aplikace ontologií, nejpoužívanější jazyky a obecná metodika pro jejich tvorbu.

V páté kapitole Aplikace ontologií v oblasti výtvarného umění budou popsány jak nejdůležitější ontologie používané pro popis uměleckých děl (Iconclass, ULAN, AAT a CIDOC CRM), tak ontologie pro vizuální popis uměleckých děl i významná metadatová schémata (Dublin Core, CDWA, ISO 21127 a VRA).

V závěru se zaměříme na otázku vzájemné kompatibility mezi jednotlivými ontologiemi popsány v rámci kapitoly 5.

Práce byla ve své teoretické části (kapitoly 1-4) vystavěna na interpretaci informací z mnoha různým zdrojů, uvedených v seznamu použité literatury (citováno dle normy ISO 690 a ISO 690-2; odkazy k použitému zdroji jsou uvedeny vždy v hranatých závorkách za textem, spolu s příslušnou stranou ve zdroji, pokud taková je – u online zdrojů není strana uvedena).

V praktické části (kapitoly 5 a 6) byly zkoumány možnosti ontologií aplikovaných v oblasti výtvarného umění, jak na základě oficiálních manuálů uvedených na webu, tak na

základě osobní zkušenosti s interaktivním rozhraním těchto systémů. Pokud je to možné, jsou uvedeny jsou i ukázky výstupů ve formě obrázků.

V závěru byla nastíněna představa o možné integraci ontologií do jednoho celku a to na základě vyhodnocení informací načerpaných z použitých zdrojů a vlastní představy o funkčnosti této teoretické ontologie.

1.2. definice domény výtvarného umění

Zbývá snad ještě definovat pojem výtvarného umění jakožto znalostní domény, na kterou se praktická část práce zaměřuje. Za výtvarné umění budeme tedy považovat veškeré artefakty lidské kultury vizuálního charakteru, vytvořené v jakémkoli minulém, současném či budoucím časovém období, jejichž tvorba byla motivována jedinečnou a originální myšlenkou autora a snahou tuto myšlenku přenést co nejautentičtějším způsobem k divákovi napříč časem i prostorem (na rozdíl od řemesla, které povětšinou neobsahuje absolutně originální ideu). Obory, spadající pod výtvarné umění, které budou relativní pro tuto práci, jsou: malířství, sochařství, architektura a fotografie.

Pro zamyšlení o šíři, různorodosti a záběru oblasti výtvarného umění viz přílohu č. 2 [McCLOUD, 2008, s. 164-165].

2. ZÁKLADNÍ KONCEPCE ONTOLOGIÍ

2.1. definice a termíny

2.1.1. ontologie

Termín ontologie v původním smyslu slova pochází z řečtiny. Do dnešní doby nese význam filozofické disciplíny, jejímž jádrem je *učení o bytí, o jeho nejobecnějších určeních a pojmech* [Kraus, 2007, s. 572], které jsou nastoleny v teoretické rovině. Už klasičtí myslitelé, Platón i Aristotelés byli oba přesvědčeni, že *“mezi jsoucnem a naším poznáním je shoda, že naše poznání a řeč jsou s to uchopit, obrážet strukturu jsoucna: že logika a ontologie se takřka kryjí nebo že je lze k sobě nepochybně přiřadit* [Störig, 2000, s. 138]“.

V posledních dvou dekadách se význam přenesl i do oblasti ryze praktické, do informatiky. Od původního se sice výrazně odchytil, stejně je ale možné vycházet ze skutečnosti, že ontologie se snaží co nejlépe a nejefektivněji definovat a popsat entity, třídy do kterých přísluší, vztahy mezi nimi nebo také jejich vlastnosti tak, aby tyto entity mohly být reprezentovány v informačních znalostních systémech.

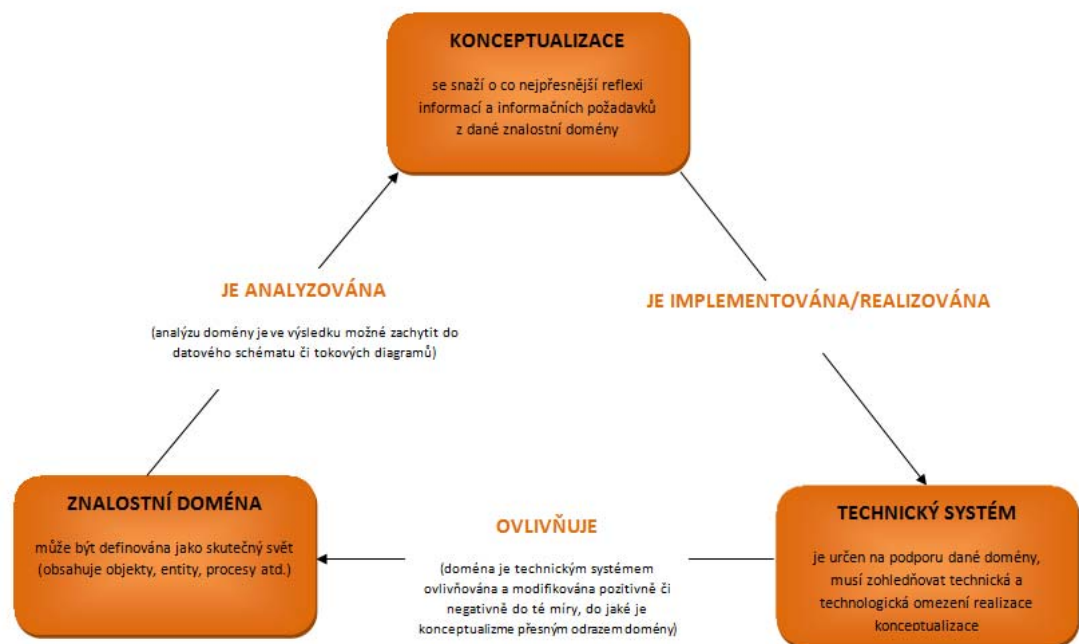
První z obecně uznávaných formulovaných definic pochází od R. Nechese: *“ontologie definují základní termíny a vztahy, obsahují slovní zásobu z oboru, zahrnují pravidla pro kombinování termínů a vztahů, za účelem definování rozšíření slovníku”* [Gómez-Pérez, 2004, s. 2].

Jedna z nejcitovanějších definic pojmu (autorem je T. R. Gruber) pak zní, že *“ontologie je explicitní specifikace konceptualizace”* [Svátek, 2002, s. 2]. Ta byla později obohacena W. N. Borstem – ontologie je *“formální specifikace sdílené konceptualizace”* [Fensel, 2004, s. 3]. *Konceptualizace* (viz obrázek 2) znamená vytváření abstraktních modelů vybraných fenoménů, a to identifikací relevantních pojmů daného fenoménu v určité oblasti vědění. *Formální* pak proto, že tvorba ontologií je závislá na jazyku s přesně definovanou syntaxí, aby mohla být strojově čitelná; *sdílená* proto, že ontologie zachycuje konsenzuální vědění uznávané ne jednotlivcem, ale velkou skupinou lidí [Gómez-Pérez, 2004, s. 2].

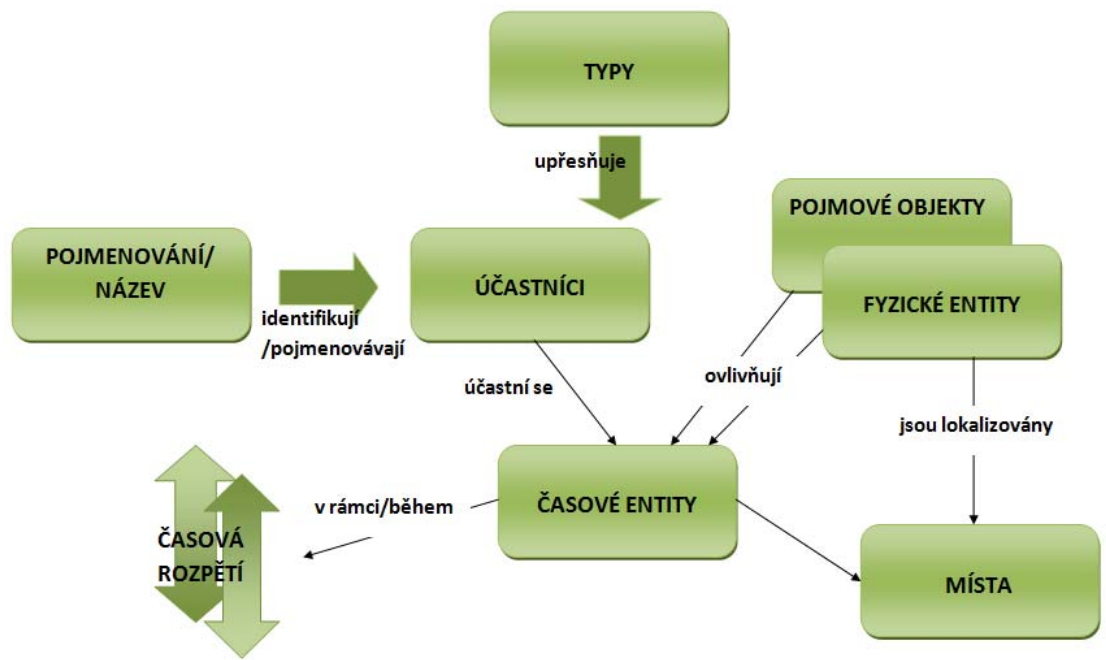
Pojem je ale také možné chápat v mnohem širší perspektivě, například z úhlu pohledu ontologického inženýrství, jež pracuje s ontologiemi jako se *“znalostními modely, tedy abstraktními popisy (určité části) znalostního systému”* [Svátek, 2002, s. 2]

Ještě jednou jinak řečeno, jedná se o soubor přesně definovaných termínů z určitých oblastí lidského poznání a přesně definovaných vztahů mezi nimi, který je reprezentován informačními artefakty – tezaury, schémata databází nebo modely. Stojí za zdůraznění, že ”přesně definovaný” termín nebo vztah neznamená nic jiného, než přesné ”slovní vymezení (obsahu) pojmu, výklad významu slova (termínu, výrazu) uvedením jeho základních, typických znaků [Kraus, 2007, s. 154]“ – důležité je to z toho důvodu, že každá ontologie v podstatě stojí na přesném a formálním definování pojmů a vztahů. Každá ontologie je tudíž metadatový systém, ve kterém o určité entitě říkáme, že znamená zároveň i něco jiného (viz obrázek 3).

Nasazení metadat v umělé inteligenci by mohlo usnadnit sdílení a opětovné využití lidských znalostí. Tvorba metadat byla v minulosti ale natolik roztržštěná, že bylo nutné vytvořit jednotný rámec pro další vývoj, jímž se stal RDF.



Obrázek 2 Teoretický rámec informací; [Crofts, 2003]



Obrázek 3 Obecná struktura ontologie; [Nordbotten, 2008]

2.1.2. RDF a XML

RDF (Resource Description Framework) je systém pro metadatový popis zdrojů. Vznikl z iniciativy konsorcia W3C¹ v roce 1999 jako doporučení způsobu reprezentace metadat na WWW. Představuje cestu, jakou lze vystihnout a posléze zachytit podstatná tvrzení o různých zdrojích informací umístěných na webu, které jsou identifikovány výše zmiňovaným URI. RDF je zároveň i formálním jazykem [Breitman, 2006, s. 57].

RDF je jako jazyk založen na poměrně triviální lingvistické struktuře, které je blízká i běžně mluvené řeči, totiž na struktuře tvrzení vyjádřené jedním subjektem, jedním objektem (zdroj může zastávat jak roli objektu, tak roli subjektu) a jedním predikátem (který určuje vlastnosti objektů a subjektů) - tedy tzv. "**trojicí RDF**" (RDF triple), která představuje jeden výrok (RDF statement). Predikát koresponduje s vlastnostmi subjektů, objekt s hodnotami vlastností. Tato trojice dostatečně vystihuje sdílenou sémantiku spojení mezi trojicemi subjekt-predikát-objekt, nebo také subjekt-predikát-hodnota (subject-predicate-object nebo subject-predicate-value) [Svátek, 2003, s. 4].

RDF se používá na reprezentaci metadat webových zdrojů, ale i na reprezentaci informací o různých objektech vyskytujících se na webu. Podporuje interoperabilitu mezi aplikacemi, jež zajišťují výměnu strojem čitelných informací na webu – RDF lze vložit i XML – viz obrázek 3.

¹ W3C – World Wide Web Consortium – je mezinárodní společenství několika členských organizací a veřejnosti, společně pracujících na vývoji webových standardů za účelem využití plného potenciálu webu. Jeho hlavními osobnostmi jsou Tim Berners-Lee and Jeffrey Jaffe [W3C, 2009].

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:cult="http://www.icom.com/schema.rdf#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:adm="http://www.oclc.org/schema.rdf#">
<!-- All statement -->

<rdf:Description rdf:about="http://www.w3.org/2000/01/rdf-
schema#ContainerMembershipProperty">
  <rdfs:subClassOf rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#Property"/>
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Class"/>
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource"/>
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource"/>
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-
schema#ContainerMembershipProperty"/>
</rdf:Description>

<rdf:Description rdf:about="http://www.museum.es/woman.qti">
  <rdf:type rdf:resource="http://www.icom.com/schema.rdf#Painting"/>
  <cult:technique xml:lang="en">oil on canvas</cult:technique>
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource"/>
  <rdf:type rdf:resource="http://www.icom.com/schema.rdf#Artifact"/>
</rdf:Description>

<rdf:Description rdf:about="http://www.w3.org/2000/01/rdf-schema#label">
  <rdfs:domain rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource"/>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Literal"/>

```

Obrázek 4 Příklad RDF [SMILE, 2004]

Na rozdíl od XML RDF nezprostředkovává sémantickou informaci jako vedlejší produkt, vznikající během definování struktury dokumentu, na kterou je XML přednostně zaměřen. Respektive, původním cílem jazyka XML bylo oddělit obsah webového dokumentu od prezentace obsahu webového dokumentu. Proto se dá říci, že XML představuje syntax a RDF přímou sémantiku metadatových modelů. [Breitman, 2006, s. 58]. To by samozřejmě nemohlo fungovat bez slovníku termínů definovaných v různých aplikacích stejným způsobem tak, aby jej aplikace sdílely a mohly být na jejich základě společně propojené do většího znalostního systému. Data a elementy obsažené v XML je možné převést do schématu RDF, nazývaného RDFS. Pro rozšíření RDF je potřeba pracovat se sdílenými terminologiemi, tezaury, slovníky výrazů, názvoslovími a kódovacími systémy [Goméz-Peréz, 2004, s.52/54].

Jazyk XML umožňuje tvůrci popsat i složitější stromovou strukturu dokumentu lineární syntaxí a to zvolením vlastních tagů. Tyto tagy ale neobsahují žádnou sémantiku, která by určovala, jak dokumenty prezentovat přes webový prohlížeč.

XML byl vytvořen pro popis strukturovaných dokumentů. XML dokument se skládá z čistého textu a značek, které jsou interpretovány aplikačními programy – viz obrázek 5. Specifikace XML schématu obsahuje sadu datových typů, které si vypůjčily právě některé ontologické formální jazyky [Breitman, 2006, s.58].

```
<?xml version="1.0" encoding="UTF-8" ?>
<painting>
  
  <caption>This is Raphael's "Foligno" Madonna, painted in
    <date>1511</date>–<date>1512</date>.
  </caption>
</painting>
```

Obrázek 5 Příklad XML; [XML - Wikipedia]

2.1.3. deskripční logika

Synonymem zastupujícím pojem deskripční logiky je terminologická logika – řada příbuzných logických formalismů [Svátek, 2002, s. 31]. Stala se přirozenou základnou pro tvorbu ontologií. Pracuje s třídami objektů a vzájemnými relacemi mezi nimi – stejně jako ontologie určitým způsobem reprezentuje systém znalostí. Liší se ale vymezením relací mezi třídami – ty deskripční logika nezahrnuje. Relace jsou dány, jak sám název disciplíny napovídá, dynamickým popisem tříd objektů. Jde vlastně o logiku pojmů (konceptů) a jejich vzájemných vztahů. Pojem je v deskripční logice chápán jako obecný výraz (např. "freska", "kvaš", "olej", "plátno" atp.), který se ale odlišuje od vlastních jmen (např. Raffaello Santi) označujících konkrétní individuum. Koncepty lze sémanticky chápat jako množiny individuí patřících pod určitý koncept (např. Umělec(Raffaello Santi) znamená, že Raffaello Santi je umělec). Ze základních konceptů se dají konstruovat koncepty složené.

Druhou podstatnou částí kromě konceptů v deskripční logice jsou role. S těmi lze z hlediska sémantiky operovat jako s uspořádanými dvojicemi (binárními relacemi), např. zápis vyjádření sémantického vztahu „jeUčitelem(Pietro Perugino, Rafael Santi)“. Díky rolím je možné konstruovat i další koncepty.

Deskripční logika je v podstatě částí predikátové logiky prvního řádu, ale používá jinou syntaxi. Oproti predikátové logice je také méně expresivní ale především se liší v tom, že je rozhodnutelná [Svátek, 2007, s. 22].

2.1.4. další pojmy

Pro kompletní popis teoretické koncepce ontologií není možné opomenout další významné pojmy, se kterými je možné se v souvislosti s tímto tématem setkat. Vzhledem k zadanému rozsahu práce však nepovažuji za praktické řešení podrobnější vymezení následujících pojmů. Proto jsou u nich uvedeny pouze stručné definice s odkazem na zdroj podrobnějších informací: [Svátek, 2002, s. 6/9].

- třída (class): tvoří základ pro tvorbu ontologií; sestává se z množiny konkrétních objektů; je definována řadou vlastností; větší množiny tříd mohou být organizovány do hierarchické struktury
- podtřída (sub-class): podmnožina objektů se společnými vlastnostmi, náležící určité třídě, tudíž se musí vyznačovat minimálně jednou shodnou vlastností s kmenovou třídou
- nadtřída (upper-class): inverzní entita k podtřídě
- koncept (concept): v některých formalismech odpovídá termínu třída, popřípadě kategorie
- kategorie (category): viz třída, koncept
- rámec (frame): základní konstrukt systémů AI (artificial intelligence - umělá inteligence)
- individuum (individual): reprezentuje konkrétní reálný objekt ve světě
- instance (instance): stejně jako individuum reprezentuje konkrétní reálný objekt ve světě, ale vázaný k určité třídě
- relace (relationship): vztah mezi objekty nebo třídami definované vybranými logickými podmínkami
- funkce (function): zvláštní typ relace, u nichž je hodnota n-tého argumentu jednoznačně určena n - 1
- slot (slot): vlastnost, binární relace, není závislá na žádné třídě
- vlastnost (property): viz slot
- role (role): slot ve vztahu k objektu, který je jeho hodnotou
- atribut (attribute): funkční slot, definovaný pro všechny instance třídy
- meta-slot (meta-slot): sloty s definovaným vlastnostmi
- fazeta (facet): lokální omezení hodnoty slotu, jež se váže na určitou třídu

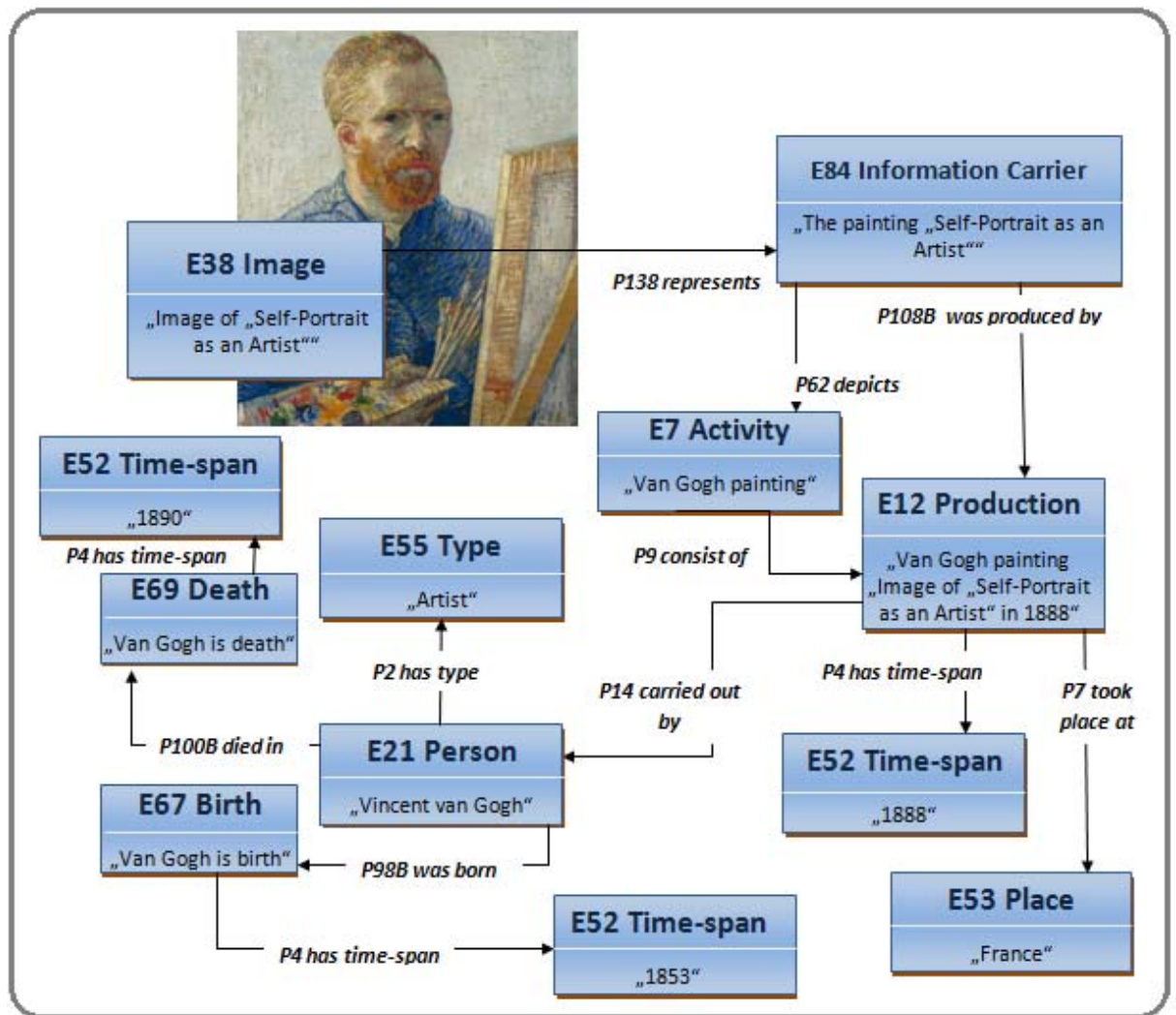
- primitivní datové hodnoty a datové typy: argumenty relací, které neodpovídají žádnému objektu
- axiom (axiom): logická formule používané pro vyjádření podřízenosti, disjunkce nebo ekvivalence tříd (popř. relací)
- definiční obor (domain): vlastnost slotu, omezení slotu
- obor hodnot (range): vlastnost slotu, omezení slotu
- dědičnost (heredity): přenos vlastností z nadtřídy na podtřídu

2.2. proč jsou ontologie vyvíjeny a používány v prostředí WWW

Možnost sdílení informací, přes vztahy mezi jevy napříč různými oblastmi lidského poznání, je lidskou nejpřirozenější schopností. Informace se uložením do naší paměti stávají znalostmi a jsou zařazeny do sítě dalších znalostí, získaných již dříve. Jednotlivé znalosti vytvářejí v mozku síť vztahů mezi konkrétními jevy. Čím lépe dokážeme ve vzájemné komunikaci vyjádřit vazby, vztahy a vlastnosti mezi jevy, tím snazší je logikou vyvodit odpovědi na kladené dotazy, které bychom jinak pravděpodobně dolovali z paměti jen velmi obtížně.

Internetová síť jako celek je sítí odkazů na jiné dokumenty nebo aplikace. Odkazy fungují pouze jako směrovače, nikoli jako obousměrná vazba definovaná vzájemným vztahem. Ty si musí doplňovat uživatel sám, tedy do alespoň okamžiku, kdy na scénu vstoupí ontologie.

Lze jimi definovat významy informací do komplexní struktury (viz obrázek č. 6), a tím pádem je učinit **počítači srozumitelné**. Dále by v budoucnu technologie ontologií mohla umožnit vytváření **interoperabilních systémů**, složených z aplikací s různými funkcemi postavených na jednotné platformě sdílených vazeb mezi subjekty a objekty. To by znamenalo **přímý, jednodušší, efektivnější a rychlejší přístup k žádaným informacím a jejich komunikaci**, kterým by v budoucnu mohl disponovat web generace 3.0 – **web sémantický**. V něm by ontologie mohly figurovat jako platforma internetové komunikace informací.



Obrázek 6 Struktura ontologie CIDOC CRM; [Crofts, 2003]

3 . R Ů Z N Á P O J E T Í O N T O L O G I Í V O B L A S T I I N F O R M A T I K Y

3.1. formální typologie ontologií

Definice ontologie sice popisuje podstatu technologie představující platformu pro efektivnější komunikaci informací na www, avšak ani zdaleka striktně nevymezuje, o kterých z konkrétních aplikací využívajících “formální specifikace sdílené konceptualizace” dostupných na www by se dalo hovořit jako o ontologiích v pravém smyslu slova. Často bude při tomto posuzování záležet čistě na úhlu pohledu jedinice. Proto bude v závěru kapitoly vymezeno, které specifikace zahrneme do ontologií a které nikoliv. Nejprve se ale zaměříme na různé typy ontologií.

Poměrně dobře a přehledně je možné sestavit typologii ontologií podle několika různých kritérií. V zásadě se jedná o 6 nejzákladnějších aspektů, podle kterých je možné ontologie dále vyčleňovat.

1. Zaprvé je možné ontologie rozdělit podle jejich **předmětu konceptualizace**, respektive typu v nich prezentovaných informací. V rámci této kategorie se vyčleňují *ontologie obecné* (věnující se obecným konceptům jako jsou procesy, stavy, události) a jejich specifická část, *ontologie základní* (zachycují nejobecnější pojmy). Dále pak *ontologie doménové* (obsahují koncepty specifické pro danou vědní oblast), *úlohové* (jedná se o generické modely znalostních úloh), *doménově-úlohové* (ty kombinují předešlé dvě kategorie a vyjadřují generické modely pro konkrétní vědní obory), *aplikační* (vhodné k vytváření modelů pro určité aplikace) a konečně *metodické* (zachycují definice pro pojmy a vztahy relevantní k daným procesům) [Breitman, 2006, s. 28].
2. Za druhé lze ontologie členit podle **typu a míry jejich strukturovanosti** na *terminologické* (specifikují termíny využitě pro reprezentaci znalostí v dané doméně), *informační* (specifikují struktury záznamu v databázi a jeho vazeb) a *znalostní* (specifikují konceptualizaci vědění) [Heijst, 1997].
3. Za třetí rozdělujeme ontologie podle **míry formalizace** na formální (jež jsou vyjádřeny formálními jazyky) a neformální (ve kterých jsou pojmy vysvětlovány

- přirozeným jazykem – např. glosáře, řízené slovníky a tezaury) [Svátek, 2002, s. 4].
4. Dále dělíme podle míry **operacionalizace** na ontologie hrubé (používající jazyk s malou expresivitou) a jemné (vysoce expresivní modelovací jazyk) [Svátek, 2007, s. 4].
 5. Dalším významným aspektem pro rozdělení ontologií jsou tzv. **ontologické závazky**. Ontologický závazek podmiňuje uchopení problematiky a přístup k tvorbě znalostního systému. Například podle toho, zda se vytvářená ontologie bude či nebude vázat na přirozený jazyk (deskriptivní proti revizionistické); bude popisovat to, co reálně existuje nebo i to, co může existovat (posibilistická oproti aktualistické); modelovat pouze to, co v daný okamžik reálně existuje nebo i to, co existovalo dříve či teprve existovat bude (presentistická proti eternalistické). Dále pak zda bude ontologie obsahovat jednotliviny nebo obecniny; vycházet pouze z malé skupiny základních konceptů, nebo naopak z jejich vysokého počtu (multiplikační proti redukcionistické ontologii); či zda bude běžné předměty vázat ke každému časovému okamžiku zcela či nikoliv (3D oproti 4D) [Svátek, 2007, s. 5/6].
 6. Poslední významný aspekt je obsahová **expresivita**, podle které rozdělujeme ontologie na těžké (ontologické závazky jsou explicitně vyjádřeny) a lehké (ontologické závazky se nevyjadřují) [Svátek, 2007, s. 5].

Shrnutí typologie je k dispozici ve formě tabulky v příloze č.3.

Někdy není prakticky možné vytyčit hranici vedoucí mezi různými typy ontologií, jako je to patrné v případě prolínání obecných, základních, doménových a jádrových ontologií. My se v dalších kapitolách práce přednostně zaměříme na několik ontologií doménových, přičemž doménou našeho zájmu bude výtvarné umění definované v úvodu práce.

3.2. co je a co není ontologie

Za ontologii je v podstatě možno považovat jakýkoli systém reprezentace znalostí, v němž jsou mezi jednotlivými objekty nebo celými třídami objektů definovány sémantické vztahy. Někdy je ale sémantika natolik slabá, že hovořit o ontologii není zcela přesné či vhodné. Proto se nyní pokusím vymezit hranici mezi ontologiemi a ostatními hierarchickými

či indexačními systémy a rovněž nastínit nuance mezi glosáři, tezaury, taxonomiemi a ontologiemi.

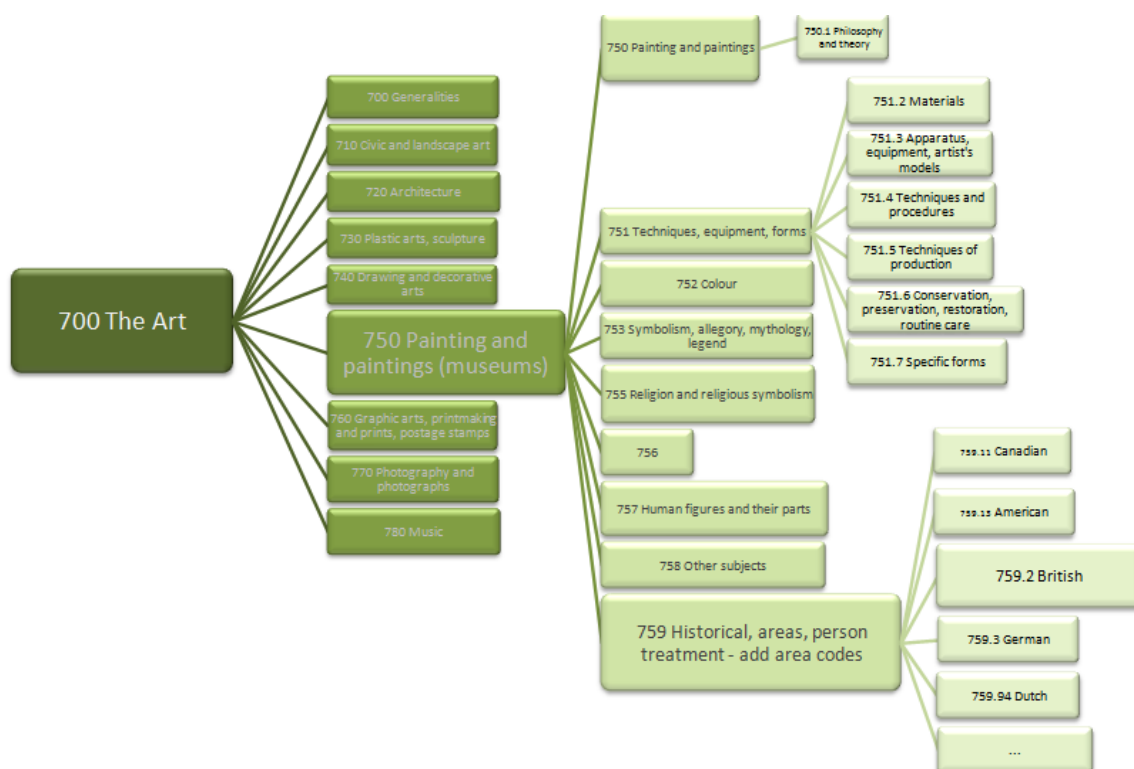
3.2.1. řízené slovníky, glosáře, katalogy

Nejjednoduššími systémy jsou bezpochyby glosáře a řízené slovníky či katalogy. Jejich sémantika je velice slabá a neformální. V případě řízeného slovníku se jedná pouze o konečný seznam termínů, který je v případě glosáře rozšířen o popis významu, ovšem pouze v přirozeném jazyce.

3.2.2. taxonomie a tezaury

Dalšími typy systémů organizace znalostí jsou taxonomie a tezaury. Přestože jak v taxonomii, tak v tezauru můžeme zpravidla nalézt hierarchické vztahy mezi objekty a třídami (často vyjádřenými sémantickým popisem – proto je možno v některých případech mluvit o ontologiích), v některých aspektech se od sebe odlišují. Na tyto aspekty se zaměříme podrobněji a pokusíme se vymezit, co pro účely této práce bude možné považovat za ontologie.

Taxonomie hierarchicky klasifikuje termíny. Užívá výlučně dědičných vztahů, žádné jiné typy vztahů v ní nenajdeme (tedy vztahy typu „je typem“ („is-a“, „type-of“)). Typický příklad taxonomie je uveden na obrázku 7. Jde o oblast Deweyho desetinného třídění věnovanou právě výtvarnému umění [Breitman, 2006, s. 20].



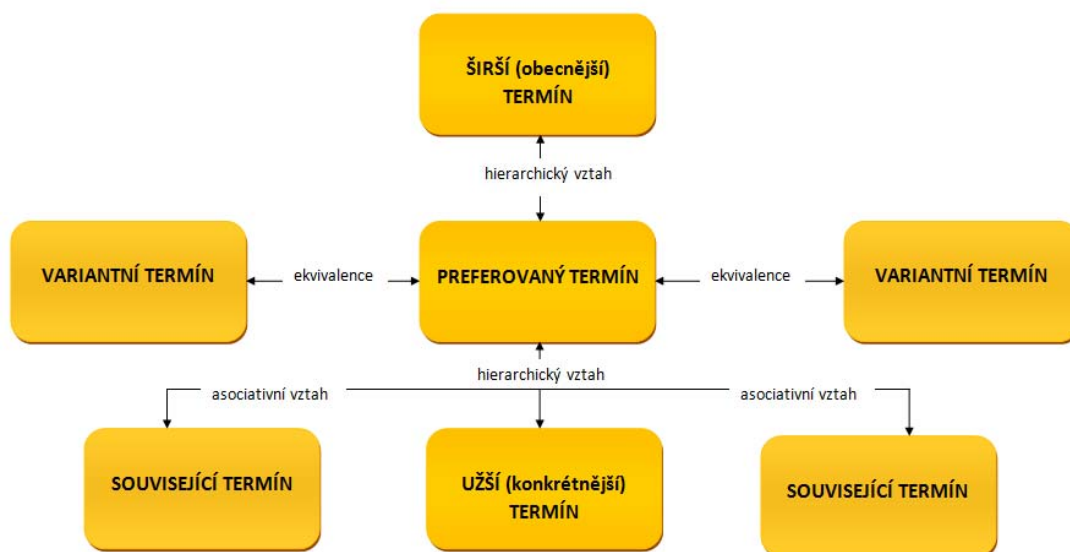
Obrázek 7 Členění větve „The Art“ v Deweyho desetinném třídění; [Grainger, 1997]

Sémantické vazby jsou v případě taxonomií poměrně slabé, a proto je nebudeme pro účely práce považovat za plnohodnotné ontologie.

Tezaurus obsahuje již více druhů vztahů mezi pojmy. V podstatě jde také o konečný seznam termínů a jejich definic, které tyto termíny standardizují často pro indexační účely – na tezaurus je možné nahlížet jako na taxonomii obohacenou o řadu sémantických vazeb jako ekvivalence, inverze nebo asociace.

Příkladem je velmi využívaný tezaurus WordNet², ze kterého pochází i příklad znázornění obecné struktury tezaurů na obrázku 8 [Breitman, 2006, s. 22/23].

² WordNet je rozsáhlá lexikální databáze anglického jazyka. Seskupuje substantiva, verba, adjektiva a adverbia do souborů kognitivních synonym, u nichž každý vyjadřuje jiný koncept. Koncepty jsou vzájemně propojeny sémantickými vztahy. Je užitečným nástrojem pro počítačovou lingvistiku a zpracování přirozeného jazyka. [WordNet, 2010].



Obrázek 8 Obecná struktura tezauru; [Breitman, 2006, s. 22/23]

Tezaury jsou často natolik bohaté na sémantiku, že je budeme v této práci považovat za ontologie, jak uvidíme dále.

Za ontologie tedy budeme považovat jakékoliv systémy reprezentace znalostí (v oblasti výtvarného umění) vyjadřující sémantické vztahy mezi objekty či třídami objektů, které poskytují natolik široké možnosti organizace, aby na jejich základě bylo možné vytvořit informační síť obsahující minimálně dvě heterogenní třídy objektů s více než dvěma typy vztahů - samozřejmě musí být zohledněna sémantika (ať už množinová nebo lexikální), neboť pouhá hierarchická struktura pojmů je pro označení informačního systému za ontologii málo.

Taxonomie, které pouze vydělují menší a užší homogenní podtřídy z jediné (homogenní) třídy nebudeme za ontologie považovat. Na druhé straně na některé tezaury, které definují vztahy mezi heterogenními třídami objektů (například osoby – role, jako je tomu v ULANu), budeme nahlížet jako na plnohodnotné ontologie, jimiž také z tohoto úhlu pohledu jsou (jak se budeme moci přesvědčit v kapitole o aplikacích ontologií do znalostních systémů výtvarného umění).

4 . O N T O L O G I C K É I N Ž E N Ý R S T V Í

V předchozích kapitolách byly ontologie představeny a v základu definovány nejdůležitější pojmy, termíny a vlastnosti. Rovněž jsme specifikovali co je a co není pro účely této práce vhodné považovat za ontologii. V poslední teoretické kapitole tedy zbývá ještě přiblížit paradoxně jednu z nepraktičtějších oblastí spojených s ontologiemi – ontologické inženýrství.

Podle definice Svátka a Vacury je ontologické inženýrství „*moderní oblastí informatiky, které se zaměřuje na návrh, implementaci a aplikaci ontologií [...], které se původně vyvíjelo v rámci umělé inteligence, postupně se však posouvalo k praktickému využití při návrhu aplikací a zpřístupňování datových zdrojů, a v současné době hraje ústřední roli v iniciativě tzv. sémantického webu* [Svátek, 2007, s. 1]“.

Jedná se tedy o veškeré procesy a činnosti s nimi spojenými, které jsou zapotřebí při navrhování, vývoji a aplikaci ontologií do praxe. Zahrnuje metodiky, jazyky i nástroje.

Ontologie jsou svým možným využitím spojeny s celou škálou odvětví umělé inteligence. Jedno toto odvětví – znalostní inženýrství – obsahuje mimo jiné i inženýrství ontologické. Pro jasnější představu činností zahrnutých v ontologickém inženýrství je uveden následující výčet [Husáková, 2008].

- a) studie proveditelnosti: uvedení možných problémů, příležitostí, řešení a ekonomických důsledků vývoje ontologií a jejich aplikace
- b) analýza domény: motivace pro doménu, existující řešení
- c) konceptualizace: konstrukce ontologického modelu, integrace a rozšíření existujícího řešení
- d) implementace: implementace formálního modelu v daném jazyku reprezentace
- e) údržba: přizpůsobení ontologie novým požadavkům
- f) užití: využití ontologického modelu v praxi

Jak je možné odvodit z výčtu činností a aktivit, je ontologické inženýrství přímo spojeno s umělou inteligencí, znalostním inženýrstvím, managementem znalostí, znalostními a expertními systémy, sémantickým webem, strojovým učením, zpracováním přirozeného jazyka, reprezentací znalostí a softwarovým inženýrstvím [Husáková, 2008].

Všechny tyto oblasti spojuje jedna z nejpodstatnějších složek ontologického inženýrství – umělé formální jazyky používané pro reprezentaci znalostí, bez nichž by nebylo možné ontologie vytvářet ani aplikovat. Proto se těm nejvýznamnějším z nich budeme podrobněji věnovat v následující podkapitole.

4.1. jazyky pro reprezentaci znalostí

První známější jazyk pro reprezentaci znalostí byl vyvinut v 90. letech ve Stanfordu. Jde o **Ontologuu**, která měla na rozdíl od modernějších jazyků široké možnosti definování pojmů, což ale vedlo k častým problémům při implementaci nástrojů. Alternativa se objevila v jazyce **LISp**, který je obecně více sepnatý než Ontologua. Tento programovací jazyk se ve své době stal syntaktickým vzorem pro téměř veškeré existující jazyky pro reprezentaci znalostí (např. OCML). Menším počtem vývojářů je používán jazyk **CyCL**, který je spojen spíše s komerčními aktivitami.

Od druhé poliny 90. let se začalo prosazovat silnější propojení ontologií s novými webovými standardy jako HTML, XML a poté i RDF. Prvním webovým ontologickým jazykem se stal **SHOE**, v němž byly ontologie vytvářeny pomocí speciálních nestandardních značek HTML, které byly navrženy speciálně pro tento účel. Poté začaly vznikat i jazyky stojící na základech XML - např. **XOL** a **OML**. Na konci 90. let byla konsorciem W3C publikována specifikace **RDF** jakožto doporučení. Od té doby se vedly diskuze o základech syntaktického rámce - zda má stát na bázi XML nebo RDF. RDF se nakonec ukázalo jako vhodnější volba s ohledem na svoji modularitu a formálně-logickou interpretovatelnost. Jazyk RDF, sám serializován pomocí XML, se tak stal základním stavebním kamenem pro ontologickou reprezentaci [Svátek, 2007, s. 12/13].

Na úrovni specifikace je s RDF propojen i jednoduchý ontologický jazyk **RDF Schema** (RDFS). RDFS definuje hierarchické vztahy mezi třídami a jejich vlastnostmi a také definiční obory a obory hodnot.

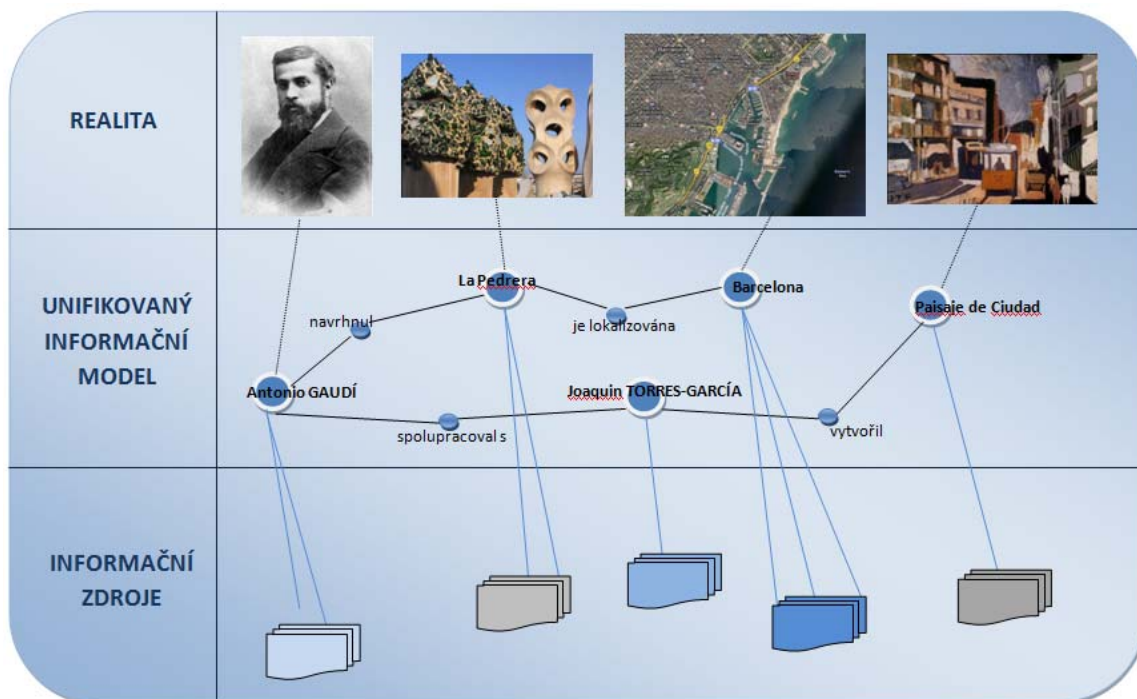
Poněkud sofistikovanějším jazykem je **OWL**, který plně využívá možností různých variant deskripční logiky. OWL byl vyvinut jako specifikace W3C (rok 2004) jakožto následovník chronologicky starších jazyků **OIL**, **DAML-ONT** a **DAML+OIL**.

OWL je používán velkou většinou ontologií minimálně jako alternativa jejich zápisu. Umožňuje na rozdíl od RDFS definování lokálních omezení, disjunkčnosti tříd a odlišení nutné a dostačující podmínky příslušnosti objektu ke třídě. Vyjadřovací síla jazyka OWL je

nicméně slabší než u predikátové logiky prvního řádu. Existují i jazyky jejichž vyjadřovací síla je stejná či dokonce vyšší - např. **GOL**. Vyjadřovací síla jazyků jde zvýšit kombinací jazyka s tzv. jazykem pravidlovým, poskytujícím doplňkové možnosti odvozování - pro tvorbu sémantického webu se typicky používá od roku 2004 **SWRL** (Semantic Web Rule Language).

Blízkým příbuzným OWL je **WSML**, dobře se hodící pro modelování komplikovaných webových služeb [Svátek, 2007, s. 13].

Variantou k ontologickým jazykům je technologie **Topic Maps**, jež umožňuje vytvářet a spravovat tzv. tematické mapy. I zde je na místě hovořit o ontologickém jazyku, i když se v principu od ostatních popsaných jazyků dost liší - tematické mapy jsou zapisovány přímo ve formě XML. Tematické mapy (XTM - XML Topic Maps; od roku 2000 ISO standard, ISO 13250) se skládají ze tří základních prvků: tématu (topic), jakožto označení pro jakýkoliv koncept nebo entitu; asociace (association), jakožto popis vztahů mezi tématy a výskytu (occurrence). Oproti metadatům jsou tematické mapy zvýhodněny tím, že definují i vazby k okolním dokumentům nebo tématům. Díky spojení značkovacího jazyka a možností vizualizace nabízí tematické mapy uživateli velmi přehledné uspořádání informací, jak znázorňuje obrázek 9 [Boldiš, 2008, s. 110-112].



Obrázek 9 Příklad struktury tematické mapy

Všechny uvedené jazyky mají přes různé odlišnosti několik prvků společných - nejvýraznější z nich je asi množinová sémantika, která se vyznačuje předpokladem, že třídám ontologie lze přiřazovat instance. Pojmy se považují za sémantické značky pro slova obsažená v textech.

Velkou otevřenou otázkou ontologických jazyků zůstává jejich vzájemná kompatibilita. Pro účely kombinace s jinými ontologiemi vznikl jazyk SKOS (Simple Knowledge Organisation System), jež je sám definovaný jako ontologie na základě RDFS [Svátek, 2007, s. 14].

4.2. obecná metodika pro tvorbu ontologií

Metodik pro tvorbu ontologií existuje hned několik (za zmínku bezpochyby stojí např. podrobně propracovaná Methoontology nebo OntoClean), ale téměř všechny se v základních bodech a postupech shodují. Proto budou zmíněny i v této podkapitole. Jejichž výčet podle Svátka [Svátek, 2007, s. 15/17] postupuje po následujících fázích:

1. rozsah a cíl ontologie

V prvním nutném kroku je třeba specifikovat, jaká znalostní doména nebo její část se bude modelovat a v jaké šíři. Často se objevuje tzv. "problém rozsahu" znamenající, že relevantních pojmů k jakékoli doméně by mohlo být za určitých okolností příliš velké množství. Efektivní strategií je aplikace tzv. "kompetenčních otázek" - tj. konkrétní příklady otázek, na které by systém s vyvíjenou ontologií měl odpovědět. Není zcela vhodné budovat ontologii pouze za jedním účelem, mělo by se počítat i s jejím znovupoužitím pro jiné úlohy ze stejné znalostní domény.

2. specifikace terminologie

Mnoho metodologií doporučuje začít s vytvořením seznamu termínů, které by měla ontologie zahrnovat. Termíny se v tomto kroku nerozřazují do tříd ani se mezi nimi nedefinují žádné typy vztahu, často je ale žádoucí opatřit je stručnými definicemi nebo vysvětlením. Tím de facto vznikne glosář, jenž se později může stát součástí dokumentace.

3. vytvoření ontologických typů

Jak již bylo vysvětleno, aby se z glosáře mohla stát ontologie, je nutné objekty, které obsahuje rozčlenit do odpovídajících tříd a obohatit je o relace, atributy a případně instance.

4. specifikace taxonomie

Taxonomii většina ontologických jazyků chápe jako základní stavební kámen, na který se pak navazují další a další vztahy. Neexistuje žádné pravidlo na to, zda je lepší budovat taxonomii "shora" (tedy od pojmů nejobecnějších ke konkrétním) nebo "zdola" (od pojmů nejkonkrétnějších k obecným). Často se ale doporučuje začít "od středu", tedy zvolit středně obecné pojmy a termíny, které jsou často relevantní pro více aplikací a postupně je rozšiřovat oběma směry. V této fázi je třeba zohlednit již existující ontologie, které by mohly být do nové ontologie integrovány.

5. tvorba netaxonomických prvků

V této fázi se mezi třídami navazují další netaxonomické relace, proto je třeba dbát na volbu vhodné hierarchické úrovně, ke které se relace vážou (totéž platí i pro atributy). Instance se do ontologií zahrnují, jen pokud mají v dané doméně důležité postavení.

6. nasazení a údržba

V posledním kroku je možné použít různé metody testování a verzování.

5. APLIKACE ONTOLOGIÍ V OBLASTI VÝTVARNÉHO UMĚNÍ

5.1. ontologie pro popis uměleckých děl



5.1.1. Iconclass

Výčet používaných ontologií otevřeme chronologicky nejstarším systémem Iconclass, který je ale v současné době také jedním ze systémů nejkompexnějších. Je navržený pro ikonografii – tzn. pro popis, klasifikaci a interpretaci různých výtvarných děl – a je navržený natolik sofistikovaně, že je zároveň široce používaným a oceňovaným nástrojem nejen pro popis, ale především pro vyhledávání subjektů vyobrazených na uměleckých dílech.

O Iconclassu je možné hovořit jako o ontologii díky jeho hierarchičnosti a také díky vyhledávání založeném na významu subjektů (ve smyslu motivů) zobrazených v konkrétních uměleckých dílech, ať už se jedná o malby na plátně, iluminace v rukopisných knihách či fotografie.

Iconclass nyní spravuje a vyvíjí nizozemský Institut pro historii umění v Haagu (Rijksbureau voor Kunsthistorische Documentatie, dále jen RKD). Na svém webovém sídle je systém tvůrci charakterizován jako „vícejazyčný klasifikační systém pro kulturní obsah“. Je to objektově-orientovaný klasifikační systém sestávající se z kolekce hierarchických definicí objektů, osob, událostí a abstraktních pojmů uspořádaných po způsobu řízeného slovníku.

Jako primární zdroj informací v následujících podkapitolách byly použity webové stránky projektu [RKD, 2010].

5.1.1.1. vývoj a charakteristika

Za vznikem Iconclassu stojí historik umění Henri van de Waal (1910-1972), který působil na univerzitě v nizozemském Leidenu. Ten v 50. letech realizoval svoji představu systematického pořádní subjektů, témat a motivů vyskytujících se v umění západní kultury do systému, který se postupem času stal rozsáhlým projektem Iconclass. Již od samého počátku spolupracoval Waal s RKD. Ze začátku byly do systému zahrnuty anotované fotografie klasických nizozemských uměleckých děl. V této době se projekt nazýval DIAL (Decimal Index of the Art of the Low Countries). Původní klasifikace DIALu je stále k dispozici právě v RKD, ovšem ne v digitální podobě. Projekt byl dokončen až po Waalově

smrti skupinou odborníků. Byl publikován Nizozemskou královskou akademií pro umění a vědu (Royal Netherlands Academy of Arts and Science - KNAW), jejímž členem byl i Waal.

Druhou fází projektu, jejímž cílem byl převod systému do počítačové podoby, si vzala na starost univerzita v Utrechtu a pracovala na ní od roku 1990 do roku 2001. Poté byla správa systému opět předána Nizozemské královské akademii, kde Iconclass získal nové jazykové varianty a od roku 2004 je volně přístupný přes online vyhledavač.

Zatím poslední změnou v managementu Iconclassu bylo předání jeho správy RKD, kde byl vyvinut i Iconclass 2100 Browser, přes který je možné od konce roku 2009 systém volně využívat. Díky popisu objektů v Iconclassu bylo mnoho děl holandských mistrů od 14. do 19. století zpřístupněno i přes databázi RKDimages. Předmětová hesla z Iconclassu jsou rovněž využita i v databázi knih a periodik RKDlibrary.

5.1.1.2. obsah

Základním stavebním kamenem Iconclassu jsou jeho alfanumerické klasifikační kódy, které se nazývají **notace**. Ty začínají na nejvyšší úrovni hierarchie vždy číslem od 0 do 9. Iconclass tedy zahrnuje deset základních tříd, které se následně rozdělují na nižší podtřídy.

Notace se používají pro popis a indexaci subjektů a motivů znázorněných v obrazových dokumentech. Pro co nejkompletnější popis subjektu může být samozřejmě použito více notací. V tabulce je uveden celkový výčet tříd notací na nejvyšší úrovni.

<i>číslo třídy</i>	<i>originální název třídy</i>	<i>český ekvivalent</i>
0	Abstract, Non-representational Art	abstrakce
1	Relegion and Magic	náboženství a magie
2	Nature	příroda
3	Human being, Man in general	lidské bytosti, lidé obecně
4	Society, Civilization, Culture	společnost, civilizace, kultura
5	Abstract Ideas and Concepts	abstraktní pojmy
6	History	historie
7	Bible	Bible
8	Literature	literatura
9	Classical Mythology and Ancient History	antická mytologie a dějiny starověku

V těchto **10 kategoriích** můžeme nalézt více než 28 000 hierarchicky řazených definic. Kromě alfanumerického kódu každá notace obsahuje ještě ikonografický popis subjektu, který tvůrci nazývají tzv. textovou korelací (korelace = „vzájemná závislost; souvztažnost jevů pojmech [Kraus, 2007, s. 572]“).

Kromě samotného klasifikačního systému obsahuje Iconclass ještě dvě součásti. První z nich je abecední index, v současnosti obsahující na 14 000 předmětových hesel používaných pro lokaci notace a textové korelace. Umístění notace je rovněž potřebné pro popis a/nebo indexaci obrazu.

Poslední částí Iconclassu je rozsáhlá bibliografie se 40 000 odkazy na nejruznější knihy, články a jiné textové zdroje s ikonografickým zaměřením. V době psaní této práce není bibliografie Iconclassu online dostupná.

Iconclass zahrnuje kromě notace a textové korelace v některých případech u notací ještě pomocné nástroje, kterými jsou: text v závorkách, tzv. klíče, zdvojení písmen či tzv. strukturální číslice. Všechny tyto nástroje slouží ke zpřesnění vyhledávaného pojmu.

Text v závorkách zjednodušuje orientaci uživatele v alfanumerických notacích tím, že uvádí osobní jméno nebo název objektu rovnou s notací – např. podtřída 25G41 patří skupině objektů „květiny“ ale pro lepší orientaci uživatele jsou u podtřídy uvedeny v závorkách konkrétní pojmenování květin, např. 25G41(LILY) pro lilie, 25G41(ROSE) pro růže atp.

Klíče se využívají pro určitý rozsah notací. Vytváří se pomocí řetězců číslic, kterým vždy předchází závorky a znaménko „+“. Např. notace pro objekt „lev“ je 25F23(LION). Pokud ale chceme vyhledat obraz lva, který se vyskytuje pouze v erbu nějakého rodu, přidáme klíč +12 (notace 12 značí třídu objektů „heraldic animals“ – zvířata v heraldice). Výsledná notace obohacená o klíč bude vypadat 25F23(LION)(+12) a systém bude vyhledávat pouze obrazy s motivem lva v rámci heraldiky.

Díky zdvojení písmen a strukturálním číslicím v notaci je také možno modifikovat význam hledaného objektu, popis těchto nástrojů ale přesahují rozsah práce a pro zájemce odkazují na manuál Iconclassu dostupný přes web (viz seznam literatury).

5.1.1.3. struktura a princip

Jak již bylo řečeno, každá hlavní třída Iconclassu se logicky rozvětňuje do menších podtříd, které jsou typické zvýšenou specifičností v porovnání s třídou hlavní. Počet podtříd v každé hlavní třídě nesmí být vyšší než 9. Je to z části proto, že podtřída je značena přidáním číslice od 1 do 9 k číslici označující hlavní třídu. Jako příklad uvedeme rozdělení třídy číslo 2

– příroda (nature).

<i>číslo podtřídy</i>	<i>originální název podtřídy</i>	<i>český ekvivalent</i>
21	the four elements, and ether, the fifth element	čtyři elementy a éter, pátý element
22	natural phenomena	přírodní jevy
23	time	čas
24	the heavens (celestial bodies)	nebe (nebeská tělesa)
25	earth, world as celestial body	Země, svět jako nebeské těleso
26	meteorological phenomena	meteorologické jevy
29	surrealia, surrealistic representations	surreálně, surrealistická zobrazení

Třetí hierarchická úroveň představuje rozdělení podtříd na podtřídy nižšího stupně, které jsou opět specifitější skupinou. Označení podtříd druhého stupně se značí přidáním velkého písmene anglické abecedy. Rozdělení podtřídy 22 – přírodní jevy (natural phenomena), vypadá v systému následovně:

<i>číslo podtřídy</i>	<i>originální název podtřídy</i>	<i>český ekvivalent</i>
22(+0)	(+variant)	jiná varianta
22A	matter	hmota, látka
22B	electricity and magnetism	elektřina a magnetismus
22C	light	světlo
22D	sound	zvuk
22E	temperature	teplota

Další hierarchické členění se značí opět přidáním číslice od 1 do 9. Například podtřída 22D – zvuk (sound) se dále rozděluje na:

<i>číslo podtřídy</i>	<i>originální název podtřídy</i>	<i>český ekvivalent</i>
22 D(+0)	(+variant)	jiná varianta
22 D 1	loud sound, noise	hlasitý zvuk, hluk

22 D 2	soft sound	měkký zvuk
22 D 3	sound-barrier	zvuková bariéra
22 D 4	echo	ozvěna
22 D 9	silence	ticho

Postup od jedné ze základních skupin hierarchicky ke skupinám nižším je pouze jednou z možností, jak se dobrat hledaného pojmu. Další možností je přímý dotaz zadaný do browseru Iconclassu. Pokud bychom hledali například termín „zvuk (sound)“, tedy podtřídu s notací 22 D, browser nám kromě celé skupiny 22 D (včetně jejích podtříd) nabídne i všechny třídy a podtřídy, které v názvu také obsahují hledaný pojem nebo pojem příbuzný – např. 46 C 28 4 fathoming, sounding; 22 D 11 0 'Armonia' (Ripa); 49 E 24 acoustics; 46 E 65 phonograph, gramophone; playing recorded sound or video; 95 A (MEMNON); 71 the speaking statue of Memnon: a seated stone figure uttering musical sound every day at sunrise atd.

Pro zajímavost je níže zobrazen výstup vyhledávání v třídě 48 C 73:

The screenshot shows the Iconclass website interface. At the top, there is a search bar and navigation links like 'Options - Help - About'. Below the search bar, there is a gallery of art images. A sidebar on the left lists various classes, with '48C73 musical instruments; group of musical instruments' highlighted. A detailed list of subclasses for 48C73 is shown on the right, including 48C73(+0) to 48C73(+7) and 48C731 to 48C739.

Obrázek 10 Výstup vyhledávání v Iconclass; [RKD, 2010]

5.1.1.4. použití Iconclassu v praxi

Využití Iconclassu v praxi se sestává ze tří klíčových momentů práce s informacemi – jedná se o jejich produkci, vyhledávání a zobrazování. O produkci informací, tedy o popis a indexaci děl na základě jejich obsahu, se stará zaštiťující agentura, již zmiňovaná RKD.

Systém Iconclass je za účelem vyhledávání přístupný přes Iconclass 2100 Browser. Možnost vyhledávání informací přes tento prohlížeč využívají nejen historici umění, rešeršéři a jiní pracovníci kulturních institucí po celém světě, ale i široká veřejnost. V systému se kromě procházení hierarchické struktury dá využít i přímé vyhledávání podle klíčových slov. Zajímavý příklad vyhledávání je uveden na webovém sídle Iconclassu – jde o spojení slov „lifting“ a „hat“. Po jejich spojení se nám zobrazí následující hierarchický kontext a s ním spojená umělecká díla – viz obrázek 11.

0 Abstract, Non-representational Art
1 Religion and Magic
2 Nature
3 Human Being, Man in General
33 relations between individual persons
33A non-aggressive relationships
33A1 saluting
33A11 baring the head, lifting one's hat
friendliness · hat · head · human being · non-aggressive · relationship · saluting · uncovering

The image shows a screenshot of the Iconclass hierarchy. On the left, a list of hierarchical levels is shown, with the final level, '33A11 baring the head, lifting one's hat', highlighted in a grey box. Below this level, a list of associated terms is provided: 'friendliness · hat · head · human being · non-aggressive · relationship · saluting · uncovering'. Two arrows originate from the highlighted level: one points to a painting of a man in a red and blue robe lifting his hat to a woman in a yellow and blue dress, and the other points to a painting of a man in a black coat and hat saluting a woman in a yellow coat and hat.

Obrázek 11 Příklad vyhledávání pojmů „lifting“ a „hat“; [RKD, 2010]



The Union List of Artist Names je stejně jako AAT tezaurus (o němž bude pojednáno v následující kapitole) součástí databázi Getty. Getty databáze jsou řízeny slovníkovým programem Getty (Getty Vocabulary Program).

Slovníkovým z toho důvodu, že ULAN i AAT jsou propracovanými oborovými řízenými slovníky. tzn. tezaury, vytvořenými podle ISO standardů. Hlavním cílem tohoto unifikovaného seznamu uměleckých jmen je zlepšení přístupu k informacím o umění, architektuře a hmotné kultuře.

Jako ontologie může být nahlížen kvůli několika charakteristickým aspektům. Tím nejpodstatnějším je to, že ohnisko této znalostní báze – jednotlivé záznamy umělců – osob i celých uměleckých institucí, jakožto korporací – jsou vzájemně propojeny vazbami. Ty sémanticky popisují vztahy mezi osobami umělců, které záznamy reprezentují. Použití této relativně jednoduché sémantické sítě může výrazným způsobem zefektivnit vyhledávání informací v tomto oboru, zvláště pokud je k ní přistupováno jako k pomůcce pro vyhledávání v databázích informačních systémů.

ULAN zahrnuje vlastní jména a související informace o umělci. Umělci mohou být buď fyzické osoby, nebo spolupracující skupiny jednotlivců (právnických osob). Umělci v ULANu obecně představují tvůrce, kteří se podílejí na koncepci a tvorbě výtvarného umění a architektury. Obsahuje i významné dárce nebo dokonce repozitáře (organizace odpovědné za uchovávání digitálního materiálu).

Jde o složený zdroj, který bohužel není komplexním nástrojem, přestože se neustále rozrůstá o jména umělců současných i dřívějších.

Jako primární zdroj informací v následujících podkapitolách byly použity webové stránky projektu [Getty ULAN, 2010].

5.1.2.1. vývoj a charakteristika

Práce na ULANu započaly v roce 1984. Jeho vznik byl iniciován potřebou rozvoje starších projektů běžících pod záštitou The Getty Research Institute v Los Angeles. Postupem času se ale ukázalo, že ULAN je schopný uspokojovat informační potřeby i širší komunity zájímaví se o katalogizaci a vyhledávání. Proto začala společnost Getty ULAN distribuovat stejným způsobem (včetně konstrukce a správy slovníku ULAN), jako svůj předcházející projekt AAT, o kterém bude pojednáno dále.

Původně byl tento tezaurus konstruován jakožto abecední unifikovaný seznam jmen

umělců s připojenou bibliografií. Ke konci 90. let byl ale uveden do souladu s mezinárodními standardy pro tvorbu tezaurů a jeho působnost byla rozšířena i na právnické osoby, jako jsou například architektonické firmy a repozitáře umění, které mohou mít i hierarchické úrovně.

Jádro ULANU, jak už bylo naznačeno, tvoří osobní či korporativní jména nutná pro katalogizaci a vyhledávání informací ohledně uměleckých děl vizuálního umění a architektury. Jádro stojí na současné terminologii, garantované užitím autoritativních literárních zdrojů a ověřené odborníky na kunsthistorii.

Během let byl ULAN intenzivně vylepšován jak komunitou uživatelů, tak mnohými specializovanými editory (v editačním systému speciálně vyvinutém pro potřeby databáze ULAN). V roce 1994 byl publikován v tištěné formě, i ve formě strojem čitelných souborů. V roce 1997 bylo zrušeno vydávání tištěné podoby a v současnosti je publikován pouze digitálně ve dvou formách: (a) jako prohledavatelné online webové rozhraní a (b) jako datové soubory dostupné na základě licence.

5.1.2.2. obsah

Časově pokrývá ULAN období od starověku po současnost. Geograficky pak není pole působnosti nijak omezeno. Neustále se rozrůstá o záznamy umělců nejen ze západních kultur (např. umělci z Číny, Japonska nebo rodilí Američané) ale i o nové typy umělců, které nebyly v ULANu doposud zahrnuty (jako např. designéři, ilustrátoři, tvůrce tzv. Electronic Mail Art, kaligrafové etc.).

Minimální záznam má přiděleno unikátní identifikační číslo, obsahuje jméno osoby či korporace, její roli, národnost a data s ní spojená.

Většina záznamů je psaná v angličtině, ale přesto ULAN podporuje i vícejazyčnost. Pokud jsou jména přejímána z jazyků psaných jiným písmem než latinkou, jsou pro ULAN transliterována podle příslušných ISO standardů.

Jestliže je v záznamu umělce uvedeno více jmen, jsou všechna prolinkována na jeden společný identifikátor zvaný „Subject ID“ a jsou považována za ekvivalenty (synonyma). Za synonyma k vybrané preferované formě jsou považována i obrácená pořadí jména a příjmení, přezdívký, variantní jména, pseudonymy, rodná ženská příjmení či jinak změněná jména, nebo jména s lišící se ortografií. Preferovaná forma jména je v zásadě uvedena v anglickém jazyku (americká angličtina). Níže je uveden příklad:

Names:

Kandinsky, Vassily (preferred, V, index)


Vassily Kandinsky (V,display)
Kandinskii, Vasilii Vasilevich (V)
Kandinskij, Vasilij Vasil'evic (V)
Kandinsky, Vasily (V)
Kandinskij, Vasilij (V)
Kandinski, Vasilij (V)
Vasilij Kandinskij (V)
Kandinsky, Wassili (V)
Kandinsky, Wassily (V)
Kandinsky, Wassily Wassiljewitsch (V)
Kang-ting-ssu-chi, Wahsili (V)
Kandinsky, Vasili Vasilevich (V)
Vasili Vasilevich Kandinsky (V)
Wassily Kandinsky (V)

V době psaní této práce zahrnoval ULAN okolo **293 000 unikátních jmen** (z toho 120 000 osobních jmen) a jiných informací o umělcích. Kromě variantních jmen umělce je ULAN koncipovaný tak, aby provázal se jménem umělce informace o datech, pojících se s umělcovým jménem (pokud taková jsou – myšleno je tím datum, od kterého umělec používá jméno, pokud před tím byl známý pod jménem jiným), jeho roli (kterých může být více) a době, ve které tvořil, originálním jazyku jeho jména, datech narození a případně úmrtí, národnosti a pohlaví. Důležitými informačními poli pojícími se se jménem jsou dále místa narození a případně úmrtí, významné události (jimiž jsou myšleny např. udělení ocenění, přestěhování atp.) a data a místa, která se s těmito událostmi pojí. Dále je umělcovo jméno provázáno s biografií, označením (v originále pole zvané „label“ – jde o stručnou a výstižnou textovou identifikaci autora), dalšími jmény osob či korporací, mezi kterými existuje určitá vazba a s dalšími informacemi ve formě poznámky.

Pokud jde o korporaci, je často její záznam provázán s informací o její pozici v hierarchii nebo polyhierarchii. U osobních jmen je nutné definovat vztah tohoto jména (části) k celku (hierarchie) – např. William Turner byl osobou a proto pojem „Turner, William“ je prvkem třídy „Person“ a jeho vztah k této třídě je definován jako „child of“. Tímto způsobem je vytvářena sémantika ULANU – viz obrázek 12.

Click the  icon to view the hierarchy.

ID: 500003526 Record Type: **Person**


 **Turner, William** (British painter, 1789-1862)

Names:
 Turner, William (**preferred**,V,index)
 William Turner (V,display)
 Turner, William, of Oxford (V)
 Turner of Oxford (V)
 Turner, Of Oxford (V)

Nationalities:
 British (**preferred**)

Roles:
 artist (**preferred**)
 painter

Gender: male

List/Hierarchical Position:
 ... Person
 Turner, William (I)

Obrázek 12 Příklad hierarchie v ULANu; [Getty ULAN, 2010]

Z výše jmenovaných polí stojí za zdůraznění některé z nich, jež jsou uvedeny níže v tabulce i s krátkým popisem [Breitman, 2006, s. 248]:

<i>pole</i>	<i>pole originál</i>	<i>popis</i>	<i>příklad</i>
Identifikační číslo objektu	Subject ID	unikátní číslo přidělené každému jednomu záznamu	500006031
Typ záznamu	Record Type	blíže charakterizuje typ záznamu	Person, Corporate Body
Označení	Label	stručná textová identifikace umělce	(Venetian painter and draftsman, ca. 1488-1576) [500031075]
Jména	Names	slova či fráze odkazující na reálnou osobu/korporaci	Manet, Édouard (preferred,V,index,LC) Édouard Manet (V,display) Manet, Édouard (V) Manet, Edouard (V) Edouard Manet (V)
Národnost	Nationality	odkaz na národnost, kulturu nebo etnickou skupinu	Brazilian (preferováno), South American
Role	Roles	slova či fráze charakterizující způsob	artist (preferováno), painter, printmaker, ceramist, studio,

		umělcovy tvorby	manufactury, workshop
Seznam/Hierarchická pozice	List/Hierarchical Position	u individuálních umělců jde většinou o vazbu ke třídě „Osoba“, u korporací může pozice nabývat větší hloubky	...Person ...Warhol, Andy – ...Corporate BodyNational Gallery of ArtFoundation for Documents of Architecture
Zdroje a přispěvatelé	Sources and Contributors	instituce nebo projekty, které do ULANU přispěli informacemi	J. Paul Getty Museum, collections online (2000-)
Poznámka (Popisná poznámka)	Note (Descriptive Note)	popis kariéry umělce, jeho vztahů k jiným umělcům nebo použití jeho jména	He was a successful caricaturist in his native Le Havre, but after studying plein-air landscape painting, he moved to Paris in 1859. He soon met future Impressionists Camille Pissarro and Pierre-Auguste Renoir...

5.1.2.3. struktura a princip

Přestože je ULAN zobrazen jako seznam, je strukturován jako tezaurus a proto obsahuje hierarchické, ekvivalentní a asociativní vztahy. Jádrem každého záznamu je jméno nebo název. Každý takový záznam je určen jedinečným identifikačním číslem. Záznam je prolinkován se souvisejícími jmény, zdroji dat a poznámkami.

I když se může zdát, že struktura ULANu je relativně plochá, obsahuje ULAN několik hierarchických úrovní. Ty se rozvětvují ze společného kmenu nazvaného „Top of the ULAN hierarchies“ (Subject_ID: 500000001) do tří hlavních větví. Jde o osoby, korporace a neznámé autory. Korporace lze někdy dále rozdělovat do menších větví, výjimečně se dále člení i větve osob. Za korporaci se považují i dva (a více) umělců, jež spolupracují například v rámci workshopu nebo jakékoli jiné kolektivní tvorby. Vzhledem k tomuto faktu lze označit strukturu ULANu jako polyhierarchickou.

To, co z ULANu ale dělá ontologii v pravém smyslu slova, nejsou pouze vztahy hierarchické, ale také **vztahy relační**. Těch nalezneme v ULANu celkem 6, nejčastěji se vyskytují mezi osobními jmény umělců – viz tabulka (kódy vlastností jsou koncovému

uživateli při práci s databází skryty) [Breitman, 2006, s. 246]:

<i>kód</i>	<i>vlastnost aj</i>	<i>vlastnost čj</i>	<i>spojený/inverzní kód</i>
1302	associate of	spojený (s kým) – jde o profesní vztah mezi dvěma entitami	1302
1311	partner of	partner (koho)	1311
1512	parent of	rodič (koho)	1511
1511	child of	potomek (koho)	1512
2828	teacher was	učitel (koho)	2829
2829	student was	student (koho)	2828

příklad:

Titian

colleague of Verdizotti, Giovanni Maria (Italian painter, draftsman, and writer, 1525-1600) [500012500]

influenced Padovanino (Italian painter, 1588-1649) [500025136]

member of Vecellio family (Italian artists, documented from the 13th century) [500115634]

parent of Vecellio, Orazio (Venetian painter, born ca. 1515, died before 1576) [500006561]

patron was Este, Isabella d' (Italian patron and collector, 1474-1539) [500115132]

related to (familial) Tizianello (Venetian painter and author, ca. 1570-ca. 1650) [500005710]

sibling of Vecellio, Francesco (Italian painter, born 1475, died 1559 or 1560) [500005079]

student of Bellini, Giovanni (Venetian painter, born 1431-1436, died 1516) [500019244]

teacher of Rosa, Pietro (Italian painter, ca. 1541-1577) [500094217]

5.1.2.4. použití ULANu v praxi

Využití tezauru ULAN už bylo naznačeno: indexátoři a katalogizátoři ho můžou použít jako řízený slovník nebo autority – zprostředkovává preferované formy jmen, dále může být použit jako výzkumný nástroj, a to díky bohatým informacím a kontextovým znalostem, které obsahuje.

Nejtypičtějším uživateli ULANu jsou především muzea, umělecké knihovny,

repozitáře, archivy, katalogizátoři vizuálních zdrojů, účastníci bibliografických projektů zaměřených na oblast výtvarného umění, teoretici a historici umění, informační specialisté, ale i studenti a širší veřejnost.

Slovník je možné využít hned dvěma způsoby. Za prvé je možné jej implementovat přímo do řídicího systému databáze nejrůznějších uměleckých sbírek, či za druhé využívat jím poskytované informace z on-line databáze na webových stránkách Getty. Pro koncového uživatele může být výhodná i varianta, kdy je slovník transformován do formátu XML nebo relační tabulky, a posléze vydáván jednou ročně. Data zprostředkovávaná přes webové služby jsou aktualizována každé 2 týdny. Licencované soubory neobsahují žádné uživatelské rozhraní.

Výstup ze systému vypadá následovně:

The screenshot shows the 'Full Record Display' for Mark Rothko in the Getty ULAN. Key sections include:

- Header:** Research, Union List of Artist Names® Online, Full Record Display.
- Basic Info:** ID: 500014869, Record Type: Person.
- Note:** A paragraph describing Rothko as a primary artist of Abstract Expressionism, born in Russia, moved to Portland in 1913, and later to New York, where he was a founding member of the 'Ten' group.
- Names:** Rothko, Mark (preferred), Mark Rothko, Rothkowitz, Marcus (variant).
- Nationalities:** American (preferred), Russian, Jewish.
- Roles:** artist (preferred), painter, abstract artist.
- Gender:** male.
- Birth and Death Places:** Born in Daugavpils (Latvia), died in New York City (USA).
- Events:** Immigration in 1913.
- Related People or Corporate Bodies:** Lists colleagues like Clifford Still, parent Kate Rothko, and teachers like Max Weber and Viola Frey.
- List/Hierarchical Position:** Shows the path from Person to Rothko, Mark (I).
- Biographies:** Lists various controlled forms of the name (e.g., [VP Preferred], [WLCI], [WCP], [GRLPSC], [BHA], [AVERY]).
- Sources and Contributors:** Lists sources like Getty Vocabulary Program rules and AVERY files.
- Subject:** Lists subject terms like [AVERY, BHA, GRLPSC, VP, WCI, WCP, WL-Courtauld] and related works.
- Note:** Additional references to art contexts and museum collections.

Obrázek 13 Výstup z tezauru ULAN na dotaz „Rothko“; [Getty ULAN, 2010]



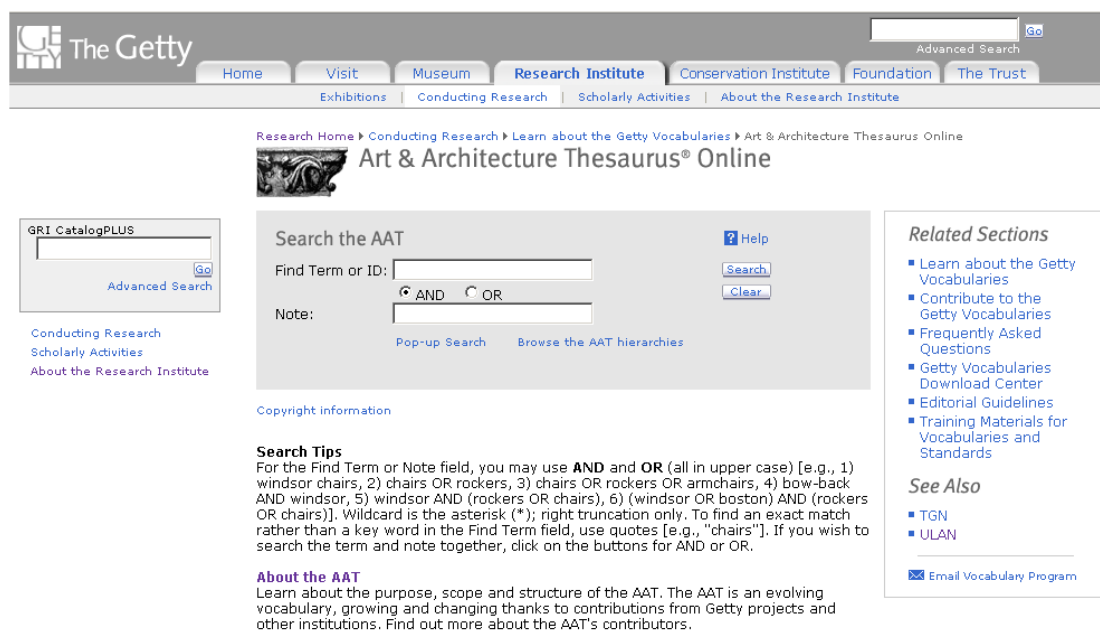
AAT (Art & Architecture Thesaurus) je po ULANu a TGN (Thesaurus of Geographic Names), který nebude jako jediný vzhledem ke svému zaměření detailněji popsán, další významnou databází společnosti **Getty**. Stejně jako ULAN i AAT staví na sémantické síti znalostí, proto je považován za oborovou ontologii.

Zaměřuje se na informace o umění, architektuře a materiální kultuře a jeho hlavním cílem je vylepšit přístup k těmto druhům informací. Na rozdíl od ULANu neobsahuje žádné konkrétní subjekty (tedy ikonografické subjekty nebo vlastní jména), ale pouze obecně použitelnou terminologii a jiné informace o objektech, umělcích, pojmech a místech, s nimiž se lze na poli umění a architektury setkat. Jinými slovy, každý pojem obsažený v AAT lze použít k označení více uměleckých entit (např. obsahuje termín „katedrála“, ale již ne další specifické vlastní názvy spadající pod tento pojem jako je např. „Sagrada Família“).

Všechny tyto obecné termíny jsou navzájem sémanticky propojeny, proto lze i v tomto případě hovořit o formální doménové ontologii. Ta ve srovnání s ULANem ve většině případů zasahuje definovanými pojmy do více úrovní, je tedy také polyhierarchická, nicméně termíny člení hlouběji. Stejně jako v případě ULANu jde o složený zdroj, který prozatím zcela nevyčerpává danou tematiku.

Je nasnadě, že přístup k databázi i její design je shodný s databází ULAN, protože se jedná o jedno uživatelské rozhraní – viz obr. 14. Přesto je jejich obsah diametrálně odlišný – domnívám se, že ve vzdálenější budoucnosti nelze vyloučit jejich vzájemné propojení do jediného zdroje, který by svou komplexností pokrýval téměř dvojnásobnou síť informací a znalostí ve výtvarném umění.

Jako primární zdroj informací v následujících podkapitolách byly použity webové stránky projektu [Getty AAT, 2010].



Obrázek 14 Vyhledávací rozhraní tezauru AAT; [Getty AAT, 2010]

5.1.3.1. vývoj a charakteristika

Na počátku vzniku AAT stála informační potřeba uměleckých knihoven a indexačních agentur časopisů o umění spojená s rozšiřující se automatizací – šlo o podporu a zefektivnění získávání informací – tezaurus se stal nástrojem, jež by mohl v tomto ohledu výrazně pomoci.

Jeho počáteční jádro bylo zaměřeno na různé skupiny uživatelů, tudíž se na něm podíleli i odborníci z různých oborů. Na AAT byly kladeny vcelku náročné požadavky, které by umožnily i vyhledávání v odlišných zdrojích dat o uměleckých objektech.

Základní principy ustanovili v roce 1981 spolu s dalšími specialisty jeho tvůrci Toni Peterson, Dora Crouch a Molholt. Terminologie jádra AAT zahrnuje pojmy potřebné pro katalogizaci a vyhledávání informací o vizuálním umění a architektuře. AAT je konstruován na základě národních a mezinárodních standardů pro tvorbu tezaurů. Jeho hierarchie byla inspirována stromovou strukturou MeSH (Medical Subject Headings Thesaurus). Veškeré termíny jsou validovány renomovanými odborníky na oblast výtvarného umění.

AAT se ve své historii dočkal publikování i v papírové podobě, konkrétně v letech 1990 a 1994 – současně byl samozřejmě distribuován i ve stroji čitelné podobě. V roce 1997 se od publikování nepraktické tištěné podoby upustilo, vzhledem k lepším možnostem nabízených webovými technologiemi. Stejně jako ULAN je v současnosti AAT dostupný buď přes licenci v datových souborech anebo přes online webové rozhraní. Celý tezaurus je upravován i ve stejném redakčním systému jako ULAN, který dovoluje kompilovat data

různých přispěvatelů, měnit hierarchické úrovně, slučovat, přesouvat a publikovat je v různých formátech. Podporuje také snadný překlad databáze do cizích jazyků.

5.1.3.2. obsah

Tezaurus AAT vymezuje časový rozsah od starověku až po současnost. Pojmy tvořící základnu AAT jsou globálního charakteru. Pojmový rámec, do něhož je zasazena charakteristika tříd a jejich hierarchie, je navržen tak, aby dovolil pouze obecné klasifikační schéma pro oblast umění a architektury – neoznačuje specifické pojmy. V praxi to znamená, že i když v tezauru najdete termíny „Baroque“ (baroko) a „paintings“ (malby), žádná část AAT přesně nespecifikuje „Baroque paintings“ (barokní malby). Přesná charakteristika tohoto spojení je obsažena v odlišných hierarchických úrovních AAT.

Záznam, který tvoří základní stavební kámen AAT se také označuje jako **subjekt**. Jeho minimální podoba obsahuje unikátní číselné identifikační číslo záznamu, pojem (který je jádrem záznamu) a jeho umístění v hierarchii AAT. V současnosti obsahuje AAT na 34 000 pojmů, které jsou popsány 131 000 termíny, deskriptory, bibliografickými citacemi a dalšími doplňujícími informacemi. Termín pro každý pojem by měl obsahovat jak svůj název v singuláru, tak i v plurálu a pokud jde o termín víceslovný, tak by měl zahrnout i veškeré jiné pořadí slov než přirozené včetně variantního přepisu odlišné ortografie. Tyto variantní termíny by měly být označeny jako synonyma nebo ekvivalenty k preferovanému termínu, který, pokud je to možné, je v preferované formě v plurálu.

Původní anglická verze byla kompletně přeložena do španělštiny, nyní se pracuje na překladu do čínštiny, němčiny, holandštiny a současně se do systému integruje kolem 3000 termínů v italštině a francouzštině. V případech transliterace z originální latinky do jiných abeced se AAT podřizoval příslušným ISO standardům.

5.1.3.3. struktura a princip

Hlavní pododdělení hierarchické struktury v AAT tvoří tzv. **fazety** (facets - aspekty). Fazety obsahují homogenní třídy objektů, jež se v rámci jedné fazety odlišují svojí charakteristikou od fazety jiné. Dohromady fazety dokladují heterogenní povahu tezauru. Například pojem „marble“ (mrarmor) odkazuje na fazetu „Materials“ (materiály) a pojem „Impressionist“ (impresionista) na fazetu „Styles and Periods“ (Styly a období).

Fazety jsou organizovány do systému, který přechází od pojmů abstraktních k pojmům konkrétním. V tabulce níže je uveden kompletní výčet všech fazet, které jsou v AAT obsaženy:

<i>název fazety</i>	<i>název fazety v originále</i>	<i>charakteristika fazety</i>	<i>příklad</i>
spřátelené pojmy	Associated Concepts	obsahuje abstraktní pojmy a jevy spojené se studiem a výkonem lidských myšlenek a aktivit zahrnujících umění a architekturu ve všech médiích a disciplínách	beauty (krása), balance (rovnováha), freedom (svoboda)
fyzické atributy	Physical Attributes	týká se vnímatelných nebo měřitelných vlastností materiálů a uměleckých artefaktů, jako jsou velikost, tvar, chemické vlastnosti, tvrdost, textura, vzor a barva	borders (hranice), round (kulatý), brittleness (křehkost)
styly a období	Styles and Periods	zprostředkovává běžně akceptované termíny pro styly a rozdílná časová období	Louis XIV (Ludvík 14.), Black-figure (černofigurová), Abstract Expressionist (abstraktní expresionismus)
agenti	Agents	obsahuje podmínky pro označení osob nebo skupin osob a organizací, identifikovaných podle povolání nebo činnosti	printmakers (grafici), landscape architects (zahradní architekti), corporations (korporace)
aktivity	Activities	zahrnuje oblast fyzické a duševní činnosti, systematické sekvence akcí, metod a procesů	exhibitions (výstavy), drawing (kreslení), corrosion (koroze)
materiály	Materials	zabývá se fyzickými a fyzikálními materiály, ať již přirozenými nebo uměle vyrobenými	iron (železo), clay (hlína), ivory (slonovina)
objekty	Objects	jsou nejrozsáhlejším aspektem AAT – zahrnuje neživé, hmotné viditelné artefakty, jež jsou produktem lidské činnosti a zahrnuje i prvky krajiny	paintings (obrazy), cathedrals (katedrály), gardens (zahrady)

Kromě pojmů obsažených v jednotlivých fazetách obsahuje každý záznam v AAT

informace o jazyku (respektive vícejazyčnosti záznamu), diakritice, identifikačním čísle záznamu, typu záznamu, dále stručnou textovou popisnou informaci o pojmu (tj. label), fazety nebo hierarchickou pozici, hierarchický typ vztahu, termíny etc.

5.1.3.4. použití AAT v praxi


AAT lze efektivně využít jako řízený slovník nebo autoritu při katalogizaci nebo indexaci uměleckých objektů. Na druhou stranu může být nápomocný během vyhledávání v (nejen Getty) databázích se zaměřením na tematiku výtvarného umění. Je také dobrým rešeršním nástrojem díky bohatým základním i kontextovým informacím, jež poskytuje. Nejvíce jej pravděpodobně využijí informační pracovníci v kulturních institucích, jako jsou galerie, muzea, umělecké knihovny, archivy nebo specificky katalogizátoři sbírek složených z vizuálních zdrojů nebo zaměstnanci bibliografických projektů týkajících se umění či jiní odborní pracovníci. Obecně je AAT velmi užitečným i pro studenty, kurátory, historiky, architekty, knihovníky a veřejnost.

Příklad výstupu z AAT je uveden níže na obrázku 15.

Research
 Research Home > Conducting Research > Art & Architecture Thesaurus > Full Record Display

Art & Architecture Thesaurus® Online
 Full Record Display

[New Search](#) [Previous Page](#) [Help](#)

Click the  icon to view the hierarchy.

ID: 300014078 **Record Type: concept**

canvas (<textile materials by process or technique>, <textile materials>, ... Materials (Hierarchy Name))

Note: Closely woven textile made in various weights, usually of flax, hemp, jute, or cotton, used especially for sails, tarpaulins, awnings, upholstery, and as a support for oil painting. Also used for a loosely woven, latzelike mesh made of similar material, used as a needlepoint foundation.

Terms:

- canvas** (preferred, C,U,I,C,English-P,D,U,N)
- canvass** (C,U,English,U,F,U,N)
- doek** (C,U,Dutch-P,D,U,N)
- toile (canvas)** (C,U,French-P,D,U,N)
- Nessel** (C,U,German-P,D,U,N)
- Leinwand** (C,U,German,AD,U,N)
- tela (canvas)** (C,U,Italian-P,D,U,N)
- (canvas)** (Spanish-P,D,U,N)
- lona** (C,U,Spanish,AD,U,U)
- lieno** (C,U,Spanish,AD,U,U)
- duk** (C,U,Swedish-P,D,U,PN)

Facet/Hierarchy Code: M,MT

Hierarchical Position:

- Materials Facet
- Materials (Hierarchy Name) (G)
- materials (matter) (G)
- <materials by form> (G)
- <materials by physical form> (G)
- <fiber and fiber products> (G)
- <fiber products> (G)
- <textile materials> (G)
- <textile materials by process or technique> (G)
- canvas (G)

Additional Parents:

- Materials Facet
- Materials (Hierarchy Name) (G)
- materials (matter) (G)
- <materials by functions> (G)
- artists' materials (G)
- supports (artists' materials) (G)
- canvas (G)

Additional Notes:

Spanish Úsese para un textil a talar apretado fabricado en diferentes pesos, habitualmente de flax, cáñamo, yute o algodón, utilizado fundamentalmente para confeccionar velas de barco, tela encerada, toldos, tapicería de muebles y como base para la pintura al óleo. Se utiliza también para trama tipo enrejado hecha de material similar, utilizada como base para bordados de aguja.

Sources and Contributors:

- canvas..... [VP Preferred]
- Avery Index (1963-)
- Brady and Clausner, Materials Handbook (1977) 229
- CDMARC Subjects: LCSH (1988-)
- Clabburn, National Trust Book of Furnishing Textiles (1988) 242
- Clabburn, Needleworker's Dictionary (1976) 50
- IFLA Glossary for Art Librarians (1984)

Subject: [CDBP-DIBAM, CHIN, VP]

- CHIN database
- RIBA, Architectural Keywords (1982) Textiles
- Swift, Larousse Encyclopedia of Embroidery Techniques (1984) 35; canvas grounds


Note:

- English [VP]
- IFLA Glossary for Art Librarians (1984)
- Spanish..... [CDBP-DIBAM]
- TAA database (2000-)

[New Search](#)

[Back to top](#)

[Printer-friendly version](#)

 The J. Paul Getty Trust

© 2004 J. Paul Getty Trust
 Terms of Use / Privacy Policy / Contact Us

Obrázek 15 Výstup z AAT na dotaz „canvas“ („plátno“); [Getty AAT, 2010]

5.1.4. CIDOC CRM



CIDOC představuje jeden z pojmových referenčních modelů, označovaných CRM (*Conceptual Reference Model*). Nejedná se o nic menšího, než o propracovanou formální oborovou ontologii, určenou k podpoře integrace, zprostředkování a výměny různorodých informací z oblasti kulturního dědictví, potažmo tedy i výtvarného umění. Původní model CIDOC ale nebyl z počátku ontologií v pravém slova smyslu – představoval model databázového schématu, který byl vyvíjen na základě potřeby zajištění společného rámce pro komunikaci informací, už od svého vzniku zaměřeného na oblast kulturního dědictví.

Datový model CIDOC byl od svého vzniku vyvíjen na základě předpokladu, že pro komunikaci informací je nutné společné datové schéma. Tento názor byl nicméně postupně nahrazen novou vizí, založené na mediaci – zprostředkování – informací mezi systémy, které umožňují správu dat z heterogenních zdrojů. Datový model CIDOC se v průběhu svého vývoje posunul z návrhu databázového schématu k sofistikovanému relačnímu modelu konceptuálních definic. Jak je tomu i u jiných formálních ontologií, primárním současným cílem CIDOC CRM je poskytování sémantických definic, často nutných pro transformaci heterogenních decentralizovaných zdrojů informací do homogenních zdrojů, ať už typu intranetu nebo globálního Internetu. Jinak řečeno, v budoucnosti by mohl pomoci k vytvoření kompaktní globální sítě pro informace z oblasti kulturního dědictví a to zajištěním lepší komunikace informací a vyvarováním se někdy dosti nákladným nedorozuměním.

Nový sofistikovanější přístup k modelu CRM by měl zajistit společný základ pro vývoj interoperabilních informačních systémů, ale zároveň by neměl předdefinovávat žádná řešení otázek na úrovni implementace. Snaží se poskytnout základní jazyk, který usnadní úkoly spojené se sémantickou integrací, ale nevytváří žádné předpoklady o obchodních procesech nebo pravidlech institucí, neobsahuje metodiku činností, ani nedefinuje validační pravidla a omezení, datové formáty nebo prvky uživatelského rozhraní. Použití CRM nezaručuje kompatibilitu mezi odlišnými systémy na úrovni bytů či úrovni dat, ale pomáhá zajistit konceptuální kompatibilitu [Crofts, 2003].

Koncepce počítá i s tím, že uživatelé CRM budou vytvářet vlastní datové struktury (respektive jejich části, jako vztahy nebo obsah) v souladu s CRM a posléze se tyto data budou zpřístupňovat i v jiném prostředí bez ztráty významu. To CRM umožňuje, pokud je nalezen deterministický logický algoritmus, na jehož základě jsou libovolná zakódovaná data

převáděna do kompatibilního tvaru. Tento algoritmus může CRM spojovat s jinými formálními ontologiemi mapujícími obecné znalosti (např. tezaury). CRM dokáže strukturované informace i interpretovat tak, aby velké objemy obsahu dat byly zprostředkovávány automaticky. Informační systémy mohou nabízet nástroje pro export dat do tvaru kompatibilního s CRM, nebo naopak poskytovat přístup k datům prostřednictvím pojmů CRM formou dotazu.

CRM je určen zejména na pokrytí teoretických, historických a geografických kontextuálních informací, využívá dostupné moderní technologie a zároveň umožňuje komunikaci i s vývojově staršími systémy. Doposud bylo nevyhnutelné na různé informační zdroje nahlížet odděleně, bez využití propojení přes tzv. cross-systémy, které jsou stále v celku raritou. CIDOC CRM ale přijímá a interpretuje dotazy a distribuuje je všem zúčastněným systémům. Ty zpětně zašlou odpověď zprostředkovateli (mediátor – v tomto případě CIDOC CRM), který konsoliduje výsledky pro koncového uživatele. Aby tento proces dobře fungoval, je podmínkou aby zprostředkovatelský systém byl schopný komunikovat se všemi zúčastněnými systémy tak, aby dokázal správně pochopit a interpretovat poskytnuté výsledky dotazu – to je komplikované s přihlédnutím k okolnosti, že zúčastněné systémy nebudou pravděpodobně disponovat stejnými datovými schématy a můžou poskytovat různě podrobné informace o objektech, takže zprostředkovatelský systém musí být jakýmsi sémantickým polyglotem, kterým naštěstí CIDOC CRM je. [Crofts, 2003].

Pokud jsou data rozdělena mezi několik různých systémů, použití jednoho referenčního zprostředkovatelského formátu je jednoduchým, účinným a elegantním způsobem, jímž lze vyřešit problém heterogenity zdrojů.

CIDOC CRM se stal ISO standardem v roce 2006 a je označen jako ISO/PRF 21127 (ISO 2006).

Jako primární zdroj informací v následujících podkapitolách byly použity webové stránky projektu [ICOM, 2006].



Obrázek 16 Logo ICOM; [ICOM, 2006]

Současně je systém používán v informační vědě a v mnoha standardních

analytických a vývojových metodách pro IS. Jeho základem je rozlišení mezi konceptuálním zpracováním a technickou realizací IS s podporou specifické znalostní domény.

5.1.4.1. vývoj a charakteristika CRM – informace vs. data

Historie vzniku sahá do roku 1996, kdy na vývoji CIDOCu začaly spolupracovat dvě mezinárodní komise – Mezinárodní komise pro dokumentaci CIDOC (*Comité International pour la Documentation*) a Mezinárodní rada muzeí ICOM (*International Council of Museums*) s podporou RLG's Cultural Materials Initiative. Pracovní skupina zahrnuje na 50 členů z Evropy, Severní Ameriky i Asie a Oceánie. Na vývoji modelu se podílejí kurátoři muzeí různého zaměření, manažeři informačních databází, informatici i informační specialisté, knihovníci, a prodejci systému. Za hlavního autora modelu je považován Patrick LeBoœuf (*National Library of France*).

Podstatu již zmíněné změny přístupu k modelu CRM v průběhu jeho vývoje, lze dobře demonstrovat rozdílem mezi informacemi a daty. Informace může být definována jako smysl sdělení, který je pro různé formy vyjádření společný: *Edvard Munch je autorem obrazu Výkřik, Obraz Výkřik namaloval Edvard Munch, Edvard Munch pintó el cuadro El grito* vyjadřují různými způsoby jeden stejný společný význam. Oproti tomu data mohou být definována jako znaky: slova, písmena, symboly nebo bity, které mají jedinečný specifický význam z hlediska sémiotického kontextu.

CIDOC se nejprve pokoušel definovat konkrétní reprezentaci informací a proto představoval datové schéma. CRM je naproti tomu úroňovým informačním modelem – je určen k upřesnění a vyjasnění pojmů potřebných pro výměnu informací o kulturním dědictví. V rámci tohoto modelu je umožněno definovat různé reprezentace a různá datová schémata. Proto lze označit CIDOC CRM za oborovou ontologii informací kulturního dědictví – je určena k pokrytí specifické oblasti zájmu. Jako ontologie je CIDOC CRM reprezentován objektově orientovaným modelem složeným z tříd organizovaných do hierarchické struktury a propojených navzájem svými vlastnostmi. Uspořádání tříd a vlastností poskytuje rámec i pro složitý popis vzájemných vztahů existujících mezi objekty, osobami, událostmi, místy a pojmy v oblasti výtvarného umění [Crofts, 2003].

Významným rysem CIDOCu CRM je, že anticipuje charakteristické vztahy a relace mezi termíny, přestože ze samotné terminologie nedefinuje nic. Je současně referenčním dokumentem, který vytváří společný pojmový základ napříč jednotlivými disciplínami a

doménami. Používá bohatou hierarchii, širokou sadu dědičných vlastností a křížových odkazů. Zaměřuje se pouze na podkladovou sémantiku schémat databází a struktury dokumentu, respektive logiku vazeb mezi dokumentovanými entitami a pojmy, vedoucí k sémantické interoperabilitě. Zároveň slouží jako návod a metodologický nástroj pro efektivní strukturování a propojování souborů informací pojmového modelování v kulturní oblasti a společný formální jazyk pro identifikaci významového obsahu v různých datových formátech. Vlastní struktura CRM umožňuje vytvářet specifická rozšíření pro konkrétní aplikace.

Prioritním cílem této ontologie je v první řadě pokrytí veškerých aspektů dokumentace na poli kulturního dědictví, potřebných pro výměnu informací v globálním kontextu pro potřeby vzájemného porozumění a dialogu. V druhé řadě dovoluje integraci, archivaci, správu a výměnu tohoto typu informací bez hrozící ztráty sémantiky mezi relativně bohatšími a chudšími schémata a dále poskytuje jasně definovaný, lehce rozšířitelný rámec použitelný pro budoucí vývoj jako motivaci pro jasnou a jednoznačnou komunikaci, jež je zásadní pro všechny technologické projekty na poli kultury. Pravděpodobně nejambicióznější aplikací CRM vývoj integrovaných dotazovacích nástrojů, zprostředkování systémů a datových skladů.

5.1.4.2. pole působnosti a použití CRM

Pro přesné vymezení rozsahu působnosti CIDOC CRM je nejprve vhodné stejným způsobem vymežit i jeho oblast zájmu, tedy pojem kulturního dědictví jakožto objektů dokumentace uchovávaných znalostí v muzeích.

Podle pracovní definice ICOM zaštiťuje pojem veškeré hmotné výtvořiny lidské kultury mající formu objektů, staveb, památek i celých krajín, stejně tak jako „živou“ kulturu existující ve formě hudby, řemesla, divadelního umění, literatury, ústní tradice i jazyka [Duer, 1998].

Kulturní dědictví je dle definice nutno brát jako dynamickou rozvíjející se oblast, ve které je s ohledem na budoucnost dobré zohledňovat tyto dynamické změny, adaptaci a rozvoj kulturních materiálů i forem, stejně jako v mnoha případech zachovávat prvky kulturního dědictví v původním stavu. Výtvarné umění tedy v současné chvíli zahrnuje v drtivé většině hmotné objekty (malby, kresby, fotografie, plastiky, sochy etc.), ale do budoucna není vyloučeno, že budou přidány i nové podoby formy výtvarného umění ve formě digitálních bitů počítačem převoditelných do nových uměleckých vjemů, se kterými by se mělo v modelu CRM počítat.

Samotné pole působnosti je teoreticky rozděleno na dvě oblasti. První je zamýšlené pole působnosti (Intended Scope), které obecně definuje principy, zásady použití a informace nutné pro správu, řízení, integraci, archivaci a výměnu vědeckého (tzn., že hloubka a kvalita informací musí být dostatečně vážná pro akademický výzkum) popisu jednotlivých položek uvnitř muzejních sbírek, respektive vědeckých dat.

Druhou oblastí je praktické pole působnosti (Practical Scope), tvořené současnými platnými standardy pro muzejní dokumentaci.

Při vývoji CIDOC CRM nebyly opomenuty další významné datové struktury, jejichž formální mapování proběhlo v roce 2002 a posléze se stalo motivací pro rozšíření modelu (tzv. „Agios Pavlos Extensions“) o některé kategorie popisu informačních objektů a jeho vlastností. Veškeré elementy z níže uvedených datových struktur, spadajících pod zamýšlené pole působnosti, byly CIDOCem CRM následně pokryty. Při formálním mapování byly zahrnuty tato datová schémata: *Dublin Core*, *Art Museum Image Consortium (AMICO)*, *MDA SPECTRUM*, *National Museum of Denmark GENREG*, *Natural History Museum (London)*, *John Clayton Herbarium Data Dictionary*, *Research Libraries Group (RLG) Cultural Materials XML Schema and Document Type Definition*, *Archival Description (EAD)*, *International Federation of Library Associations and Institutions (IFLA) Functional Requirements for Bibliographic Records (FRBR)*, *Association of American Museums Nazi-era Provenance Standard*, *ABC Harmony*, *OPENGIS* a mnoho dalších [Crofts, 2003].

5.1.4.3. použitá forma, struktura a principy modelování

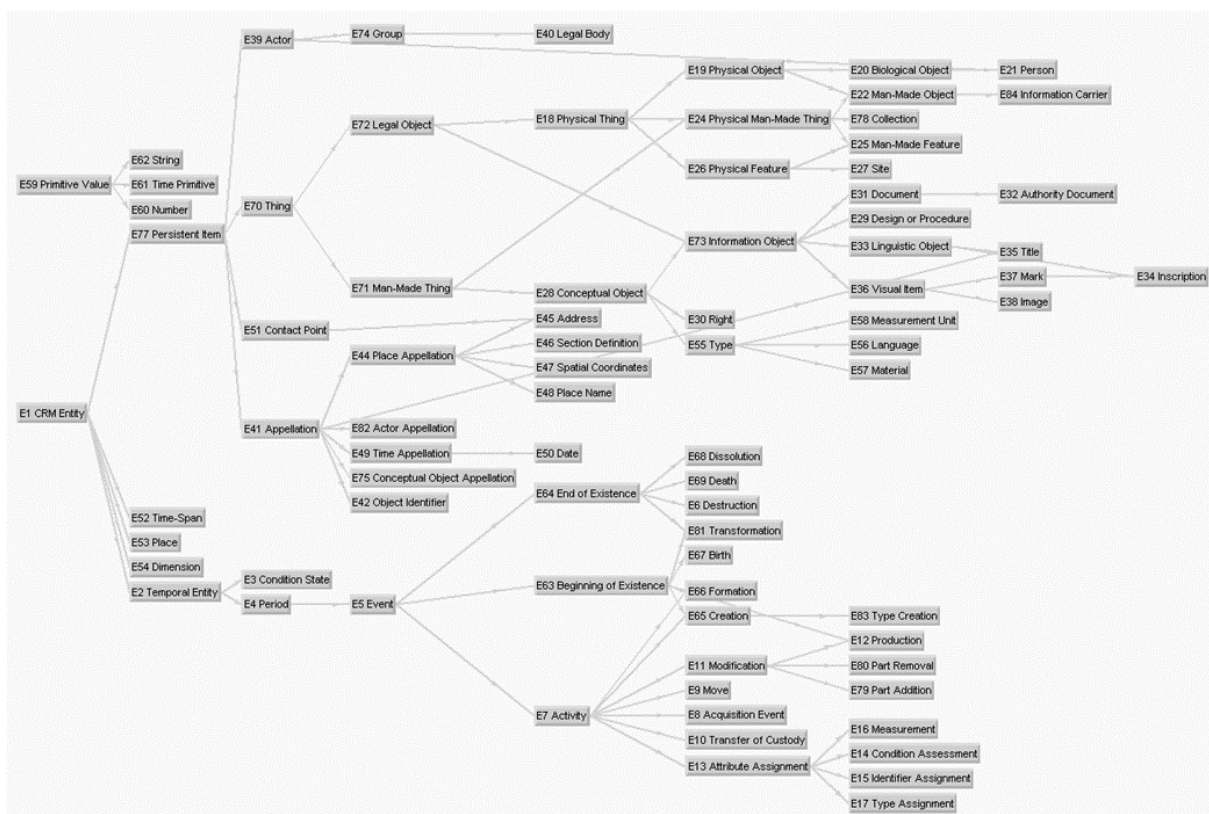
Ve vývoji CRM byl jako referenční systém použit datový model TELOS, nicméně nebyl použit jeho výrokový jazyk. Ten, stejně jako ostatní jazyky pro reprezentaci znalostí, komplexní znalosti rozkládá na elementární tvrzení jako deklarace tříd, individuí, unárních a binárních vztahů a tím vytváří ontologie. Jelikož podobné vlastnosti jako má jazyk TELOSu nalezneme i u RDF a RDFS a s přihlédnutím k okolnosti, že RDF je doporučením konsorcia W3C pro aplikace, na něž se CIDOC CRM zaměřuje, byla přijata terminologie bližší RDF (která je lépe známá členům komunity webových technologií). V plánu je produkce CRM ve více různých reprezentacích, ku příkladu v XML DTD či RDFS (v souladu s použitím TELOSu jakožto primárního zdroje informačního systému), se společným ontologickým zájmem na prvním místě [Crofts, 2003].

CRM je formulován jako typ objektově orientovaného sémantického modelu. V době jeho vývoje nebyl žádný tehdejší formalismus klasifikovaný jako standard pro tvorbu sémantických modelů, ale i přes to mezi jednotlivými sémantickými modely nalezneme jen

drobné sémantické odchylky. To samo o sobě zajišťuje relativně snadnou převoditelnost do počítačem čitelných formátů a nejen to. CRM může být prakticky zaveden do libovolného objektově nebo relačně orientovaného schématu, jako jsou sady trojic v RDFS, DAML+OIL, OWL, STEP aj. (jednotlivé instance CRM mohou být kódovány opět v RDF, XML, DAML+OIL nebo OWL). Kromě této nesporné výhody byl objektově-orientovaný model zvolen i s ohledem na některé své další vlastnosti. Poskytuje mechanismy např. pro skrytí nepotřebných detailů nebo rozšíření (z toho důvodu, že některé aspekty modelů by byly pro realizaci modelů nadbytečné a některé by bylo nutné rozšířit pro potřeby konkrétních institucí), dále také sofistikované nástroje a jazyky pro vývoj a řízení objektově-orientovaných schémat a obsahuje známé prvky z tzv. entity-relationship modelování³. Velmi podstatný je fakt, že CRM lze implementovat pomocí široké škály technických platforem, včetně relačních databází.

Strukturu CIDOCu nejlépe vystihuje kompletní seznam tříd a výčet použitých vlastností (celkem 211 prvků – 81 tříd a přes 130 vlastností), které s ohledem na omezený rozsah práce zde není možné uvést, ale pro všechny zájemce je k dispozici na URL a v příloze č. 4, kde naleznete i detailnější definice základních pojmů a vazeb mezi nimi, které stručněji popisuje obrázek 17.

³ **Entity-relationship model** (ERM) se v softwarovém inženýrství používá pro abstraktní a konceptuální znázornění dat. ERM je metoda datového modelování, která vytváří jeden z typů konceptuálních schémat či sémantických datových modelů systému (obvykle relační databáze) a požadavků na něj stylem shora dolů [Entity Relationship Model (Wikipedia), 2010].



Obrázek 17 Struktura CIDOC CRM, [Nordbotten, 2008]

Základní struktura je zde alespoň nastíněna výčtem meta-vlastností CRM, jimiž jsou: identifikace objektů reálného světa pomocí reálných názvů; klasifikace objektů reálného světa; částečná dekompozice a strukturální vlastnosti pojmových a fyzických objektů; časových období, osob a míst; účast trvalých objektů v přechodných entitách (vytvářející pojetí dějin jakožto střet objektů s časoprostorovou linií), lokalizace období v časoprostoru a fyzických objektů v prostoru; vliv objektů na činnosti a produkty, odkazy na informační objekty jakýchkoliv reálných entit.

5.1.4.4. CIDOC CRM a FRBR

Spolu s komunitou muzeí a galerií započala práci na vývoji vlastního pojmového modelu pro oblast potřebných informací i knihovnická komunita pod záštitou IFLA (*International Federation of Library Associations*). Její snahou je sjednocení konceptualizace muzejních a bibliografických informací a dokonce i odvážná vize budoucnosti – propojení modelů CIDOC CRM a modelu FRBR (*Functional Requirements for Bibliographic Records*).

Model FRBR je vyvíjen od roku 1991 a jeho definitivní verze se stále očekává. Jeho pole působnosti je vymezeno informacemi předávanými v bibliografickém záznamu (včetně

záhlaví), z nichž jsou vyloučeny informace z autoritních záznamů i sémantika předmětových hesel a klasifikace indexů.

Bylo nutné vymezit rozdíl mezi muzejní a knihovnickou informací a shodnout se na postupu při vývoji jednotného modelu pro popis unikátních i neunikátních položek. Rozhodlo se, že CIDOC CRM bude použit jako referenční bod, a že jeho formalismus bude cílovou podobou, do níž bude FRBR přetransformováno. Pracovní skupina podrobně zanalyzovala veškeré atributy a vztahy vykázané v definici FRBR a pokusila se z nich co nejpřesněji „vydestilovat“ sémantiku. Dále bylo nevyhnutelné shodnout se na jednom významu a všech implikacích definice každého jednoho atributu a vztahu, který byl posléze vyjádřen jako formalismus použitý v CIDOCu CRM. Tzn. že struktura CIDOC CRM byla znovupoužita vždy, pokud to bylo jen trochu možné. Pokud ne, bylo nevyhnutelné vytvořit novou podtřidu nebo podvlastnost jako rozšíření CIDOCu. Každý z těchto „atributů“ deklarovaných v FRBR se tak přemění ve výsledném objektově orientovaném modelu na jednu, nebo více vlastností.

Původní model FRBR se tak sice poněkud zkomplikoval, ale na druhou stranu bylo dosaženo větší logické přesnosti. Model CIDOC CRM musel být v některých případech také přizpůsoben potřebám rozšíření konceptualizace modelu FRBR. Jeho současná objektově orientovaná verze obsahuje 39 tříd a 55 vlastností a znovu-použitých 44 tříd a 45 vlastností z modelu CIDOC. Záměrem je, aby původní model FRBR i jeho nová varianta mohly společně koexistovat – oba formalismy jsou stejně užitečné, byť každý pokrývá jiné potřeby. Objektově orientovaný model FRBR je vhodný i pro aplikace RDF, jelikož je také vyjádřen po vzoru CIDOC CRM jako řetězec sémantických trojic [Bæuf, 2007].

5.1.4.5. použití CIDOCu CRM v praxi

CIDOC CRM byl již aplikován do téměř dvou desítek projektů, z nichž v této práci budou stručně představeny ty nejvýznamnější.

Projekt **CLAROS** (*Classical Art Research Online Service*) zpřístupňuje informace o umění starých Řeků a Římanů (převážně 3D objektů), jehož reprezentanty najdeme v muzeích po celém světě. Používá CIDOC CRM a RDF jako základů pro kombinování více než 2 milionů záznamů uložených v 5 databázích. Na adrese <http://www.clarosnet.org> lze v těchto 5 sloučených databázích vyzkoušet vyhledávání, přístup k datům, fasetové vyhledávání CLAROS Explorer (<http://www.clarosnet.org/claros-screencast.mov>) a vyhledávání obrazových informací.

Jako rámec pro definování vztahů mezi objekty, osobami, místy, kategoriemi a tématy použilo model CIDOC i *Novozélandské muzeum Te Papa Tongarewa*, jehož

uživatelské rozhraní je znázorněno na obrázku č. 18 (<http://collections.tepapa.govt.nz/>). Výsledek implementace je možné posoudit prohledáváním sekce „Collections Online“ – viz obrázek 18. Dále CIDOC využívá i **Germánské národní muzeum v Norimberku** (Germanische Nationalmuseum Nuremberg – URL: <http://www.gnm.de/>) nebo **Finská národní galerie** (*Finnish National Gallery* – URL: <http://www.fng.fi/fng/rootnew/en/vtm/etusivu.htm>). I portál **MusInfo** (URL: <http://www.ville-ge.ch/musinfo/>), který spojuje informace ze 4 významných vědeckých institucí v Ženevě (včetně Muzea umění a historie) stojí na této technologii.

Projekt neziskového konsorcia RLG (Research Libraries Group – zahrnuje 160 institucí z 15 různých zemí) nazvaný **RLG's CULTURAL MATERIALS** zprostředkovává informace ze 54 kulturních institucí pomocí CIDOCu CRM [WorldCat, 2007].

Pro evropský projekt **SCULPTEURWEB** (URL: <http://www.sculpteurweb.org/html/approach.htm>) zajišťuje CIDOC CRM sofistikované sémantické vrstvy pro multimediální informační management a znalostní strukturu spojující nízko-úrovňové a vysoko-úrovňové multimediální reprezentace.

Velmi zajímavé uživatelské rozhraní s rozšířenou realitou (tzv. augmented reality interface⁴ nabízí díky CIDOCu CRM i projekt **ec(h)o** (URL: <http://echo.iat.sfu.ca/>), s využitím prostorových zvukových palet a sémantickému přístupu ke znalostem – CIDOCem standardně popisuje muzejní artefakty.

⁴ Rozšířená realita (*anglicky augmented reality*) představuje mezistupeň mezi realitou skutečnou (není-li toto spojení samo o sobě pleonasmem) a realitou virtuální. Rozšířená realita je doplněním obrazu skutečnosti o uměle doplněné obrazce či jiné informace. V současnosti je nejčastějším provedením rozšířené reality zobrazení skutečného obrazu na displeji a jeho doplnění o počítačem dodané informace. To vše samozřejmě v reálném čase [Zandl, 2009].

The screenshot shows the Museum of New Zealand Te Papa Tongarewa website. At the top, there is a navigation bar with links for 'About Te Papa', 'Research', 'Sponsors', and 'Friends of Te Papa'. Below this is a search bar and a menu with categories like 'Home', 'What's on', 'Plan your visit', 'Education', 'Collections Online', and 'Ship'. The main content area displays search results for 'ship', with a focus on the topic 'Captain Cook's arrival at Ship Cove, Queen Charlotte Sound'. A detailed record for the painting 'Ship Cove, Queen Charlotte Sound' by John Webber is shown, including a description of the scene, related people and places, and a map of the location. The record also includes references and related topics.

Obrázek 18 Uživatelské rozhraní muzea Te Papa Tongarewa užívající CIDOC CRM; [Museum Te Papa, 2004]

Detailní výčet aplikací CIDOCu CRM naleznete na [ICOM, 2006] v sekci THE CIDOC CRM / Applications.

5.2. ontologie pro vizuální popis uměleckých děl

Umělecká díla a veškeré subjekty, které zobrazují, je možné ontologicky popisovat rovněž na základě obsahu, jak bylo možné vidět v případě Iconclassu. Pro efektivní popis je důležité obsáhnout absolutní i relativní pozici objektu v rámci uměleckého díla. Relativní pozice operuje s pojmy jako popředí a pozadí obrazu, pro absolutní pozici se používají standardizované výrazy převzaté z již existujících ontologií. Každý kompletní vizuální popis díla se skládá z popisu prostorového a z popisu hlavních vizuálních rysů díla, jako jsou dominantní barva nebo textura [Breitman, 2006, s. 249].

5.2.1. prostorový popis

Pro určení přesné pozice objektu na obraze se využívá specifikace převzatá z tezauru WordNet, jejímiž nejvýznamnějšími body jsou sever, jih, východ a západ: North (N), South (S), East (E), West (W) a jejich kombinace: Northeast (NE), Southeast (SE), Northwest (NW) a Southwest (SW). Plocha obrazu je na základě těchto indikátorů rozdělena na 9 polí využívajících značení - viz obrázek 15. WordNet nicméně nabízí celkově 32 takovýchto prostorových indikátorů, jež mohou být využity pro přesnější a jemnější popis.

NW	N	NE
W	center	E
SW	S	SE

Obrázek 19 Specifikace pro prostorový popis 2D uměleckých děl; [Breitman, 2006, s. 249]

Prostorový ontologický popis mimo to obsahuje sadu jednoduchých relativních prostorových vztahů vypůjčených z ontologie SUMO. Jde o atributy Right, Left, Above, Below, Near, Far (jako jediný atribut je převzatý z WordNet) a Contains. Pro každý vztah ontologický popis indikuje, zda jde o vztah symetrický nebo tranzitivní a jaký vztah je jeho

inverzí.

Každá ontologie pro vizuální popis tedy obsahuje třídu viditelných objektů (VisibleObject) a poziční a relační vlastnosti těchto objektů - spatialRelation, hasPosition [Breitman, 2006, s. 249].

5.2.2. ontologický popis vizuálních rysů

Pro popis vizuálních rysů díla je efektivní částečně využít deskriptory MPEG-7⁵. K nejčastěji popisovaným rysům patří barva, textura a tvar. Na základě MPEG-7 lze popisovat i pohyb, který se ale vzhledem ke klasickému charakteru uměleckých děl nevyužívá. V tabulce níže jsou uvedeny základní vizuální rysy a jejich deskriptory [Breitman, 2006, s. 251].

<i>rys</i>	<i>deskriptory</i>
barva	DominantColor, ScalableColor, ColorLayout, ColorStructure, GoFGoPColor (rozpětí barevné struktury)
textura	HomogeneousTexture, TextureBrowsing, EdgeHistogram
tvar	RegionShape, ContourShape, Shape3D

Díky ontologii pro vizuální popis je možné vyhledávat obrazy podle barvy. To si lze vyzkoušet například na webovém sídle slavné petrohradské galerie Ermitáž [Hermitage, 2006] v sekci Digital Collection / QBIC Colour and Layout Searches. Vyhledávání podle barev, tvarů nebo vrstev objektů stojí na principech, které nejsou daleko od ontologického vizuálního popisu uměleckých děl. [Breitman, 2006, s. 249/252].

5.3. metadatová schémata pro popis uměleckých děl

V poslední podkapitole se ve stručnosti zaměříme na metadatová schémata, na nichž

⁵ MPEG7, formálně pojmenovaný "Multimedia Content Description Interface", je standard (ISO/IEC JTC1/SC29/WG11) pro popis multimediálního obsahu dat, který podporuje několik úrovní interpretace významu informace, které mohou být přeneseny nebo zpřístupněny přes počítačový kód. Není zaměřen na jednu aplikaci, ale elementy MPEG7 standardizují podporu celé řady aplikací [Martinez, 2004].

některé dříve popsané ontologie stojí.

5.3.1. Dublin Core

Jedná se patrně o nejznámější metadatové schéma a patří také k těm nejstarším a nejvšestrannějším. Vzniklo roku 1995 v americkém Dublinu v Ohiu a jako ISO standard 15836 platí od roku 2003. V současné době obsahuje patnáct prvků uvedených v příloze č. 5. Popis je možné aplikovat na libovolný zdroj na libovolném médiu, je tedy využitelný i v oblasti výtvarného umění. Existuje v 25 jazykových verzích a do budoucna usiluje o zlepšení harmonizace popisu napříč různými zeměmi, institucemi i tematickými oblastmi. Taktéž je zřetelná snaha o sémantickou integraci s jinými metadatovými modely [Synková, 2008].

5.3.2. CDWA The Categories for the Description of Works of Art

AAT i ULAN využívají vlastní metadatové schéma CDWA, vyvinuté přímo v institutu Getty Research. Jde o kategorie a subkategorie pro popis uměleckých děl, kterých je v současné době 381. Tzv. jádrové kategorie představují soubor minima informací, které je potřebné pro unikátní popis a identifikaci uměleckého či architektonického díla.

CDWA zohledňuje rozdíl mezi informacemi, jež se zobrazují uživateli a daty, nutnými pro automatické vyhledání informace – viz obrázek 16. Informace na zobrazení by měly zohledňovat veškeré jazykové nuance pro odstranění nejistoty a dvojznačnosti. Na druhou stranu některé klíčové metadatové prvky by měly být formátované pro zjednodušení vyhledávání [Breitman, 2006, s. 253/254].

Classification*	Paintings	<i>Authority/ Controlled Vocabulary</i>
Object/Work Type*	paintings	<i>Authority/ Controlled Vocabulary</i>
Title or Names*	Irises	
Creation-Creator/Role*	Vincent van Gogh painter: Gogh, Vincent van (Dutch painter, 1853-1890)	
Creation-Date*	1889	
	earliest: 1889	latest: 1889
Subject Matter*	irises	regeneration
	Iridaceae	soil
	nature	<i>Authority/ Controlled Vocabulary</i>
Context-Historical/Cultural	exhibited at Salon des Indépendents, September 1889	
Measurements	71 x 93 cm (28 x 36 5/8 in.)	
	height: 71 cm	width: 93 cm
Materials and Techniques	oil on canvas, applied with brush and palette knife	
	oil paint	brush
	canvas	impasto
	palette knife	<i>Authority/ Controlled Vocabulary</i>
Descriptive Note	This work was painted when the artist was recuperating from a severe attack of mental illness, and it depicts the garden at the asylum at Saint-Rémy. It is influenced by the work of Gauguin and Hokusai, and is remarkable for the contrasts of color ...	
Current Location-Repository Name*	J. Paul Getty Museum	
Current Location-Repository Location	Los Angeles (California, USA)	
Current Location-Repository Numbers	90.PA.20	
Image Credits: <i>Irises</i> (painting), 1889; artist: Vincent van Gogh (1853-1890, active in Holland); oil on canvas, 71 x 93 cm; J. Paul Getty Museum (Los Angeles, California), 90.PA.20. © The J. Paul Getty Trust, 2000. All Rights Reserved.		

Obrázek 20 Tabulka s kategoriemi CDWA; [Breitman, 2006, s. 254]

5.3.3. ISO 21127

V podstatě se nejedná o nic jiného než o normativní označení metadatového schématu CIDOC CRM, jež byl podrobně popisován výše. Jako ISO standard bylo schéma přijato v roce 2006.

Zajímavým tématem je sémantická integrace popisu sbírek uměleckých děl, právě mezi CIDOC CRM a Dublin Core, nebo rozšiřování CIDOC o elementy MPEG-7. Více o něm bude pojednáno v závěru práce [Breitman, 2006, s. 255].

5.3.4. VRA – The Visual Resources Association Core Categories

Kategorie multioborové organizace Visual Resources Association jsou novou interpretací metadatového schématu Dublin Core – viz obrázek 17. Schéma může být bez problémů v případě potřeby obohaceno o dodatečná pole a každý prvek schématu se může opakovat tolikrát, kolikrát je potřeba. Komise VRA zaobírající se datovými standardy

doporučuje při tvorbě metadat využívat tezaurů, částečně vzniklých v institutu Getty a další dostupné autority práce [Breitman, 2006, s. 256/259].

6 . I N T E R O P E R A B I L I T A O N T O L O G I Í

Jak je možné odvodit z výčtu výše uvedených ontologií, je na internetu v současné době dostupných více sémantických databází, které jsou ovšem vytvářeny mnoha různými způsoby podle ISO norem a standardů a obsahují různé sémantické vztahy mezi různými objekty. Všechny ale spojuje oblast jejich bezprostředního zájmu: výtvarné umění a také jejich účel: usnadnit uživateli (ať už se jedná o jednotlivce nebo celé odborné instituce) vyhledávání a zpracování informací, co nejvíce to jde.

Uživatel nechce složitě pátrat po žádaných informacích v mnoha zdrojích, což je časově velmi náročné, a ani se nepotřebuje rozumět tomu, na jakých principech jsou hledané termíny mezi sebou spojeny. Naopak je pro něj výhodné moci si vybrat z výsledků prohledávání několika heterogenních zdrojů předložených v jediném výstupu. Z těchto důvodů je více než žádoucí učinit odlišné systémy sémantické organizace znalostí - ontologie - vzájemně interoperabilními a prohledavatelnými přes jediný vyhledávací aparát [Zeng, 2004, s. 377-395].

Pro účely sdílení informací v prostředí sítě se rýsují dvě cesty, kterými by se dalo teoreticky dosáhnout interoperability. Zaprvé tvorba nových znalostních systémů a ontologií a za druhé mapování a vzájemná integrace již existujících systémů.

Tvorba nových ontologií je náročná po všech stránkách a proto je žádoucí pokusit se vytěžit co nejvíce z již existujících a zavedených systémů. Pokusit se je integrovat, je ale také velmi složité - je to z toho důvodu, že jsou založeny na odlišných metadatových schématech.

6.1. teoretický návrh integrace existujících ontologií do jednoho systému

Vezměme v úvahu všechny ontologie popsané v této práci a jejich jednotlivé možnosti. CIDOC CRM poskytuje velmi sofistikované zázemí pro práci s informacemi o výtvarném umění jakožto metadatové schéma – model, které je pro tvorbu ontologií základní a výchozí. Podle metodologie pro tvorbu ontologií (viz kapitola 3) by dobře definoval cíl i rozsah ontologie, která by potenciálně mohla vzniknout propojením informací obsažených v AAT, ULAN a Iconclass.

Na základě AAT společně s ULANem by mohla být kvalitně specifikována terminologie. AAT obsahuje pojmy obecné, ULAN konkrétní osobní jména a Iconclass má

velké možnosti popisu konkrétních uměleckých děl. Každá z ontologií obsahuje jiný typ informací ze stejné domény, což přímo vybízí k pokusu o alespoň teoretickou imaginaci integrace těchto stávajících systémů do jediné velké ontologie - znalostního systému.

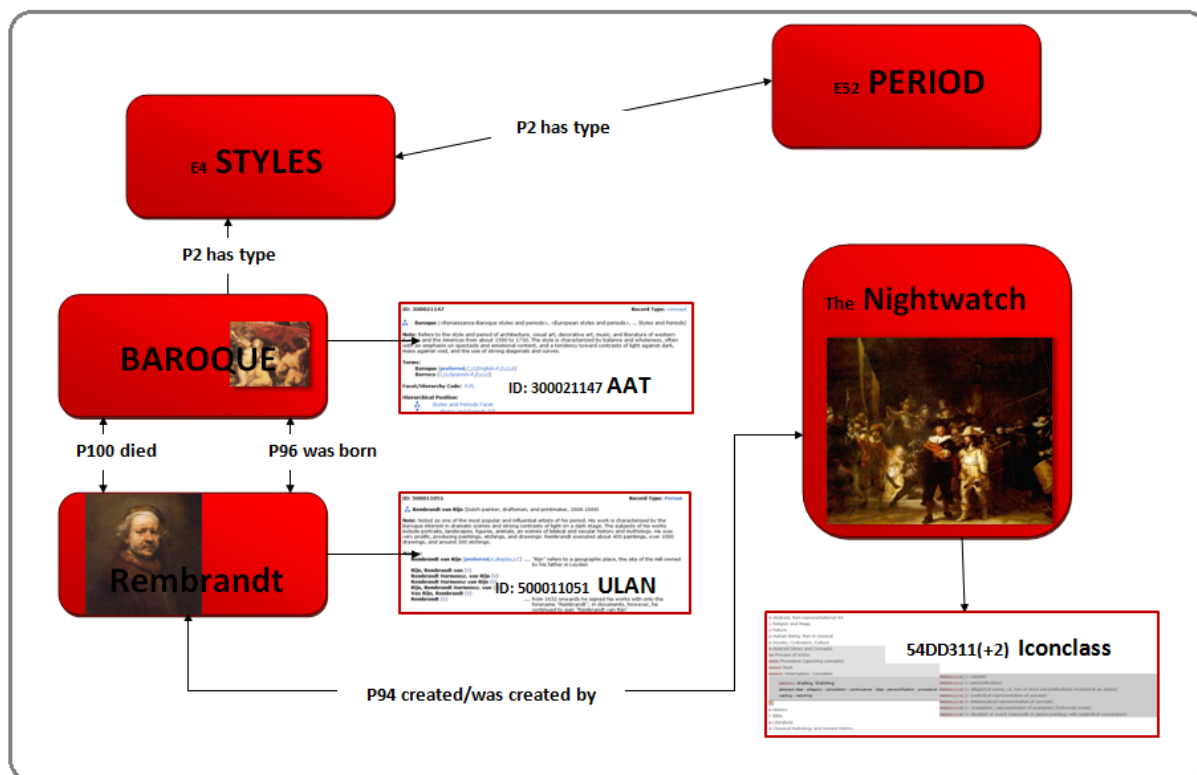
Představme si vybraný pojem z AAT "Baroque" (ID 300021447) - baroko - jako výchozí třídu pro tvorbu integrované ontologie metodou "od středu". V hierarchii by pojem spadal pravděpodobně do natřídý "Styles and Periods" - Styly a období - a vztah by se dal jednoduše popsat vztahem "je typem" (v CIDOCu P2 "has type (is type of)". Fazeta "Styly a období" odpovídá v CIDOCu poli E4 "Period". V tomto okamžik by bylo možné poskytnout uživateli obecnou informaci o existenci dělení výtvarného umění do různých období a poskytnout mu o konkrétních z nich informace, které o nich podává AAT.

Obecný pojem "Baroko" by se mohl hypoteticky sémanticky provázat s již konkrétními osobními jmény umělců tvořících v tomto období a to hned přes několik vazeb (vlastností), jež CIDOC nabízí: přes data narození a úmrtí osob - "veden/a do života (se narodil/a)" (P98 "brought into life (was born)") a "byl/a úmrtím (zemřel/a)" (P100 "was death of (died in)" - nebo přes další možné vlastnosti - např. "odkazuje na (je odkazován z)" (P67 "refers to (is referred to by)" nebo "se časově shoduje s" (P114 "is equal in time to"). Je komplikací, že CIDOC prozatím neobsahuje vlastnost "tvořil v" s definičním oborem E4 "Období" a oborem hodnot E21 "Osoba".

Dále vyberme konkrétního barokního umělce a uvedme na scénu možnosti ULANu. Pro příklad zvolíme Rembrandta. V našem ukázkovém příkladě bychom tedy vytvořili dvě trojice typu *subjekt(Rembrandt van Rijn)-predikát(se narodil)-objekt(Baroko)* a zároveň *subjekt(Rembrandt van Rijn)-predikát(zemřel)-objekt(Baroko)*, z čehož jasně plyne, že byl barokním umělcem. Pokud by se u některého umělce lišila období narození a smrti, logicky by žil na přelomu obou a uživatel by mohl potřebovat informace o obou obdobích, které by získal z báze AAT. Informace o Rembrandtovi, provázané například se jmény jiných osob, by pocházeli z báze ULAN.

Zkusme pouze hypoteticky propojit pojem "Rembrandt van Rijn" CIDOCovskou vlastností P94 "vytvořil/a (byl vytvořen)" ("has created (was created by)"). Touto cestou můžeme podat uživateli výčet všech známých Rembrandtových děl – např. *subjekt(Rembrandt van Rijn)-predikát(vytvořil)-objekt(Noční hlídka)*.

Zbývá pouze popsat známé dílo Noční hlídka, ať už z hlediska fyzického (rozměry, technika, materiál) tak v ideálním případě i na základě obsahu, jak to umí systém Iconclass – výsledný systém by mohl být schematicky reprezentován způsobem na obrázku 19.



Obrázek 21 Model integrované ontologie pro doménu výtvarného umění

Ve shrnutí jsme využili maximálně veškeré možnosti ontologií uvedených v této práci a to pod křídly pojmového modelu CIDOC. Jednalo se samozřejmě pouze o hypotetický popis možné integrace. Realita je bohužel poněkud jiná.

6.2. možný budoucí vývoj ontologií

Již na první pohled je jasné, že systémy určitě nebudou vzájemně kompatibilní, a že by případná integrace by určitě nebyla jednoduchá. Přestože je sémantika u všech 4 schémat v mnoha ohledech podobná, doposud neexistuje převodní mechanismus mezi vícerymi datovými modely. I drobné rozdíly ve správě modelů jednotlivých institucí činí téměř nemožné sloučit již existující systémy do jediné sofistikované ontologie. Několik možností do budoucnosti se ale přece jen na obzoru rýsuje. Mohl by jím být tzv. mediační systém. Podle Honga et al. [Lourdi, 2009] "*fyzické kombinování dat do jednoho systému může být nemožné z technických, organizačních nebo ekonomických důvodů. Proto se používají mediační systémy namísto spojování informačních zdrojů do jednoho, což umožňuje zodpovídat distribuované*

dotazy bez fyzického slučování informací do jediné monolitické databáze. Typický mediační systém se chová jako jednotné uživatelské rozhraní. Akceptuje, interpretuje dotazy a distribuuje je do zúčastněných systémů. Tyto systémy odpovídají mediátoru, který sjednotí výsledky pro koncového uživatele."

Podle Zeng a Chan [Zeng, 2004, s. 377-395] existuje 7 způsobů zajištění kompatibility: derivace/modelování, překlad/adaptace, uzlové linkování, přímé mapování, mapování společného výskytu, přepínání, přepínání přes dočasný seznam jednotek a linkování tezauru přes serverový protokol. Jelikož podrobnější popis přesahuje rámec práce, základní možnosti a charakteristiky metod jsou uvedeny v příloze č. 6.

Ontologie jsou v současnosti tak trochu otázníkem a zároveň velkou výzvou pro budování nové sémantické organizace znalostí v rámci webu generace 3.0. Jejich architektura i tvorba je velice náročným procesem, který vyžaduje lidské zdroje i odpovídající technologii a techniku. Je trochu otázkou, nakolik se v budoucnosti vývoj a vytváření ontologií vyplatí a hlavně v jakém směru. Cílem práce bylo poukázat možnosti využití a šíři záběru informací, které by mohly integrované ontologie poskytnout například v oblasti kunsthistorie, ikonografie i muzejnictví. Význam by mohl hypoteticky spočívat v možném propojování ontologií i napříč různými znalostními obory. Pro koncového uživatele by efekt měl spočívat především v usnadnění vyhledávání a možnost dotazů formulovaných jednodušším způsobem, třeba i v podobě otázek kladených v přirozeném jazyce a nejen ve formě klíčových slov a pojmů (záleží na přesnosti definic a popisu vazeb mezi slovy a termíny).

V práci byla představena základní koncepce a účel webových ontologií, vysvětleny nejdůležitější termíny, otázky sémantiky a metadat, stejně jako principy tvorby ontologií a nástrojů, které jsou předmětem ontologického a znalostního inženýrství. Zaměřili jsme se na doménu výtvarného umění a popsali princip fungování, možnosti nejvýznamnějších webových ontologií v této oblasti lidského vědění. Směr budoucího vývoje ontologií ukazuje potřebu kompatibility metadat, integrace schémat a slučování menších systémů do větších ontologických celků, které by díky své propracované sémantice mohli uživatelům, zajímajícím se o doménu výtvarného umění, at' už z profesních či osobních důvodů, co nejvíce zjednodušit vyhledávání a práci s informacemi na webu.

Je nadějně, že se nám v blízké budoucnosti, díky této technologii, při zadání i složitějšího dotazu na oblíbeného umělce či obraz do Vašeho vyhledávače (at' to bude Google nebo jakýkoliv jiný browser) dostane přesně takové odpovědi, jakou si představujeme. Možná, že i vyhledávače budou díky ontologiím totiž skutečně chápat význam toho, co hledáme.

Seznam použité literatury

- BÆUF, Patrick Le;DOERR, Martin. Harmonising CIDOC CRM and FRBR. *International Cataloging & Bibliographic Control*. 2007, vol. 36, issue 4, s.90-92 Dostupný také komerčně z databáze EBSCO HOST/LISTA: <<http://web.ebscohost.com>>. ISSN 1011-8829.
- BOLDIŠ, Petr. Pořádání informací a znalostí na internetu [online]. Verze 1.0, Praha : Ústav informačních studií a knihovnictví FF UK v Praze, 2008 [2010-06-14]. 153 s. (PDF). Dostupný z WWW: <http://texty.jinonice.cuni.cz/studijni-texty/boldis-petr/boldis_01.pdf/view>.
- BRATKOVÁ, Eva. (zprac.). Metody citování literatury a strukturování bibliografických záznamů podle mezinárodních norem ISO 690 a ISO 690-2 : metodický materiál pro autory vysokoškolských kvalifikačních prací [online]. Verze 2.0, aktualiz. a rozšř. Praha : Odborná komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací, Asociace knihoven vysokých škol ČR, 2008-12-22 [2010-05-03]. 60 s. (PDF). Dostupný z WWW: <<http://www.evskp.cz/SD/4c.pdf>>.
- BREITMAN, Karin Koogan; CASANOVA, Marco Antonio; TRUSZKOWSKI, Walter. Semantic Web: Concepts, Technologies and Applications. 1st ed. New York : Springer, c2006. 314 s. NASA Monographs in Systems and Software Engineering. [cit. 2010-07-26] Dostupný také komerčně ze SpringerLink (DOI): <http://dx.doi.org/10.1007/978-1-84628-710-7_9>. ISBN 978-1-84628-710-7.
- BRŮCHA, Filip. Google přichází se sémantickým vyhledáváním. *Computerworld* [online]. 25.03.2009 , roč. XX, č. 6, [cit. 2010-07-26]. Dostupný z WWW: <<http://computerworld.cz/internet-a-komunikace/Vyhledavac-Google-prichazi-se-semantickym-vyhledavanim-3733>>. ISSN 1210-9924.
- CROFTS, Nick; DOERR, Martin, GILL, Tony The CIDOC conceptual reference model: a standard for communicating cultural contents. *Cultivate Interactive* [online]. 2003, issue 9 [cit. 2010-07-28]. Dostupné z WWW: <<http://www.cultivate-nt.org/issue9/chios/index.html>>. ISSN 1471-3225. DOI 10.1016/j.websem.2008.09.002.
- DUER, Tia; STERN, Stephen. *ICOM* [online]. 1998 [cit. 2010-07-28]. Cultural Heritage and Development Action Network: Working Group Meeting. Dostupné z WWW: <<http://icom.museum/bank.html>>.
- Entity-relationship model. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg

- (Florida) : Wikipedia Foundation, 30. 12. 2008, last modified on 1. 6. 2010 [cit. 2010-07-30]. Dostupné z WWW: <http://cs.wikipedia.org/wiki/Entity-relationship_model>.
- FENSEL, Dieter. *Ontologies : a silver bullet for knowledge management and electronic commerce*. 2nd rev. and enlarg. ed. Berlin : Springer, c2004. s. 3. ISBN 354-0-00302-9.
 - GÓMEZ-PÉREZ, Asunción; FERNÁNDEZ-LÓPEZ, Mariano; CORCHO, Oscar. *Ontological engineering : with examples from the areas of knowledge management, e-commerce and the Semantic Web*. 2nd ed. London : Springer, c2004. 404 s. *Advanced information and knowledge processing*. ISBN 1-85233-551-3.
 - HEIJST, G. van; Schreiber, A. T.; Wielinga, B. J. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies* [online]. 1997, vol. 46, issue 2-3, s.183-292, [cit. 2010-07-26]. Elektronická verze ve formátu dostupná také komerčně z databáze ScienceDirect (DOI): <<http://dx.doi.org/10.1006/ijhc.1996.0090>>. ISSN 1071-5819.
 - HUSÁKOVÁ, Martina. *LIDE.UHK.CZ* [online]. 2008, 2009 [cit. 2010-07-27]. Znalostní technologie I. Dostupné z WWW: <http://lide.uhk.cz/fim/ucitel/fshusam2/lekarnicky/z1/z1_kap02.html>.
 - ICOM. *The CIDOC Conceptual Reference Model* [online]. 2006 [cit. 2010-07-28]. Dostupné z WWW: <<http://www.cidoc-crm.org/index.html>>.
 - J. Paul Getty Trust . *Getty Union List of Artist Names : Art & Architecture Thesaurus Online* [online]. Los Angeles (CA) : The Getty Research Institute, 2010 [cit. 2010-07-28]. Dostupné z WWW: <http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html>.
 - J. Paul Getty Trust . *Getty Union List of Artist Names : Union List of Artist Names Online* [online]. Los Angeles (CA) : The Getty Research Institute, 2010 [cit. 2010-07-28]. Dostupné z WWW: <http://www.getty.edu/research/conducting_research/vocabularies/ulan/about.html>.
 - KASHYAP, Vipul; BUSSLER, Christoph; MORAN, Matthew. *Semantic Web Standards*. IN *The semantic web : semantics for data and services on the web*. 1st ed. New York : Springer, c2008. 404 s. *Data-centric systems and applications*. ISBN 978-3-540-76451-9.
 - KRAUS, Jiří, et al. *Nový akademický slovník cizích slov : A-Ž*. Vzd. 1. Praha : Academia, 2007. 880 s. ISBN 978-80-200-1351-4.
 - LOURDI, Irene; PAPTAEODOROU, Christos; DOERR, Martin. *Semantic Integration of Collection Description*. *D-Lib Magazine*. 2009, vol. 15, issue 7/8. Dostupný také komerčně z databáze EBSCO HOST/LISTA: <<http://web.ebscohost.com>>. ISSN 1082-9873.

- MARTÍNEZ, José M. . *MPEG* [online]. Verze 10. Palma de Mallorca : International Organisation for Standardisation Organisation , 2004 [cit. 2010-07-30]. MPEG-7 Overview. Dostupné z WWW: <<http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>>.
- McCLOUD, Scott. *Jak rozumět komiksu*. Praha : BB Art, 2008. 216 s. ISBN 978-80-7381-419-9.
- NEGROPONTE Nicholas. *Digitální svět: Being Digital*. Praha : Management Press, c2001. 194 s. ISBN 80-7261-046-5.
- PEŠEK, Jiří. Internetový provoz se do roku 2013 zvýší téměř čtyřikrát. *Lidové noviny* [online]. 2009, roč. 22, č. 140, [cit. 2010-07-26]. Dostupný z WWW: <http://www.lidovky.cz/internetovy-provoz-se-do-roku-2013-zvysi-temer-ctyrikrat-pjs-/ln-svet-techniky.asp?c=A090714_234918_ln-svet-techniky_pse>. ISSN 1213-1385.
- Rijksbureau voor Kunsthistorische Documentatie. *Iconclass* [online]. [Haag] : RKD, c2010 [cit. 2010-07-28]. Dostupné z WWW: <<http://www.iconclass.nl/>>.
- STÖRIG, Hans Joachim. *Malé dějiny filosofie*. 7. přeprac. a rozšířené vyd. Kostelní Vydří : Karmelitánské nakladatelství, 2000. 630 s. ISBN 80-7192-500-2.
- SVÁTEK, Vojtěch. Ontologie a WWW. In *DATAKON 2002 : sborník databázové konference*. 1. vyd. Brno : Masarykova univerzita, 2002. s. 1-35. ISBN 80-210-2958-7.
- SVÁTEK, Vojtěch; LABSKÝ, Martin. Objektové modely a znalostní ontologie – podobnosti a rozdíly [online]. Praha: Katedra informačního a znalostního inženýrství, Vysoká škola ekonomická v Praze. 12 s. [cit. 2010-07-26]. Dostupný z WWW: <<http://nb.vse.cz/~svatek/obj03fi.pdf>>.
- SVÁTEK, Vojtěch; VACURA, Miroslav. Ontologické inženýrství. In POPELÍNSKÝ, Lubomír. *DATAKON : sborník databázové konference*. 1. vyd. Brno : Masarykova univerzita, 2007. s. 1-32. ISBN 978-80-7355-0.
- SYNKOVÁ, Veronika. Vybrané aplikace metadatového formátu Dublin Core. *IKAROS : elektronický časopis o informační společnosti* [online]. 2008, roč. 12, č. 5, [cit. 2010-07-29]. Dostupný z WWW: <<http://www.ikaros.cz/node/4715>>. ISSN 1212-5075.
- *The State Hermitage Museum* [online]. [St. Petersburg] : c2006 [cit. 2010-07-29]. Dostupné z WWW: <http://www.hermitagemuseum.org/html_En/index.html>.
- *W3C* [online]. c2009 [cit. 2010-07-30]. About W3C. Dostupné z WWW: <<http://www.w3.org/Consortium/>>.
- *WordNet : A lexical database for English* [online]. Princeton University, c2010 [cit. 2010-07-30]. About WordNet. Dostupné z WWW: <<http://wordnet.princeton.edu/>>.

- *WorldCat* [online]. [Mountain View (CA)] : RLG Cultural Materials Alliance, 2006 [cit. 2010-07-29]. Cultural Materials Initiative. Dostupný z WWW: <<http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file2813.html>>
- ZANDL, Patrick. Příští fenomén: rozšířená realita je budoucnost webu. *Lupa* [online]. 2009 [cit. 2010-06-10]. Dostupný z WWW: <<http://www.lupa.cz/clanky/rozsirena-realita-augmented-reality/>>.
- ZENG, Marcia Lei; CHAN, Lois Mai. Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. *Journal of the American Society for Information Science and Technology*. 2004, vol. 55, issue 5, s. 377-395. Dostupný také komerčně z databáze EBSCO HOST/LISTA: <<http://web.ebscohost.com>>. ISSN 1532-2882.

Seznam a zdroje obrázků:

1, s.13:

ClickforArt [online]. c2010 [cit. 2010-07-29]. Art Prints & Canvas. Dostupné z WWW: <http://www.clickforart.com/styles/Art_Prints_Canvas/>.

2, s.17:

CROFTS, Nick; DOERR, Martin; GILL, Tony. The CIDOC Conceptual Reference Model A Standard for Communicating Cultural Contents. *Cultivate Interactive* [online]. 2003, issue 9, [cit. 2010-08-06]. Dostupný z WWW: <<http://www.cultivate-int.org/issue9/chios/>>. ISSN 1471-3225.

3, s.18:

NORDBOTTEN, Joan C. *SVEIN NORDBOTTEN & ASSOCIATES - BERGEN -NORWAY* [online]. 2008 [cit. 2010-07-29]. Multimedia Information Retrieval Systems . Dostupné z WWW: <http://nordbotten.com/ADM/ADM_book/F4-5-CIDOC-CRM.htm>.

4, s.20:

SMILE [online].Massachusetts Institute of Technology, c2008 [cit. 2010-07-29]. Semantic Interoperability of Metadata and Information in unLike Environments. Dostupné z WWW: <<http://simile.mit.edu/repository/sesame/branches/1.2.1-ops/test/files/museum/fullclosure.rdf>>.

5, s.21:

XML. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, 27 September 2001 , last modified on 4 August 2010 [cit. 2010-08-04]. Dostupné z WWW: <<http://en.wikipedia.org/wiki/XML>>.

6, s.24

CROFTS, Nick; DOERR, Martin; GILL, Tony. The CIDOC Conceptual Reference Model A Standard for Communicating Cultural Contents. *Cultivate Interactive* [online]. 2003, issue 9, [cit. 2010-08-06]. Dostupný z WWW: <<http://www.cultivate-int.org/issue9/chios/>>. ISSN 1471-3225.

7, s.28:

GRAINGER, Gail Shea. *Dewey Browse* [online]. c1997, 2010 [cit. 2010-08-04]. Dewey Browse 700's. Dostupné z WWW:

<http://www.deweybrowse.org/Dewey_Browse_700.html>.

8, s.29:

BREITMAN, Karin Koogan; CASANOVA, Marco Antonio; TRUSZKOWSKI, Walter. *Semantic Web: Concepts, Technologies and Applications*. 1st ed. New York : Springer, c2006. 314 s. NASA Monographs in Systems and Software Engineering. [cit. 2010-07-26] Dostupný také komerčně ze SpringerLink (DOI): <http://dx.doi.org/10.1007/978-1-84628-710-7_9>. ISBN 978-1-84628-710-7.

8, s.32

10, s. 39; 11, s.40:

Rijksbureau voor Kunsthistorische Documentatie. *Iconclass* [online]. [Haag] : RKD, c2010 [cit. 2010-07-28]. Dostupné z WWW: <<http://www.iconclass.nl/>>.

12, s.44; 13, s.47:

J. Paul Getty Trust . *Getty Union List of Artist Names : Union List of Artist Names Online* [online]. Los Angeles (CA) : The Getty Research Institute, 2010 [cit. 2010-07-28]. Dostupné z WWW:

<http://www.getty.edu/research/conducting_research/vocabularies/ulan/about.html>.

14, s.49; 15, s.53:

J. Paul Getty Trust . *Getty Union List of Artist Names : Art & Architecture Thesaurus Online* [online]. Los Angeles (CA) : The Getty Research Institute, 2010 [cit. 2010-07-28]. Dostupné z WWW: <http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html>.

16, s.55:

ICOM. *The CIDOC Conceptual Reference Model* [online]. 2006 [cit. 2010-07-28]. Dostupné z WWW: <<http://www.cidoc-crm.org/index.html>>.

17, s.60:

NORDBOTTEN, Joan C. *SVEIN NORDBOTTEN & ASSOCIATES - BERGEN -NORWAY* [online]. 2008 [cit. 2010-07-29]. Multimedia Information Retrieval Systems . Dostupné z WWW: <http://nordbotten.com/ADM/ADM_book/F4-5-CIDOC-CRM.htm>.

18, s.63:

Museum of New Zealand Te Papa Tongarewa [online]. Wellington : c2004 [cit. 2010-07-29]. Collections Online. Dostupné z WWW: <<http://collections.tepapa.govt.nz/>>.

19, s.64; 20, s.67:

BREITMAN, Karin Koogan; CASANOVA, Marco Antonio; TRUSZKOWSKI, Walter. Semantic Web: Concepts, Technologies and Applications. 1st ed. New York : Springer, c2006. 314 s. NASA Monographs in Systéme and Software Engineering. [cit. 2010-07-26] Dostupný také komerčně ze SpringerLink (DOI): <http://dx.doi.org/10.1007/978-1-84628-710-7_9>. ISBN 978-1-84628-710-7.

21, s.71

Seznam příloh:

Příloha 1: **MIME TYPY**

Příloha 2: **CO JE MOŽNÉ POVAŽOVAT ZA UMĚNÍ**

Příloha 3: **TYPOLOGIE ONTOLOGIÍ**

Příloha 4: **KATEGORIE CIDOC CRM**

Příloha 5: **PRVKY DUBLIN CORE**

Příloha 6: **ZPŮSOBY INTEGRACE ONTOLOGIÍ**

MIME TYPY

Způsoby kódování, které standard MIME zavádí, umožňují "protlačit" skrz přenosový mechanismus v podstatě libovolná data, která mají být přenesena. S vlastními daty však musí být vždy přenesena i informace o tom, co jsou tato data zač - zda například jde o obrázek ve formátu GIF, dokument ve formátu MS Word, text ve formátu jazyka HTML apod. Právě podle této informace se totiž příjemce bude rozhodovat, co s příslušným datovým objektem učiní (například tehdy, když mu uživatel dá kliknutím najevo, že si jej přeje zobrazit). Je dobré si uvědomit, že tato informace o typu dat je nezávislá na způsobu jejich kódování - fakt, že určitá data jsou pro potřeby přenosu zakódována např. pomocí kódování Base64, ještě neříká nic o tom, co jsou tato data zač a jak má být s nimi dále naloženo.

Pro potřeby identifikace datového obsahu používá standard MIME dvouúrovňovou "klasifikaci", obecně zapisovanou jako dvojice "typ/podtyp" (anglicky: type/subtype). Na vyšší úrovni se rozeznává celkem sedm možných hlavních typů (které jsou dále upřesňovány konkrétními podtypy):

<i>typ</i>	<i>podtyp</i>	<i>charakteristika</i>	<i>příklady</i>
text	čistý ASCII	bez jakéhokoli formátování; pro něj se používá podtyp "plain"	text/plain
	text v jazyku HTML		text/html
	text ve formátu "richtext" či "enriched"	který je definován v dokumentech RFC 1341, resp. 1896	
image		obrázky	image/gif, image/jpeg
audio		zvuková data	
video	obrazová data	obrazová data	video/mpeg
application		obecně pro všechny ostatní druhy dat, které musí být pro potřeby zpracování "předhozeny" nějakému konkrétnímu programu (aplikaci, odsud označení hlavního typu), nebo mají samy "spustitelný" charakter	application/msword, application/pdf skripty v jazyce JavaScript
multipart		vychází vstříc potřebě připojovat ke zprávám více příloh, resp. vytvářet takové zprávy, jejichž obsah (tělo) bude rozdělen na několik částí, které mohou mít zcela samostatný obsah - se samostatným způsobem kódování i samostatným datovým typem	multipart/mixed
message		používá se v případě, kdy obsahem	message/RFC822

		jedné zprávy je jiná zpráva	
--	--	-----------------------------	--

Na celém konceptu tzv. MIME typů je zajímavý ještě jeden fakt - jde o řešení, které se prosadilo i jinde, mimo rámec elektronické pošty, v situaci kdy také bylo třeba přesně vyspecifikovat, co jsou určitá data zač. Konkrétně se jedná v komunikaci mezi WWW serverem a jeho klientem, uskutečňovanou prostřednictvím protokolu HTTP (HyperText Transfer Protocol). Také zde je totiž nutné, aby server klientovi explicitně řekl, jaký je typ dat, která mu právě posílá. No a k tomu se využívá již jednou zavedená dvouúrovňová hierarchie MIME typů.

PETERKA, Jiří. Co může Internet nabídnout? : Co je a jak funguje MIME?. *Softwarové noviny* [online]. 1997, č. 10, [cit. 2010-07-29]. Dostupný z WWW: <<http://www.earchiv.cz/a710s200/a710s232.php3>>. ISSN 1210-8472.

CO JE MOŽNÉ POVAŽOVAT ZA UMĚNÍ



McCLOUD, Scott. *Jak rozumět komiksu*. Praha : BB Art, 2008. s.164-165. ISBN 978-80-7381-419-9.

TYPOLOGIE ONTOLOGIÍ			
ASPEKT	ONTOLOGIE	POPIS	POZNÁMKA
předmět konceptualizace - typ prezentovaných informací	reprézentací znalostí <i>knowledge representation</i>	specifikují základní elementy modelování znalostních systémů, nabízí modelové konstrukty (např. třídy, podtřídy, hodnoty, sloty, axiomy)	snaže o neutralitu vůči subjektům
	obecné (generické) <i>generic</i>	věnují se obecným konceptům přesahujícím jednotlivé domény (např. procesy, události, akce, stavy); mohou být použité v rozdílných vědních oborech	vyšší nároky na odvozování
	základní (<i>foundational, top-level nebo upper-level</i>)	jsou součástí obecných ontologií; zachycují nejobecnější pojmy a vztahy	např. SUMO nebo CYC
	doménové <i>domain</i>	obsahují koncepty specifické pro danou doménu, klade se omezení na strukturu a obsah domény znalostí	nejfrekventovanější, pod doménové ontologie spadají ontologie jádrové (<i>core</i>) – modelují centrální koncepty určité domény
	úlohové <i>task</i>	jde o generické modely znalostních úloh a metod jejich řešení	zaměřují se na procesy odvozování
	doménově-úlohové <i>domain-task</i>	jde o generické modely znalostních úloh a metod jejich řešení v jedné specifické vědní oblasti	zaměřují se na procesy odvozování
	aplikační <i>application</i>	obsahují veškeré definice, potřebné k vytvoření znalostního modelu požadovaného pro konkrétní aplikace; typicky jde o směsici doménových a obecných ontologií	Nespecifičtější
	metodické <i>method</i>	zachycují definice pro pojmy a vztahy relevantní k daným procesům	
míra a typ struktury konceptualizace	terminologické (lexikální)	specifikují termíny využitě pro reprezentaci znalostí v oboru	taxonomický charakter (např. sémantické sítě v Unified Medical Language System)
	informační znalostní	specifikují strukturu záznamu v databázi specifikují konceptualizaci vědní	bohatší vnitřní struktura
míra formalizace	neformální (semi-formální)	řízené slovníky, glosáře (pojmy vysvětlené přirozeným jazykem), thesaury, taxonomie	
	formální	vycházejí formálními jazyky – formální taxonomie,	svým způsobem v sobě obsahují i neformální ontologie
operacionalizace	hrubé (sálené/provozní) <i>coarse (stared/run-time)</i>	jazyk s malou expresivitou a malým počtem axiomů	
	jemné (referenční/vývojové) <i>fine-grained (development-time)</i>	vysoce expresivní modelovací jazyk s velkým množstvím axiomů	vyšší nároky na odvozování
ontologické závazky (předpoklady přístupu k modelování reality – <i>ontological commitments</i>)	obsahující jednotliviny	entitty nemohou mít instance	
	obsahující obecniny	entitty mohou mít instance	
	obsahující jednotliviny i obecniny		
	deskriptivní	založené na vztazích vyplývajících z běžného jazyka	vychází z běžného chápání skutečnosti (<i>common sense</i>) a studie jazyka
	revizionostické	nevědí se na přirozený jazyk	vychází z vědeckých a odborných přístupů
	multiplikační	vychází z malé skupiny základních konceptů – ostatní koncepty se rekonstruují z této malé skupiny	připouští kolokalizaci entit
redukcionistické	snaže o co nejvyšší expresivitu – velký počet základních konceptů	poměrně složitě	

	posibilistické	popisuje i to, co může existovat	
	aktualistické	popisuje pouze to, co aktuálně existuje	
	presentistické	modeluje jen to, co v daný okamžik aktuálně existuje	
	eternalistické	modeluje vše, co existovalo v minulosti, existuje nyní, či bude existovat v budoucnu	
	3D	běžné předměty jsou plně přítomny v každém časovém okamžiku	předměty jsou proměnné (v různých okamžicích mají různé vlastnosti)
	4D	běžné předměty jsou chápány jako „časoprostorová červí“ – jsou částečně přítomny v každém okamžiku	předměty se skládají z fází, v rámci kterých mohou mít různé vlastnosti
expresivita (týká se obsahu)	těžké <i>heavy-weight</i>	ontologické závazky jsou vyjádřeny explicitně; větší množství axiomů	
	lehké <i>light-weight</i>	neobzahuji vyjádření ontologických závazků	soustředí se jen na relevantní oblasti

SVÁTEK, Vojtěch; VACURA, Miroslav. Ontologické inženýrství. In POPELÍNSKÝ, Lubomír. DATAKON : sborník databázové konference. 1. vyd. Brno : Masarykova univerzita, 2007. s. 1-32. ISBN 978-80-7355-0.

P94	--vytvořila (byl vytvořen)	E65 Tvorba	E28 Pojmový objekt
P13	---vytvořilo typ (byl vytvořen)	E83 Vytvoření typu	E55 Typ
P95	--zformovalo (byla zformována)	E66 Zformování	E74 Skupina
KATEGORIE CIDOC CRM	--uvedlo do života (se narodila)	E67 Narození	E21 Osoba
	--vyprodukovala (byl vyprodukován)	E12 Výroba	E24 Fyzický umělý Objor hodnot výtvor
P123	--rezultovala do (byla výsledkem)	E81 Proměna	E77 Trvalá položka
P93	-ukončil existenci (byla zbavena existence)	E64 Konec existence	E77 Trvalá položka
Id			
P13	je identifikován jako (identifikuje)	E6 CRM Entita	E18 Fyzický výtvor
P99	je rozpuštěna (rozpuštěna)	E68 Rozpuštění	E74 Skupina
P100	--byla úmrtím (zemřela)	E69 Smrt	E21 Osoba
P124	--přeměnila (byla přeměněna) (je preferovaným identifikátorem)	E81 Proměna	E77 Trvalá položka
P15	byla ovlivněna (ovlivnila)	E7 Akce	E1 CRM Entita
P78	je identifikována (identifikuje)	E52 Doba	E49 Označení času
P16	byla použita (byl použit)	E7 Akce	E70 Výtvor
P87	je identifikována jako (identifikuje)	E53 Místo	E44 Označení místa
P192	byla vytvořena (byla vytvořena)	E71 Umělý výtvor	E17 CRM Entita
P331	byla použita (byl použit)	E71 Modifikace	E56 Návrh/postup
P2	je typu (je typem)	E1 CRM Entita	E55 Typ
P134	byla pokračována (měla pokračování)	E71 CRM Entita	E62 Řetězec
P79	počátek je vymezen	E52 Doba	E62 Řetězec
P96	byla opora (byla oporou při)	E83 Vyhoření typu	E62 CRM Entita
P49	byla použita (byla využita)	E7 Akce	E57 Umělý výtvor
P5	je součástí (představuje část z)	E3 Fyzický stav	E3 Fyzický stav
P70	je součástí (je součástí)	E4 Období	E53 Místo
P26	je součástí (je součástí)	E9 Řádek	E53 Místo
P27	je součástí (je součástí)	E9 Řádek	E53 Místo
P8	je součástí (je součástí)	E4 Období	E19 Fyzický objekt
P90	je součástí (je součástí)	E40 Období správce	E48 Období výtvor
P30	je součástí (je součástí)	E40 Modifikace	E50 Období
P43	je součástí (je součástí)	E50 Výtvor	E54 Roznáší položka
P44	je součástí (je součástí)	E58 Místní výtvor	E39 Místní stav
P45	je součástí (je součástí)	E78 Místní výtvor	E59 Místní stav
P28	je součástí (je součástí)	E58 Místní výtvor	E39 Místní výtvor
P49	je součástí (je součástí)	E18 Fyzický výtvor	E39 Aktér
P30	je součástí (je součástí)	E58 Místní výtvor	E39 Aktér
P51	je součástí (je součástí)	E18 Fyzický výtvor	E39 Aktér
P38	je součástí (je součástí)	E18 Fyzický výtvor	E39 Aktér
P39	je součástí (je součástí)	E18 Fyzický výtvor	E39 Aktér
P55	je součástí (je součástí)	E19 Fyzický objekt	E53 Místo
P54	je součástí (je součástí)	E67 Narození	E53 Místo
P99	je součástí (je součástí)	E68 Rozpuštění	E74 Skupina
P56	je součástí (je součástí)	E79 Fyzický objekt	E70 Fyzický útver
P25	je součástí (je součástí)	E79 Fyzický objekt	E60 Fyzický objekt
P48	je součástí (je součástí)	E18 Modifikace	E46 Určování umělý
P59	je součástí (je součástí)	E18 Fyzický výtvor	E53 Místo
P68	je součástí (je součástí)	E24 Fyzický umělý výtvor	E17 CRM Entita umělý
P67	odkazuje na (je odkazována z)	E73 Informační objekt	E17 CRM Entita
P70	je součástí (je součástí)	E37 Dokument	E17 CRM Entita umělý
P71	je součástí (je součástí)	E32 Autoritní dokument	E55 Typ
P129	je součástí (je součástí)	E73 Informační objekt	E17 CRM Entita umělý
P138	je součástí (je součástí)	E36 Vizuální položka	E17 CRM Entita
P33	je součástí (je součástí)	E11 Modifikace	E29 Návrh/postup
P68	je součástí (je součástí)	E29 Návrh/Postup	E57 Materiál
P69	je součástí (je součástí)	E29 Návrh/Postup	E29 Návrh/postup
P92	je součástí (je součástí)	E63 Počátek existence	E77 Trvalá položka
P72	je součástí (je součástí)	E33 Jazykový objekt	E56 Jazyk

P34	přidání (byl přidán k)	E39 Aktér	Přiřazení	E53 Místo	Identifikátor
P75	je držitelem (je drženo)	E39 Aktér	Identifikátoru	E39 Právo	
P38	mačkou (byl mačkáno)	E39 Aktér	Přiřazení	E41 Kontaktní	Identifikátor
P81	po celou dobu	E52 Doba	Identifikátoru	E60 Časový	údaj
P62	roztáhl rozměr (byl zjištěn při)	E52 Doba	Měření	E61 Časový	údaj
P43	trvala delší dobu (byla trvanlivější)	E52 Doba	Přiřazení typu	E54 Rozměr	
P84	trvala nejvíce (byl největším trváním)	E52 Doba		E54 Rozměr	
P86	spadá do (zahrnuje)	E52 Doba		E52 Doba	
P88	se skládá z (představuje část z)	E53 Místo		E53 Místo	
P89	spadá do (zahrnuje)	E53 Místo		E53 Místo	
P90	má hodnotu	E54 Rozměr		E60 Číslo	
P91	má jednotku (je jednotkou pro)	E54 Rozměr		E58 Měrná	jednotka
P97	z otce (byla otcem)	E67 Narození		E21 Osoba	
P101	měl jako obecné použití (byl využitím)	E70 Výtvar		E55 Typ	
P103	byl zamýšlen pro (byl zamýšleným cílem)	E71 Umělý výtvar		E55 Typ	
P104	je předmětem (vztahuje se na)	E72 Objekt práva		E30 Právo	
P105	právně podléhá (má právo na)	E72 Objekt práva		E39 Aktér	
P106	se skládá z (představuje část z)	E73 Informační objekt		E73 Informační	objekt
P107	měla/má jako člena (byl/je členem)	E74 Skupina		E39 Aktér	
P109	měla/má jako kurátora (byl/je kurátorem)	E78 Sbírka		E39 Aktér	
P111	přidalo (byl přidán při)	E79 Přidání části		E18 Fyzický	výtvar
P113	odstranilo (byl odstraněn při)	E80 Odstranění části		E18 Fyzický	výtvar
P114	se časově shoduje s	E2 Dočasná entita		E2 Dočasná	entita
P115	ukončuje (je ukončena)	E2 Dočasná entita		E2 Dočasná	entita
P116	startuje (je odstartována)	E2 Dočasná entita		E2 Dočasná	entita
P117	nastává v průběhu (zahrnuje)	E2 Dočasná entita		E2 Dočasná	entita
P118	překrývá časově (je časově překryta)	E2 Dočasná entita		E2 Dočasná	entita
P119	bezprostředně předchází (bezprostředně následuje)	E2 Dočasná entita		E2 Dočasná	entita
P120	je časově před (je časově po)	E2 Dočasná entita		E2 Dočasná	entita
P121	se překrývá s	E53 Místo		E53 Místo	
9122	hraničí s	E53 Místo		E53 Místo	
P125	využívala objekt typu (byl typem objektu použitého při)	E7 Akce		E55 Typ	
P126	použila (byl použit při)	E11 Modifikace		E57 Materiál	
P127	je pokryt širším výrazem (zahrnuje užší výraz)	E55 Typ		E55 Typ	
P128	nese (je na nosiči)	E24 Fyzický umělý výtvar		E73 Informační	objekt
P65	-ukazuje vizuální položku (je předváděna)	E24 Fyzický umělý výtvar		E36 Vizuální	položka
P130	ukazuje vlastností (vlastností jsou též na)	E70 Výtvar		E70 Výtvar	
P73	-má překlad (je překladem)	E33 Jazykový Objekt		E33 Jazykový	objekt
P132	se překrývá s	E4 Období		E4 Období	
P133	je odděleno od	E4 Období		E4 Období	
P137	je doloženo příkladem (dokládá)	E55 Typ		E1 CRM Entita	
P139	má alternativní tvar	E41 Označení		E41 Označení	
P140	přidalo atribut k (byla popsána)	E13 Přiřazení atributu		E1 CRM Entita	
P34	-se týkalo (byl ohodnocen)	E14 Zjištění stavu		E18 Fyzický	výtvar
P36	-zaregistrovalo (byl registrován)	E15	Přiřazení	E19 Fyzický	objekt
P39	-změřilo (byl změřen)	E16 Měření		E70 Výtvar	
P41	-klasifikovalo (byla klasifikována)	E17 Přiřazení typu		E1 CRM Entita	
P141	přiřadilo (byla označena)	E13 Přiřazení atributu		E1 CRM Entita	
P35	-zjistilo (byl zjištěn)	E14 Zjištění stavu		E3 Fyzický	stav

Příloha 5

PRVKY DUBLIN CORE

prvek	kvalifikátor	schéma zápisu
Datum (date)	Dostupný (available) Vytvořen (created) Datum schválení (dateAccepted) Datum copyrightu (dateCopyrighted) Datum odevzdání (dateSubmitted) Vydán (issued) Modifikovaný (modified) Platný (valid)	DCMI období W3C-DTF
Formát (format)	Rozsah (extent) Médium (medium)	IMT
Identifikátor zdroje (identifier)	Bibliografická citace (bibliographicCitation)	URI
Jazyk (language)		ISO 639-2 RFC 1766 RFC 3066
Název (title)	Alternativní (alternative)	
Pokrytí (coverage)	Prostorový (spatial)	DCMI Box ISO 3166 DCMI bod TGN
	Časový (temporal)	
Popis (description)	Abstrakt (abstract) Obsah (tableOfContents)	
Předmět a klíčová slova (subject)		DDC, LCC, LCSH, MeSH, NLM, UDC
Příspěvatel (contributor)		
Správa autorských práv (rights)	Přístupová práva (AccessRights) Licence (license)	
Tvůrce (creator)		
Typ (type)		Slovník typů DCMI
Vydavatel (publisher)		
Vztah (relation)	Odpovídá čemu (conformsTo) Má formát (hasFormat)	URI

	Obsahuje část (hasPart) Má verzi (hasVersion) Je formátem (isFormatOf) Je částí (isPartOf) Je odkazován (isReferencedBy) Je nahrazen (isReplacedBy) Je požadován (isRequiredBy) Je verzí (isVersionOf) Odkazuje (references) Nahrazuje (replaces) Vyžaduje (requires)	
Zdroj (source)		URI

Kromě základního souboru patnácti prvků Dublin Core je možné použít několik dalších prvků. Jsou uvedeny v tabulce 2 (DCMI, 2006a; Masarykova univerzita, 2005). Obvykle prvek navrhla pracovní skupina DC nebo určitá komunita, která potřebovala evidovat některé údaje nad rámec základních prvků DC.

SYNKOVÁ, Veronika. Vybrané aplikace metadatového formátu Dublin Core. *IKAROS : elektronický časopis o informační společnosti* [online]. 2008, roč. 12, č. 5, [cit. 2010-07-29].

ZPŮSOBY INTEGRACE ONTOLOGÍ

Derivation/Modeling

In this approach, a specialized or simpler vocabulary is developed with an existing, more comprehensive vocabulary as an initial model, as illustrated in Figure 1. Examples include MeSH and *Sears List of Subject Headings*, both initially followed the structural pattern of LCSH (Wiggins, 1976; Sinkankas, 1979). The *Canadian Subject Headings* (CSH), used by Canadian libraries primarily for English language materials, was developed as a supplementary list to LCSH. Ferrari (1999) reported that four art museum libraries (Toledo Museum of Art Library, the Rhode Island School of Design Library, the Museum of Modern Art Library, and the John and Mable Ringling Museum of Art Library) have used or still use "alternative classification systems" derived from both DDC and LCC. These alternative systems exist primarily for the more specific needs of the librarians and patrons of these collections. A more recent example is the FAST (Faceted Application of Subject Terminology) scheme under development by OCLC. FAST derives subject terms from LCSH and modifies the syntax to enable a post-coordinate approach (Chanet al., 2001).

Translation/Adaptation

Some controlled vocabularies consist of terms translated from a controlled vocabulary in a different language. A number of recent non-English lists of subject headings have been developed initially with terms translated from LCSH. For example, *Repertoire de vedettes-matière* (RVM), developed by the Université Laval in Canada, is a French language adaptation of both LCSH and CSH (Schweitzer, 1995). *Rameau*, another vocabulary in the French language maintained by Bibliothèque Nationale in Paris, is in turn an adaptation of RVM (Jouguet, 1995). Other subject vocabularies that began with translations of headings from LCSH or that used LCSH as syntactic models include ones in Belgium, Lithuania, Malaysia, and, more recently, in the Czech Republic and Sweden. To meet the needs of their intended audiences, these vocabularies often incorporate modifications. There are also translations between non-English vocabularies. The *Yleisien suomenalaisten asiasaisto* (YSA), a general thesaurus in Finnish maintained by the National Bibliographic Services, was translated into a Swedish general thesaurus *Allmän teaurus på svenska* (Alliirs). They are both available online and can be searched simultaneously through one Web site at: <http://vesa.lib.helsinki.fi/index.html> (IFLA, 2001:21). Although, in some cases, the original reason for translation or adaptation may not have been to achieve interoperability, the results are often interoperable vocabularies.

Satellite and Leaf Node Linking

Traditionally, in the satellite thesauri approach, specialized thesauri are treated as satellites of a superstructure. Specialized vocabularies are often created to meet the needs for managing specialized materials or areas, as in the case of LCSH and the LCSH-based thesauri, including LIV, TGM, and GLIN (*Library of Congress Thesauri* Web site, available at: <http://www.loc.gov/lexico/servlet/lexico/tgnil/brsearch.html>). In this case, LCSH functions as a superstructure. Other examples are the Getty vocabulary databases, which provide a search interface for three separate online thesauri developed under the Getty Vocabulary Program: *Art and Architecture Thesaurus* (AAT), *The Union List of Artist Names* (ULAN), and *The Getty Thesaurus of Geographic Names* (TGN) (Getty Research Institute, 2000, available at: <http://www.getty.edu/research/tools/vocabularies/>). Currently, there are few superstructures in existence. A more recent development is the English Heritage

Project's *National Monuments Thesauri* (English Heritage, 1999), composed of eight separate online thesauri, including *Monument Types Thesaurus*, *Archaeological Objects Thesaurus*, *Building Materials Thesaurus*, *Defense of Britain Thesaurus*, *Evidence Thesaurus*, *Maritime Cargo Thesaurus*, *Maritime Craft Type Thesaurus*, and *Maritime Place Name Thesaurus* (available at: <http://www.english-heritage.org.uk/thesaurus/thesapiash.htm>). This project does not include a physical superstructure. The thesauri are displayed in an integrated space through a frame-based Web site. This could be considered as presenting a virtual superstructure. Terms are grouped by class rather than by the broadest term (Top Term) so that terms included from thesauri can be linked together. Leaf nodes in a tree structure can be used to link to an extended scheme for sub-topics of the node. When a leaf node in one thesaurus is linked to a high-level (e.g., wetlands) and more specific sub-topics of that concept exist in a specialized vocabulary or classification system (e.g., wetlands classification scheme), then the leaf node can refer to the specialized scheme.

Direct Mapping

The process of mapping vocabularies basically consists of establishing equivalence between terms in different controlled vocabularies or between verbal terms and classification numbers. Figure 4 illustrates direct mapping. Mapping generally requires considerable intellectual effort, as exemplified by MACS and Merimee. or, to a lesser degree, by OCLC's LCSH/DDC mapping and Northwestern University's LCSH/MeSH project. In recent years, computer-assisted mapping has been investigated, as in the case of the mapping between LCSH and MeSH conducted at the Northwestern University Libraries, and between LCSH and DDC in *WebDewey*. In the Renardus project, the mapping is thus far an intellectual effort. But in the field of Engineering, an experiment with automatically mapped classes (between the DDC and the EI Classification) has been carried out. There have been a number of efforts for identifying candidates for automatic mapping, both in the classification-mapping field and in the schema matching field. Such systems usually rely on dictionaries for determining synonyms, evaluating common substrings, and identifying concepts whose documentation shares many non-common words. These approaches do not take into account the internal structure of concept representation, the structure of a classification itself, or the steps users take when merging search terms. ChNaera, which is an interactive merging tool based on the Ontolingua ontology editor and PROMPT, exemplifies an approach that incorporates a conceptual and relational structure. PROMPT is a tool for semi-automatic, guided ontology merging mechanism developed by the Medical Informatics group at Stanford University. It identifies candidates for merging as pairs of matching terms, terms from different source ontologies representing similar concepts. The second generation of PROMPT system, Anchor-PROMPT, uses relations among the terms in an ontology and a set of anchors—pairs of similar terms—to determine which other terms in the ontology are similar. The algorithm was tested by using unrelated source ontologies developed by different research groups. Promising results were achieved, with a precision rate of between 61 and 100% depending on the size of the initial anchor set and the maximum length of the path that was traversed (Noy & Musen, 2001). The algorithm has been implemented as a plugin to the Protégé-2000 (<http://protege.stanford.edu/>) ontology-editing and knowledge acquisition tool developed at Stanford Medical Informatics.

Co-occurrence Mapping

The co-occurrence mapping method works at the application level, i.e., in metadata records, where the group of subject terms can actually result in loosely-mapped terms. Instead of preparing a completely mapped work at the source vocabulary level (as discussed above), the group of subject terms can actually result in loosely-mapped terms. When a metadata record includes terms from more than one KOS source, the co-occurrence of terms from different schemes provides an automatic, loose mapping between vocabularies. As a group, these loosely-mapped terms can answer a particular search query or a group of questions. Figure 5 illustrates the co-occurrence mapping approach: In cataloging, the MARC 21 format allows subject terms from various sources to co-exist in one record, as indicated by fields 650 (topical term, subject added entry), 653 (index term, uncontrolled subject added entry), and 654 (faceted topical term, subject added entry) (Library of Congress Cataloging Distribution Service, 1999). Metadata standards developed in recent years also designate more than one controlled subject vocabulary as the recommended source of subject terms for subject-related elements. For example, the art- and image-related metadata standard VRA (Visual Resource Association) Core Categories requires using AAT for the TYPE, MATERIAL, and STYLE/PERIOD elements, while recommending the use of AAT, LCSH, TGM, ICONCLASS (an international classification system for iconographic research and the documentation of images), and *Sears Subject Headings* for the data values for the CULTURE and SUBJECT elements (VRA, 2002). Metadata records usually include both controlled terms and uncontrolled keywords. If used effectively, these metadata records can provide a large pool of loosely mapped terms from various controlled vocabularies as well as from freetext. Such sets may be used in mapping between vocabularies or directly for retrieval.

Switching

In translating equivalent terms in different vocabularies, a switching language or scheme may be used as an intermediary. This is the approach adopted by H.W. Wilson Company's Megathesaurus, project Renardus, and a number of other projects introduced in Establishing Interoperability Among KOS in Different Languages section in this paper. FIG.

Such a switching system can be a new system (e.g., UMLS' *Metathesaurus*) or an existing system (e.g., DEXT). The use of an anchor term or a notation around which terms from various vocabularies can be gathered facilitates link management. The anchor terms can reside in an Interlingua (e.g., HEREIN) or in one of the vocabularies designated as the "master language" (e.g., the Polish Project). Existing classification schemes, such as DDC, UDC, and LCC, have also been used as gathering devices. The use of a classification structure as the backbone enables hierarchical browsing and the display of thesaurus terms.

Linking Through a Temporary Union List

Terms that are not conceptual equivalents but are closely related linguistically may be linked to enhance retrieval. A project for CAMEd, a comprehensive resource for complementary and alternative medicine, developed software to link four thesauri used by the participating databases. Terms are linked according to their word matching, whole or partial, when being displayed in a temporary union list generated by the software in response to a query (Zeng & Chen, 2003).

It should be noticed that new products are not created in this approach. The temporary union list functions like a concordance of various vocabularies, but, in fact, no concordance or mapping table is produced and saved for future use. There is no "source" or "target" thesaurus either. The temporary union term list corresponds to the term(s) from a query, rather than to the terms of a "source" thesaurus. This approach is illustrated in Figure 7.

Other examples can be found in the projects introduced in Establishing Interoperability Among KOS in Different Languages section, such as MACS and Merimee, which employ linking as their fundamental approach.

Linking Through a Thesaurus Server Protocol

Another approach that does not lead to new products includes establishing a linked environment through a thesaurus server protocol. The Alexandria Digital Library, (ADD Project at the University of California, Santa Barbara, developed both a Gazetteer and a Thesaurus Service Protocol to support programmatic searching and retrieval of distributed online resources. The Thesaurus Protocol is based on the ANSI/NISO (1993, R2C)03 Z39.19 thesaurus model and supports downloading, querying, and navigating thesauri (Fig. 8).

Through the use of the thesaurus service protocol, thesaurus services register or otherwise make themselves known. They accept queries in the format specified by the protocol, run the queries against their local thesaurus, and return results to the application that initiated the query in a standard format or in a customized extended format (Janice, Ikeda, & Hill, 2002).

ZENG, Marcia Lei; CHAN, Lois Mai. Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. *Journal of the American Society for Information Science and Technology*. 2004, vol. 55, issue 5, s. 377-395. Dostupný také komerčně z databáze EBSCO HOST/LISTA: <http://web.ebsco>

Evidence výpůjček

Prohlášení:

Dávám svolení k půjčování této bakalářské práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

V Praze, 23. 7. 2010

Martina Čulíková

Jméno	Katedra / Pracoviště	Datum	Podpis