

METADATA JAKO NOVÝ NÁSTROJ PRO KOMUNIKACI WEBOVSKÝCH INFORMAČNÍCH ZDROJŮ

Eva Bratková
ÚISK FF UK, Praha
E-mail: brt@cuni.cz

1 Co jsou metadata

Aktuálním a často diskutovaným tématem v odborných publikacích a na konferencích jak počítačových odborníků, tak i informačních pracovníků a knihovníků je problematika popisu digitálních zdrojů, zejména síťově dostupných, označovaná v češtině přejatým výrazem „**metadata**“. Jde o téma, které se původně objevilo již v polovině 80. let v souvislosti s budováním kolekcí či archivů digitalizovaných textů. V 90. letech na ně pak navázalo budování komplexnějších digitálních knihoven, které však již obsahují také jiné typy digitálních zdrojů. Problematika metadata se však stala atraktivní, a to pro podstatně větší okruh odborníků i uživatelů informací, až v momentu, kdy začaly být digitální zdroje dostupné přes síť Internet. Metadata hluboce souvisejí se vznikem a rozvojem sítě Internet a jejích služeb, především pak WWW. Zdá se, že právě v **síťovém prostředí metadata** nabírají zcela nové rozměry a význam, a právě tomu je věnován následující příspěvek. Problematika metadata se týká celého modelu komunikace informačních zdrojů (informací) v počítačových sítích. Je problematikou celých 90. let. Knihovníci - tedy zatím zejména zahraniční - se jí pak intenzivně zabývají posledních 4-5 let.

Knihovnické pracovníky, zejména pak katalogizátory problematika metadata zajímá proto, že síťově dostupné elektronické dokumenty byly a jsou také, i když zatím jenom v malé míře, předmětem jejich zpracovatelské činnosti. Dodejme, že činnosti v klasickém modelu, kdy záznamy o dokumentech jsou oddělené od nich samých, hromadí se v dnes rozsáhlých bázích katalogů a slouží koncovým uživatelům pro vyhledávání těchto dokumentů. Příslušné části katalogizačních předpisů včetně formátů pro strojové zpracování zahrnují kapitoly, které takové informace a instrukce běžně obsahují. Základním odborným termínem, který doposud pro označení zpracování dokumentů užívali a stále užívají, je katalogizační popis či záznam, resp. bibliografický popis či záznam. Podotkněme, že jde o výrazy užívané v moderní katalogizaci nejméně 150 let [23, Introduction]. Proto mnohé překvapilo, že se v posledních letech začalo i v odborných knihovnických kruzích hovořit a psát o metadatach jakožto nástrojích či prostředcích popisu elektronických dokumentů, především pak síťově dostupných. Těmi, kdo se těmito problémy prioritně zabývá, nejsou však knihovníci, nýbrž počítačovní odborníci, kteří rozvíjejí další technologie i celkovou novou architekturu komunikace informací na

WWW pro 21. století. Přípravuje se budování „**sémantického webu**“ (Semantic Web) [5]. Do oblasti knihovnictví vnesli uvedený výraz právě počítačovní odborníci v této oblasti působící.

Zhruba v polovině 90. let se začaly mezi knihovníky objevovat první definice a výklady metadata, které byly velmi jednoduché a nepostihovaly zcela podstatu, funkce a význam metadata zejména v síťovém prostředí. V citacích se často opakoval výklad, že metadata jsou data o datech, z čehož bylo usuzováno, že katalogizační záznamy jsou vlastně také metadata. V řadě studií, připravených v rámci významných výzkumných projektů v zámoří i v Evropě, se dokonce v přehledech, analýzách a hodnoceních metadata objevilo zařazení tradičních bibliografických formátů typu MARC (USMARC, UKMARC, UNIMARC, PICA aj.) jako jednoho z typů metadata [16, Content]. Někteří odborníci přesněji navrhli i základní kategorie (skupiny) metadata, například z hlediska typů organizací, které metadata používají. V jedné ze studií evropského projektu BIBLINK, věnovaného problematice propojování sféry vydavatelů a národních knihoven v prostředí počítačových sítí, jsou zmíněné formáty zařazeny ve skupině 4.2.2 [21, kap. 4.2]. Mnohé knihovnické odborníky závěry výzkumů překvapily, někteří je nechápali a ptali se, proč že mají knihovníci své bibliografické formáty nazývat metadata a výsledek své katalogizační činnosti záznamy metadata [22, část 1]. Někteří se však v diskusích k problému zamysleli hlouběji nad závěry výzkumů a jejich prezentací a po vyhodnocení informací přicházejících také z oblasti informačních služeb provozovaných na Internetu ho pochopili také jako výzvu k zamýšlení se nad pradávnou a palčivou otázkou vztahu mezi popisnými a selekčními údaji v popisu dokumentů [20]. Objevují se otázky, jestli mají informace o webovských zdrojích, které jsou dnes přenášeny na webu současně se zdroji, stejný charakter a podstatu jako záznamy o dokumentech ukládaných do katalogů či bibliografií. Domnívám se, že nikoliv.

Pokud chceme lépe pochopit problematiku současných metadata, musíme vyjít za hranice poznatků a zkušeností získaných ve vlastní dílčí specializaci oboru tradičního knihovnictví či bibliografie a snažit se analyzovat a vyhodnocovat nové cesty a procesy komunikace informací, o které se starají dnes také jiní odborníci. Komunikace informací již dávno není doménou jenom knihovníků. Svět WWW je světem informací a nové paradigma komunikace informací (včetně informací o informacích) je realitou. Pokud zůstaneme v zajetí tradičního modelu zpracování dokumentů, byť elektronických, budou nám pojmy „metadata“ a katalogizační/bibliografický záznam připadat stejné. Jejich významy však stejné nejsou.

Autor posledně citované práce [20] Stefan Gradmann se pokusil porovnat a odhalit podstatné rozdíly mezi oběma pojmy. Poukazuje na fakt, že tvůrci dnes nejnámější specifikace (sémantiky) metadata „Dublinského jádra“ (Dublin Core, dále též DC - viz výklad ve 3. části) při jeho navrhování jednoduše nepřevzali a neupravili existující formát MARC, ale navrhli zcela nový soubor údajů k po-

pisu digitálních dokumentů. Pravdou je, že pro současnost ale zejména budoucí architekturu komunikace informací na WWW je formát typu MARC nevhodný co do struktury (syntaxe) i co do obsahu (sémantiky). Autor zdůraznil, že při porovnávání obou pojmů (metadata a katalogizační/bibliografický záznam) nejde v žádném případě jen o rozdíl v počtu údajů obsažených v záznamech [20, Introduction]. Kdo by chtěl považovat proces tvorby metadata za nějaký typ zjednodušené katalogizace, bude se hluboce mýlit. Podstatné rozdíly jsou funkční i strukturní povahy, vězí v **celém kontextu produkce a užití metadata** v rámci síťové digitální komunikace informací [20, Who does it, and How is it done?]. Konkrétněji autor ukazuje na:

- otázku předmětu zpracování a popisu (v rámci komunikace na WWW může být popisován jakýkoliv digitální objekt; každý digitální objekt, který je přenášen sítí, musí být opatřen jistým minimem informací o sobě)
- otázku účelu tvorby údajů o zdrojích (u metadata se podtrhuje účel zjišťování a vyhledávání zdrojů v síťovém prostředí, u bibliografických či katalogizačních záznamů hraje podstatnou roli stránka deskriptce; deskriptivní katalogizace dokáže vyloučit i důležitý údaj pro vyhledávání jenom proto, že není obsažen na titulní stránce, přestože dokument takový znak má)
- otázku tvůrců záznamů (u metadata na WWW se předpokládá primární tvorba ze strany autorů, editorů či vydavatelů)
- nezávislost či samostatnost jednotlivých metadata a jejich účelného shromáždění, přeskupování, propojování a zejména zpracování a nové užití pro nejrůznější formy výstupů a služeb včetně služeb v oblasti znalostních systémů
- otázku užití metadata, která je klíčová: v rámci WWW jsou primárními uživateli metadata speciální inteligentní programy zvané „agenti“; bibliografické/katalogizační záznamy jsou určeny koncovým uživatelům.

Při odhalování rozdílů se S. Gradman odvolává na výroky autority v této oblasti nanejvýše povoláné - Tima Berners-Leea, zakladatele WWW (toho času ředitele Konsorcia World Wide Web, dále jen W3C) a jednoho z tvůrců současné architektury WWW. T. Berners-Lee ve své práci „Architektura metadata“ [4] podal několik základních pregnantně vyjádřených axiomů (předpokladů) a definic s komentáři, které se týkají metadata jakožto podstatné, imanentní součásti celého prostoru WWW. Základní charakteristikou WWW je fakt, že **[informační] zdroje** (v pojetí autora je výraz „zdroj“ základním formálním pojmem, pod nějž spadá výraz dokument jakožto zdroj textové povahy), popřípadě obecněji **objekty**, o které při webovské komunikaci v režimu klient-server žádáme a které dostáváme, jsou vždy v rámci komplexu protokolů TCP/IP doprovázeny určitým množstvím informací o nich samotných, aniž je uživatel vnímá či vidí (datum poslední manipulace se souborem, majitel zdroje, formát aj.). Jde však o **informace o informacích**, které počítačová odborníci běžně nazývají „**metadata**“ [4, Documents, Metadata ...]. Na

obrazovkách nejsou na první pohled vidět, protože jsou součástí tzv. „hlaviček“ (headers) přenášených dokumentů (informací) nebo je doprovázejí při přenosu po síti.

Pokud jde o český jazyk, bylo by možné pro neologismus „metadata“ použít jistě ne nesprávný výraz „metaúdaj(e)“ (Francouzi například užívají vlastní výraz „méta-données“), řada dalších národních jazyků však preferuje výraz pocházející z angličtiny. Je věcí další diskuse, kterému výrazu bude dána přednost v blízké budoucnosti u nás, prozatím je v rámci tohoto příspěvku většinou v řadě kontextů upřednostňován výraz „metadata“. V některých případech se může objevit výraz „metaúdaj(e)“ - v českém výrazu lze užívat i singuláru. Výrazy „data“ (pouze v plurálu) a „údaje“ jsou synonyma. Pokud jde o rozdíl mezi obecním pojmem „data“, resp. „údaje“, a pojmem „informace“, pak uveďme, že jde o složitější problematiku, ke které se jistě vyjádří i v souvislosti s rozvojem digitální informační komunikace mnoho odborníků, v daném okamžiku snad bude postačovat jednoduchá známá charakteristika, uváděná v mnoha variantách, že údaj je znakový projev uložený na nosiči či přenášený v počítačové síti, který se v procesu užití stává informací, resp. poznatkem [30, s. 192]. Již citovaný T. Berners-Lee ve svém výkladu rovněž nečiní podstatnější rozlišování obou pojmů [4]. Volně lze dále parafrázovat: metadata (metaúdaje) se v procesu užití stávají metainformacemi. Protože výraz „užití“ je vztahován v první instanci na inteligentní programy-agenty, bude vhodné v rámci zcela automatizovaných složitých cest a procesů zpracování a využívání údajů o zdrojích, které v prostoru WWW technologií již existují a dále se rozvíjejí, preferovat pojem „metadata“.

T. Berners-Lee uvádí ve výše citované práci základní definici pro metadata: jde o „**stroji srozumitelné informace o webovských zdrojích nebo dalších věcech**“ [4, Documents, Metadata ..., Definition]. Podtrhla bych, že v definici se skutečně nevyskytuje výraz „strojem čitelné informace“, na který jsme byli zvyklí u bibliografických formátů typu MARC, nýbrž výraz „stroji srozumitelné informace“ (angl. „machine understandable information“). Jde o zásadní rozdíl obou výrazů. Definice vypadá na první pohled velmi jednoduše, ale skrývá v sobě řadu podstatných momentů, které byly již zmíněny výše v textu v rámci charakteristiky metadata S. Gradmana. Zdůrazněna je klíčová charakteristika metadata, totiž, že jde o stroji srozumitelné informace. Předpokládá se jejich zpracování v/pro nejrůznější aplikace pomocí inteligentních programů označovaných výrazem „**agenti**“. Znamená to, že informační jazyk uložených metadata musí být naprosto formálně logicky správný a jednoznačný. V budoucnu, až se podaří zdokonalit informační jazyky metadata i programy, které je budou zpracovávat a využívat, vytvoří se základna pro **web strojům srozumitelných informací o čemkoliv**: o lidech, věcech, pojmech, faktech, myšlenkách atd. Předpokladem ovšem je v daném okamžiku vybudování **systému pro informace o informacích** (informačních zdrojích) [4, Documents, Metadata ..., Definition].

Pro metadata platí podle T. Berners-Lee několik předpokladů, o kterých není nutné diskutovat. Předně platí, že

metadata jsou data. Obecně jde o informace o informacích, a proto musejí být metadata považována za informace ve všech jejich aspektech [4, Documents, Metadata ..., Definition]. Metadata mohou být uložena jako každá jiná data v nějakém zdroji, to znamená, že nějaký zdroj může obsahovat informaci o sobě samém nebo o jiných zdrojích. V současné praxi WWW tedy existují tři způsoby existence a cest metadat:

- 1) údaje o webovském dokumentu jsou obsažené v něm samotném v hlavičce v tagu <HEAD>; jde o způsob základní existence a předpokládá se, že bude uplatněn v řadě aplikací metadat (v této chvíli se využívá omezených možností jazyka HTML). Příkladem mohou být v současné době provizorní aplikace metadat Dublinského jádra v dokumentech HTML, v budoucím provozu se počítá s jazykem XML
- 2) údaje o webovském dokumentu, které doprovázejí komunikaci typu „klient-server“; po přenosu dokumentu je možné údaje pomocí příslušné funkce prohlížeče získat; tento způsob je předmětem značného zájmu počítačových odborníků
- 3) údaje o nějakém webovském dokumentu je možné získat z jiného webovského dokumentu (jsou jeho součástí); tato cesta je velmi perspektivní pro budoucí efektivní komunikaci na WWW (konkrétně půjde již ale zejména o XML dokumenty).

Metadata jako taková mohou být sama předmětem popisu jako svébytné digitální objekty. Pak by se dalo hovořit o „meta-metadatech“. Z praktických důvodů se však tento termín spíše nepoužívá a o takových datech se také hovoří jako o metadatech.

Pokud jde o **formu metadat**, je tvořena množinou nezávislých **výroků**, které reprezentují **údaje o zdroji**. V počítačovém systému nabývají výroky formu **jména** či **typu výroku** a souboru dalších **parametrů**. Například:

Jméno výroku	Autor zdroje
Parametr 1	Jméno autora zdroje
Parametr 2	Afiliace autora zdroje
Parametr 3	E-mail autora zdroje

O dvou výrocích týkajících se stejného zdroje platí, že jsou nezávislé a mohou existovat samostatně. Jde o významnou a pro komunikaci na WWW důležitou vlastnost metadat. Pokud se vyskytují dohromady na jednom místě, označujeme je jako „**kombinovaný výrok**“. Množiny výroků jsou považovány za neuspořádané seznamy.

Výroky o zdrojích korespondují s jejich příslušnými vlastnostmi (znaky). To znamená, že typ výroku je výrokem o tom, že zdroj má pojmenovanou vlastnost (např. autor, název, datum apod.). Parametrem se rozumí buď dílčí typ vlastnosti (např. autor-fyzická osoba, autor-korporace apod.), nebo dílčí vlastnost (např. jméno autora, afiliace autora, e-mail autora apod.). Například:

Autor zdroje	
Jméno autora zdroje	Tim Berners-Lee
Afiliace autora zdroje	World Wide Web Consortium
E-mail autora zdroje	timbl@w3.org

Ke každé vlastnosti náleží konkrétní hodnota, obecně pak hovoříme o modelu dvojice (páru) **vlastnost/hodnota**.

Na obecné úrovni je výrok o zdroji celkově tvořen následujícími komponentami:

- URI (Uniform Resource Identifier) zdroje
- Identifikátorem typu výroku
- Dalšími parametry k typu výroku.

Implicitně nebo explicitně musí být dále součástí výroku:

- Strana, která ho učinila
- Datum a čas učiněného výroku.

Aby byla tvorba metadat na WWW důvěryhodná a spolehlivá, budou v celkové budoucí architektuře hrát významnou roli prostory (dokumenty s příslušným URI), které budou obsahovat **slovníky se jmény či typy vlastností**, jež jsou definovány podle stanovené metodiky v rámci příslušných specifikací metadat. Předpokládá se jejich hypertextové propojení s tvořenými záznamy metadat jednotlivých webovských zdrojů. V tomto případě půjde o specifické výroky typu **vztahu mezi dvěma zdroji**, které budou realizovány přes hypertextové odkazy. Celý takový výrok bude tvořen typem výroku a dvěma identifikátory URI.

Uvedená základní charakteristika metadat, jejich prvků a principů, je východiskem pro další a podrobnější popis jejich konkrétních aplikací, jimž budou věnovány další části textu. Pozornost bude nejdříve věnována charakteristice metadat z hlediska sémantiky. Vedle některých vybraných příkladů metadat bude zvláštní místo věnováno projektu „Dublinského jádra“ s ohledem na jeho mezinárodní význam. V závěru je nastíněna problematika syntaxe metadat, kterou dnes reprezentuje především model označovaný zkratkou RDF.

2 Metadata pro digitální zdroje v zahraničních systémech a službách

Formát metadat označovaný výrazem „Dublinské jádro“, jehož základní charakteristika je rozvedena v následující 3. části, nebyl první svého druhu, který byl přímo navržen a implementován v provozu Internetu. Již před ním (a také po něm) byly navrženy a do praxe uvedeny jiné formáty. První metadata se začala využívat v lokálních systémech digitálních fondů plných textů, později napojovaných i na Internet. Později přibýly další aplikace navržené přímo pro komunikaci na WWW. Vybrané příklady nejznámějších metadat ze světové praxe dokládají

jejich postupný vývoj, specifika jejich sémantik ovlivněných funkcemi, které plní, ale i různorodost syntaxí, která brání jejich vzájemné součinnosti.

2.1 Formáty TEI, EAD a CIMI

První tři formáty metadat jsou specifické především svým zaměřením na historické texty či rukopisy, archiválie nebo objekty muzejních sbírek. Společné mají i to, že vznikly v rámci výzkumných projektů budování digitálních archivů textů a digitálních knihoven. Všechny mají specifikaci metadat založenou na obecném značkovacím jazyce SGML (Standard Generalized Markup Language), který je normou ISO 8879-1986. Všechny aplikace mají definované své specifické DTD (Document Table Definition) [8, část 1.].

TEI (Text Encoding Initiative) Independent Headers

Formát tzv. nezávislých hlaviček TEI je hlavním reprezentantem této skupiny, další dva z něho vycházejí. Formát TEI, jenž vznikl v rámci výzkumného projektu v letech 1987-1994 v USA <<http://etext.lib.virginia.edu/TEI.html>>, je jako celek komplexním formátem pro kódování úplných textů všech typů se zaměřením na detailní textový rozbor. Dokumentace k celému formátu, který je určen především pro provozy lokálních systémů, čítá více než 1400 stran. Jeho povinnou součástí jsou však hlavičky (Headers), které jsou formátem metadat, určeným k popisu digitalizovaného textu. Formát předpokládá, že hlavičky s metadaty mohou být buď součástí textu samotného (proto ho může tvořit i autor nebo vydavatel), nebo mohou být vytvářeny a ukládány odděleně do databází záznamů (třeba pro využití v knihovnách). Hlavičky lze využít i pro popis síťově dostupných zdrojů, které nejsou kódovány ve formátu TEI.

Základní specifikace metadat hlaviček TEI popisu textových zdrojů je co do množství údajů velmi bohatá (výchozí základnou jsou pravidla AACR2), je však přípustné definovat i množinu menší (podle potřeb aplikace). Struktura hlavičky může mít celkově 4 části: 1. Popis celého souboru, 2. Popis kódování textu, 3. Popis věcného charakteru a 4. Údaje o revizi textu. Základní specifikace nemá definován údaj pro URI. Konverze do formátu USMARC je možná. Na obrázku 1 je připraven podle dostupné dokumentace [8, část 2.1.2] ilustrativní záznam metadat ve formátu hlavičky TEI nikoliv historického textu, nýbrž běžného (českého) textu dostupného v Internetu (ačkoliv záznam nemůže mít URL). Ze čtyř oblastí popisu dokumentu jsou uplatněny první (tag FILEDESC) a poslední (tag REVISIONDESC).

EAD (Encoding Archival Description)

Formát metadat EAD, který vznikl v roce 1993 na Kalifornské univerzitě v Berkeley, je pro potřeby Společnosti amerických archivářů (Society of American Archivists) udržován Kongresovou knihovnou ve Washingtonu <<http://www.loc.gov/ead/>>. Je určen pro fondy archiválií a rukopisů [8, část 2.2]. Základním účelem bylo zpřístupnění jejich inventářů a registrů. Základní popis dokumentů je velmi detailní, dlouhá je ale i specifikace selekčních údajů. Obsah údajů má relace na popis formátu hlaviček TEI, ale je také v souladu s pravidly ISAD(G) (International Standard Archival Description). Celý záznam metadat má podobné členění na úseky jako hlavičky TEI, ale obsahuje i údaj pro URI (kód AEDID i kód DAO pro digitální archivní objekty).

Obrázek 1

```
<TEIHEADER>
<FILEDESC>
<TITLESTMT><TITLE>Syrské pahorky</TITLE>
<AUTHOR>Nováková, Nea</AUTHOR>
<RESPSTMT><RESP>Konverze</RESP>
<NAME>Michal Jeřábek</NAME></RESPSTMT></TITLESTMT>
<EXTENT>Soubor ve formátu TEI; 148680 bytů</EXTENT>
<PUBLICATIONSTMT>
<PUBLISHER>Ústav starého Předního východu FF UK Praha</PUBLISHER>
<DATE>1997</DATE>
<AVAILABILITY><P><ADDRESS>
<ADDRLINE>Copyright, 1997, Ústav starého Předního východu FF UK Praha, využívání
publikace podléhá Autorskému zákoníku. Přístup v rámci veřejného webovského
serveru ENLIL</ADDRLINE>
<ADDRLINE> Ústav starého Předního východu FF UK Praha</ADDRLINE>
<ADDRLINE>Celetná 20, Praha 1</ADDRLINE>
<ADDRLINE>Česká republika</ADDRLINE>
<ADDRLINE>URL: http://enlil.ff.cuni.cz/</ADDRLINE></ADDRESS>
</P></AVAILABILITY>
</PUBLICATIONSTMT>
<SERIESSTMT>
<TITLE>Scripta Electronica. Starý Orient. Seria monographica</TITLE>
</SERIESSTMT>
</FILEDESC>
<REVISIONDESC>
<CHANGE><DATE>1998-01-22</DATE>
<RESPSTMT><NAME>Michal Jeřábek</NAME>
<RESP>Údržba</RESP></RESPSTMT></CHANGE>
</REVISIONDESC>
</TEIHEADER>
```

Jde o formát metadat velmi významného a dynamicky se rozvíjejícího systému mezinárodního charakteru (USA a Velká Británie), který je řízen Konsorciem CIMI <<http://www.cimi.org/>>. Formát vznikl v roce 1988 a byl rozvíjen zejména v letech 1990-1993 [8, část 2.3]. Záznamy metadat prezentují jednak texty (katalogy výstav aj.), jednak záznamy objektů muzejních exponátů a také obrazové zdroje. Základní budovaný systém, který je dostupný přes Internet, se jmenuje CHIO (Cultural Heritage Information Online). Specifikace formátu vychází z hlaviček TEI, je však doplněna řadou dalších údajů, zejména věcné povahy (s ohledem na předmět zpracování). Jako URI se uplatňuje kód FPI (Formal Public Identifier). Systém CIMI úzce spolupracuje s formátem Dublinského jádra (testování vzájemné součinnosti), výzkumné práce jsou v současné době zaměřeny zejména na aplikaci protokolu pro vyhledávání informací Z39.50 a také na aplikaci syntaxe RDF (viz 4. část textu).

Formát metadat systému **GILS (Government Information Locator Service)** je reprezentantem metadat z oblasti informací státní správy USA. Vznikl v roce 1994 z iniciativy a za podpory federální vlády USA <<http://www.gils.net/>>. Je rozvíjen i na bázi mezinárodní spolupráce s Kanadou a Austrálií. Informace registrované v systému jsou k dispozici přes WWW rozhraní. Předmětem zpracování jsou především textové materiály administrativního charakteru, řada z nich je dostupná v plném textu na WWW [16, GILS].

Specifikace metadat je velmi obsáhlá, záznamy obsahují značné množství základních popisných údajů, věcných údajů předmětového charakteru (uplatňují se známé řízené slovníky a tezaury), administrativních údajů včetně kontaktních informací. Záznamy obsahují údaje o copyrightu, údaje URI (URL) i údaje pro vazby mezi jednotlivými digitálními zdroji. Popis vychází z instrukcí AACR2, a proto je také možné bez potíží provádět konverze do formátu USMARC a zpět.

Na obrázku 2 je zobrazen ilustrativní záznam stejného webovského dokumentu jako u obrázku 1 ve struktuře vybraných údajů metadat GILS. Záznam byl připraven na základě konverze z původně připraveného formátu DC do formátu GILS pomocí britského konvertoru „DC-dot“, který je k dispozici zdarma na WWW [10].

Obrázek 2

```

<gils>
  <Title>
    Syrské pahorky
  </Title>
  <Originator>
    Nea Nováková
  </Originator>
  <Abstract>
    Scripta Electronica 1997, monographica
  </Abstract>
  <Availability>
    <Distributor>
      <Distributor-Name>
        Ústav starého Předního východu FF UK Praha
      </Distributor-Name>
    </Distributor>
    <Available-Linkage>
      <Linkage-Type>
        (SCHEME=IMT) text/html
      </Linkage-Type>
      <Linkage>
        http://enlil.ff.cuni.cz/cz/scripta/monog/syrpah.htm (SCHEME=URN)
      </Linkage>
    </Available-Linkage>
  </Availability>
  <Date-of-Publication>
    (SCHEME=ISO8601) 1997-03-14
  </Date-of-Publication>
  <Medium>
    Text.Monograph
  </Medium>
  <Language-of-Resource>
    (SCHEME=ISO639-1) cs
  </Language-of-Resource>
  <Cross-Reference>
    <Cross-Reference-Linkage>
      <Linkage>
        (SCHEME=URL)
        http://enlil.ff.cuni.cz/cz/scripta/monographica_cz.htm
      </Linkage>
    </Cross-Reference-Linkage>
  </Cross-Reference>
  <Use-Constraints>
    (SCHEME=URL)
    http://enlil.ff.cuni.cz/cz/scripta/scr_el_cz.htm
  </Use-Constraints>
</gils>

```

2.3 Formuláře IAFA / ROADS

Tzv. formuláře IAFA pro popis různých typů digitálních zdrojů předložila internetové komunitě v roce 1995 <<http://info.webcrawler.com/mak/projects/iafa/iafa.txt>> Pracovní skupina „IAFA (Internet Anonymous FTP Archive) Working Group“, která patří pod organizaci IETF (Internet Engineering Task Force) [16, IAFA/WHOIS++ Templates]. Formuláře byly připraveny přímo pro popis různých typů dokumentů ukládaných do FTP archivů na Internetu: textové dokumenty, obrazové a zvukové zdroje, služby diskusních skupin, databáze, archivy USENET skupin, soubory dat, programy aj. Jde o velmi pestrou a zajímavou typologii zdrojů, a to byl jeden z důvodů, proč tyto formuláře převzal k adaptaci jeden z nejvýznamnějších projektů virtuálních knihoven na WWW britský systém „ROADS“ (Resource Organisation and Discovery in Subject-based Services). Formuláře ROADS jsou k dispozici na webovské adrese <<http://www.ukoln.ac.uk/metadata/roads/templates/>>.

Formuláře IAFA představují co do tvorby dat velmi jednoduché, ale účelné záznamy metadat pro popis internetových zdrojů (staly se též východiskem pro přípravu formátu Dublinského jádra). Na rozdíl od Dublinského jádra ale nejde v tomto případě z formálního hlediska o definici jednoho formátu (formuláře) pro mnoho typů dokumentů, ale o řadu formulářů pro různé typy zdrojů, které mají jádro společných metadat. Záznam zahrnoval jak popisné údaje, tak údaje věcné. Povinným údajem je URI zdroje. Formuláře IAFA se staly součástí některých internetových norem, např. RFC 1835 pro architekturu WHOIS služeb.

2.4 Internetová norma RFC 1807

V krátkém přehledu nejvýznamnějších formátů metadat uvedme ještě známý formát pro popis technických zpráv, který byl zveřejněn v roce 1992 jako internetová norma RFC 1807 <<ftp://ftp.vse.cz/pub/docs/rfc/rfc1807.txt>>. Formát byl určen technické komunitě v USA pro tvorbu a přenos metadat přes Internet. Připravená specifikace je jednoduchá a účelná. Záznam zahrnoval základní údaje o dokumentu, důležité věcné údaje, nezbytné údaje administrativního charakteru pro přenos dat a také URI dokumentů (dostupných z lokálních systémů).

3 Formát Dublinské jádro (DC)



V dalším textu je věnována detailní pozornost nejznámějšímu formátu metadat, který byl navržen pro popis webovských informačních zdrojů. Jde o formát, který je již pátým rokem středem pozornosti v celosvětovém měřítku, a to proto, že se jeho

obsah týká podstatných vlastností informačních zdrojů z hlediska potřeby jejich zjišťování a vyhledávání v rámci rozvíjení nové architektury celého systému WWW. For-

mát má univerzální záběr a v současné chvíli je aplikován v systémech mnoha zemí světa. Je jedním ze základních formátů, které přispěly k vytváření syntaktické struktury metadat v projektu RDF (viz část 4). Má značný význam a perspektivu pro komunikaci informací po roce 2000.

3.1 Vznik formátu DC

„Dublinská iniciativa“ (její přesné současné jméno je Dublin Core Metadata Initiative, dále také DCMI) se zákonitě zrodila v technologicky příznivých podmínkách rozvoje sítě Internet a jejích informačních služeb. Postupný nárůst síťově dostupných zdrojů (v letech 1991-1995) na jedné straně, ale jejich jen velmi obtížné zjišťování či vyhledávání na straně druhé, zejména pokud jde o obsah poskytovaných informací, přimělo řadu odborníků z oblasti počítačové vědy, informační vědy, knihovnictví a dalších příbuzných oblastí k intenzivním úvahám a krokům k řešení tohoto důležitého problému. Na pořad dne se, vedle jiných, dostala v historickém kontextu mnohokrát opakovaná otázka účinné kontroly, ale i nových způsobů komunikace informačních zdrojů - tentokrát síťově dostupných - přes informace, které jsou v současné době označovány výrazem „metadata“.

Za oficiální začátek Dublinské iniciativy je pokládáno jednání pracovního semináře, které se uskutečnilo v březnu 1995 v americkém městě Dublin (Ohio). Není náhodou, že šlo právě o toto město. Je totiž sídlem centra největší americké i světové knihovnické sítě OCLC a hlavní aktéři semináře i celé iniciativy jsou zaměstnanci jejího výzkumného oddělení. Mezi přední osobnosti patří Stuart Weibel a Eric Miller. Vedle OCLC se sponzorsky na semináři podílelo i Národní centrum pro počítačové aplikace NCSA (National Centre for Supercomputing Applications). Cílem semináře bylo společné posouzení potřeb, předností, nedostatků a stávajících řešení daného problému ze strany majitelů či provozovatelů informačních fondů na WWW a zejména dosažení shody (konsensu) při vytváření **základního souboru údajů (metadat) k popisu zdrojů**. Kritériem výběru jednotlivých údajů (prvků) byly podstatné vlastnosti (znaky) síťových digitálních zdrojů z hlediska jejich vyhledávání, vyloučeny byly v dané chvíli všechny další vlastnosti (znaky), které směřovaly k plnění jiných funkcí práce s těmito zdroji (technické, technologické, archivační, obchodní aj.). Za účelem rychlého dosažení pozitivních výsledků byly v úvahu v prvním okamžiku vzaty pouze digitální zdroje textové povahy, tj. dokumenty.

Při navrhování základní množiny údajů (prvků) k popisu zdrojů se zajisté, jak bylo krátce uvedeno již v 1. části textu, nabízela i cesta odvodit ji např. ze známých stávajících modelů knihovnických katalogizací (formátů typu MARC s příslušnými pravidly). K „derivační“ cestě však nemohlo dojít. Profesionální a složitá tvorba záznamů informačních zdrojů aplikovaná v oblasti knihoven byla vyloučena z několika důvodů. Formáty typu MARC jsou jenom „strojem čitelné“, nikoliv však „stroji srozumitelné“, tj. jejich struktura i sémantika nevyhovují požadavkům na budoucí architekturu komunikace informací včetně

ně informací o informacích (metadat) na WWW. Podstatným důvodem je i ohled na budoucí tvůrce záznamů, jimiž mohou být i samotní **autoři (tvůrci)** nebo **vydavatelé (zprístupňovatelé)** zdrojů. Záměrem bylo, jak deklarují materiály ze semináře [37, část 1.3], vytvoření **zcela jednoduchého strukturovaného záznamu (popisu) nového typu**, který by byl doplňkem jednak k popisům knihovnických katalogizací typu MARC apod., resp. popisům typu metadat složitějšího obsahu (např. formát TEI), jednak k zatím nevyhovujícím či nedostatečným popisům realizovaným v rámci zcela automatizované indexace plných textů webovských dokumentů přes internetové systémy typu „search engine“. Z hlediska dnešních, nově formulovaných cílů dalšího rozvoje webu se metadata Dublinského jádra jeví spíše jako základ nebo východisko pro všechny další typy záznamů o webovských zdrojích. Při vytypování jednotlivých údajů nicméně profesionální katalogizační popis sehrál jistou pozitivní úlohu také, a to především proto, že předmětem byly textové digitální dokumenty, jejichž vlastnosti (znaky) se v řadě případů shodují s tradičními. Nebylo tedy nutné při zjišťování a poznávání podstatných vlastností začínat od stavu „nula“, ale mohly se uplatnit již dřívější znalosti a zkušenosti. Na semináři byl deklarován také požadavek na tzv. „promítání“ (mapping) metadat do profesionálních struktur typu MARC, které však narazilo na některé problémy, v jejichž pozadí jsou formy katalogizačních údajů, ale také nové specifické vlastnosti dynamických digitálních zdrojů. Pozitivním momentem tvorby nového typu popisu informačních zdrojů je z jistého aspektu také **zintenzivnění jejich poznávání** s cílem přispět k rozvoji teorie společenské informační komunikace.

Navržená množina údajů pro popis digitálních dokumentů, označená výrazem „**Soubor prvků metadat Dublinského jádra**“ (**Dublin Core Metadata Element Set**, ve zkratce **DC**), zahrnovala v době svého vzniku celkem 13 prvků (údajů). Je potřeba uvést, že nešlo o definici komplexního formátu, jak jsme zvyklí vídat u formátů typu MARC, zdůrazňována byla v první fázi pouze stránka **sémantiky** (obsahu). Syntax (struktura) potřebná zejména ke komunikaci metadat v reálných sítích, stejně jako jejich bližší specifikace ve formě různých průvodců, byly ponechány stranou. Z pozdějšího výčtu základních 15 údajů (viz dále v části 3.3) chyběl údaj pro anotační popis (Description), který byl zaveden až později zejména pro potřeby zpracování obrazových zdrojů, a dále speciální údaj pro ochranu autorských práv (Rights), jehož potřeba se ukázala později rovněž jako důležitá.

Jako hlavní a v podstatě jediný údaj věcné povahy, pokud nebereme v potaz údaj o názvu digitálního dokumentu či objektu a specifický údaj o geografickém a časovém **pokrytí** (Coverage) dokumentu/objektu, byl v souboru uveden údaj o **předmětu/tématu** (Subject), a to v pořadí jako první z důvodu jeho nejdůležitějšího postavení z hlediska vyhledávání. Následovaly údaje dobře známé z oblasti tradičního popisu dokumentů: **název** (Title), dále trojice tzv. „činitelů“, resp. „agentů“ (z angl. Agents) procesu informační, resp. dokumentové komuni-

kace, tedy **autor** (Author), **nakladatel/vydavatel** (Publisher), u něhož se ovšem v krátké definici zdůrazňovala novodobá funkce instituce, která dokument v síti zpřístupňuje, a sekundární **další činitelé** (Other Agents) - z hlediska knihovnické katalogizace bychom mohli říci, že jde o tzv. další původce dokumentů. V kontextu komunikace digitálních zdrojů jim však mohou být přisuzovány další role, resp. jejich role se i mění. K dalším vytypovaným údajům patřily **datum** (Date), jehož uplatňování v popisu dynamických digitálních zdrojů je značně komplikované a stále diskutované, **typ objektu** (Object Type), který patří co do dílčích hodnot k jednomu z nejsložitějších [7], **forma** (Form), představující reprezentaci (uspořádání) dat na počítačovém médiu, specifický **identifikátor** (Identifier), kterým je pro webovské dokumenty především jejich adresa uložení (URI, URN, URL), údaje o **vztahu/vazbě** k jiným dokumentům (Relation), údaj o původním **zdroji** (Source) digitálního dokumentu a konečně **jazyk** (Language) dokumentu.

Specifickou a stále diskutovanou otázkou bylo (a je do dnešních dnů) deklarované zaměření formátu DC na **autory a vydavatele** dokumentů. Jde o velmi zajímavý (v této chvíli ještě hypotetický) prvek v rámci společenské dělby práce v informační komunikaci. Předpokladem však bude, a již první seminář tuto vizi nastínil [37, část 3.0], vytváření účinných interaktivních **pracovních formulářů** a dalších nástrojů dostupných na WWW a v jejich rámci zejména automatické využívání **profesionálních souborů metaúdajů (metadat)** typu číselníků, souborů identifikátorů, řízených heslářů, tezaurů, identifikátorů apod., které už existují. V experimentálních provozech jsou však na webu k dispozici také nové jiné pomůcky budované v rámci současných projektů, jako je například experimentální nástroj k automatickému generování třídniců Deweyho desetinné klasifikace pro webovské dokumenty včetně tvorby struktury RDF na Univerzitě ve Wolverhamptonu, Velká Británie [1].

Aktéři prvního semináře se rovněž shodli na zásadách dalšího rozvíjení formátu. Zdůrazněno bylo zejména:

- zahrnutí pouze podstatných údajů v popisu, a to s ohledem na vyhledávání dat na WWW
- rozšiřitelnost formátu v konkrétních lokálních aplikacích či systémech v mezích stanovené vzájemné budoucí součinnosti (interoperability)
- volitelnost a opakovatelnost všech údajů
- modifikovatelnost vybraných údajů zejména pro potřeby speciálních systémů přes tzv. kvalifikátory (schémata/modely hodnot údajů a typy údajů).

3.2 Další rozvíjení DC

Výsledky prvního semináře Dublinské iniciativy uvítala řada institucí v USA ale i v zahraničí, které se rozhodly připravit první projekty zpracování webovských dokumentů ve svých automatizovaných provozech. Patřily k nim například významný Severský projekt metadat (**The Nordic Metadata Project**) zajišťovaný a řízený informační sítí NORDINFO [27], v jehož rámci byl připraven na

WWW dostupný generátor unifikovaných jmen zdrojů URN (Uniform Resource Names) pro účastníky ze Skandinávie, dále pohodlný interaktivní formulář k tvorbě metadat podle Dublinského jádra [18] a také experimentální konvertor pro převod metadat do formátů typu MARC [9]. Prověřování možností převodu („mapování“) metadat Dublinského jádra do knihovnického formátu USMARC v rámci výzkumných aktivit zahájila také Kongresová knihovna ve Washingtonu. Již výše v textu zmiňované problémy, které z toho vzešly, jsou především povahy sémantické. V zájmu realizace konverze bylo dokonce v roce 1996 do formátu USMARC doplněno nové pole 720 pro nekontrolované jméno pro potřebu převodu údaje o autorech/tvůrcích digitálních zdrojů DC, které nejsou řízeny přes soubory autorit [17, část 2.3].

V průběhu příprav na realizaci prvních projektů využívajících specifikace údajů DC vyvstala před zpracovateli řada problémů. Po zhruba roční elektronické diskusi byly jednotlivé sporné a otevřené otázky předloženy účastníkům 2. pracovního semináře DC, který se konal již za početné účasti odborníků z dalších zemí ve Velké Británii na Univerzitě ve Warwicku v dubnu 1996 [17]. Seminář má v historii Dublinské iniciativy mimořádný význam, protože na něm byl předložen k posouzení významný návrh modelu komunikace metadat v síti Internet, který dostal jméno „**Warwick Framework**“ (Warwickský Rámec).

Jednání konkrétně projednalo následující okruhy problémů:

- **Specifikace syntaxe metadat** pro potřeby jejich přenosu v prostoru WWW. Ukázalo se totiž, že při realizaci projektů zpracování a využívání informací o digitálních zdrojích tvůrci začali navrhovat a uplatňovat různou navzájem neslučitelnou syntax a strukturní vazby k přenosu metadat v prostoru WWW. Chyběl definovaný společný model takového přenosu. Příslušná pracovní skupina Dublinské iniciativy proto připravila návrhy k vytvoření jednoduchého nástroje na vnoření metadat přímo do dokumentů. V rámci jazyka SGML byla pro metadata navržena speciální tabulka pro definici dokumentu DTD, která byla promítnuta do formátu HTML (v roce 1996 ve verzi 2.0), a to v rámci jeho hlavičky, tj. tagu <HEAD>. Zároveň byl připraven návrh na propojení metadat na externí zdroj, který obsahuje jejich specifikaci (definici). Shodou okolností byl v květnu 1996 konán i seminář Pracovní skupiny pro distribuovanou indexaci a vyhledávání konsorcia W3C za účasti představitelů všech významných vyhledávacích systémů a služeb na WWW, na kterém byl přijat konsensus, v rámci něhož byly jako dočasné řešení návrhy Dublinské iniciativy (v rámci WF - viz dále v textu) přijaty.
- **Příprava pracovních manuálů a interaktivních formulářů** k tvorbě metadat. Tvorba takových pomůcek se stala v rámci realizovaných projektů velmi náležitou, protože je měli vytvářet autoři dokumentů nebo vydavatelé. Příklady dobře fungujících interaktivních formulářů k tvorbě metadat včetně

doplňkových elektronických průvodců je možno dnes najít na WWW mnoho. Jmenovala bych alespoň formulář Severského projektu metadat [18], pomocí něhož byly připraveny pro tento text i některé obrázky s příklady metadat.

- **Způsoby rozšíření základní skupiny metadat Dublinského jádra** představovaly a představují dodnes velmi palčivý problém. Řada systémů apelovala na doplnění různých dalších údajů, které by zajišťovaly další potřebné funkce při využívání zdrojů. Velký tlak byl vyvíjen ze strany představitelů užívajících formáty typu MARC. Základní myšlenkou tvůrců DC ovšem je, že údaje potřebné pro zajišťování lokálních a specializovaných funkcí pro práci s digitálními zdroji jsou v kompetenci jejich systémů. DC představuje skutečně jenom základní společné jádro pro popis zdrojů - v rámci WWW bude plnit základní funkci poznávací (nástroj popisu/zobrazení zdroje) a komunikativní (nástroj uložení, přenosu a využití obsahu popisovaného zdroje).
- **Specifikace rámce (architektury)** pro vzájemnou komunikaci metadat v prostoru WWW. Tým amerických odborníků, kteří měli zkušenosti již z budování digitálních knihoven, navrhl a jednání semináře předložil k posouzení návrh modelu pro komunikaci metadat pocházejících z různých systémů, který dostal jméno „Warwick Framework“ (dále též WF). Pro jeho přípravu byly do úvahy kromě Dublinského jádra vzaty ještě formát RFC 1807 a formáty (formuláře) IAFA (informace o nich viz v části 2). Warwickský rámec se stal vedle jiných předchůdcem tzv. „Rámce pro popis zdrojů“ (RDF, Resource Description Framework), který je dnes rozvíjen konsorciem W3C (viz část 4).

Warwickský rámec byl návrhem **architektury**, který měl vyhovovat jednotlivým modelům metadat v prostoru WWW a navzájem je spojovat. Šlo o architekturu pro vzájemnou výměnu jednotlivých souborů (balíčků) metadat nejrůznější povahy: záznamy zdrojů (DC, IAFA aj.), soubory definic, struktur a sémantik jednotlivých metadat, různé číselníky, tezaury apod. I proto důležité místo ve WF hrála navržená **typologie jednotlivých balíčků metadat** jakožto objektů. Byly navrženy tři základní typy: jednoduchý soubor metadat (např. záznam jednoho zdroje), nepřímý soubor metadat (tj. odkaz na externí jiný zdroj, který je obsahuje) a kolekce více balíčků metadat najednou (tzv. kontejner) [25, The Warwick Framework architecture]. V reálném provozu se pak předpokládal zcela automatizovaný sběr a směna metadat v konzistentní formě, seskupování jednotlivých typů metadat nebo jejich využívání pro specifické skupiny uživatelů (agenti/klienti).

Extenzí návrhu WF byla i jeho konkrétní, byť limitovaná, implementace pro HTML, verzi 2.0 s podmínkou, že bude transparentní pro existující prohlížeče WWW, indexační a vyhledávací služby a další HTML nástroje. Implementace byla důležitá pro realizaci řady projektů. Konkrétně byla pro HTML realizována následující syntax:

META tag pro uložení vnořených metadat v rámci hlavičky <HEAD> HTML dokumentu. V každém META tagu byla specifikována dvojice „jméno/hodnota“, která byla kódována pomocí atributu „NAME“ a atributu „CONTENT“. V hlavičce mohly být obsaženy vícenásobné údaje. Jednoduchý příklad jednoho údaje:

```
<META NAME="title" CONTENT="Metadata  
Architecture">
```

Pro hodnotu atributu NAME byla navržena zvláštní tečkovací notace, pomocí níž se udávalo jméno souboru metadat a (po tečce) jméno údaje. Později se tečky uplatnily i pro vyjádření typologie jmen údajů. Dnes je tento systém označován výrazem „pseudo-hierarchická tečkovací notace“ [14, část 3.1] (s největší pravděpodobností v budoucnosti nebude využívána). Metadata Dublinského jádra obdržela jméno „DC“. Konkrétní příklad dvou údajů:

```
<META NAME="DC.Title" CONTENT="Metadata  
Architecture">  
<META NAME="DC.Autor" CONTENT="Tim Berners-Lee">
```

Konkrétní úplný záznam o textovém zdroji - článku z časopisu [7], je zobrazen na obrázku 3. Má pouze ilustrativní funkci, metadata nejsou reálně obsažena v textovém dokumentu, jenž ale má své reálné URL. Soubor údajů obsahuje také další parametr definovaných údajů (SCHEMA, schéma), který Dublinská iniciativa schválila až později. Záznam byl připraven pomocí interaktivního formuláře Severského projektu metadat [18] v souladu s verzí 1.0 DC.

LINK tag byl navržen pro potřeby uvedení odkazu na webovský zdroj, v němž se nachází specifikace (definice) daného použitého souboru metadat. Syntax obsahovala typ schématu souboru metadat (atribut REL) a URL zdroje na WWW (atribut HREF). Například definice údaje „název“ ve specifikaci DC bude odkazována následujícím způsobem:

```
<LINK REL=SCHEMA.dc  
HREF="http://purl.org/metadata/  
dublin_core_elements#title">
```

Příklad reálné množiny metadat o webovském dokumentu (domovské stránce ÚISK FF UK) včetně dílčích odkazů na definice údajů specifikace metadat DC obsahuje obrázek 4. Metadata byla rovněž připravena pomocí interaktivního formuláře Severského projektu metadat [18] a byla uložena do prostoru hlavičky <HEAD> HTML dokumentu (součástí obrázku nejsou další značky jazyka HTML). Z reálné webovské stránky je možné si pomocí speciálního odkazu, umístěného na ní, prohlédnout výstupní záznam generovaný přes jeden ze serverů Severského projektu (jde ovšem pouze o vnější efekt určený pro koncového uživatele). Zajímavá je možnost konverze metadat do formátu typu MARC pomocí speciálního programu-konvertoru „d2m“, jehož rozhraní je dostupné na WWW [9]. Zatím je však pochopitelně propracována konverze především vůči severským formátům MARC. K získání formátu MARC postačí zápis URL webovského dokumentu, který obsahuje metadata v hlavičce <HEAD>, do interaktivního formuláře konvertoru.

Obrázek 3

```
<META NAME="DC.Title" CONTENT="K otázkám pojmu, třídění a typologie internetových  
a webovských informačních zdrojů">  
<META NAME="DC.Creator.PersonalName" CONTENT="Bratková, Eva">  
<META NAME="DC.Creator.PersonalName.Address" CONTENT="brt@cuni.cz">  
<META NAME="DC.Subject" CONTENT="Webové informační zdroje">  
<META NAME="DC.Subject" CONTENT="Třídění">  
<META NAME="DC.Subject" CONTENT="Typologie">  
<META NAME="DC.Publisher" CONTENT="Národní knihovna ČR">  
<META NAME="DC.Date" SCHEME="ISO8601" CONTENT="1998-09-30">  
<META NAME="DC.Type" CONTENT="Text.Article">  
<META NAME="DC.Format" SCHEME="IMT" CONTENT="text/html">  
<META NAME="DC.Identifier" CONTENT="http://www.nkp.cz/start/publikace/k_revue/5.htm">  
<META NAME="DC.Source" CONTENT="Národní knihovna, 1998, roč. 9, č. 5, s. 262-276">  
<META NAME="DC.Language" SCHEME="ISO639-1" CONTENT="cs">  
<META NAME="DC.Date.X-MetadataLastModified" SCHEME="ISO8601" CONTENT="1999-08-03">
```

```

<META NAME="DC.Title" CONTENT="Institute of Information Studies and Librarianship (IISL)
Home Page">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">
<META NAME="DC.Creator.CorporateName" CONTENT="Ústav informačních studií a knihovnictví">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.Subject" CONTENT="Informační věda">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" CONTENT="Knihovnictví">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Subject" CONTENT="Vysokoškolské vzdělávání">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.Publisher" CONTENT="Ústav informačních studií a knihovnictví">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#publisher">
<META NAME="DC.Contributor.PersonalName" CONTENT="Skenderija, Saša">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">
<META NAME="DC.Contributor.PersonalName" CONTENT="Bratková, Eva">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#contributor">
<META NAME="DC.Date" CONTENT="(SCHEME=ISO8601) 1995-01-10">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">
<META NAME="DC.Type" CONTENT="Text.Homepage.Organizational">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#type">
<META NAME="DC.Format" CONTENT="(SCHEME=IMT) text/html">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#format">
<LINK REL=SCHEMA.imt HREF="http://sunsite.auc.dk/RFC/rfc/rfc2046.html">
<META NAME="DC.Identifier" CONTENT="http://www.cuni.cz/ffuisk">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">
<META NAME="DC.Language" CONTENT="(SCHEME=ISO639-1) cs">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">

```

Větší zásah do souboru (množiny) Dublinského jádra byl proveden po konání v pořadí 3. semináře Dublinské iniciativy (opět v Dublinu) v září 1996 [40]. Jeho ústředním tématem byly **obrazové informační zdroje** zpřístupňované v prostředí Internetu a WWW. Diskuse byla zaměřena na řadu aspektů těchto zdrojů ve vztahu k jejich popisu v rámci specifikace DC. Metadata pro tento typ webovských zdrojů, která mají jiný typ formátu než HTML, vyžadují jiný prostor, než je samotný obrazový zdroj. Záznamy se ukládají zpravidla do interaktivních databázových souborů budovaných jako součást digitálních knihoven, které mohou být přes příslušné rozhraní na WWW propojeny.

Jeden ze závěrů semináře byl podstatný: 13 dosavadních prvků nepostačovalo k popisu obrazových zdrojů. Podstatný údaj, který po diskusi přibyl do sestavy, byl údaj **popis** (Description), který je nutný k vyjádření jeho obsahu pro potřeby vyhledávání. Poslední údaj dnes známé „patnáctky“ Dublinského jádra, který přibyl taktéž v souvislosti s obrazovými zdroji, byl údaj **práva** (Rights), týkající se informací o autorských právech a různých omezeních využívání zdroje.

Celá základní množina metadat Dublinského jádra (viz úplný seznam DC v části 3.3) byla po doplnění dvou nových údajů i nově uspořádána. Definitivní verze DC, která nesla označení **DC, verze 1.0**, byla publikována na WWW v prosinci 1996. Během dalšího roku byly prove-

deny dílčí změny (soubor byl aktualizován). Poslední aktualizace byla provedena 2. října 1997 [11]. V září 1998 byla tato verze přijata jako internetová norma **RFC 2413**, jež je jako text rovněž k dispozici na WWW [28]. Podle plánu bude DC předloženo také organizaci NISO k přijetí za americkou normu a organizaci CEN k přijetí za evropskou normu [36, Standardization ...].

Čtvrtý pracovní seminář Dublinské iniciativy se konal v Austrálii (Canberra) v březnu 1997. Zcela naplno se na tomto semináři projevil dvě tendence formování DC. Zastánci první tendence, označovaní jako „**minimalisté**“, prosazovali zachování stávající množiny definovaných prvků. Jejich požadavkem bylo zachování maximální jednoduchosti formátu, tj. minimální počet prvků bez dalších dílčích specifikací (kvalifikátorů), protože jen tak lze vyhovět principům tvorby a užití metadat v prostoru WWW. Zastánci druhé tendence, označovaní jako „**strukturalisté**“, naopak, s vědomím komplikací, které mohou nastat při tvorbě a užití metadat v provozu na WWW včetně velkých nákladů, požadovali rozšíření definované množiny DC jednak co do počtu prvků, jednak co do kvalifikátorů. Řešení dané situace nebylo a není jednoduché a bude dáno až v budoucnu po získání dalších zkušeností z provozu prvních systémů [39].

Nicméně, canberrský seminář projednal návrh na podrobnější specifikaci DC pomocí tzv. kvalifikátorů (pa-

parametrů blíže určujících definované prvky). Šlo konkrétně o kvalifikátory:

- **Jazyk (Language)**, ve kterém je hodnota údaje zapsána a uložena. Jde o významný parametr vzhledem k mezinárodnímu charakteru budování webovských systémů a služeb. Jak uvidíme později, tento parametr nemusí být součástí specifikace sémantiky (viz možnosti kódování takové hodnoty v navrhovaném jazyce XML [31, část 2.12], jak je vidět i na obrázku 8).
- **Schéma (Scheme)**, který udává model či způsob zápisu nebo i tvorby údaje. Užívání tohoto parametru je zpravidla vázáno na různé normy (např. ISO 8601 pro formu dat jako časových údajů, ISO 639 pro kódování jazyků, internetovou normu MIME pro formáty souborů přenášejících v rámci Internetu aj.) nebo známé řízené předmětové hesláře, tezaury či klasifikační schémata). V rámci komunikace metadat se předpokládá odkazování takových externích zdrojů metadat, pokud budou dostupné na WWW, pomocí příslušné syntaxe formátu. Několik kvalifikátorů tohoto typu je vidět v příkladech na obrázcích 3 a 4. Za povšimnutí stojí v obr. 4 v pořadí druhý „LINK“ odkaz u údaje DC.Format realizovaný v podmínkách jazyka HTML, který míří k webovskému dokumentu - internetové normě RFC 2046, jež definuje formáty MIME. Takové dokumenty jsou považovány za speciální digitální objekty metadat a v rámci specifikace budoucího jazyka XML se označují anglickým výrazem „namespace“ („prostor jmen“).
- **Typ (Type)**, který udává buď další parametr údaje (e-mail autora apod.) nebo dílčí typ údaje (autor-fyzická osoba, autor-korporace aj.). Pro potřeby systémů užívajících specifikaci DC byla navržena provizorní pseudo-hierarchická tečkovací notace, pomocí níž jsou oddělovány dílčí údaje (v dokumentaci DC jsou označovány také anglickým výrazem „subelement“). Například:

```
<META NAME="DC.Creator.CorporateName.Address"  
CONTENT="uiskff-@ff.cuni.cz">
```

Jak je vidět z předchozího textu a ukázek příkladů, zabýval se australský seminář podrobněji i syntaxí zápisu kvalifikátorů. K výše uvedenému komentáři připojme ještě malou ukázkou jednoho údaje zahrnujícího najednou všechny 3 typy kvalifikátorů:

```
<META NAME="DC.Subject" CONTENT="(SCHEME=LCSH  
(LANG=EN) Metadata">
```

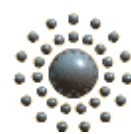
Problematika ze semináře v Austrálii byla dále rozvedena na 5. semináři konaném v tomtéž roce (v říjnu 1997) v Helsinkách. Účastníci hodnotili již početné výsledky z provozu aplikací DC v prostoru WWW (např. viz výše komentovaný úspěšný Severský projekt metadat). Diskutovány byly opět otázky kvalifikátorů a zvláštní místo zaujaly dílčí specifikace údajů **datum**, **pokrytí** a **vztah** [38]. Problematika dat jako časových údajů (např. jejich typologie, forma zápisu apod.) je s ohledem na dynamický charakter webovských zdrojů velmi obtížná a řeší se do dnešní doby i v návaznosti na řešení konsorcia W3C. Speciální údaj o prostorovém a časovém pokrytí zdroje také

doznává mnoho změn, v současné chvíli jsou navrhovány další dílčí specifikace. V rámci údaje o vztazích se řeší komplikovaná otázka vztahu digitálních zdrojů mezi sebou navzájem. Jak známo ze současného webu, identifikace vztahů je nesmírně náročná i pro profesionála, natož pro běžného uživatele, který má potřebu třeba jen citovat určitý dokument nebo jenom dílčí stránku. Jinak jde samozřejmě také o otázku popisné jednotky. V terminologii DC se objevil tento problém pod výrazem „**princip 1:1**“. Konkrétně se problém týká zejména popisu kolekcí webovských stránek sdružených v rámci webovských sídel, verzí dynamických digitálních dokumentů, multimediálních zdrojů apod. V rámci semináře v Helsinkách byly navrženy základní typy dílčích vztahů (relací).

Výsledkem helsinského semináře a následné diskuse příslušné pracovní skupiny bylo publikování materiálu ke kvalifikátorům jednotlivých prvků DC [15] v únoru 1998, který je zatím poslední verzí této dílčí specifikace. Nově byla navržena (zatím pouze pro testovací potřeby) dokonce formalizovaná definice kvalifikátorů DC pro potřeby tvorby metadat DC v rámci schématu RDF [14, Appendix 2].

Jednou z důležitých otázek diskutovaných na 6. semináři DC, který se konal v listopadu 1998 ve Washingtonu, byla formalizace procesů, organizace a řízení Dublinské iniciativy (DCMI). DCMI se stala významným mezinárodním virtuálním společenstvím, jejímž posláním je rozvoj, standardizace a podpora souboru prvků DC. Od počátku roku 1999 bylo proto zahájeno postupné formování organizační struktury, které by mělo být potvrzeno na 7. semináři v říjnu 1999 ve Frankfurtu nad Mohanem. Ve stávající chvíli je podle návrhu DCMI celkově řízeno Ředitelstvím DCMI (DCMI Directorate), které sídlí v rámci hostitelské organizace OCLC. Kromě jiného zajišťuje také webovskou prezentaci <<http://purl.org/DC/>>. Řízení a koordinaci normalizačních prací zajišťuje Výkonná skupina DC (DC Executive Group, DC-EG). Poradní výbor Ředitelství DCMI (DC Advisory Committee, DC-AC), který je tvořen z reprezentantů různých systémů a služeb z celého světa, řídí práci tematických pracovních skupin DC, schvaluje jejich návrhy a spolupracuje s jinými mezinárodními aktivitami a organizacemi (W3C, IETF aj.). Důležitá činnost se odehrává v rámci 16 pracovních skupin (DC Working Groups, DC-WG), které projednávají dílčí problémy rozvoje DC v rámci elektronických diskusních skupin i na seminářích. Základním produktem skupin jsou „**pracovní návrhy**“ (Working Draft), jež mohou po dalším projednávání v celkové diskusní skupině (DC General mail) a schválení v Poradním výboru nabýt povahy „**návrhu doporučení**“ (Proposed Recommendation) a konečného „**doporučení**“ (Recommendation).

3.3 Aktuální stav formátu DC



V červenci roku 1999 předložila DCMI odborné veřejnosti jako **návrh doporučení** revidovanou verzi základního souboru prvků Dublinského jádra, která nese označení „**verze 1.1**“ (**Version 1.1**) [12]. Nejde

o verzi vyššího řádu (verze 2.0 bude předložena pravděpodobně koncem roku 1999 po konání v pořadí již 7. pracovního semináře DC ve Frankfurtu nad Mohanem), ale pouze o podstatnou formální úpravu, v rámci níž došlo ke zlepšení a zpřesnění definic jednotlivých prvků (údajů). Předkládaná verze je „oděna do nového kabátu“ - celý soubor je definován podle společné mezinárodní normy ISO/IEC 11179 „Specifikace a standardizace datových prvků“. Formalizace tohoto typu je velice důležitá pro budoucí využívání metadat DC v reálném automatizovaném provozu na WWW. Návrh definic základních prvků je v současné chvíli k dispozici na WWW v rámci návrhu základního modelu DC-RDF [14, Appendix 1]. Definice prvků musejí podle zmíněné normy obsahovat celkem 10 položek:

1. Jméno údaje
2. Identifikátor údaje jako jeho jedinečný kód pro automatizované zpracování
3. Verze definice údaje
4. Úřad pro registraci údaje
5. Jazyk, v němž je údaj vyjádřen
6. Vlastní definice údaje
7. Povinnost uvádění údaje
8. Typ údaje
9. Maximální počet výskytů údaje
10. Komentář k využití údaje

Ačkoliv od konání posledního 6. semináře DC v prosinci 1998 byla vedena řada diskusí k možnostem rozšíření nebo přeuspořádání dosavadní sestavy údajů Dublinského jádra, odpovědní pracovníci v průběhu jara ujistili, že se žádné změny v tomto směru konat nebudou [36]. Veřejnosti předkládaná verze 1.1 je toho rovněž dokladem. Co do počtu definovaných údajů je soubor totožný s verzí 1.0. Změny nastávají v definicích. O které podstatnější změny jde? V následujícím přehledu jsou uvedeny v rámci komentářů k jednotlivým údajům (prvkům). Jako první je uveden vždy český překlad jména údaje, v kulaté závorce je jméno v anglickém originálu (podle verze 1.1 z července 1999); fráze uvedené v uvozovkách v kulatých závorkách jsou úplnými nebo dílčími citáty ze zveřejněného nového dokumentu [12]:

1. **Název (Title)**

Aktuální definice je oproti původní zkrácená („jméno dané zdroji“); stanovení činitele (původně autor a vydavatel) procesu přidělování jména, tj. názvu zdroje, bylo zrušeno, nevyskytuje se ani v komentáři.

2. **Tvůrce (Creator)**

V rámci tohoto údaje došlo především ke změně jména údaje (původně „Autor nebo tvůrce“); po diskusích byl výraz „autor“ vyloučen též s ohledem na řadu nových aspektů, které autorská role u digitálních informačních zdrojů nabírá (platí zejména pro nově vzniklé typy); nová - a velmi progresivní je i nová definice údaje („entita primárně odpovědná za provedení obsahu zdroje“); z původní definice bylo vyloučeno slovo „intelektuální“, které bylo převzato z katalogizačních instrukcí, které však dnes již ne

zcela koresponduje s novou situací; v komentáři přibývá ve výčtu tvůrců, vedle fyzické osoby a organizace, zcela nově položka „služba“ (služby, systémy apod.); kdo zná současný web, jistě ocení zařazení této potřebné položky; v komentáři se objevuje nová obecnější charakteristika tvůrců jakožto „entit“, ovlivněná též současnými novými náhledy na autorství v materiálech IFLA [19].

3. **Předmět (Subject)**

Původní definice byla přeformulována a podstatně zpřesněna („téma obsahu zdroje“), tj. přibýlo velmi potřebné slovo „obsahu“, které tam původně nebylo.

4. **Popis (Description)**

Nová definice je více zobecněna („výčet obsahu zdroje“) a komentář podává více příkladů k užití tohoto údaje.

5. **Vydavatel (Publisher)**

Nová definice je oproti původní zkrácena („entita odpovědná za zpřístupnění zdroje“); komentář opět, jako v případě údaje „tvůrce“, nově upřesňuje, že vydavatelem může být vedle fyzické osoby a organizace také služba (služby, systémy apod.).

6. **Příspěvatel (Contributor)**

Stejně jako v případě údaje „tvůrce“ došlo i u tohoto údaje ke změně jména, i když jenom formální - byl vypuštěn zbytečný výraz „další“; definice údaje je však přepracována podstatně („entita odpovědná za provedení příspěvku k obsahu zdroje“); výraz o „významném intelektuálním příspěvku“ příspěvatele byl zrušen; komentář k definici rovněž doplňuje ve výčtu příkladů službu (služby, systém), která také může být v této roli.

7. **Datum (Date)**

Původní definice je zcela přepracována, a to na základě rozsáhlé diskuse v pracovních skupinách („datum spojené s nějakou událostí v životním cyklu zdroje“); v definici se odráží charakteristický rys digitálních zdrojů, které jsou u řady typů velmi dynamické (včetně písemných dokumentů); doporučení pro užívání normy ISO 8601 zůstává nezměněno.

8. **Typ (Type)**

Definice v nové verzi dosáhla sice jistého pokroku - je více zobecněná („povaha nebo žánr obsahu zdroje“), je však stále diskutabilní; určitě by mohla být dále zpřesňována; jak napovídá elektronická diskuse v příslušné pracovní skupině, došlo během posledního roku k dalším změnám ve výčtu jednotlivých typů (stav z roku 1998 byl komentován v časopise Národní knihovna [7, část 4]).

9. **Formát (Format)**

Rovněž v případě formátu digitálních zdrojů došlo k upřesnění definice („fyzická nebo digitální manifestace zdroje“); formulace definice je, kromě jiného, ovlivněna současnými náhledy a závěry IFLA k problematice informačních zdrojů a dokumentů,

kteřé byly často v diskusích brány v potaz; komentář k definici doplňuje některé další údaje, jako jsou např. rozměry potřebné pro popis některých typů informačních zdrojů (obrazové, trojrozměrné apod.).

10. Identifikátor (Identifier)

V definici tohoto údaje, který hraje v rámci architektury současné i budoucí webovské komunikace velkou úlohu, došlo ke zpřesnění definice („jednoznačný odkaz na zdroj v rámci daného kontextu“); komentář doplňuje příklady o identifikátor DOI (Digital Object Identifier).

11. Zdroj (Source)

Ačkoliv se kolem tohoto údaje v poslední době rozproudila velká diskuse - zdálo se, že údaj bude sloučen s údajem „vztah“, nakonec k tomu nedošlo a zůstává i nadále v celé sestavě. Jeho definice byla mírně upravena („odkaz na zdroj, ze kterého byl popisovaný zdroj odvozen“).

12. Jazyk (Language)

Definice údaje je nezměněna, komentář ovšem upřesňuje odkazy na normy, které jsou doporučeny k využívání; zpřesněny jsou dílčí instrukce.

13. Vztah (Relation)

Definice tohoto údaje je upřesněna („odkaz na příbuzné zdroje“), tj. původní výraz „souvislost s dalšími zdroji“ byl nahrazen přesnějším výrazem „odkaz na zdroj“.

14. Pokrytí (Coverage)

Definice tohoto speciálního údaje byla zobecněna („rozsah nebo záběr zdroje“), rovněž tak komentář obsahuje přesnější specifikaci a příklady.

15. Práva (Rights)

Základní definice tohoto posledního údaje byla také upřesněna („informace o právech udržovaná ve zdroji nebo mimo něj“).

Pracovní návrh nové verze specifikace Dublinského jádra (verze 1.1 z 2. 7. 1999) byl po krátké diskusi dne 9. září 1999 beze změn přijat Poradním výborem DCMI jako definitivní **doporučení**. Plné znění tohoto dokumentu je k dispozici na adrese: <http://purl.org/dc/documents/rec-dces-19990702.htm>. Dá se očekávat, že na říjnovém 7. pracovním semináři DC bude, podle předběžných zpráv, předložena již pravděpodobně verze vyšší 2.0. Ta by měla ale již být v korespondenci s nově navrženým datovým modelem DC, který byl připraven v souladu s Rámcem pro popis zdrojů (RDF) a jazykem XML. Této podstatné otázce je věnována další část textu.

4 Dublinské jádro a Rámec pro popis zdrojů (RDF)

Specifikace sémantiky Dublinského jádra je jednou z mnoha, které v 90. letech vznikly a i nadále se rozvíjejí. Jednou z největších překážek, které stojí před systémy a službami, jež se zabývají popisem a zpracováním digitál-

ních zdrojů, je multiplicita navzájem nekompatibilních norem či směrnic pro syntax metadat i jazyky definic schémat. Je tak prakticky znemožněno vzájemné využívání metadat v různých aplikacích v celosvětovém měřítku. Situaci by měl v blízké budoucnosti radikálně řešit tzv. **„Rámec pro popis zdrojů“** (Resource Description Framework, dále také jako RDF), který je dílem společného úsilí řady odborníků z různých organizací celého světa. Práce na RDF jsou vedeny a koordinovány konsorciem W3C a jeho výsledky jsou pro veřejnost k dispozici na hostitelském serveru Technologického institutu v Massachusetts, USA <http://www.w3.org/RDF/>.

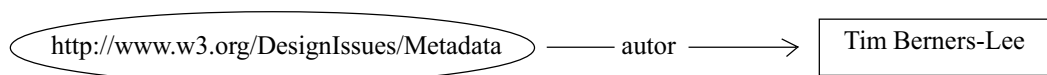
RDF je založen na webovské technologii a je navržen jako aplikace nového jazyka XML (**eXtensible Markup Language**) [31], jenž je derivátem základního značkovacího jazyka SGML. Jazyk XML je považován za nástupce jazyka HTML, který je používán v současnosti jako hlavní formát webovských dokumentů. Překonává řadu jeho limitů a zřejmě se v budoucnosti stane i hlavním přenosovým formátem mezi různými aplikacemi. XML přichází s řadou vynikajících vlastností (např. autoři mohou definovat vlastní tagy), textové dokumenty v tomto formátu budou snadno čitelné, srozumitelné, zpracovatelné a konvertibilní. Zobrazování dat bude realizováno pouze přes styly. Podstatné změny jsou připraveny v oblasti propojování XML dokumentů pomocí speciálních jazyků „XLink“ a „XPointer“, jež umožní nejen tvorbu odkazů na části stránek, které nejsou předem označeny, nebo obousměrné odkazy či odkazy na několik zdrojů najednou, ale i odkazy uložené mimo samotný dokument.

RDF jako aplikace XML má definovanou svoji vlastní standardní DTD (Document Type Definition). Je významnou obecnou **specifikací modelu a syntaxe** [34] a **specifikací schématu** [35] metadat jakéhokoliv zaměření a charakteru. Jeho cílem je zajištění vzájemné součinnosti (interoperability) jednotlivých implementací metadat na mezinárodní úrovni, která zahrnuje jak popis webovských digitálních zdrojů a tvorbu metadat, tak jejich výměnu a zejména užití realizované v první fázi pomocí inteligentních programů-agentů. Jednou z oblastí, která by mohla bohatě využívat metadat k plnění svých úkolů, zajisté budou i knihovny včetně knihoven národních. Bude to však i řada dalších specializovaných institucí, řídicích orgánů státních správ apod. Konečným cílem RDF je zcela automatická kontrola a komunikace webovských zdrojů v celosvětovém měřítku, která je předpokladem k budování webovských informačních systémů a služeb vyššího řádu.

RDF nezahrnuje specifikaci sémantik pro popis zdrojů jednotlivých metadat. Jeho základní model je založen na koncepci **popisu zdrojů prostřednictvím souboru jejich vlastností zvaného „popis RDF“**. Základní model RDF zahrnuje následující tři typy objektů [34, část 2.1]:

1. **zdroje** (webová stránka, část stránky, kolekce stránek nebo celé webovské sídlo) identifikované jednoznačným identifikátorem zdroje URI (URL, URN aj.); zdrojem podle RDF ale může být i tradiční dokument.

Obrázek 5



2. **vlastnosti zdrojů** (specifický aspekt, atribut nebo vztah užívaný k jejich popisu), které mají specifické významy a v rámci kterých jsou definovány jejich přípustné hodnoty i typy zdrojů, jež jsou popisovány.
3. **RDF údaje (výroky)**, které jsou tvořeny třemi komponentami (srovnej také s úvodním výkladem v 1. části textu):
 - **Zdrojem** (ve specifikaci syntaxe RDF označovaný jako subjekt výroku)
 - **Pojmenovanou vlastností** (označovaná jako predikát výroku)
 - **Hodnotou pojmenované vlastnosti** (označovaná jako objekt výroku); hodnotou může být jednoduchý řetězec znaků (viz obrázek 5), popř. jednoduchý údaj definovaný v rámci specifikace jazyka XML, nebo jí může být jiný zdroj (entita), který může být rovněž popsán svými vlastními vlastnostmi (viz obrázek 7).

Příklad:

Výrok v přirozeném jazyce:

*Tim Berners-Lee je autorem zdroje s adresou
http://www.w3.org/DesignIssues/Metadata*

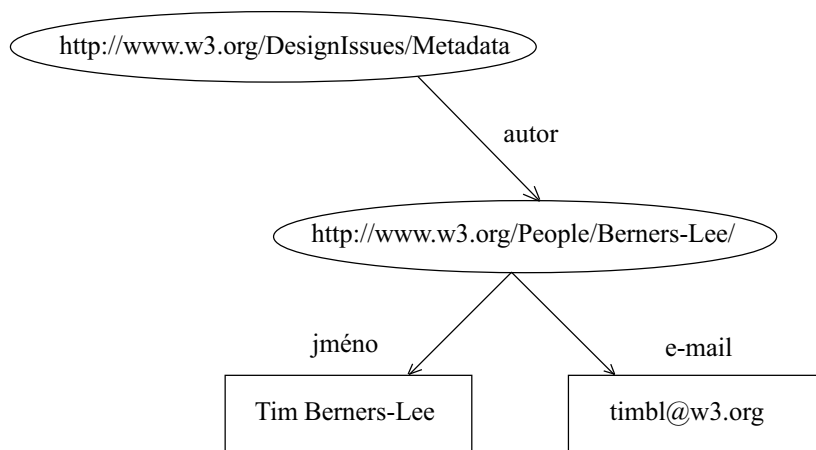
je možné vyjádřit pomocí jednoduchého diagramu tak, jak je uvedeno na obrázku 5 (zdroj je vždy uveden v elipse, šipka prezentující vlastnost musí směřovat ze zdroje na hodnotu pojmenované vlastnosti, jež je v tomto případě v rámečku).

Prezentace vztahu zobrazeného na obrázku 5 bude prostředky jazyka XML/RDF vyjádřena tak, jak je uvedeno na obrázku 6. Je-li hodnotou vlastnosti výchozího zdroje jiný webovský zdroj, bude údaj RDF vyjádřen pomocí diagramu tak, jak je uvedeno na obrázku 7. Takový zdroj může mít sám o sobě své vlastní vlastnosti (údaje) s příslušnými hodnotami.

Obrázek 6

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  <rdf:Description rdf:about="http://www.w3.org/DesignIssues/Metadata">
    <author> Tim Berners-Lee </author>
  </rdf:Description>
</rdf:RDF>
```

Obrázek 7



RDF je jednou z aplikací nově vytvářeného jazyka XML, a proto vychází z jeho modelu i syntaxe a nabízí také velmi významný prostředek k zajištění vzájemné součinnosti různých aplikací metadat, který se v anglické terminologii označuje výrazem „namespace“ (prostor jmen). **Prostor jmen v XML** je v základní dokumentaci definován jako kolekce jmen identifikovaných URI, jež jsou užívána v XML dokumentech jako typy prvků a jména vlastností [33, část 1]. Jednotlivým systémům se umožňuje, aby při popisu zdrojů deklarovaly své vlastní způsoby vyjadřování popisu zdrojů. Při popisu zdroje v jednom systému je možné využít některé prvky jiného systému. Prostory jmen představují vlastně specifické webovské dokumenty obsahující definice specifikací metadat (syntaxe i sémantiky), které zajišťují kontext jakéhokoliv prvku použitého v popisu zdroje pomocí odkazu na tyto dokumenty.

Konkrétní mechanismus uplatnění prostorů jmen znamená, že prvky (údaje) popisu RDF, které nejsou součástí základního jazyka XML, jsou opatřeny prefixem identifi-

kujícím příslušný prostor jmen jejich původu. Např. Dublinské jádro má podle výše citované směrnice navržený prefix „dc:“. K zajištění odkazu mezi prefixem „dc:“ a příslušnou definicí Dublinského jádra, který by byl srozumitelný příslušnému programu, je v rámci popisu RDF uvedena deklarace s užitými prostory jmen, například:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-
  rdf-syntax-ns#">
```

V příkladu je v rámci deklarace RDF uveden prostor jmen vlastní syntaxe RDF. Jeho formalizovaný zápis obsahuje kvalifikované jméno prostoru jmen „xmlns:rdf“ a příslušné URI dokumentu se specifikací syntaxe. Znak „#“ na konci URI je důležitý, je používán pro kombinaci jména prostoru jmen s lokálním jménem, aby bylo možné získat úplné URI jednotlivých typů vlastností (například <http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag>). Na obrázku 8 jsou v rámci RDF deklarace uvedeny také dva

Obrázek 8

```
<?xml version="1.0" encoding="windows-1250"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.0/"
  xmlns:dcq="http://purl.org/dc/qualifiers/1.0/">
  <rdf:Description rdf:about="http://www.cuni.cz/ffuisk/">
    <dc:title>
      <rdf:Alt>
        <rdf:li xml:lang="en">Institute of Information Studies and
          Librarianship (IISL) Home Page </rdf:li>
        <rdf:li xml:lang="cs">Ústav informačních studií a knihovnictví (ÚISK)
          [Domovská stránka]</rdf:li>
      </rdf:Alt>
    </dc:title>
    <dc:creator> Ústav informačních studií a knihovnictví </dc:creator>
    <dc:description>
      <rdf:Alt>
        <rdf:li xml:lang="en">Main Home Page of the Institute of
          Information Studies and Librarianship Web site. </rdf:li>
        <rdf:li xml:lang="cs">Hlavní domovská stránka webovského sídla
          Ústavu informačních studií a knihovnictví.</rdf:li>
      </rdf:Alt>
    </dc:description>
    <dc:subject> Informační věda; Knihovnictví; Vysokoškolské
      vzdělávání </dc:subject>
    <dc:publisher> Ústav informačních studií a knihovnictví </dc:publisher>
    <dc:contributor>
      <rdf:Bag>
        <rdf:li>Saša Skenderija</rdf:li>
        <rdf:li>Eva Bratková</rdf:li>
      </rdf:Bag>
    </dc:contributor>
    <dc:date>
      <rdf:Description>
        <dcq:dateScheme> WTN8601 </dcq:dateScheme>
        <rdf:value> 1995-01-10 </rdf:value>
      </rdf:Description>
    </dc:date>
    <dc:format>
      <rdf:Description>
        <dcq:formatScheme> IMT </dcq:formatScheme>
        <rdf:value> text/html </rdf:value>
      </rdf:Description>
    </dc:format>
    <dc:language>
      <rdf:Description>
        <dcq:languageScheme> RFC1766 </dcq:languageScheme>
        <rdf:value> cs </rdf:value>
      </rdf:Description>
    </dc:language>
  </rdf:Description>
</rdf:RDF>
```

prostory jmen jednoduchého i komplexního DC. Pokud se tedy v záznamu objevuje např. údaj o tvůrci dokumentu zapsaný ve formě <dc:creator>, pak příslušný program (parser) pro potřeby jeho dalšího zpracování načte a případně zkontroluje jeho plnou formu (<http://purl.org/dc/elements/1.0/creator>).

Výše uvedený text této části je jen velmi stručným uvedením do komplexu celé specifikace RDF. Z důvodu omezeného rozsahu článku nebude tento velmi významný materiál charakterizován podrobněji, zájemce lze odkázat zatím na řadu původních zdrojů [31, 33, 34, 35 aj.]. Ve zbylém prostoru textu proto bude dále následovat pouze komentář k nejdůležitějším principům a prvkům specifikace RDF realizované v prostředí jazyka XML, a to v návaznosti na příklad souboru metadat na obrázku 8, jenž byl připraven podle nově **navrhované pracovní směrnice aplikace RDF pro Dublinské jádro** Pracovní skupinou pro datový model DCMI [14]. Citovaný dokument, který byl zveřejněn v červenci tohoto roku pro potřeby připomínkování, je velmi významným dokumentem, který prezentuje výhodné spojení pět let rozvíjené sémantiky Dublinského jádra a syntaxe RDF. Materiál má povahu technické zprávy, která je jednak prostředkem k prověření aplikace modelu DC v rámci syntaxe RDF, jednak návrhem mechanismu k vyjádření jednoduchého i komplexního Dublinského jádra prostředky jazyka XML. Využití jazyka XML v citovaném materiálu ale neznamená, že by model DC musel být vyjádřen jenom prostřednictvím něho samotného.

Na obrázku 8 je soubor metadat reálné domovské stránky (ÚISK FF UK) s URL: <<http://www.cuni.cz/ffuisk/>>. Soubor obsahuje údaje specifikované jak základní sestavou Dublinského jádra (dle verze 1.0), tak provizorní rozšířené sestavy s kvalifikátory (rovněž dle verze 1.0). Obě specifikace jsou v souladu s RDF korektně odkazovány prostřednictvím adresy URL jako prostory jmen vlastností DC (viz 4. a 5. řádek (hodnoty dc a dcq) v záznamu na obrázku 8 s návěštím „xmllns“). Základní definice syntaxe RDF (hodnota rdf) je podobným způsobem odkazována ve 3. řádku. Uvedený soubor metadat DC ve struktuře RDF/XML, který je připraven podle základní (serializační) syntaxe [34, část 2.2.1], není součástí reálné webovské stránky, a to z toho důvodu, že by prozatím nebyl korektně zpracován pomocí některých současných prohlížečů. Je proto dostupný odděleně jako sálový textový XML soubor s adresou URL: <<http://www.cuni.cz/ffuisk/rdf.xml>>. Dodejme, že v současné chvíli je podle citované směrnice [14, část 2.3] možné „vnořit“ záznam s metadaty DC ve struktuře RDF také do stávajících HTML dokumentů - hlavíčky <HEAD>, ovšem musí být užito pouze tzv. „zkrácené syntaxe RDF“ [34, část 2.2.2], kdy jsou všechny údaje v podstatě uvedeny za sebou v rámci jediného tagu/ značky <rdf:Description>.

K syntaxi záznamu z obrázku 8 uvedme ještě následující základní vysvětlení:

- Kontrolu správnosti syntaxe ukázkového záznamu je možné prověřit pomocí speciálního jednoduchého

programu (parseru a kompilátoru) „SiRPAC“, který je k dispozici na WWW v rámci produktů konsorcia W3C [29]. Záznam je po načtení a kontrole rozložen do logicky uspořádaných trojic (anglicky „triple“) reprezentujících jednotlivé výroky/údaje o dokumentu v pořadí: 1. Pojmenovaná vlastnost zdroje, 2. Popisovaný zdroj a 3. Hodnota pojmenované vlastnosti. Příklad rozkladu jednoho z „jednoduchých“ údajů <dc:publisher> Ústav informačních studií a knihovnictví </dc:publisher> je vidět na obrázku 9.

Obrázek 9

```
triple ('http://purl.org/dc/elements/1.0/publisher',  
       'http://www.cuni.cz/ffuisk/',  
       'Ústav informačních studií a knihovnictví').
```

- V první řádce záznamu je uvedena povinná deklarace verze jazyka XML (verze 1.0) včetně užitého kódování. Na druhé řádce je taktéž povinná deklarace užití struktury RDF v značce <rdf>, která musí mít na konci párovou značku </rdf> (v rámci striktních pravidel XML musejí být všechny značky párové).
- Jména údajů Dublinského jádra jsou uvedena malými písmeny (stanoveno jako optimální dle požadavků modelu) s návěštím dc: (např. dc:title, dc:description apod.).
- Údaje o názvu dokumentu <dc:title> a anotace <dc:description> jsou ve dvou variantách (anglicky a česky), a proto je v rámci opakovatelných hodnot <rdf:li>, které jsou uvedené ve skupině alternativních údajů <rdf:Alt>, využito přímo vlastnosti jazyka XML - je uvedena deklarace užitého jazyka v kódovaném tvaru s hodnotami podle mezinárodní normy ISO 639 (xml:lang="en", xml:lang="cs").
- Opakovatelné hodnoty přispěvatelů <dc:creator> jsou rovněž uvedeny pomocí prostředků jazyka RDF jako skupina <rdf:Bag> - v tomto případě neuspořádaných hodnot; pro uspořádané množiny je v RDF připravena značka <rdf:Seq>, tj. sekvence.
- Údaje o datu <dc:date>, formátu <dc:format> a jazyku dokumentu <dc:language> obsahují hodnoty podle zatím neschválených kvalifikátorů-schémat Dublinského jádra, které jsou stále diskutovány (viz informace v části 3.2 tohoto textu). Jejich potřeba je však stále více evidentní. Jméno zdrojové specifikace „dcq:“ (dcq:dateScheme, dcq:formatScheme a dcq:languageScheme) je zatím užíváno pouze pro testovací potřeby. Výše zmiňovaný materiál Pracovní skupiny pro datový model DC zahrnuje přílohu 2 [14, Appendix 2], která kvalifikátory zatím jen navrhuje.

Závěr

Několikaletá komunikace informací na WWW vstupuje v současné době do své další významné vývojové fáze. Na scénu vstupuje, kromě řady nových technologií, nový značkovací jazyk pro tvorbu webovských dokumentů XML, který bude sám o sobě znamenat velký zvrat v oblasti využívání informací v nich obsažených. Vyhledávání a využívání digitálních informací by mělo být v dalším

období zdokonaleno také díky specifickým strukturovaným údajům, které nesou označení „metadata“, jež by se měly stát součástí dokumentů nebo je v procesu komunikace na WWW doprovázet. Úsilí odborníků míří k vybudování dokonalejší struktury webu, v němž informace budou pomocí programů nejen čitelné, ale i srozumitelné. Tento příspěvek je stručným přehledem i diskusním materiálem k tematice metadat webovských dokumentů, sumarizuje úsilí, které bylo v této oblasti až doposud vykonáno v zahraničí. Oprávněně největší místo zaujal v přehledu rozbor formátu „Dublinské jádro“ a jeho implementace v syntaktické struktuře popisu RDF. Jde o významné mezinárodní aktivity, které v dohledné době najdou i větší uplatnění také v rámci budování domácích systémů. První kroky pro to jsou již učiněny (české webovské sídlo pro formát „Dublinské jádro“ se základní definicí z 2. 10. 1997 a dalšími informacemi bylo zřízeno na adrese <http://www.ics.muni.cz/dublin_core/DC-czech.html>), další budou jistě následovat.

Použité a citované informační zdroje

1. *Automatic RDF Metadata Generator* [online]. Designed and implem. by Ch. Jenkins. Wolverhampton (UK) : Wolverhampton Univ., [cit. 1999-07-27]. Přístup z: <<http://www.scit.wlv.ac.uk/~ex1253/metadata.html>>
2. BAKER, Thomas. Languages for Dublin Core. *D-Lib Magazine* [online]. December 1998 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/december98/12baker.html>>
3. BEARMAN, D.; MILLER, E.; RUST, G.; TRANT, J.; WEIBEL, S. A Common Model to Support Interoperable Metadata : Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. *D-Lib Magazine* [online]. 1999, vol. 5, no. 1 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/january99/bearman/01bearman.html>>
4. BERNERS-LEE, Tim. *Metadata Architecture* [online]. W3C, 1997, last edited 1998-12-30 [cit. 1999-07-27]. Přístup z: <<http://www.w3.org/DesignIssues/Metadata.html>>
5. BERNERS-LEE, Tim. *Semantic Web Road map* [online]. W3C, 1998, last mod. 1998-10-14 [cit. 1999-07-27]. Přístup z: <<http://www.w3.org/DesignIssues/Semantic>>
6. BERNERS-LEE, Tim. *Why RDF model is different from the XML model* [online]. W3C, 1998, last mod. 1998-10-14 [cit. 1999-07-27]. Přístup z: <<http://www.w3.org/DesignIssues/RDF-XML.html>>
7. BRATKOVÁ, E. K otázkám pojmu, třídění a typologie internetových a webovských informačních zdrojů. *Národní knihovna : knihovnická revue*. 1998, roč. 9, č. 5, s. 262-276. Přístup také z: <http://www.nkp.cz/start/publikace/k_revue/5.htm>
8. BURNARD, L.; LIGHT, R. *Three SGML metadata formats : TEL, EAD, and CIMI : A Study for BIBLINK Work Package 1.1* [online]. Bath (UK) : UKOLN, December 1996, last updated 1998-05-14 [cit. 1999-07-27]. BIBLINK - LB 4034, Work Package D1.1. Přístup z: <<http://hosted.ukoln.ac.uk/biblink/wp1/sgml/>>
9. *d2m : Dublin Core to MARC converter* [online]. Ole Husby. Trondheim : BIBSYS, 1998-05-18 [cit. 1999-07-27]. Nordic Metadata Project. Přístup z: <<http://www.bibsys.no/meta/d2m/>>
10. *DC-dot : Dublin Core Generator* [online]. Maintained by Andy Powell. Bath : UKOLN, last updated 1999-05-10 [cit. 1999-07-27]. Přístup z: <<http://www.ukoln.ac.uk/metadata/dcdot/>>
11. DCMI. *Dublin Core Metadata Element Set : Reference Description* [online]. [Version 1.0]. Dublin : DCMI, 1996, last mod. 1997-10-02 [cit. 1999-07-27]. Přístup z: <http://purl.org/dc/about/element_set.htm>
12. DCMI. *Dublin Core Metadata Element Set Reference Description : Proposed Recommendation* [online]. Version 1.1. Dublin : DCMI, 1999-07-02 [cit. 1999-07-27]. Přístup z: <http://purl.org/dc/documents/proposed_recommendations/pr-dces-19990702.htm>
13. DCMI. *Dublin Core Metadata Initiative : Home Page* [online]. Dublin : DCMI, c1999 [cit. 1999-07-27]. Přístup z: <<http://purl.org/dc/>>
14. DCMI. *Guidance on expressing the Dublin Core within the Resource Description Framework (RDF) : Draft Proposal* [online]. Ed. E. Miller, P. Miller and Dan Brickley. Dublin : DCMI, 1999-07-01 [cit. 1999-07-27]. Přístup z: <<http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/WD-dc-rdf-19990701.html>>
15. DCMI. *Subelement Working Draft* [online]. Dublin : DCMI, 1998-02-11 [cit. 1999-07-27]. Přístup z: <http://purl.org/dc/documents/working_drafts/wd-subelements-current.htm>
16. DEMPSEY, L.; HEERY, R. aj. *Specification for resource description methods. Part 1, A review of metadata : a survey of current resource description formats* [online]. Bath (UK) : UKOLN, 1996-12-12 [cit. 1999-07-27]. DESIRE - RE 1004, D3.2 (1). Přístup z: <<http://www.ukoln.ac.uk/metadata/desire/overview/>>
17. DEMPSEY, L.; WEIBEL, S. The Warwick Metadata Workshop : A Framework for the Deployment of Resource Description. *D-Lib Magazine* [online]. July/August 1996 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/july96/07weibel.html>>
18. *Dublin Core Metadata Template* [online]. CGI-programming T. Koch and M. Borell; Javascript by M. Berggren. Lund : Lunds universitetsbibliotek, 1997-09-26, last updated 1998-03-17 [cit. 1999-07-27]. Nordic Metadata Project. Přístup z: <http://www.lub.lu.se/metadata/DC_creator.html>

19. *Functional Requirements for Bibliographic Records : Final Report*. IFLA Study Group on the Functional Requirements for Bibliographic Records. München : Saur, 1998. 136 s. Přístup také z: <<http://ifla.inist.fr/VII/s13/frbr/frbr.pdf>>
20. GRADMANN, Stefan. Cataloguing vs. Metadata : old wine in new bottles? In *64th IFLA General Conference, Amsterdam, Netherlands, August 16 - August 21, 1998* [online]. Vandoeuvre-les-Nancy, last mod. 1999-06-29 [cit. 1999-07-27]. Přístup z: <<http://ifla.inist.fr/IV/ifla64/007-126e.htm>>
21. HEERY, R. aj. *Metadata Formats* [online]. Bath (UK) : UKOLN, 1996-12-12 [cit. 1999-07-27]. BIBLINK - LB 4034, Work Package D1.1. Přístup z: <<http://hosted.ukoln.ac.uk/biblink/wp1/d1.1/>>
22. HOPKINSON, A. UNIMARC and Metadata : Dublin Core. In *64th IFLA General Conference, Amsterdam, Netherlands, August 16 - August 21, 1998* [online]. Vandoeuvre-les-Nancy, last mod. 1999-06-29 [cit. 1999-07-27]. Přístup z: <<http://ifla.inist.fr/IV/ifla64/138-161e.htm>>
23. CHAPMAN, A.; DAY, M.; HIOM, D. Metadata : Cataloguing practice and Internet subject-based information gateways. *Ariadne : The Web Version* [online]. December 1998, issue 18 [cit. 1999-07-27]. Přístup z: <<http://www.ariadne.ac.uk/issue18/metadata/>>
24. IANNELLA, R. *An Idiot's Guide to the Resource Description Framework* [online]. Brisbane : University of Queensland, DSTC, 1998-09-03, updated 1999-01-25 [cit. 1999-07-27]. Přístup z: <<http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/>>
25. LAGOZE, C.; LYNCH, Clifford A.; DANIEL, Ron, Jr. *The Warwick Framework : A Container Architecture for Aggregating Sets of Metadata* [online]. Ithaca (NY) : Cornell University, NCSTRL, 1996-06-12 [cit. 1999-07-27]. Přístup z: <<http://cs-tr.cs.cornell.edu/Dienst/Repository/2.0/Body/ncstrl.cornell/TR96-1593/html>>
26. MILLER, E. An Introduction to the Resource Description Framework. *D-Lib Magazine* [online]. May 1998 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>
27. *The Nordic Metadata project* [online]. Funded by NORDINFO. 1996, last updated 1998-03-05 [cit. 1999-07-27]. Přístup z: <<http://linnea.helsinki.fi/meta/>>
28. RFC 2413. *Dublin Core Metadata for Resource Discovery* [online]. S. Weibel, J. Kunze, C. Lagoze, M. Wolf. 1998-09-25 [cit. 1999-07-27]. Přístup z: <<http://www.ietf.org/rfc/rfc2413.txt>>
29. *SiRPAC - Simple RDF Parser & Compiler* [online]. W3C; Software Janne Saarela. Last updated 1999-05-21 [cit. 1999-07-27]. Přístup z: <<http://www.w3.org/RDF/Implementations/SiRPAC/>>
30. STRAKA, Josef. *Sociální informatika : terminologický a výkladový slovník pro posluchače katedry vědeckých informací a knihovnictví*. Praha : Karolinum, 1990. 217 s.
31. W3C. *Extensible Markup Language (XML) 1.0 : W3C Recommendation 10-February-1998* [online]. Editors Tim Bray, Jean Paoli, C. M. Sperberg-McQueen. Last mod. 1998-04-07 [cit. 1999-03-26]. REC-xml-19980210. Přístup z: <<http://www.w3.org/TR/1998/REC-xml-19980210>>
32. W3C. *Frequently Asked Questions about the Extensible Markup Language : The XML FAQ* [online]. Maintained by Peter Flynn. Version 1.5 (1 June 1999). Cork (IR) : University College Cork, last mod. 1999-07-19 [cit. 1999-07-27]. Přístup z: <<http://www.ucc.ie/xml/>>
33. W3C. *Namespaces in XML* [online]. Editors Tim Bray, D. Hollander, A. Layman. Last mod. 1998-01-14 [cit. 1999-07-27]. REC-xml-names-19990114. Přístup z: <<http://www.w3.org/TR/1999/REC-xml-names-19990114/>>
34. W3C. *Resource Description Framework (RDF) Model and Syntax Specification : W3C Recommendation 22 February 1999* [online]. Editors Ora Lassila, Ralph R. Swick. c1999, last updated 1999-02-24 [cit. 1999-07-27]. REC-rdf-syntax-19990222. Přístup z: <<http://www.w3.org/TR/REC-rdf-syntax/>>
35. W3C. *Resource Description Framework (RDF) Schema Specification : W3C Proposed Recommendation 03 March 1999* [online]. Editors Dan Brickley, R.V. Guha. c1999, last updated 1999-03-04 [cit. 1999-07-27]. Přístup z: <<http://www.w3.org/TR/PR-rdf-schema/>>
36. WEIBEL, S. The State of the Dublin Core Metadata Initiative : April 1999. *D-Lib Magazine* [online]. 1999, vol. 5, no. 4 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/april99/04weibel.html>>
37. WEIBEL, S. aj. *OCLC/NCSA Metadata Workshop Report* [online]. Dublin : DCMI, 1995-12-05 [cit. 1999-07-27]. Přístup z: <<http://purl.org/dc/workshops/dc1conference/index.htm>>
38. WEIBEL, S.; HAKALA, J. DC-5 : The Helsinki Metadata Workshop : A Report on the Workshop and Subsequent Developments. *D-Lib Magazine* [online]. February 1998 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/february98/02weibel.html>>
39. WEIBEL, S.; IANNELLA, R.; CATHRO, W. D. The 4th Dublin Core Metadata Workshop Report : DC-4, March 3 - 5, 1997, National Library of Australia, Canberra. *D-Lib Magazine* [online]. June 1997 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/june97/metadata/06weibel.html>>
40. WEIBEL, S.; MILLER, E. Image Description on the Internet : Summary of the CNI/OCLC Image Metadata Workshop, September 24 - 25, 1996, Dublin, Ohio. *D-Lib Magazine* [online]. January 1997 [cit. 1999-07-27]. Přístup z: <<http://www.dlib.org/dlib/january97/oclc/01weibel.html>>