



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost



Slezská univerzita v Opavě

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato inovace předmětu *Digitální knihovny* je spolufinancována Evropským sociálním fondem a Státním rozpočtem ČR, projekt č. CZ.1.07/2.3.00/09.0197, "Posílení konkurenceschopnosti výzkumu a vývoje informačních technologií v Moravskoslezském kraji".



Slezská univerzita v Opavě
Filozoficko-přírodovědecká fakulta v Opavě

Ústav informatiky
Oddělení informační vědy

Digitální knihovny

Studijní opora

PhDr. Jindra Planková, Ph.D.

jindra.plankova@fpf.slu.cz

Opava 2017

Předmluva

Tento text je určen studentům kurzu *Digitální knihovny* navazujícího magisterského studia oboru „Informační a knihovnická studia“ na Ústavu informatiky Slezské univerzity v Opavě. Předpokládá základní orientaci v oboru, znalosti technologického vývoje informačních a komunikačních technologií a praktické dovednosti při vyhledávání informací na Internetu.

Kurz Digitální knihovny obsahuje objasnění základních pojmů informačních služeb digitálních knihoven. Postihuje okolnosti a historii vzniku prvních digitálních knihoven v 90. letech 20. století. Příklady prvních digitálních knihoven, definici digitální knihovny, sektory, kterými je digitální knihovna tvořena, technické protokoly, standardy a systémy tvorby digitálních knihoven. Jednotlivé kapitoly jsou věnovány funkcím a úkolům digitálních knihoven, technickým aspektům, propojování. Stručně jsou nastíněny také oblast autorskoprávní ochrany budování digitálních knihoven a dále uživatelským a sociálním dopadům budování digitálních knihoven. Výrazně je specifikována oblast metadat a jejich užití v digitálních knihovnách. Součástí kurzu jsou také stručné prezentace českých a světových digitálních knihoven na internetu.

Pro oblast digitálních knihoven existuje v současné době mnoho materiálů, článků i konferenčních příspěvků, či ucelených textů, většinou v angličtině. Účelem tohoto textu není pokrýt celou problematiku budování a fungování digitálních knihoven v prostředí internetu. Mnohé oblasti jsou pouze stručně nastíněny, některé informace naznačeny a uvedeny do souvislostí. Pro hlubší seznámení s problematikou proto odkazují na použitou literaturu, která byla podkladem pro tvorbu daného textu. Z důvodu další propojitelnosti textu je výklad doprovázen odkazy na informační zdroje a vysvětlivkami pod čarou.

Vzhledem k rozsáhlosti problematiky, jejímu neustálému vývoji a odkazům na mnohé cizojazyčné zdroje je možné, že se v textu budou vyskytovat chyby. Proto přivítám upozornění na případné chyby a také připomínky k samotnému textu.

Jindra Planková

V Opavě, 31. 03. 2017

Obsah

ÚVOD DO PROBLEMATIKY	1
VSTUPNÍ ZAMYŠLENÍ	1
1. DEFINOVÁNÍ POJMU „DIGITÁLNÍ KNIHOVNA“	4
1.1. POHLED INFORMAČNÍCH PROFESIONÁLŮ	6
1.2. POHLED POČÍTAČOVÝCH SPECIALISTŮ	8
1.3. NEJPODSTATNĚJŠÍ DEFINICE	9
1.4. PRACOVNÍ VYMEZENÍ TERMÍNŮ	11
2. HISTORIE, ČINNOSTI A FUNKCE DIGITÁLNÍCH KNIHOVEN	15
2.1. HISTORIE VÝVOJE DK	17
2.1.1. <i>Důvody vzniku a podpory digitálních knihoven</i>	19
3. TYPOLOGIE DIGITÁLNÍCH OBJEKTŮ (ZDROJŮ)	21
3.1. TYPOLOGIE ELEKTRONICKÝCH ZPRÁV V RÁMCI INTERNETOVÉ NORMY MIME	24
3.2. TŘÍDĚNÍ A TYPOLOGIE INTERNETOVÝCH A WEBOVÝCH ZDROJŮ V KNIHOVNICKÝCH SYSTÉMECH	27
3.2.1. <i>USMARC v návaznosti na AACR2R</i>	28
3.2.2. <i>MARC 21 v návaznosti na AACR2R</i>	30
3.2.3. <i>UNIMARC v návaznosti na AACR2R a ISBD (ER)</i>	32
3.3. TYPOLOGIE INTERNETOVÝCH A WEBOVÝCH ZDROJŮ V INFORMAČNÍCH SYSTÉMECH INTERNETU A WWW	35
3.3.1. <i>Kanadská typologie webových zdrojů „VW96“</i>	36
3.3.2. <i>Typologie britského systému ROADS</i>	38
3.3.3. <i>Typologie formátu „Dublin Core“</i>	43
3.4. ZÁVĚR TYPOLOGIÍ	50
4. ARCHITEKTURA DIGITÁLNÍ KNIHOVNY	53
4.1. IDENTIFIKÁTORY A SYSTÉMY PRO JEJICH SPRÁVU A ŘÍZENÍ	54
4.1.1. <i>Identifikátor URN</i>	57
4.2. DIGITÁLNÍ OBJEKT	58
4.3. SKUPINA (AGREGÁT) DIGITÁLNÍCH OBJEKTŮ	59
4.4. DIGITÁLNÍ SKLADIŠTĚ A REPOZITÁŘE	59
4.5. IDENTIFIKÁTORY INFORMAČNÍCH ENTIT	60
4.5.1. <i>Identifikátory tvůrců děl</i>	60
4.5.2. <i>Identifikátory tvůrčích děl</i>	60
4.5.3. <i>Identifikátory vyjádření děl</i>	60
4.5.4. <i>Identifikátory zhmotnění děl</i>	60
5. METADATA	61
5.1. METADATA PRO DIGITÁLNÍ ZDROJE V ZAHRANIČNÍCH SYSTÉMECH A SLUŽBÁCH	67
5.1.1. <i>Formáty TEI, EAD a CIMI</i>	67
5.1.2. <i>Formát GILS</i>	69
5.1.3. <i>Formuláře IAFA / ROADS</i>	70
5.1.4. <i>Internetová norma RFC 1807</i>	70
5.2. FORMÁT DUBLINSKÉ JÁDRO (DC)	71
5.2.1. <i>Vznik formátu DC</i>	71
5.2.2. <i>Další rozvoj DC</i>	74
5.2.3. <i>Aktuální stav formátu DC</i>	82
5.3. DUBLINSKÉ JÁDRO A RÁMEC PRO POPIS ZDROJŮ (RDF)	86
5.4. SHRNUTÍ	91
6. INTEROPERABILITA	93
6.1. ÚVOD DO PROBLEMATIKY A STRUČNÝ PŘEHLED	93
6.2. PROTOKOL Z39.50	96

6.3.	OPEN ARCHIVES INITIATIVE (OAI)	99
6.4.	STANFORDSKÝ INFOBUS	100
6.5.	OPEN URL A SFX.....	102
7.	GLOBÁLNÍ VYHLEDÁVÁNÍ ZDROJŮ	106
7.1.	ÚVODNÍ POZNÁMKY.....	106
7.2.	DK A INTERNETOVÉ VYHLEDÁVACÍ SYSTÉMY	108
8.	SOUVISEJÍCÍ OBLASTI S PROBLEMATIKOU DK.....	111
8.1.	COPYRIGHT, INTELEKTUÁLNÍ VLASTNICTVÍ.....	111
8.2.	DLOUHODOBÉ UCHOVÁVÁNÍ DIGITÁLNÍCH INFORMACÍ	112
9.	PROGRAMY A PROJEKTY DK.....	113
9.1.	DIGITAL LIBRARY INITIATIVE – PHASE 1	113
9.2.	DIGITAL LIBRARY INITIATIVE – PHASE 2	114
9.3.	ELECTRONIC LIBRARY PROGRAMME (ELIB).....	115
9.4.	NATIONAL DIGITAL LIBRARY PROGRAM (NDLP)	116
9.5.	DALŠÍ PROJEKTY	116
10.	ZÁVĚREM	118
	POUŽITÁ LITERATURA:	119

Úvod do problematiky

Na sklonku 20. století byly tématem dne samotné prognózy a plány pro přechod do třetího tisíciletí, jejichž dominantou byla především vize **informační společnosti**, jako dalšího stupně vývoje lidstva. Byly vypracovávány mnohé koncepce **informatizace**¹ společnosti, významné projekty a programy na zvládnutí a vytvoření **globální informační infrastruktury**². Někteří vizionáři viděli novou informační společnost jako „domnělý univerzální prostředek“³ na vyřešení světových problémů a na vybudování lepšího světa. Zatímco jiní vyjadřovali obavy z nebezpečí, které mohou s sebou přinést nové technologie, a z jejich negativního vlivu na jedince i celou lidskou společnost. Dnes je zřejmé a domnívám se, že i oprávněně tvrdit, že informační technologie jsou pouze nástrojem, kterého lze použít k dobrým či špatným účelům. Přitom však je nutno si uvědomit, že tzv. „informační revoluce“ si žádá vážnou diskusi především o schopnostech a možnostech moderního člověka vnímat, utřídit, vyhodnotit a zapracovat do svého intelektuálního či poznatkového fondu nepřeborné informační bohatství, které nám mnohá komunikační a informační média nabízí. Z tohoto úhlu pohledu se potom jeví jako samozřejmé úvahy nad budoucností, novým posláním a funkcí oblastí spojených s knihovnicko-informační profesí a snahy o jejich formulaci.

Vstupní zamyšlení

Není pochyb o tom, že se nacházíme v době převratných technologických a společenských změn. Vývoj **digitálních a informačních technologií** určených k vytváření, zpracování, šíření a užívání informací, závažně přispěl k formování nové informační společnosti. Snad jednou z nejvýznamnějších charakteristik současné informační společnosti je exponenciální nárůst nových informací, dokonce nových vědeckých poznatků v digitální podobě a jejich zpřístupnění komunikačními a informačními technologiemi, často v reálném čase, bez ohledu na místo jejich výskytu. Průnik nových technologií tedy ovlivňuje rychlý přenos digitálních informací do všech sfér lidské společnosti. Vzhledem k tomu je třeba si uvědomit, že příliš mnoho informací, podobně jako jejich nedostatek může škodit. Typickým

¹ např. Informační koncepce EU; Státní informační a komunikační politika ČR, atd.

² ang. Global Information Infrastructure – viz. Brown, D.H.; Dostupné z WWW:

<http://www.itu.int/newsarchive/press_releases/1997/np-08-fr.html>

³ jakousi „panaceu“

příkladem selhání způsobeného přemírou informací, jsou celkem nedávné americké události z 11. září 2001, kdy obranné složky a vláda v USA nedokázaly ničivé útoky předpovědět ne z nedostatku informací, ale protože jich bylo příliš mnoho. Vládní představitelé nedokázali zvládnout velké množství varovných signálů a podnětů, jejich správné vyhodnocení a interpretování⁴.

Role, kterou zastávají informační technologie v každodenním životě lidské společnosti, tedy způsobuje, že je stále větší pozornost věnována otázkám mezi přijatelného využívání počítačů a počítačového zpracování informací. Při běžné práci s informačními technologiemi dochází vlastně pouze k určité manipulaci s bity⁵. Přitom veškeré dopady těchto manipulací zprostředkovaně závisí na interpretaci zpracovávaných posloupností bitů, jako dat, informací, či jednotlivých akcí s nimi. Tato zkušenost je přirozenému lidskému souhrnu znalostí do jisté míry vzdálená. Lehkost provádění některých operací na úrovni práce s počítačovými daty potom způsobuje, že člověk snáze ztrácí zábrany k některým druhům chování. Mnohem snadněji podléhá negativním vlivům, které způsobují jeho překračování bariér v „běžných“ situacích normálního života. Jestliže se ale informace v dnešní společnosti stávají ekvivalentem materiálních statků, musí také nakládání s nimi podléhat přirozeným omezením, morálním a etickým zásadám a je také nutné stanovit právní meze, které jsou běžné i pro jiné oblasti lidského konání.

Z tohoto pohledu pak hesla „informace znamenají moc“ a „kdo má informace, má moc“ znějí už téměř jako fráze, přitom je však jasné, že přístup k informacím často umožňuje ovládat myšlení jiných lidí. Bez účinného **pochopení souvislostí** nových trendů v knihovnictví a jisté úrovně profesionálního vzdělávání, které je příhodně definováno jako „umění společenského přežití“⁶, mohou být informace zneužity k manipulaci s lidmi, dokonce proti jejich vůli a často v protikladu k jejich zdravému úsudku. Právě proto je v dnešní době velmi důležité mít stále na paměti, že bez stanovení jisté úrovně etických a morálních norem; výchozího konceptu učení a vzdělávání, rozvíjejícího náš talent a tvořivý potenciál, bychom se mohli ocitnout na prahu nové generace, která nebude schopna dostát naplnění základního úkolu informační společnosti, tj. zprostředkování dosavadních lidských zkušeností a nashromážděných výsledků poznání v síťovém prostředí. V této souvislosti je

⁴ Sinai, J., nestr.

⁵ ang. binary digit – jednotka elementární informace

⁵ Bernier, Ch., S. 211

⁶ Bernier, Ch., S. 211

nutné ještě poznamenat, že i když je každá nová generace jistým způsobem změnou vitální senzibility generace předchozí, vždy jsou v ní přítomny její rysy nejcennější i nejvšednější a jejím úkolem je přijetí toho, co žila generace předešlá (názory, ideje, hodnoty, atd.) a nechat plynout svou vlastní živelnost⁷.

Myslím si, že blízká budoucnost se rodí z nás a spočívá především v prodloužení toho, co je podstatné a ne nahodilé; normální a ne nejisté. Proto chceme-li zachytit záblesk jisté budoucnosti nových principů zpracování a zprostředkování znalostí v informační společnosti, musíme vnímat stále se rozvíjející informační sektor⁸ a jeho obecné rysy. Svým způsobem se totiž v informačním sektoru pohybuje každý z nás; každý se účastníme sběru, správy a přenosu informací. Problémy komunikace informací a jejich osvojování si v rámci vzdělávacího procesu jsou proto přirozenou součástí dilemat dnešního běžného života. Informační teorie⁹ a praxe je vlastně plná různých problémů a sporných otázek, které se týkají informací a jejich komunikování. Ty jsou aktuální nejen v přítomnosti, ale s některými se informační společnost bude muset potýkat a vyrovnat i budoucnosti.

Domnívám se, že teprve zajištění volného a svobodného přístupu všech lidí na celém světě k rozsáhlému souboru informací a potažmo svobodný přístup k vzdělávání (tj. informacím, znalostem a dovednostem) je nepostradatelným nástrojem a zárukou dosažení ideálů demokracie, svobody a sociální spravedlnosti. Přitom nelze chápat vzdělávání jako zázračný lék či kouzelný klíč otevírající dveře do světa, ve kterém bude dosaženo všech ideálů, ale jako jeden ze základních prostředků k podpoře hlubší a vyváženější formy rozvoje lidstva a tedy i k potlačení chudoby, segregace jedince, nevědomosti, útisku a války.

⁷ Ortega y Gasset, J.

⁸ dnes snad nejrychleji rostoucí profesní obor

⁹ především Informační věda

1. Definování pojmu „digitální knihovna“

Svět se mění rychleji, než si myslíme, nevyvíjí se jenom technika a ekonomika, ale i naše postoje a hodnoty, způsoby chápání, poznávání a vzdělávání. Po padesáti letech prakticky existuje již nová společnost, nové prostředí. A lidé, kteří se v tomto světě narodili, si těžko dovedou představit svět minulý, ve kterém žili jejich prarodiče a do kterého se narodili jejich rodiče. Světem v němž se nacházíme, už nehýbou jen tradiční ekonomické faktory – práce, kapitál, zdroje, ale také poznatky a znalosti čili *informace*. V období vědeckotechnického rozvoje představují významnou strategickou surovinu, která urychluje vědecký a technologický pokrok, hospodářský rozvoj a také informovanost a vzdělanost celé společnosti. Na rozdíl od předcházejících rozhodujících surovin jsou informace zdrojem, který se používáním nevyčerpává. Narůstají do objemu i kvality, v posloupnosti **data-informace-znalost-poznání-moudrost**. Současná doba se potom snaží řešit problémy spojené s racionalizací získávání, zpracování a poskytování adekvátních informací. Dimenze tvorby, zpracování a zprostředkování informací jsou těžištěm nového, tzv. informačního sektoru ekonomiky – *informačního průmyslu*¹⁰. Další nakládání se zprostředkovanými daty, fakty a informacemi je potom spojeno s **oblastí tvorby vhodných nástrojů** pro jejich přímé zpřístupnění uživateli. Při podrobném prozkoumání nabízených možností a současných trendů se budeme pohybovat v intencích pojmu **digitální – virtuální – elektronická knihovna**. V odborné literatuře potom mají pojmy stejný význam a především anglicky psaná literatura striktně používá pojmu digitální knihovna.

Termín **digitální knihovna**¹¹ zahrnuje širokou myšlenkovou základnu, a přestože patří v poslední době k velmi frekventovaným pojmům, panuje řada nejasností, co vlastně tento termín obnáší. Jednou z možných příčin daného stavu je fakt, že obsah pojmu digitální knihovna se průběžně vyvíjí a to v souvislosti s jeho technologickou základnou¹². Jiným důvodem daného stavu je fakt, že problematikou digitálních knihoven se zabývá mnoho různorodých odborných komunit, z nichž každá vytváří vlastní obsahové chápání pojmu, v souladu s vlastním zaměřením. Tak např. *knihovníci* chápou digitální knihovnu jako další krok v *automatizaci* na cestě od knihovny s tištěnými dokumenty (analogová) přes

¹⁰ Japonec Joneji Masuda sem zahrnuje také výzkum, vývoj, vědu, vzdělávání, umění a etiku – Naisbitt, J.; Aburdenová, P., S. 11-52.

¹¹ ang. digital library

¹² ICT – výpočetní a telekomunikační technikou

automatizovanou či hybridní (fyzické sbírky doplněné automatizovaným katalogem) až po digitální (zahrnující většinu či veškerou nabídku informační zdrojů a služeb v elektronické podobě); z pohledu *informatika* (databázového specialisty) je digitální knihovna synonymem pro *informační systém* využívající architekturu federativních databází; *odborníci* zabývající se *hypertextem* a *šířením informací* ji vnímají jako jednu z nadstavbových *aplikací webu*. V neposlední řadě přispívá ke zmatení pojmů fakt, že pojmem digitální knihovna jsou někdy označovány systémy, které (při nejmenším z pohledu informačního specialisty) představují úplně jiné entity (např. soubory algoritmů a procedur, systémy na správu dokumentů, apod.).

V souvislosti s masovým rozšířením internetu a po nástupu webových technologií se ještě objevily představy, že celý prostor *internetu*, resp. *dimenze webu* jsou vlastně jednou digitální knihovnou. Vůči tomuto tvrzení se ovšem postavili odborníci¹³ z oblasti informační vědy¹⁴, kteří připomínají, že služba WWW nebyla vůbec navržena pro podporu organizovaného publikování a vyhledávání informací. Výstižně charakterizoval a shrnul příslušnou problematiku Carl Lagoze¹⁵, který ve svém příspěvku říká: „*Ačkoli internet poskytuje přístup k nesmírnému množství informací, současný neaktuálnější stav dalece postrádá, to co je běžně chápáno jako služba knihovny – to je, relativně jednoduché navigování a přístup k souboru dokumentů, které jsou nedílnou součástí sbírky (fondu knihovny). Idea sbírky (kolekce dokumentů) je důležitá v tom, že naznačuje, že soubor dokumentů nebyl vybrán náhodně, ale spolehlivým zprostředkovatelem (mediátorem/knihovníkem/informačním specialistou). Současní uživatelé internetu stojí před obrovským informačním prostorem, kde kvalita dokumentů je vzdálená od hodnověrnosti a možnosti pro objevení a nalezení dokumentů jsou jednoduché (v prvotním stádiu vývoje). Přístup ke konkrétnímu dokumentu obvykle znamená bloudění skrze „babylonskou věž“ specifické struktury závislostí a formátů souborů*“¹⁶.

Přestože od zveřejnění této myšlenky uplynulo již více než deset let a vývoj v oblastech spojených s webovými technologiemi (např. sémantický web) dosáhl pozoruhodných výsledků, domnívám se, že má výše uvedená charakteristika stále svou platnost.

¹³ např. Clyford Lynch – americký informační specialista

¹⁴ především informační specialisté a knihovníci

¹⁵ významný americký odborník v oblasti DL (Cornell University)

¹⁶ Lagoze, C., nestr.

Co se tedy skrývá pod oním označením **digitální knihovna**? V současné době se na Internetu vyskytuje celá řada definicí pojmu digitální knihovna (jen zběžným rešeršováním ve volně přístupných zdrojích povrchového webu se mi podařilo nashromáždit cca 1600 různých definicí pojmu „digital library“). Z uvedeného spektra se budeme zabývat pouze těmi definicemi, které odrážejí základní principy tvorby, struktury a fungování digitální knihovny, a to tak, jak je chápou odborníci z oblasti informační vědy a počítačové specialisté.

1.1. Pohled informačních profesionálů

Vytvoření a aplikování jednotné definice termínu digitální knihovna je velmi obtížné také pro odborníky z oblasti informační vědy a knihovnictví. Většina dnes publikovaných terminologických slovníků nemůže poskytnout (ani v ČR ani v zahraničí) vhodnou jednovětou (jeden výrok) definici. Částečně je tato neurčitost způsobena tím, že pojem „digitální knihovna“ má velmi odlišný význam pro různé odborníky, což souvisí s jejich vnímáním vlastního pojetí digitální knihovny. Například pro informačního pracovníka je DK síťovým nebo distribuovaným informačním systémem a naopak knihovník definuje DK jako doplněk fyzické knihovny. Aktuálně ovšem pojem DK zahrnuje oba z uvedených pohledů a přidává další charakteristiky, které dokreslují situaci DK z hlediska knihovnicko-informačního, technického, technologického, koncepčního, atd. Nyní se tedy můžeme podívat na možnosti definování pojmu DK, které se nám nabízí:

- ♥ běžná definice DK říká, „že se jedná o řízenou sbírku informací, spojovanou se službami, kde informace jsou uloženy v digitálních formátech a jsou přístupné pomocí sítě“¹⁷. Rozhodující částí této užívané definice je to, že informace jsou řízeny. Zároveň je nutné chápat fakt, že ohromné množství dat posílaných přes satelitní či jiné komunikační zařízení není knihovna. Ovšem stejná data, když jsou systematicky organizována, se stávají sbírkou digitální knihovny. Je zřejmé, že většina lidí nepovažuje databázi obsahující finanční záznamy jedné společnosti za digitální knihovnu, ovšem uznává, že sbírka takovýchto informací z více společností je její součástí. Důležité tedy je, že digitální knihovny obsahují odlišné informace, které jsou určeny k použití mnoha různým uživatelům. Digitální knihovny z tohoto pohledu pak

¹⁷ Whaley, T., nestr.

zahrnují informační rozsah od maličností k rozsáhlým znalostem. Pro práci mohou využívat různé typy počítačového vybavení a odpovídající software. Jednotící myšlenkou této definice je, že chápe informace, jako základ digitální knihovny, který musí být počítačově uspořádán a zpřístupněn pomocí sítě, s dodržáním postupu – výběr materiálu (informací/dokumentů) pro sbírku, uspořádání, zpřístupnění uživatelům a jejich uložení a archivování.

- ♥ DK je *„sbírka sbírek elektronických zdrojů elektronických znalostí vyvíjená a udržovaná za účelem uspokojení souhrnu informačních potřeb určité uživatelské populace“*¹⁸.
- ♥ DK je *„organizovanou sbírkou vybraných digitálních zdrojů, vytvářená k podpoře vědy, výzkumu a výuky“*¹⁹. Díky využití vhodných technologických standardů je DK vytvářena k ulehčení permanentního přístupu k vybraným digitálním zdrojům a jako prostředek přístupu k objevům a vynálezům. V této definici už vidíme příklon k praktickému použití a je zde kladen obrovský důraz na možnosti využití DK.
- ♥ DK je *„integrováný systém zahrnující soubor elektronických informačních zdrojů a služeb umožňující získávání, zpracování, vyhledávání a využívání informací v tomto systému uložených“*²⁰. Nutnou podmínkou fungování je zpřístupnění DK prostřednictvím dostupných počítačových sítí. Účelem budování je poskytnout uživatelům možnost jednotného přístupu k digitálním nebo digitalizovaným dokumentům, příp. i k sekundárním informacím o tištěných primárních zdrojích, uložených ve fondu knihovny.
- ♥ DK není pouze ekvivalentem pro digitalizovanou sbírku s nástroji informačního řízení. Spíše se jedná o *„množinu aktivit, které sjednocují sbírky, služby a lidi, s podporou úplného cyklu tvorby, šíření, užití a ochrany dat, informací a znalostí“*. Výzvy a příležitosti, které podněcují rozvoj výzkumu a iniciativ v oblasti DK jsou spojeny s širokým nazíráním na digitální knihovní prostředí a digitální svět obecně. DK by měla být permanentní nadstavbou knihovny, která umožňuje vědcům přístup k informacím v mnoha formátech. Informace jsou hodnoceny, utříděny, uloženy a archivovány. Přístup ke stále se rozvíjející sbírce digitálních informací je

¹⁸ MacCall, S.L.; Cleveland, A.D.; Gibbon, I.E., 1999

¹⁹ Whaley, T., nestr.

²⁰ Česká, nestr.

zprostředkován nejen pomocí přizpůsobených systémů, ale i jako služeb informačních profesionálů. DK přispívá ke zvýšení hodnoty a šetří čas posunutím času zpřístupnění informací. Snižuje potřebu blízké potřeby informačních zdrojů, ale stále zdůrazňuje jejich kvalitu. DK je potom knihovnou, která může být individuálně přizpůsobena a konečně bude jednoduše použitelná.

- ♥ Poslední z námi uvedených definicí říká, že DK je:
 - sbírkou služeb
 - a sbírkou informačních objektů
 - to vede uživatele k zabývání se informačními objekty
 - a k utřídění a prezentování těchto objektů
 - přístupná přímo nebo zprostředkovaně
 - míněno elektronicky nebo digitálně

1.2. Pohled počítačových specialistů

Zde je nejprve velmi důležité poukázat na zcela odlišné intence vnímání pojmu digitální knihovna. Většina definicí vychází ze základních znalostí databázových technologií, znalosti algoritmizace a procedur spojených se systémy správy dokumentů. Nicméně i zde můžeme najít určité podstatné rysy, které obohacují vnímání pojmu digitální knihovny, zejména otázky propojování, rozhraní jednotlivých systémů, pojmenování objektů uvnitř struktury DK, apod. Nyní opět přistoupíme k nezákladnějším definicím:

- ♠ DK jsou „organizace, které poskytují zdroje, včetně specializovaných pracovníků, k výběru, výstavbě, nabídce intelektuálního přístupu, interpretování, šíření, ochraně integrity a k zajištění stálosti sbírek digitálních prací (platnosti) v čase tak, že budou čitelné a ekonomicky přístupné uživatelům v definovaných komunitách nebo stanovených komunitami.“²¹
- ♠ Další možné chápání DK je „digitální knihovny jsou určeny elektronickými zdroji a spojenými technickými schopnostmi pro tvorbu, vyhledání a užití informací“²². V tomto smyslu jsou rozšířením a vylepšením klasického uložení informací a vyhledávacích systémů, protože zpracovávají digitální data uložená na rozličných nosičích (text, obraz, zvuk, statické či dynamické obrázky) a vyskytující se

²¹ Whaley, T., nestr

v rozšířených sítích. Obsah digitálních knihoven zahrnuje data, metadata popisující různé aspekty dat (např. reprezentaci, tvůrce, vlastníka, vlastnická práva, copyright, atd.) a metadata složená z odkazů nebo vztahů k dalším datům nebo metadatům, ať už interním nebo externím.

- ♣ Jiné nazírání na problematiku říká, že DK „*je možné pokládat za systémy zpřístupňující komunitě uživatelů s jasnou soudržností přístup k rozsáhlému, organizovanému skladišti informací a znalostí...*“²³ Dovednost a schopnost uživatele získat, přepracovat a zužitkovat toto skladiště je dáno možnostmi digitálních technologií.
- ♣ DK je „knihovna, která udržuje všechny související části své kolekce v počítačem přístupné formě jako alternativu, doplněk nebo přílohu k tradičním tištěným a mikrofilmovým materiálům, které jsou běžnou dominantou knihovních sbírek (knihovního fondu)“²⁴. Užití termínu „sbírka“ v tomto kontextu naznačuje dokumenty, které knihovna získává nebo vlastní.

1.3. Nejpodstatnější definice

Uvedenou paletu existujících definic doplníme o **dvě nejpodstatnější**. Každá odpovídá jednomu z výše jmenovaných okruhů a vychází z podmínek jim daných. První z nich je velmi obecná a pochází z počítačového prostředí.

- *Digitální knihovna je spravovaná sbírka informací spolu s odpovídajícími službami, přičemž informace jsou uloženy v digitální podobě a jsou dostupné prostřednictvím sítě*²⁵.

Klíčovými slovy v definici jsou: spravovaná sbírka informací (collection), služby, informace v digitální podobě, přístup prostřednictvím sítě. To, že jde o sbírku informací, která je nějakým systematickým způsobem spravována, řízena, má v definici zásadní význam. Proud dat zasílaný družici na Zemi není knihovnou. Avšak tatáž data, jakmile jsou

²² Bergman, Ch., S. 29

²³ Lunch, C.; Hector, G.M., nestr.

²⁴ Saffady, 224

²⁵ Arms, W.Y., 2000

systematicky uspořádána, stávají se sbírkou v digitální knihovně. Podobně málokdo bude považovat za digitální knihovnu databázi obsahující lékařské záznamy jednoho pacienta, ale soubor takovýchto záznamů od více pacientů jednoho (i více) lékařů již může být částí nějaké digitální knihovny.

Druhá charakteristika pochází z prostředí knihoven a naznačuje, že digitální knihovna v tomto chápání je především knihovnou. Vychází z tradičních knihovních funkcí, jak je výběr, získání, zpracování, zprostředkování, zpřístupnění a uchovávání materiálu, a zdůrazňuje, že digitální knihovny budou vždy budovány tak, aby sloužily konkrétní komunitě uživatelů (ovšem univerzální digitální knihovna není v praxi reálná).

- *Digitální knihovny jsou organizace, které poskytují zdroje (včetně specializovaného personálu) umožňující provádět výběr, strukturování a zpřístupnění sbírek digitálních prací, tyto práce dále distribuovat, udržovat jejich integritu a dlouhodobě uchovávat; a to s ohledem na snadné a ekonomické využití určitou komunitou nebo množinou komunit uživatelů²⁶*

Z uvedeného výčtu definic a řady projektů digitálních knihoven vyplývají určité podstatné **společné znaky digitálních knihoven**:

- ➔ pro digitální knihovnu není klíčovou otázkou digitalizace fyzického materiálu, nýbrž **organizace** elektronické sbírky za účelem lepšího přístupu
- ➔ digitální knihovna obvykle **není jedna uzavřená entita** (většinou autoři užívají množné číslo – digitální knihovny)
- ➔ informační zdroje tvořící digitální knihovnu jsou **heterogenní** (způsobem uložení (organizací, správou) objektů a použitými platformami) **dynamické** (začleňováním a vyřazováním komponent do/ze struktury digitální knihovny) a **multimediální** (povahou dat)

²⁶ Waters, D.J., nestr

- ➔ realizace digitální knihovny vyžaduje **technologie** pro **propojení** různých (autonomně spravovaných) informačních komponent
- ➔ toto **propojení** musí být pro **uživatele transparentní**
- ➔ cílem je zajistit uživateli **jednotný (koherentní) přístup** k relevantním digitálním informacím bez ohledu na jejich formu, formát, způsob a místo uložení.

Přestože se v současné době setkáváme s určitou nejednotností při definování pojmu digitální knihovna, všechny definice ukazují, že pojem nezahrnuje pouze technickou infrastrukturu, ale i množství právních a sociálních aspektů. Digitální knihovny nacházejí plné uplatnění především díky globální síti Internet. Zatímco první fungující systémy digitálních knihoven se soustřeďovaly na vyřešení technických aspektů jejich fungování, digitální knihovny budoucnosti předpokládají vyřešení souvisejících problémů. K těm nejdůležitějším patří: popis informací v digitálních knihovnách prostřednictvím metadat; odhad a vyhodnocení uživatelských potřeb v nové informační společnosti; sociální aspekty digitálních knihoven; rozhraní mezi jednotlivými digitálními knihovnami; digitalizace a konverze; pojmenování objektů a prvků vnitřní struktury digitálních knihoven; archivace digitálních informací, apod. Mnohé z problémů jsou dnes již částečně řešeny a jejich výsledky se také dostávají do definování základního termínu digitální knihovna. Nicméně nezměníme fakt, že stále neexistuje jednotné a všezahrnující terminologické vymezení pojmu digitální knihovna. Myslím, že termín není možné úplně a přesně popsat také vzhledem k jeho provázanosti s technickým a technologickým vývojem v oblasti ICT. Proto je potřeba termín digitální knihovny chápat jako neustále se měnící a aktuálně odkazující pojem, se všemi důsledky, které to přináší v jeho praktickém užití.

1.4. Pracovní vymezení termínů

V současné době slouží větší množství termínů k označování systémů, které částečně, nebo zcela automaticky zjišťují, zpracovávají, ukládají či archivují (dlouhodobě), rozšiřují a zpřístupňují informace (zejména textové povahy) přes Internet. Jednotlivé typy těchto systémů zahrnují různé funkce a procesy a mnohé z nich představují zároveň nové formy elektronického publikování (např. digitální knihovny vydavatelů; archivy elektronických tisků,

tj. vědeckých preprintů²⁷ a postprintů, aj.). Většina z existujících pojmů pro označení těchto systémů zahrnuje přídavné jméno „**digitální**“, které znamená, že **předmětem jejich činnosti** jsou informace v digitální formě (tj. digitální objekty, resp. digitální zdroje informací v definovaných formátech). Čistě český ekvivalent „**číslicový**“ se ve spojení „**číslicová knihovna**“ v praxi neujal (na rozdíl od jiných jazyků, např. francouzštiny). Za synonymum (nikoli čistá synonymie) termínu „**digitální**“ může být považován termín „**elektronický**“. Další slova (podstatná jména) ze slovních spojení jsou převzata ze slovníků tradičních systémů pro zpřístupňování informací (tj. knihovna, archiv, muzeum apod.). To ovšem neznamená, že by dnes budované digitální knihovny, archivy byly budovány na institucionálním principu. Situace je mnohem složitější vzhledem k tomu, že pro počítačové odborníky například institucionální hledisko hraje minimální úlohu, jestli ji vůbec hraje. Proto je pro současné období příznačné, že na přípravě a poskytování digitální či digitalizované informace se podílí daleko širší spektrum institucí, než dříve (a to i takových institucí, které se činnostmi v tradičních podmínkách buď věnovaly málo, nebo vůbec ne). Proto dochází k návrhům a praktickému uplatnění efektivní integrace v podmínkách počítačových sítí. V zájmu uživatelů pak vznikají nadstavbové systémy a služby pracující nad „primárními“ digitálními knihovnami.

Některá spojení zahrnují také slovo „**virtuální**“, které ovšem znamená především způsob (cestu) zpřístupňování či poskytování digitálních (popř. digitalizovaných) informací prostřednictvím počítačových sítí. Uživatel se „pohybuje“ v dané knihovně, sbírce prostřednictvím obrazovky svého klientského počítače.

V současné praxi i literatuře najdeme následující odborné termíny:

- ∞ **Digitální knihovna** (Digital Library) – základní a zatím nejpoužívanější termín pro systémy, které zabezpečují zjišťování, získávání, zpracování, ukládání, archivaci, rozšiřování a zpřístupňování digitálních informací (dle povahy zpracovávaných dat).
- ∞ **Elektronická knihovna** (Electronic Library) – v podstatě synonymum k termínu „digitální knihovna“ (elektronická zdůrazňuje prostředky elektronického zpracování dat).

²⁷ Předběžný dokument zpřístupňující dílo nebo jeho zkrácenou verzi před připravovaným vydáním či zveřejněním; bývá rozmnožen a distribuován v uzavřeném okruhu zájemců.

- ∞ **Automatizovaná digitální knihovna** (Automated Digital Library) – termín pro systémy, které mají všechny procesy zabezpečované počítačovým programem (optimálně bez zásahu člověka).
- ∞ **Virtuální knihovna** (Virtual Library) – obecný termín pro systémy digitálních knihoven dostupné přes počítačové sítě (virtuální zdůrazňuje aspekty přístupu a propojování jednotlivých částí, které tvoří celek knihovny).
- ∞ **Digitální virtuální knihovna** (Digital Virtual Library) – termín pro systémy digitálních knihoven dostupné na dálku přes počítačové sítě (objevuje se v některých projektech).
- ∞ **Kybernetická knihovna** (CyberLibrary) – méně frekventovaný termín pro označení zcela automatizované digitální knihovna, která plní funkce bez zásahu člověka v prostředí počítačových sítí (dnes spíše termín pro budoucí systémy digitálních knihoven).
- ∞ **Digitální sbírka** (Digital Collection) – termín pro označení buď podstatné komponenty digitální knihovny jako systému, nebo pro označení jednoduché formy digitální knihovny, která nezajišťuje všechny její charakteristické funkce.
- ∞ **Elektronický [digitální] archiv** (Electronic [Digital] Archive) – obecný termín pro systémy, v jejichž rámci jsou ukládány digitální materiály (textové i jiné); zajišťována je prioritně archivní funkce (zejména dlouhodobá archivace); dnes jsou tyto systémy běžně budovány vydavateli publikací, národními knihovnami, vědeckými komunitami, apod.
- ∞ **Archiv elektronických tisků** (E-print Archive) – specifický termín pro systémy fungující v oblasti vědy; jsou určeny k online ukládání a vyhledávání tzv. „elektronických tisků“ (tj. preprintů nebo postprintů vědeckých článků) včetně metadat (na bázi autoarchivace ze strany autorů) v prostředí počítačových sítí; jsou také jednou z nových forem elektronického publikování a komunikace informací, často se zdůrazňuje princip jejich otevřenosti (open archives), jenž umožňuje jejich snadnou integraci.
- ∞ **Digitální [elektronické] muzeum** (Digital [Electronic] Museum) – termín pro systém, jenž zajišťuje ukládání a zpřístupňování digitalizovaných muzejních informací, tj. především exponátů muzeí v digitální obrazové, zvukové, audiovizuální aj. formě.

- ∞ **Digitální [elektronická] galerie** (Digital [Electronic] Gallery) – termín pro systém, jenž zajišťuje ukládání a zpřístupňování digitalizovaných galerijních informací, tj. především uměleckých exponátů galerií v digitální obrazové formě (galerie výtvarného umění spadají pod muzea).
- ∞ **Brána k informacím** (Information Gateway) – obecný a velmi frekventovaný termín pro systém, zajišťující přístup k různým typům a zdrojům informací včetně tradičních

Od výše uvedených termínů je třeba striktně odlišovat termín:

- **Automatizovaná knihovna** (Automated Library) – je určen pro označení běžných knihovnických institucí, které se zabývají zpracováním, ukládáním a zpřístupňováním především tradičních dokumentů. Výrazným rysem těchto knihoven je automatizace jednoho či více procesů tradiční činnosti (akvizice, katalogizace, výpůjční systém, atd.); v rámci konvenčních knihoven mohou být budovány dílčí digitální knihovny či sbírky digitalizovaných dokumentů z fondů těchto knihoven (např. staré tisky, vzácné dokumenty, jedinečné práce, apod.).

V posledních letech se často objevuje i termín:

- **Hybridní knihovna** (Hybrid Library) – který vyjadřuje „nutné“ spojení konvenční knihovny a jejích služeb s novodobými systémy digitálních informací (dokumentových i dokumentografických). Zajišťuje informace jak z tradičních informačních zdrojů, tak ze zdrojů digitálních či digitalizovaných. Rozvíjejí se softwarové prostředky, které budou uživatelům schopny zpřístupňovat informace pocházející z tradičního i digitalizovaného informačního fondu jednotným způsobem a to přes jednotné rozhraní prostřednictvím počítačové sítě. Zpřístupňování tradičního dokumentu znamená v současné době využití buď tradiční meziknihovnické služby (MVS či MMVS), nebo jedné z moderních forem tzv. „elektronického zpřístupňování dokumentů (DDS, EDD, atd.).

2. Historie, činnosti a funkce digitálních knihoven

Jak ukazují naše současné znalosti je termín „digitální knihovna“ termínem problémovým, protože skrytě zahrnuje komplex vztahů mezi **elektronickými informačními fondy** a **knihovnami jako institucemi**. Vzhledem k tomu by se mohlo lépe využít širšího termínu „*knihovna digitálních materiálů*“, který se ovšem v praxi téměř nevyskytuje. K dnešnímu dni bylo různými informačními kanály publikováno velké množství definic a vyjádření termínu, které se ovšem různí podle zaměření samotného autora definice – jak jsme již zmiňovali v úvodní přednášce. Termín je dokonce některými americkými autory označen za **oxymóron**, tedy slovní spojení zdánlivě si odporujících slov (v našem případě slov „digitální“ a „knihovna“). Jak ovšem mnozí odborníci poznamenávají, termín „digitální knihovna“ je nadále předmětem neustálých debat, definice je stále ve vývoji a bude se měnit. Co však zůstává stále stejné a neměnné jsou základní funkce digitálních knihoven, které byly rozpracovány a jsou uplatňovány v praxi (jak se ukázalo, jedná se opravdu o exaktní vyčlenění funkcí, které nejsou tolik měněny jako definování samotného pojmu DK). Digitální knihovny zajišťují následující **základní funkce**:

- **zjišťování, výběr a získávání** digitálních objektů (informací)
- **identifikace a zpracování** digitálních objektů (tvorba metadat, včetně indexace ruční povahy, popř. automatická indexace v případě úplných textů, konverze metadat, apod.)
- **uložení, archivace a údržba, ochrana a bezpečnost** digitálních objektů v digitalizovaných skladištích (repozitářích)
- **vyhledávání, rozšiřování a zpřístupňování** digitálních informací uživatelům DK, včetně zajištění přístupu k objektům; může jít o služby typu SDI, dodávání dokumentů na základě profilů uživatelů, statistiky, bibliometrické analýzy, nebo nové služby propojování informací uložených v digitální knihovně s informacemi z jiných systémů

Vedle základních funkcí může být digitální knihovna rozšířena o funkce další podle toho, v jakém prostředí působí. Například v oblasti univerzity jsou budovány digitální archivy disertací, které mají důležitou funkci automatického podávání (předkládání) prací v elektronické formě. Některé z archivů mají i další funkci, totiž funkci prodeje jejich úplných

textů, jinak je prodej informací typickým příkladem služeb digitálních knihoven komerčních vydavatelů.

V rámci budování současných digitálních knihoven se řeší řada dalších **problémových okruhů**:

- problematika **obsahu** (předmětu) zpracování, uložení či archivace objektů (zdrojů) v digitálních knihovnách, archivech; jejich typologie, formátů, netafatových formátů a trvalých identifikátorů
- problematika **technických otázek** se týká zejména zabezpečení kvalitní HW základny; k ní patří jednak *WWW server*, určený ke komunikaci a publikování informací uložených v repozitáři (databázi) digitální knihovny; jednak *databázový server* k uložení digitálních fondu; servery mohou pracovat pod různými operačními systémy (UNIX, Linux, Windows, atd.) a měly by mít dostatečnou paměťovou kapacitu, zejména pokud jde o digitální fond.
- problematika **technologie zpracování a zpřístupňování digitálních objektů**, která se týká koncepce, architektury DK, uložení a organizace digitálních objektů včetně metadat v celkové struktuře DK, protokolů pro komunikaci objektů uvnitř i vně systémů, bezpečnost dat, apod. s návazností na technické zabezpečení. Výrazným trendem současnosti z hlediska základního *SW vybavení* je využívání tzv. *otevřených zdrojů* (Open Sources, tj. programů s přístupným zdrojovým kódem) zejména těch, které jsou k dispozici zdarma (např. MySQL, Perl, Common Gateway Interface, PHP, JavaScript, Apache Web Server, Tomcat, apod.)
- složitá problematika **autorsko-právní ochrany** digitálních děl (zahrnuje otázky autorských práv k dílu v digitální podobě, vztahu mezi autory a provozovateli digitálních knihoven a eventuelně i vydavateli komerčními)
- **organizační problémy** provozu digitálních knihoven včetně otázek funkcí a otázek finančního zabezpečení
- problematika **mezinárodní kooperace** budování DK (společné budování digitálních knihoven, archivů v celosvětovém měřítku, jejich dostupné a účelné propojování – integrace v síťovém prostoru).

2.1. Historie vývoje DK

Koncepce digitální knihovny není nová a vize digitálních knihoven provází v různých podobách větší část historie výpočetní techniky. Prakticky byla koncepce digitálních knihoven vždy předurčována samotným vývojem prvních počítačů. Jedním z prvních autorů, kteří psali o možnosti vytvoření a zpřístupnění světového poznání, které by mohlo doplnit, přidat funkční závislosti a dokonce stát u nahrazení tradičních funkcí knihoven a vytvoření nových typů knihovnicko-informačních institucí byl *H. G. Wells* (cca v roce 1938). Podstatný pokrok v oblasti digitálních knihoven však nastal až počátkem 90. let minulého století. Prudký rozvoj informačních a komunikačních technologií umožnil praktickou realizaci teoretických návrhů a představ, které bylo doprovázeno zpřístupňováním prvních praktických výsledků v oblasti tvorby digitálních knihoven.

Pro historii digitálních knihoven jsou podstatné dvě postavy, které sehrály nezanedbatelnou roli při rozvoji digitálních knihoven. Jejich průkopnické práce stále provázejí většinu současných teoretiků digitálních knihoven a inspirují k novým úvahám o dalším vývoji digitálních knihoven.

Prvním představitelem je **Vannevar Bush**²⁸, který se ve svém známém článku „*As We May Think*“²⁹ věnoval problematikou efektivnějšího automatizovaného zpracování odborných informací. „naše používané metody přenosu a hodnocení výsledků výzkumu jsou staré celé generace, a proto jsou dnes pro svůj účel zcela neadekvátní“. V článku Bush analyzoval potenciální možnosti, které nabízely soudobé (analogové) technologie pro získávání, ukládání a vyhledávání informací. Nastínil vizi systému (stroje *Memex*) využívajícího fotografické postupy a kompresi dat pomocí mikrofilmů. Takto navržený systém koncepčně odpovídá dnešnímu osobnímu počítači, v němž jsou informace provázány asociativními vazbami a je tak předchůdcem hypertextu a koncepce dnešního webu. Někteří dnešní teoretici digitálních knihoven³⁰ nás potom přesvědčují, že okolo roku 2015 (tj. 70 let po publikování snu o memexu) budeme mít každý na svém stole ekvivalent významné výzkumné knihovny. Ta bude mít prohledávací možnosti nad rámec Bushových představ.

Druhým představitelem je C. R. Licklider³¹, který studoval přeměnu knihoven v návaznosti na využití digitálních počítačů. V roce 1965 publikoval knihu „*Libraries of the*

²⁸ profesor MIT a ředitel amerického Národního úřadu pro vědecký výzkum a vývoj, za 2. světové války

²⁹ Bush, V., S. 101-108

³⁰ Lesk, M., S. 270

³¹ Studoval v 60. letech 20. století na MIT

Future“, ve které identifikoval výzkum a potřebný vývoj k realizaci prakticky použitelné digitální knihovny. Nastínil přitom vizi fungující digitální knihovny po 30 letech (cca pro rok 1994), která v obecné rovině přesně odpovídá skutečnosti. Mnohé z předpokladů se vyplnily, i když ne vždy s předpokládaným výsledkem – celkově byly výrazně podceněny výsledky využití možností výpočetní techniky a přeceněny pokroky založené na umělé inteligenci, automatizovaných metodách zpracování přirozeného jazyka.

V 60. letech 20. století se objevily také první významné praktické výsledky uplatnění výpočetní techniky v knihovnách. Typickým příkladem je bezesporu oblast zpracování informací spojený s vývojem formátu MARC (Machine Readable Cataloguing) vyvinutý Kongresovou knihovnou ve Washingtonu. Tento formát standardizuje strukturu bibliografického záznamu v elektronické podobě a jeho použití pro sdílenou katalogizaci knihoven (původně pouze knihoven systému OCLC); rozvoj online knihovních katalogů (OPAC). Navzdory překážkám, které byly dány technickými omezeními, podnítily tyto první úspěchy a jejich praktické uplatnění mnohé optimistické předpovědi.

Za nezdařilejší považují prognózu A. L. Samuela³², který v roce 1964 předpověděl, že klasické knihovny zaniknou do dvaceti let. Důvody, proč se většina předpovědí ze 60. let nenaplnila, byly samozřejmě různé. Často však sehrála svou roli otázka finančního zabezpečení digitálních knihoven. Např. pro naplnění Samuelovy vize by bylo potřeba digitalizovat zhruba 100 miliónů titulů knih, přičemž aktuální čísla z amerického prostředí uvádějí cenu digitalizace v rozmezí 1,5 – 10 dolarů za stránku; mnohem podstatnější položkou by však byly finanční náklady na kompenzaci samotných autorských práv.

Skutečný a jedinečný rozmach digitální knihoven nastává počátkem 90. let 20. století. Obrovskou zásluhu na tom měla skutečnost, že technologický pokrok ve všech podstatných oblastech (*computing* – výpočetní a krátkodobá i dlouhodobá paměťová kapacita; *communications* – globální síť a přenosová kapacita; *content* – množství informací v digitální podobě) digitálních knihoven dosáhl dostatečně vysoké úrovně při zachování rozumně nízké jednotkové ceně a široké všeobecné dostupnosti. To umožnilo začít realizovat projekty reagující na skutečné potřeby uživatelů a odstartovalo prudký rozvoj v oblasti digitalizace, elektronického publikování a šíření informací. Rovněž tím byl dán nový impuls pro výzkum a vývoj v oblasti digitálních knihoven (podpořený výrazným celosvětovým rozšířením webových technologií a všeobecnou potřebou efektivnějšího sdílení vědeckých poznatků).

2.1.1. Důvody vzniku a podpory digitálních knihoven

Na základě historického vývoje a také daných vizí si můžeme položit otázku, proč by vlastně měly digitální knihovny vznikat. Je účelné podporovat tento trend v informační vědě a knihovnictví. Není nový fenomén hrozbou pro naši další existenci?

Počáteční představy digitální knihovny vycházely z koncepce klasické knihovny a byly orientovány na digitalizaci existujících sbírek, jako nástroje pro zlepšení klasických informačních služeb knihoven. Výraznými prvky byla podpora následujících oblastí:

- dokonalejší ochrana fondu (náhrada vzácných a významných fyzických objektů digitálními)
- sdílení informací mezi různými knihovnami
- vhodnější využití fondu (jednotný a souběžný přístup k jednomu konkrétnímu dokumentu)
- efektivnější metody vyhledávání (práce s indexy, plnými texty, soubory informačních objektů, atd.)
- vzdálený a nepřetržitý přístup k informacím

Velmi brzy se ukázalo, že potenciální možnosti digitálních knihoven jdou nad rámec možností klasických knihoven s fyzickými dokumenty. Výrazným projevem byly např.: možnosti neomezené globální integrace digitálních repozitářů v celosvětovém měřítku; nové možnosti permanentní aktualizace informací uložených v digitální knihovně; zcela nové typy služeb, atd. Přes tyto a další odlišnosti (provozně ani organizačně nemusí mít digitální a klasické knihovny vůbec nic společného) mají oba typy knihoven principiálně řadu shodných rysů:

1. systematicky budovanou sbírku **datových objektů**
2. obsahovou analýzu datových objektů ve sbírkách a z ní vyplývající soubory **metadatových struktur** (katalogy, rejstříky, indexy, tezaury)
3. množinu **služeb** (přístupové metody, správa dat, akvizice, vyhodnocování, referenční služby, SDI)
4. **tematické** zaměření
5. sledování **kvality**
6. dlouhodobé **uchování** materiálu

³² Samuel, A.L., S. 529-530

Metody a postupy klasických knihoven jsou za mnoho staletí svého vývoje propracována a tvoří ucelený, efektivně fungující systém. Digitální knihovny však přinášejí nové výzvy a problémy, pro jejichž řešení nelze často použít vůbec klasických postupů, nebo jen ve velmi omezené míře. Po počátečním optimismu z první poloviny 90. let 20. století se ukazuje, že problémem budování funkčních digitálních knihoven je mnohem složitější, než se na počátku zdálo. Zásadním problémem a základem všech obtíží je nedostatečně propracovaná technologie na jedné straně a nepřipravené společenské prostředí zahrnující složitý komplex navzájem provázaných problémů z oblastí ekonomické, právní, sociální a etické na straně druhé. To, s čím se klasické knihovny učily pracovat a co bylo propracováno v průběhu mnoha staletí, musí digitální knihovny vyřešit za několik málo let a za pochodu.

3. Typologie digitálních objektů (zdrojů)

Předmětem prvních digitálních knihoven, resp. sbírek digitálních informací, byly **textové informační zdroje**. Šlo především o textové soubory digitalizovaných knižních materiálů, zejména starých tisků, které se původně vyskytovaly v papírové podobě. Z tohoto pohledu označení „digitální knihovna“, se jeví jako zcela odpovídající. S vývojem informačních technologií ovšem přibývaly další typy objektů a tak v současné době může „digitální knihovna“ zpřístupňovat jakýkoli typ informace (informačního zdroje). Vzhledem k tomu je možné zvažovat, jestli pojem „digitální knihovna“ lze používat nadále bez výhrad.

V dnešní společenské informační komunikaci nabývají mimořádného významu **elektronické (digitální) informační zdroje a způsoby jejich veřejného zpřístupňování** prostřednictvím globální sítě Internet, resp. služby WWW. Z hlediska informační vědy jsou intenzivně zkoumány různé aspekty této komunikace i její jednotlivé prvky, s cílem přispět k další optimalizaci řízení toků informací, organizaci jejich zdrojů a zejména k jejich efektivnímu využívání.

Pro potřeby našeho předmětu je podstatné seznámení se složitou problematikou poznávání funkcí, charakteru a vlastností elektronických (digitálních) informačních zdrojů dostupných v síti Internet. Nejde ovšem o komplexní zkoumání dané problematiky, ale budeme analyzovat pouze jeden aspekt zdrojů, tj. **jejich místo či zařazení v celkové množině všech informačních zdrojů** a zejména jejich **další vnitřní dělení** (třídění – dané jediným znakem zdrojů; typologii – vymezenou kombinací dvou či více znaků). Východiskem rozboru je pracovní vymezení **základních pojmů**, vztahujících se k **elektronickým zdrojům**, včetně zdrojů síťových užívaných doma i v zahraničí. Celá množina zdrojů umožňuje vymezení čtyř dílčích skupin, které postupně zahrnují pojmy užšího rozsahu. Jde o následující skupiny:

- výchozí skupina zahrnuje pojmy vymezující všechny „**elektronické informační zdroje**“, tedy i zdroje přes počítačové sítě nedostupné; termín „**digitální informační zdroje**“ lze považovat za synonymum; do skupiny patří také pojem užívaný v počítačové vědě „**elektronické (digitální) objekty**“; v knihovnictví se vyskytující pojmy „**elektronické (digitální) materiály**“ nebo dnes už ne zcela vyhovující pojem „**počítačové soubory**“ a také užší „**elektronické (digitální) dokumenty**“, které pro některé odborníky představují často jen zdroje textových

informací, nebo „**elektronické (digitální) publikace**“, jejichž nově definovaný obsah je v současné době předmětem výzkumů; uvedené pojmy bývají často užívány také ve smyslu zdrojů dostupných již přes počítačové sítě.

- druhá skupina zužuje skupinu první na základě kritéria přístupu k těmto zdrojům přes počítačové sítě (jakékoliv); zahrnuje pojem „**síťové (informační) zdroje**“, popř. „**síťové elektronické zdroje**“, dále zejména v knihovnických systémech frekventovaný termín „**elektronické zdroje dostupné na dálku**“³³, standardní termín „**online (informační) zdroje**“, který může ovšem mít i jiný význam, a užší pojmy „**síťové dokumenty**“, nebo „**síťové publikace**“
- třetí skupina, která je podmnožinou skupiny předchozí, obsahuje informační zdroje dostupné veřejně³⁴ přes celosvětovou počítačovou síť Internet, tj. v rámci definovaných protokolů TCP/IP; skupina zahrnuje frekventované pojmy „**internetové (informační) zdroje**“³⁵, „internetové materiály“ a opět užší „**internetové dokumenty**“ či „**internetové publikace**“
- poslední skupina zahrnuje informační zdroje dostupné veřejně v síti Internet pouze přes protokol http³⁶; zahrnuje pojem „**webové**“³⁷ (informační) zdroje“, užší ovšem velmi frekventovaný pojem „**webové dokumenty**“, jenž míří z hlediska typu obsažených informací především ke zdrojům textové povahy uložených na webových serverech ve formátu HTML, dále jeho synonyma „**dokumenty na WWW**“, nebo přesněji označený termínem „**dokumenty využívající http protokol**“; do této skupiny dále patří konkrétní pojmy „**webová stránka**“, „**domovská stránka**“, „**personální (webová) stránka**“ nebo „**HTML stránka**“ – ty se ale také už objevují v rámci konkrétních třídění nebo typologií těchto zdrojů; doplníme ještě komplexnější pojem „**soubor webových dokumentů**“, jakož i „**webovské sídlo**“ nebo základní pojem „web“; k již uvedeným pojmům je možné dnes zařadit i další informační zdroje zpřístupňované původně v rámci Internetu samostatně přes jiné protokoly (FTP, Gopher, Usenet, atd.), protože protokol http je nyní také podporuje (jinak patří do skupiny třetí); lze konstatovat, že množina

³³ angl. remotely accessed resources

³⁴ „veřejně dostupné“ – znamená v daném kontextu, že množina zdrojů, jež jsou sice přes Internet přístupné, ale nejsou k dispozici veřejně (vstup pouze ne veřejné heslo) do pojmu internetový zdroj nepatří

³⁵ pozor na užití „internetovské“ – není doporučeno a odmítají lingvisté

³⁶ angl. hypertext transfer protocol

³⁷ možno užít i varianty „webovské“

„internetových informačních zdrojů“ se tak velmi sblížila s množinou „webových informačních zdrojů“.

Pro potřeby digitálních knihoven jsou podstatné znalosti třídění a typologie informační zdrojů prezentovaných pojmů ze třetí a zejména čtvrté skupiny. Jestliže lze množiny těchto informačních zdrojů, daných v podstatě přístupem k nim, relativně snadno a bez větších problémů pracovní vymezení, pak vymezení rozdílů mezi pojmy „internetový/webový informační zdroj“ a „internetový/webový dokument“ se zdá naopak velmi složité. Je pravdou, že zejména knihovníci, se v současné době zabývají řadou otázek spojených s vymezením obou pojmů. Nicméně je vhodné podotknout, že užití pojmu „dokument“ v síťovém digitálním prostředí je v určitém aspektu sice možné i legitimní, ale nepostačuje již k zahrnutí dalších forem informací, které jsou dnes prostřednictvím sítě Internet (jakožto média) přenášeny. Tradiční pohled na dokument jako médium statické povahy, sloužící k přenosu informací fixovaných a uspořádaných na materiálním nosiči v prostoru a čase již dnes nevyhovuje při pokusu o objasnění povahy „síťového digitálního dokumentu“. Domnívám se, že v tomto případě je nutné využít pojmu „*nedokumentové informační prameny (zdroje)*“, které zajišťují vysílání a přenos informací prostřednictvím energie. Jak ukazuje i samotná praxe, v informačních systémech provozovaných v rámci Internetu se běžně užívá již jen termín „**[informační] zdroje**“ a v rámci teorie knihovnictví se k tomu postupně přistupuje také.

Řada současných autorů rovněž správně poukazuje na zcela nové, specifické znaky či vlastnosti „internetových a webových dokumentů“. Asi nejvýstižnější charakteristiku nalezneme v článku *Lindy Schamberové*³⁸, která zdůrazňuje skutečnost, že v prostoru sítě jsou na základě elektronických zdrojů informací uložených na serverech nebo i na základě informačních zdrojů reálného světa vysílány a směrem k uživatelům přenášeny (transferovány) na jejich vyžádání vlastně jenom jejich digitální „kopie“, které uživatel může v daném okamžiku vnímat v různých podobách v závislosti na užívaném programu (prohlížeči), koncovém zařízení, včetně pořízení kopie na vlastní médium³⁹. Přes síť získané „elektronické dokumenty“ nejsou vlastně reálnými dokumenty, jejich fyzická forma (vlastnost) se ztrácí, ale to podstatné, co uživatel získává, je informační obsah. **Typ**

³⁸ Schamber, L., S. 669-671

³⁹ V případě složitějšího zdroje je situace komplikovanější. Zdánlivě jednoduchý seznam vyhledaných informací automaticky generovaný do formátu HTML a zobrazený na obrazovce může ve skutečnosti mít za sebou složitou strukturu velkého počtu db a datových souborů lokalizovaných i na různých serverech.

přenášených informací a formát (uspořádání dat) patří k hlavním a podstatným kritériím pro třídění či typologii síťových zdrojů.

Dříve, než se pustíme do analýzy vybraných reprezentantů třídění a typologií síťových digitálních informačních zdrojů, je uvedena stručná typologie a charakteristika zpráv, resp. jejich formátů, přenášených v síti Internet, která úzce navazuje na předchozí text a která do značné míry souvisí i s dalším rozbohem.

3.1. Typologie elektronických zpráv v rámci internetové normy MIME

Typologie zpráv přenášených v rámci protokolů TCP/IP stanovuje internetová norma, označovaná jako **MIME**⁴⁰. MIME⁴¹ je normou, která rozšiřuje původní normy RFC 821 - RFC 822, které byly věnovány přenosu jednoduchých textových zpráv přes elektronickou poštu. Norma MIME, zajišťuje přepravu zpráv složitějšího charakteru – texty s diakritikou, obrázky, zvuky, apod., prostřednictvím stávajícího poštovního systému. Typologie zpráv⁴² užívá terminologického označení „*typy médií*“⁴³ ve smyslu typů zpráv jako zprostředkovatelů přenosu informačního obsahu. Uvedená typologie je pro provoz moderních služeb Internetu velmi užitečná a uplatňuje se v jednotlivých informačních systémech internetových i knihovnických, které se zabývají jejich zpracováním, jak z hlediska formátu tak i typu či charakteru přenášené informace.

Hodnoty typů zpráv jsou obsaženy v části záhlaví zprávy v hlavičce (řádka označená návěštím „Content-Type“⁴⁴, které specifikuje přesně charakter obsahu zprávy pomocí definovaného typu a podtypu⁴⁵ s případnou doplňkovou informací.

Příklad zápisu informace v hlavičce zprávy:

Content-type: text/html; charset=ISO-8859-2

Content-type: image/gif

Content-type: application/msword; name="clanek.doc"

Content-type: model/vrml

⁴⁰ angl. Multipurpose Internet Mail Extensions – víceúčelová rozšíření internetové pošty

⁴¹ Nejnovější verze normy představuje soubor textových dokumentů RFC 2045 - RFC 2049.

⁴² definovaná v 2. části normy RFC 2046.

⁴³ angl. media types

⁴⁴ typ dat v obsahu zprávy

Typologie RFC 2046 definuje v současnosti celkem 6 jednoduchých a 2 složené typy zpráv, dat či informací. V jejich rámci je definována velká řada dílčích podtypů prezentujících známé i méně známé formáty přenášených zpráv. Typy a podtypy jsou schvalovány úřadem IANA⁴⁶ a registrovány na veřejných WWW serverech⁴⁷.

K jednoduchým typům zpráv patří:

- a) **text** – typ k posílání textových (grafických písemných) informací; podtyp formát „plain“ (čistý neformátovaný text), frekventovaným podtypem formát „html“ (text obohacen o značky jazyka HTML); textové informace vyžadující k percepci speciální program jsou řazeny k typu „application“
- b) **image** (obraz) – k odesílání obrazových (ikonografických nebo ideografických) informací, včetně animovaných; k zobrazení potřeba specifické zařízení; nejznámější podtypy formát „jpeg“ a „gif“ či „bmp“
- c) **audio** (zvuk) – k posílání zvukových (auditivních) informací, k výstupu je potřeba přehrávacího zařízení; základní podtyp je formát „basic“
- d) **video** (video) – k posílání pohyblivých obrazových informací, k zobrazení je třeba specifického přehrávacího zařízení; základním podtypem je formát „mpeg“
- e) **application** (aplikace) – specifický typ k posílání jiných typů informací, zpravidla buď binárních dat, nebo informací, které je nutné zpracovat pomocí nějaké aplikace (programu), aby byly čitelné pro uživatele; definovány jsou dva základní formáty – „octet-stream“ (obsahem zprávy jsou binární data) a „postscript“ (obsahem je postscriptový dokument čitelný pouze pomocí speciálního programu); další specializované podtypy – „msword“; „pdf“; „sgml“; „marc“, atd.
- f) **model** (model) – k odesílání troj- a vícerozměrných systémů, ve kterých lze zavést pravouhloú soustavu souřadnic; model se skládá z jednoho nebo více objektů, které se pak skládají z prvků, jež mají mezi sebou definovány vztahy; užívá se hovorového označení „virtuální realita“; k základním podtypům patří formát „vrml“.

⁴⁵ Podtypy prezentují konkrétní formáty zpráv a hrají důležitou roli při vyhledávání informací.

⁴⁶ The Internet Assigned Numbers Authority

Ke složeným (kompozitním) typům zpráv patří: (obsahují více dílčích zpráv)

- a) **multipart** (multipart) – obsahuje několik dílčích zpráv; k základním podtypům patří – formát „*multipart/mixed*“ (díličí nezávislé zprávy ve stanoveném pořadí); „*multipart/alternative*“ (díličí zprávy se shodnými informacemi, ale v různém tvaru); „*multipart/digest*“ (zpráva či více zpráv tvořených posloupností jiných dílčích zpráv); „*multipart/parallel*“ (díličí zprávy bez ohledu na jejich řazení); „*multipart/signed*“ a „*multipart/encrypted*“ (bezpečné zprávy složené ze základní díličí zprávy a elektronického podpisu nebo šifrované zprávy)
- b) **message** (zpráva) – umožňuje poslat zprávu jako tělo jiné zprávy – podtyp „*message/rfc822*“ nebo poslat dlouhou zprávu jako několik kratších – podtyp „*message/partial*“ či poslat informaci o zprávě uložené na nějakém serveru – typ „*message/external*“

Většina definovaných typů normy MIME, koresponduje s typologií sociálních, příp. fyzikálních informací z hlediska způsobů jejich prezentace a percepce. Nicméně, typologie má význam především z hlediska způsobů uspořádání dat při jejich přenosu v počítačové síti, a proto se také často uplatňuje v dále zmiňovaných systémech při dělení informačních zdrojů dle formátu.

V návaznosti na typologii MIME je vhodné ještě jako doplněk zmínit základní rozdělení přístupů k digitálním zdrojům v rámci sítě Internet, jak je definovala norma RFC 1738 „Jednotné lokátory zdrojů URL“. Dané rozdělení hraje důležitou roli při zpracování internetových informačních zdrojů hlavně v knihovnických systémech.

Norma definuje následující přístupy:

ftp	protokol přenosu souboru (File Transfer Protocol)
http	protokol přenosu hypertextu (Hypertext Transfer Protocol)
gopher	protokol Gopher (The Gopher Protocol)
mailto	adresa elektronické pošty (Electronic mail address)
news	novinky USENET (USENET news)

⁴⁷ Dostupný z WW: <<http://www.iana.org/assignments/media-types>>

nntp	noviny USENET užívající protokol přenosu síťových novinek NNTP (USENET news using NNTP – Network News Transfer Protocol)
telnet	odkaz na interaktivní relaci (Reference to interactive sessions)
wais	WAIS – vyhledávání informací (Wide Area Information Servers)
file	jména specifických souborů na serveru (Host-specific file names)
prospero	adresářová služba – distribuovaný systém souborů (Prospero Directory Service)

3.2. Třídění a typologie internetových a webových zdrojů v knihovnických systémech

Jednou z důležitých oblastí a podstatných reprezentantů informačních systémů, které se hlásí ke zpracování internetových a webových informačních zdrojů jsou knihovny, informační střediska a jejich sítě. V celosvětovém měřítku se již dnes řada těchto institucí různého zaměření zabývá daným procesem prakticky i teoreticky. Záznamy o online informačních zdrojích vznikají a jsou ukládány, buď integrovaně v rámci stávajícího automatizovaného katalogu, nebo v rámci samostatných katalogů či databázi s WWW rozhraním. O širší registraci elektronických zdrojů v mezinárodním nebo národním záběru se pak zcela legislativně ucházejí i velké národní knihovny, resp. národní bibliografické agentury, jejichž záměrem je rozšířit stávající systémy souběžných národních bibliografií o registraci online dostupných zdrojů. V rámci jejich zpracování se řeší celá řada problémů týkajících se také formátů pro záznamy a návazných pravidel. K problémovým otázkám pak patří i zařazování těchto online informačních zdrojů do celkové množiny všech informačních zdrojů (materiálů) a jejich třídění nebo typologie za účelem jejich efektivního vyhledávání.

V další části se pouze zmíníme o již známých faktech z problematiky třídění a typologie síťově dostupných zdrojů, v rámci nejvýznamnějších knihovnických formátů a na ně navazujících normativních dokumentů – bývalého amerického bibliografického formátu USMARC, aktuálního harmonizačního formátu MARC 21 a v evropském měřítku často aplikovaného formátu UNIMARC.

3.2.1. USMARC v návaznosti na AACR2R

Elektronické informační zdroje byl v rámci zpracování všech „materiálů“⁴⁸ v bývalém americkém formátu *USMARC* zařazovány a dále členěny dle různých hledisek odrážejících podstatné vlastnosti jich samotných, nebo dalších prvků komunikačního procesu.

Internetové a webové informační zdroje byly zařazeny společně s dalšími síťovými i nesíťovými elektronickými zdroji v rámci **základního třídění**, které mělo 14 tříd. Příslušná kódovaná hodnota jednotlivých kategorií se udávala v rámci definovaného „typu záznamu“ (pozice 06) v návěští záznamu⁴⁹.

Šlo o kategorie:

- a jazykový materiál
- c tištěná hudebnina
- d rukopisná hudebnina
- e tištěná mapa
- f rukopisná mapa
- g projekční médium (film, videozáznam, diafilm, diapozitiv, průsvitka, aj.)
- i nehudební zvukový záznam
- j hudební zvukový záznam
- k dvojrozměrná neprojekční grafika (koláže, výkresy, obrazy, malby, fotografie)
- m počítačový soubor (číselná/textová data, počítačové programy a kombinace)
- o souprava (složená z jednotek různého typu, ani jedna nebyla dominantní)
- p smíšený materiál (složený z dig.zdrojů různého typu, žádný nebyl dominantní)
- r trojrozměrný artefakt nebo předmět
- t rukopisný jazykový materiál

Všechny elektronické zdroje byly před rokem 1997 zařazovány výlučně do kategorie „*počítačový soubor*“, ovšem tento termín nevyhovoval ani americkým profesionálům, nicméně jeho záměna za jiný nebyl tehdy oficiálně kodifikována. Kód „**p**“ jako jediný z daného třídění nebyl dříve jeho součástí, přibyl jako náhradní řešení pro zařazování digitálních zdrojů složených z různých typů informací. Třídění ovšem bylo v souvislosti s rozvojem nových forem síťových elektronických zdrojů značně problematické. Prolínala se

⁴⁸ Původní anglický termín užívaný v AACR2R, v češtině volně překládáno jako „dokument“.

v něm protichůdná hlediska dělení⁵⁰; počítačovou elektronickou formu mohly mít již v podstatě všechny materiály dalších kategorií daného třídění. V roce 1997 bylo přijato doporučení, aby v kategorii „m“ byly zařazovány pouze některé typy počítačových souborů – počítačové programy, multimedia, online služby, aj⁵¹. Jiné typy počítačových souborů bylo možné zařazovat podle jejich nejvýznamnějšího znaku, daného typem informace obsažené v materiálu. Ovšem nic z toho neřešilo narůstající problém „obsah kontra nosič“ – což vyústilo do zásadního požadavku na komplexní revizi a přestavbu AACR2R. Na základě jednání vzniká významné doporučení – 1. část návrhu nového modelu „*Logické struktury Anglo-amerických katalogizačních pravidel*“ (v roce 1998).

Nový model „logických pravidel“ směřoval k zásadnímu vymezení tříd na základě **obsahu** (informací) a popř. **formy**, v jejímž rámci byl obsah vyjádřen; **rozbil** tehdejší chápání *entity* „dokument“ a byly nově definovány jeho jednotlivé komponenty (obsah, fixace a fyzický nosič) za účelem nového vymezení tříd.

Formát USMARC v návaznosti na AACR2R umožňoval další rozdělování počítačových souborů až v datových polích (v kontrolním poli 008 pozici 26, nebo poli 006 pozici 09), což mělo důležitou roli při vyhledávání. V případě množiny počítačových souborů dostupných na dálku formát USMARC jednoznačně vymezoval kódovaným údajem „r“ (remote – vzdálený zdroj) v rámci kontrolního pole 007 pozice 01.

Další ovšem pouze hrubé rozdělení počítačových souborů, bylo v souladu s AACR2 definováno v poli 256, které korespondovalo s kódy polí 008 a 007. Pole bylo povinné pro počítačové soubory dostupné na dálku. Tehdejší nabídka jednotlivých typů byla velmi malá – počítačová data, počítačové programy a kombinace obou typů.

Nejnovější rozdělování síťových elektronických zdrojů nabízelo tehdy nově vytvořené *pole 856*, jehož podstatou byly informace o přístupu k elektronickým zdrojům, jejich vztazích k variantním tradičním zdrojům a o nejrůznějších dalších, zejména technických podrobnostech. Pro nás podstatné: v prvním indikátoru pole 856 byl kódován konkrétní přístup ke zdrojům (0 elektronická pošta, 1 FTP, 2 telnet, 4 http; v podpoli „q“ pole 856 umožněno rozdělení elektronických zdrojů podle jejich formátů (možnost uplatnění typologie MIME).

⁴⁹ Dělení odpovídalo rozdělování materiálů z pravidel AACR2R.

⁵⁰ některé kategorie dány typem informace obsažené v materiálu, jiné typem nosiče informace

⁵¹ Zřetelné vymezení hlediska pro jejich zařazení ovšem chybělo.

3.2.2. MARC 21 v návaznosti na AACR2R

Společný formát USA, Kanady a přidaly se i další země (VB, ČR, aj.) nevnese do problematiky třídění žádnou výraznou změnu. Stále neřešen problém „obsah kontra nosič“. První verze kopírovaly třídění a typologie známé z formátu USMARC. Jisté změny formální podoby nastaly v roce 2005. Došlo ke změnám v *pojmenování* některých kategorií a také při *vymezení jejich obsahu*. Typickým rysem je zahrnování elektronických forem materiálů do definic téměř všech kategorií (tříd), tedy nejen do původní kategorie „m“⁵².

Aktuální třídění všech „informačních materiálů“ zachovává 14 tříd (pozice 06 v návěští záznamu), z původního USMARC zachováno i písmenné označení jednotlivých tříd:

- a Jazykový [textový] materiál (Language material)
Jméno kategorie zůstává stejné, zahrnuty pod ni však jsou již nejenom tištěné formy, ale i materiály ve formě mikrodokumentů a ve formě elektronické
- c Hudebnina (Notated music)
Jméno kategorie je nové a označuje lépe daný typ materiálu. Zahrnuty jsou také mikrodokumenty a forma elektronická
- d Rukopisná hudebnina (Manuscript notated music)
Jméno kategorie je rovněž nové a zahrnuty jsou také rukopisné hudebniny ve formě mikrodokumentu
- e Kartografický materiál (Cartographic material)
Jméno kategorie je nové a obecnější povahy, zahrnovány jsou mapy, atlasy, glóby, digitální mapy a další kartografické dokumenty
- f Rukopisný kartografický materiál (Manuscript cartographic material)
Jméno kategorie je rovněž nové, zahrnovány jsou rukopisné mapy ve formě mikrodokumentu
- g Projekční médium (Projected medium)
Jméno kategorie zůstává stejné, zahrnovány jsou filmy, videozáznamy včetně digitálního videa, diafilmy, diapozitivy, průsvitky aj.
- i Nehudební zvukový záznam (Nonmusical sound recording)
Jméno kategorie zůstává stejné, zahrnovány jsou záznamy s mluveným slovem
- j Hudební zvukový záznam (Musical sound recording)

⁵² Není to principiální řešení problému.

Jméno kategorie zůstává stejné, zahrnovány jsou záznamy na discích, kompaktních discích nebo kazetách

- k Dvojměrná neprojektovatelná grafika (Two-dimensional nonprojectable graphic)
Jméno kategorie se nemění, zahrnovány jsou grafy, koláže, počítačová grafika, výkresy, technické výkresy, obrazy, malby, obrázky, fotografie, fototisky, foto CD, pohlednice, fotomechanické reprodukce aj.
- m Počítačový soubor (Computer file)
Jméno kategorie se nezměnilo, zahrnovány jsou různé typy elektronických zdrojů: počítačové programy číselná data, počítačově orientovaná multimédia, online systémy a služby; uvedené typy zdrojů lze případně zařadit do jiné kategorie, pokud by v nich byl výrazně zastoupen nějaký aspekt z takové kategorie; jiné typy elektronických zdrojů jsou zařazovány do všech ostatních kategorií
- O Souprava (Kit)
Jméno kategorie se nezměnilo, zahrnovány jsou materiály složené ze dvou nebo více jednotek různého charakteru, z nichž ani jedna není dominantní
- p Smíšený materiál (Mixed material)
Jméno kategorie zůstává stejné, změněné je vymezení obsahu; zahrnovány jsou materiály složené z dokumentů v jedné nebo více formách, které jsou spojeny nějakou nebo kvůli nějaké osobě či korporaci. Patří sem archivní fondy a rukopisné sbírky složené z různých forem materiálů (texty, fotografie a zvukové záznamy)
- r Trojrozměrný artefakt či přírodní předmět (Three-dimensional artifact or naturally occurring object)
Kategorie se stejným jménem, zahrnovány jsou předměty vytvořené člověkem (modely, puzzle, sochy a jiná trojrozměrná umělecká díla a jejich reprodukce, stroje, oděvy, hračky aj.)
- t rukopisný [textový] jazykový materiál (Manuscript language material)
Kategorie se stejným jménem, zahrnovány jsou rukopisné textové materiály

Výše uvedené třídění je v jisté sumarizační podobě uplatněno také v rámci pole 006 a potažmo v poli 008. Specifikace kódů je definována pro následující druhy materiálů – knihy, počítačové soubory/elektronické zdroje, mapy, smíšené materiály, hudba, pokračující zdroje a vizuální materiály. Vnitřní typologie samotných počítačových či elektronických zdrojů zůstává zatím prakticky beze změn oproti bývalému americkému formátu USMARC.

3.2.3. UNIMARC v návaznosti na AACR2R a ISBD (ER)

Formát UNIMARC je pro potřeby katalogizace v knihovnických systémech aplikován stále v řadě zemí (dříve byl aplikován i v ČR), zaznamenal v průběhu vývoje řadu dílčích změn vyvolaných zejména mohutným rozvojem síťových elektronických zdrojů a přístupů k nim. Vznik a rozvoj jsou do značné míry závislé na amerických standardech, lze konstatovat, že to, co bylo uvedeno v komentáři k americkým formátům, bude platit i pro formát UNIMARC, zejména pokud jeho obsahová náplň úzce souvisí s katalogizačními pravidly AACR2R.

Počítačové soubory včetně síťově dostupných jsou také zařazeny v celkovém, opět pragmaticky vymezeném, třídění informačních materiálů. Jejich kódovaná hodnota je uvedena v návěští záznamu v pozici 06 (typ záznamu). Přehled tříd není zcela totožný s americkými formáty, proto by bylo vhodné jej připomenout.

Třídění obsahuje *kategorie*:

- a jazykové materiály, tištěné
- b jazykové materiály, rukopisné
- c hudební partitury, tištěné
- d hudební partitury, rukopisné
- e kartografické materiály, tištěné
- f kartografické materiály, rukopisné
- g projekční materiály a videozáznamy (filmy, videozáznamy, diafilmy, diapozitivy, průsvitky aj.)
- i zvukové záznamy, nehudbní nahrávky
- j zvukové záznamy, hudbní nahrávky
- k dvojrozměrná grafika (obrazy, kresby aj.)
- l elektronické zdroje
- m multimédia
- r trojrozměrné artefakty a reálie

Definice současně platné verze formátu v tomto případě již připouští, že katalogizovaný materiál, i když bude elektronický, by mohl být zařazen do jiné kategorie než do elektronických zdrojů (kód „l“), a to na základě svého nejdůležitějšího znaku či vlastnosti -

typu obsažené informace. Tento princip byl potvrzen a upřesněn i na jednom ze zasedání příslušné komise pro rozvoj formátu. Problémem je skutečnost, že řada kategorií je přesně vymezena také formou (tisk nebo rukopis), což nejde dohromady s případnou formou elektronickou. V praxi systémů, které jsou vázány v rámci formátu UNIMARC přímo na AACR2R, mohou tak nastat určité potíže, protože základní koncept těchto pravidel vychází z pravidla 0.24, které stanovuje zpracování jednotky podle základní přidělené kategorie v rámci třídění. I v případě třídění informačních zdrojů formátu UNIMARC lze konstatovat, že se v něm prolínají protichůdná kritéria – obsahu **a fyzického nosiče**, což vedlo v knihovnické praxi mnohdy k problémům. Dokud nebude zásadně vyřešena otázka základního třídění s následnou kodifikací v aktualizovaných standardech, budou pokračovat jisté potíže v katalogizační praxi. Konečným důsledkem pak budou nejasnosti, problémy a ztráty informací při procesu jejich vyhledávání ze strany koncových uživatelů.

Poměrně velké změny při zpracování elektronických zdrojů v rámci formátu UNIMARC nastaly po zveřejnění revidované normy ISBD(CF), která kromě jiného změnila i svůj název – starý termín „počítačové soubory“ byl nahrazen moderním termínem **„elektronické zdroje“**. Takové rozhodnutí lze jen uvítat, protože tento termín daleko lépe postihuje vlastnosti nejen lokálně dostupných elektronických zdrojů, ale zejména zdrojů dostupných přes počítačové sítě. S ohledem na zcela nové vlastnosti síťových zdrojů samotných (včetně těch, které zatím ani nebyly pojmenovány) a způsoby jejich rozšiřování lze konstatovat, že užití termínu „elektronický dokument“, jenž byl uplatněn také v rámci návrhu NK ČR k tvorbě záznamů speciálních druhů dokumentů pro souborné katalogy, není již zcela vyhovující pro pojmenování množiny všech typů elektronických informačních zdrojů. Termín „elektronické zdroje“ je tak doporučen k aplikaci v rámci základního třídění v návěští formátu UNIMARC (nová slovní hodnota kódu „I“), ale také v popisu jako hodnota všeobecného označení materiálu (GMD) v podpoli „b“ pole 200.

Podstatnou novinkou ISBD(ER) je nová, poměrně rozsáhlá třístupňová typologie elektronických zdrojů určená pro oblast jejich specifického popisu „Typ a rozsah zdroje“.

Typologie obsahuje:

Elektronická data

Elektronické soubory znaků (fonty)

Elektronická obrazová data

Elektronická číselná data

Elektronická statistická data ze sčítání lidu

Elektronická přehledová data

Elektronická obrazová data

Elektronické mapy

Elektronická zvuková data

Elektronická textová data

Elektronické bibliografické databáze

Elektronické dokumenty ve smyslu textu (např. dopisy, články)

Elektronické časopisy

Elektronické bulletiny

Elektronické programy

Elektronické aplikační programy

Elektronické programy pro počítačově podporovaný design (CAD)

Elektronické databázové programy

Programy pro elektronické publikování

Elektronické hry

Elektronické tabulkové procesory

Elektronické psací editory

Elektronické systémové programy

Elektronické operační programy

Elektronické programovací jazyky

Elektronické vyhledávací programy

Elektronické obslužné programy

Elektronická data a programy

Elektronická interaktivní multimédia

Elektronické online služby (např. diskusní skupiny, elektronické konference, website).

I tato typologie vychází na první úrovni z rozdělení elektronických zdrojů na zdroje obsahující data, prezentující sociální informace komunikované ve společnosti, a na

programy, prezentující informace strojové povahy sloužící především pro fungování počítačových systémů samotných. Specifickou skupinu tvoří zdroje, které jsou kombinací předchozích dílčích typů. Na druhé a třetí úrovni jsou prezentovány podtypy informačních zdrojů, jejichž výčet je v současné chvíli diskutabilní (viz např. neúplná podskupina elektronických textových dat nebo blíže nerozpracovaná podskupina online služeb). Typologie je však otevřená a předpokládá se její doplňování i další zpřesňování – zejména ve skupině, kde se objevují novodobé síťově dostupné zdroje. V každém případě znamená krok kupředu.

V rámci formátu UNIMARC se typologie ISBD(ER) uplatnila jednak v podpoli „a“ pole 135 v pozici 0 pro kódované informace, jednak v poli 230, kde se uplatňují slovní termíny dané typologie. Typologie se povinně týká síťových zdrojů. V tomto směru je zajímavé i její porovnání s typologiemi uplatněnými v systémech provozovaných přímo v prostoru sítě Internet nebo WWW. V poli 135, podpoli „a“ se nově objevuje i další znaková pozice (1) pro kódování nosičů. Množina síťových zdrojů je jednoznačně vymezena kódem „r“ (online systémy) – srovnej se stejnou hodnotou v poli 007 formátu USMARC a MARC 21.

Formát UNIMARC zavedl dle amerického vzoru také nové (shodně označené) pole 856 pro údaje o způsobech přístupu k elektronickým zdrojům v prostředí sítě a o dalších technických detailech. Definice polí 856 nejsou zcela shodné (např. v UNIMARCu není obsazen 2. indikátor). Podstatná informace o způsobech přístupu k informačním zdrojům, zejména v síti Internet k dispozici ale je, takže lze i v tomto případě pro potřeby vyhledávání případně vymezovat množiny jednotlivých zdrojů podle specifického přístupu.

3.3. Typologie internetových a webových zdrojů v informačních systémech Internetu a WWW

Velký rozvoj síťových informačních zdrojů i způsoby jejich zpřístupňování způsobily významné změny v oblastech praktického zpracování a efektivního zpřístupňování informačních zdrojů, a to zejména v jednotlivých institucích a systémech. V rámci procesů členění lidských pracovních činností a operací a vzhledem ke společenské dělbě práce vznikly a dále se rozvíjejí novodobé instituce a systémy, které se díky výhodným technologickým podmínkám této činnosti věnují. Kromě známých systémů webu typu „search engines“, nebo předmětových katalogů Internetu, které registrují informační zdroje v nestrukturované podobě, se v celosvětovém měřítku rozvíjí řada registračních systémů. Tyto systémy

přístupují ke zpracování informačních zdrojů přes strukturované záznamy obsažené někdy i ve zdrojích samotných (metadata). V rámci navrhovaných formátů je zpravidla vždy také řešen problém jejich třídění, nebo typologie. My si v následující přednášce představíme tři reprezentanty:

3.3.1. Kanadská typologie webových zdrojů „VW96“

Mezi často citované typologie internetových a webových zdrojů patřila koncem 90. let 20. stoletá typologie kanadské firmy „*Vancouver Webpages*“, zabývající se službami v oblasti Internetu. Firma provozovala webový vyhledávací systém⁵³, se zaměřením na servery v regionu Britské Kolumbie. V rámci systému byl vyvinut speciální interaktivní formulář k tvorbě údajů o webových zdrojích (metadat)⁵⁴. Typologie vyplývající z těchto aktivit je stále k dispozici na WWW⁵⁵ a byla označena jako „**typologie objektů**“⁵⁶ s přesným označením metatagu VW96.ObjectType. Předmětem registrace typologie byly pouze zdroje ve formátu html a vrm1, což znamená, že nebyla zahrnuta celá množina webových zdrojů.

Typologie byla založena na jednoduché koncepci a vycházela z rané verze známé typologie zdrojů „Dublin Core“. Představuje krátký abecedně uspořádaný seznam 21 typů se stručnými definicemi v angličtině, ovšem princip vymezení typů (model, hledisko, atd.) není uvedeno. Z uvedeného je zřejmé, že typologie je při vymezení předmětu registrace neúplná, velmi pragmatičtá a účelově orientovaná. Podstatná je nevyváženost uváděných typů, zahrnující typy obecné, ale i typy velmi specifické (např. z oblasti obchodu a počítačové vědy).

Základním typem naznačené typologie je položka „*dokument*“, která je charakterizována jako standardní (běžná) a zřejmě byla vytvořena pro účel, že uživatel ji musel použít v případě, že nenašel v seznamu potřebný jiný typ. Zařazení této položky je ovšem z hlediska logiky značně problematické⁵⁷.

Nejčtenější část typologie je věnována množině typů zdrojů, které známe z oblasti *typologií tradičních statických dokumentů*. Tvůrci stanovili typy, u nichž dominoval znak

⁵³ searchBC a vlastní robot VWbot (Vancouver Webpages Robot).

⁵⁴ Meta tag builder – dostupné z WWW: <<http://vancouver-webpages.com/META/mk-metas.html>>

⁵⁵ Dostupný z WWW: <<http://vancouver-webpages.com/META/VW96-schema.html>>

⁵⁶ angl. object type

⁵⁷ Není jasné, zda některé položky typologie byly také něčím jiným, než dokumenty.

primárnosti informace a její grafické textové prezentace: „**kniha**⁵⁸“; „**slovník**“; „**časopis**“; „**magazín**“; „**manuál s technickým zaměřením**“.

Speciálními důležitými znaky se však již vyznačovaly další typy textových zdrojů, které typologie obsahovala. Zahrnuty byly především výrazné internetové zdroje typu „**FAQ**“⁵⁹, které mají charakter instruktážní a propagační a typu „**RFC**“⁶⁰, které zahrnují standardy, protokoly a další materiály týkající se sítě Internet. Do této skupiny je možné dále zařadit typy zdrojů z oblasti *ekonomicko-obchodní*, jako je specifický typ „**linecard**“ (seznam produktů nebo obchodních značek) a „**catalog**“ (katalog jako seznam položek ve skladech či pro prodej), nebo z oblasti počítačových služeb „HowTo“ (jakési internetové kuchařky, týkající se SW a HW).

Komplexnější povahou se vyznačují následující *dva typy zdrojů*, které představují již skupiny zdrojů vytvářejících určité logické celky. Tvůrci uváděli typ „**keybank**“ (schránka pro kryptografické aplikační programy, např. známý „PGP“⁶¹) nebo známější typ „**archive**“ (archiv počítačových programů nebo souborů).

V typologii je možné dále vyčlenit *typy informačních zdrojů*, jež jsou specifické zejména pro prostředí Internetu či WWW a jejichž důležitým znakem je *interakce mezi uživatelem a systémem*. Slovo „systém“ napovídá, že užití slova „dokument“ by nebylo už asi v dané situaci zcela adekvátní a že frekventovaný termín „zdroj“, popř. „informační zdroj“ by byl vhodnější. Tvůrci uváděli typ „**database**“ (databáze, ve smyslu databázového systému jakéhokoliv zaměření), dále „**hypercatalog**“ (hyperkatalog), kterým byly míněny pouze webovské **vyhledávací systémy předmětového typu** (předmětové katalogy, jako je známý Yahoo! aj.) a základní kategorii vyhledávacích systémů typu „**search engines**“. Volněji by se dal do této skupiny ještě zařadit typ „**index**“ (index, rejstřík, seznam nějakých zdrojů v prostoru Internetu nebo WWW). *Ekonomicko-obchodní orientace* typologie měla v této skupině zastoupení ještě jednou položkou typem „**mall**“ (webovský online obchod, kde se prodávají výrobky), který lze také jen stěží označit jako dokument, minimálně z hlediska jeho hlavní funkce. Vhodnější by bylo užití slova „služba“, popřípadě i „systém“.

Zvláštní dojem vyvolávalo zařazení typu zdroje „**home page**“ (domovská stránka organizace nebo nějaké fyzické osoby). Fyzický počítačový soubor takové stránky je sice

⁵⁸ angl. book; dictionary; journal; magazine; manual

⁵⁹ frequently asked questions – často pokládané otázky

⁶⁰ request for comments – žádost o komentář

⁶¹ angl. pretty good privacy od P.R. Zimmermanna

základní registrační jednotkou např. pro vyhledávací systémy, ale je, dá se říci, vstupní bránou, „titulní stránkou“ menšího či většího komplexu webovských zdrojů, které mohou tvořit dnes i koncepčně velice složité a obsahově významné „webové informační systémy“, tedy informační zdroj institucionálního typu (v angl. se užívá frekventovaného termínu „website“). Konečně, i v předchozím odstavci jmenované „search engines“ jsou institucionální jednotky, které na WWW mají své domovské stránky. Uvedení tohoto typu v kanadské typologii bylo stejně problematické, jako uvedení typu „dokument“.

Celkový výčet zastoupených *typů informačních zdrojů* (objektů) je na závěr nutné doplnit *dvěma posledními*, které reprezentují trojrozměrné reálné objekty. Šlo o zdroje ve formátu **VRML**. Přestože byly tyto zdroje z hlediska formátu již v rámci registračního systému označeny jako zdroje VRML, stejně jako zdroje HTML (to znamená, že zde bylo uvedeno třídění všech registrovaných zdrojů, resp. objektů), byly ještě navíc v typologii vlastně zastoupeny znovu, a to ve dvojí podobě: buď jako *obecný VRML* typ „**world**“ (svět), který je zároveň typem standardním („běžným“), nebo jako typ „**real world**“ (reálný svět), jehož „.wrl“ soubor musel být opatřen dalšími potřebnými atributy (měřítko, geografická lokalizace, schéma pro geografické souřadnice aj.).

Pragmatická typologie „VW96“ patřila k jedněm z prvních, které se na WWW vyskytly (1996). Z hlediska potřeb dnes budovaných systémů pro registraci internetových či webových zdrojů nebude využívána. Je ovšem nutné vědět o její existenci, protože se stala východiskem pro zpracování řady dalších typologií a třídění běžně užívaných.

3.3.2. Typologie britského systému ROADS

Další typologie internetových a webovských zdrojů souvisí s významným projektem, který byl na webu budován ve Velké Británii v rámci rozsáhlého Programu elektronických knihoven „eLib“ (Electronic Libraries Programme). Jeho název je „**ROADS**“ (Resource Organisation And Discovery in Subject-based services, Organizace a zjišťování zdrojů v předmětově založených službách) a jeho základním posláním je registrace a vyhledávání webovských informačních zdrojů⁶². Typologie souvisí především s tvorbou záznamu o zdrojích pomocí metadat a byla postupně formována v souvislosti s rozvojem samotného systému ROADS, ale i v souvislosti s rozvojem dalších systémů, zejména mezinárodního projektu pro metadata „Dublin Core“.

⁶² Kirriemuir, J., nestr.

Východiskem typologie z roku 1997⁶³ se stala typologie „žánrů zdrojů“ systému BibTeX. Autoři sami také v úvodu hovořili o „**žánru zdroje**“ (angl. genre of the resource). Typologie obsahovala seznam celkem 36 abecedně uspořádaných položek opatřených velmi stručnou definicí, šlo-li ovšem vůbec o definici. Hlediska typologie nebyla nijak specifikována. V typologie byla propracovanější, univerzální, zahrnovala další typy zdrojů, zejména z hlediska prezentace informací, a v oblasti textových dokumentů, se snažila jít až na úroveň analytických jednotek registrovaných zdrojů. Z celkového přehledu bylo patrné, že u jejího zrodu stáli také knihovníci.

Ačkoliv abecední sestava jednotlivých typů nebyla nijak dále členěna ve smyslu vymezení dílčích skupin podle nějakého významného znaku (vlastnosti) zdroje/dokumentu, bylo možné je přesto uměle identifikovat. Největší skupinu tvořily *primární textové zdroje* prezentované dnes na WWW (dříve publikované klasickou cestou přes nakladatele). Zahrnuty byly *základní univerzální* i *specializované typy elektronických publikací*, dokonce i žánry krásné literatury:

„**book**“ (kniha), „**booklet**“ (brožura), „**collection**“ (sborník či sbírka), „**manual**“ (manuál), „**proceedings**“ (sborník z konference), „**preprint**“ (preprint vědeckého článku), „**journal**“ (vědecký časopis), „**magazine**“ (populární časopis či magazín), „**newspaper**“ (noviny), „**course material**“ (materiál k výuce, jako jsou teze, osnovy, rozvrhy apod.), „**research paper**“ (výzkumná zpráva), „**tech report**“ (technická zpráva) a „**master thesis**“ (magisterská práce), „**PhD thesis**“ (doktorská práce), „**honour thesis**“ (práce k udělení čestného titulu).

Z krásné literatury byl zastoupen pouze žánr „**poem**“ (báseň). Za zvláštní bylo možné pokládat i zařazení typu „**advertisement**“ (reklama), protože jiné formy z této oblasti už zařazeny nebyly. Oblast *elektronických sekundárních textových zdrojů* byla zastoupena typem „**bibliography**“ (bibliografie). Protože se v rámci WWW dnes zveřejňují i historické rukopisné dokumenty a také novodobé písemné dokumenty, které nebyly z různých důvodů publikovány klasickou cestou přes vydavatelství, zahrnuli autoři také typ „**unpublished**“ (nepublikovaný). Typologie v této skupině nebyla vyčerpávající, výčet však dále nepokračoval, a snad i proto byl nabídnut jeden naprosto *obecný typ* „**misc(ellaneous)**“ (ostatní díla).

⁶³ Autory byli John Knight a Martin Hamilton

Některým výše uvedeným typům textových zdrojů odpovídaly i jejich analytické ekvivalenty. Zařazeny byly „**in book**“ (část v knize), „**in collection**“ (část ve sborníku či sbírce), „**in proceedings**“ (část ve sborníku z konference), „**article**“ (článek v lektorovaném časopise) a „**unrefereed article**“ (nelektorovaný článek z časopisu, magazínu nebo novin).

Významnou funkcí typologie je, že zahrnuje zdroje s dalšími *typy prezentace informací*, které také známe z období předelektronického a které jsou ovšem dnes významnou

součástí elektronických zdrojů: „**image**“ (obraz), „**video**“ (video, videozáznamy) a „**music**“ (hudební zdroje, které ovšem nebyly v typologii blíže charakterizovány, takže nebylo možné zjistit, jestli pod ně spadaly také zvukové informační zdroje).

K *novodobým typům elektronických zdrojů* patří v této verzi typologie především jednotlivá sdělení (zprávy) v elektronických konferencích (listech) a skupinách, a to „**message on moderated mailing list**“ (zpráva v moderované elektronické konferenci), „**message on unmoderated mailing list**“ (zpráva v nemoderované elektronické konferenci), „**posting to moderated newsgroup**“ (dopis do moderované skupiny newsgroup) a „**posting to unmoderated newsgroup**“ (dopis do nemoderované skupiny newsgroup). Zařazen byl i typ „**dataset**“ (soubor dat určitého druhu). Zvláštní místo zaujal typ zdroje „**service**“ (služba), který vycházel nad rámec tradičně definovaného pojmu „dokument“, a dále poslední položky seznamu „**organisation info**“ (informace o organizaci nebo nějaké skupině, což může být jedna domovská stránka nebo i komplex všech WWW stránek) a „**personal info**“ (jedna personální domovská stránka nebo komplex stránek vztahujících se k jedné fyzické osobě).

Výše uvedený návrh typologie zdrojů systému ROADS je dnes pouze historickým dokumentem. Současně platnou, značně redukovanou typologií informačních zdrojů systém představuje *formou navržené množiny typů formulářů pro zpracování jednotlivých typů zdrojů*, resp. „žánrů zdrojů“. Seznam typů formulářů je derivovanou a zároveň doplněnou množinou formulářů pracovní skupiny IAFA (Internet Anonymous FTP Archives) patřící pod organizaci IETF (Internet Engineering Task Force). V současné době obsahuje následující *typy informačních zdrojů a typy specifických klastrů*:

Typy informačních zdrojů:

Collection (sbírka - experimentální typ)
Dataset (soubor dat)
Document (dokument)
Dublin Core
Event (událost -experimentální typ)
Image (obraz)
Mailarchive (archiv elektronické konference)
Project (projekt)
Resource (zdroj, platný v síti RDN)
Service (služba)
Software (počítačový program)
Sound (zvuk)
Trainmat (výukový materiál)
Usenet (diskusní skupina usenet/newsgroup)
Video (video)

Klastry:

Organization (organizace)
User (uživatel)
Agent (agent, kategorie užívaná v Dublin Core)

Seznam typů informačních zdrojů a speciálních klastrů má otevřený charakter, a je velice zajímavý. Jde o hrubou typologii, která také zahrnuje základní typy informačních zdrojů, (tj. dokument ve smyslu textu, obraz, zvuk aj.). Především se však zaměřuje na typická novodobá **seskupení webovských informačních zdrojů**, statické nebo dynamické povahy, jako např. elektronické sbírky či archivy, služby či systémy apod.

Oficiální typologie systému ROADS znamenala jistý kompromis, je však známo, že některé partnerské dílčí systémy registrovaly zdroje i na analytické úrovni (např. články z elektronických periodik), protože to považovaly za potřebné. Analytičnost, která byla

významným rysem typologie původní, byla jinak potlačena (části publikací či zprávy z konferencí).

Z původní typologie se v seznamu objevilo přímo pouze 5 typů – „**soubor dat**“, „**obraz**“, „**služba**“, „**video**“ a ještě „**sbírka**“ (jen experimentálně). Během výzkumu, který sledoval využívání jednotlivých typů formulářů, se zjistilo, že nejvyužívanějším typem zdroje je „služba“ (až 70 %), což vedlo tvůrce systému ROADS k úvaze zavést další typy, jako např. „archiv elektronických dokumentů či zpráv“. O něco menší využití zaznamenal nově zavedený univerzálně použitelný typ „**dokument**“, který nahradil původní širší skupinu dílčích typů elektronických publikací monografické povahy a který se však z druhé strany u některých systémů užívá takřka výlučně pro všechny registrované jednotky. Po diskusi, se pod tento typ skryl i původně samostatně navržený experimentální typ „**FAQ**“. Třetí nejuzívanější typ „**archiv elektronické konference**“ nahradil původní analytické typy zpráv z konferencí, stejně jako typ „**diskusní skupina usenet**“. Typ „**soubor dat**“ je prozatím využíván pouze u některých oborových systémů (např. sociologie) a jde především o databanky faktografických informací. Sólové obrazové zdroje a video prozatím figurují v minimální míře.

Totéž platí pro nově zavedený typ „**zvuk**“ a „**software**“. Po zvážení byl zařazen ještě typ „**projekt**“, který byl ovšem míněn širěji, než dříve samostatné zdroje typu „výzkumná či technická zpráva“, neboť na webu dnes existuje již velké množství informačně bohatých souborů výzkumných a vývojových úloh, které v sobě integrují jak prezentační webové stránky, tak zdroje zpráv či dalších informací a popř. interaktivních databází. Totéž je možné uvést o typu „**událost**“ určeného pro akce různého zaměření, který rovněž nahradil typy dílčích dokumentů „konferenční sborník“ nebo „preprint“. Jeho zařazení je experimentální. Jde v každém případě o žádané hodnoty při rešerších. Novinkou byl i typ „**výukový materiál**“ a typ „**zdroj**“, užívaný v projektu sítě RDN⁶⁴.

Speciálními kategoriemi jsou 3 typy formulářů představující záznamy o specifických entitách – fyzických osobách a korporacích, které vystupují v roli tvůrců, autorů, distributorů výše uvedených zdrojů. Připomínají známé záznamy autorit z katalogizačních systémů.

Typologie systému ROADS je velice zajímavým experimentem v této oblasti. Rozvoj samotného systému sice již nepokračuje, rozvíjeny však jsou jeho jednotlivé systémy

(předmětová gateway), které registrují a zpracovávají internetové zdroje. Poslední verze typologie zdrojů je k tomu stále hojně využívána.

3.3.3. Typologie formátu „Dublin Core“

Nejvýznamnějším formátem pro tvorbu dat o síťových informačních zdrojích je v současnosti formát označovaný zkratkou „**DC**“, pod níž se skrývá zkrácené jméno „**Dublin Core**“ (Dublin Core), v úplném tvaru pak „**Dublin Core Metadata Element Set**“ (Soubor metadatových prvků Dublin Core). U jeho zrodu v roce 1995 stáli pracovníci výzkumného oddělení OCLC, kteří ve spolupráci s dalšími odborníky z celého světa založili mezinárodní virtuální společenství **DCMI** (Dublin Core Metadata Initiative). Formát DC se během posledních deseti let stal významnou mezinárodní záležitostí. Je využíván v řadě projektů a systémů pro registraci internetových a webovských zdrojů po celém světě, a to buď přímo (tvorba metadat v rámci dokumentů HTML) nebo je zajišťována konverze do jiných formátů pro metadata či knihovnických formátů typu MARC. Specifikace „Dublin Core“ se stala základem významné aplikace nově navrženého jazyka „XML“ (eXtensible Markup Language), která je známa pod zkratkou „RDF“ (Resource Description Framework, Rámec pro popis zdrojů), jež umožňuje kódování, tvorbu a využívání strukturovaných metadat. Pro naše účely bude nutné zabývat se jen jediným momentem celého formátu, a to jeho *typologie zdrojů nahlížená v historickém kontextu*.

Typologie zdrojů byla a i nadále je významnou částí základní specifikace DC, kterou tvoří v současné chvíli celkem 15 prvků (metadat). Jde o údaj „**Resource Type**“ (Typ zdroje), jenž vychází z typologie informací obsažených ve zdrojích, a údaj „**Format**“ (Formát), který je založen na typu uspořádání dat v počítačovém souboru přenášeném v síti.

Základní typologii je věnována značná pozornost od samého vzniku formátu. Podílela se na ní a stále ještě se podílí komunita odborníků z oblasti knihovnictví i počítačové vědy, jednotlivé návrhy a problémy byly diskutovány na pracovních seminářích a v elektronické konferenci. V současné době je jí věnována péče v rámci stálé pracovní skupiny.

Třístupňová typologie informačních zdrojů systému „Dublin Core“ (verze z roku 1997):

Text

⁶⁴ angl. resource description network

Text.Abstract
Text.Advertisement
Text.Article
Text.Correspondence
Text.Correspondence.Discussion
Text.Correspondence.Email
Text.Correspondence.Letter
Text.Correspondence.Postcard
Text.Dictionary
Text.Form
Text.Homepage
Text.Homepage.Organizational
Text.Homepage.Personal
Text.Index
Text.Manual
Text.Manuscript
Text.Minutes
Text.Monograph
Text.Pamphlet
Text.Poem
Text.Proceedings
Text.Promotion
Text.Seriál
Text.Serial.Journal
Text.Serial.Magazine
Text.Serial.Newsletter
Text.Serial.Newspaper
Text.TechReport
Text.Thesis
Text.Thesis.Doctoral
Text.Thesis.Masters

Image
Image.Moving
Image.Moving.Animation
Image.Moving.Film
Image.Photograph
Image.Graphic

Sound
Sound.Ambient
Sound.Effect
Sound.Music
Sound.Narration
Sound.Speech

Software

Software.Executable
Software.Source

Data
Data.Numeric
Data.Spatial
Data.Spectral
Data.Statistical
Data.Structured-Text

Interactive
Interactive.Chat
Interactive.Games
Interactive.Multimedia
Interactive.VR

Verze z roku 1997 byla *třístupňová*. První úroveň představovala 6 základních typů informačních zdrojů, jež vycházely z typu informace v nich obsažené. Některé korespondovaly se známou typologií sociálních informací (**text, obraz, zvuk**), popř. také s typologií MIME. Reprezentantem hybridního typu informace byl typ zdroje „**program**“ (software), typ „**data**“ a zejména novodobý reprezentant síťových zdrojů „**interaktivní zdroj**“ (interactive). Druhá a třetí úroveň typologie zahrnovala podtypy informačních zdrojů, vymezených *souhrnem znaků*, k nimž náležely minimálně jejich *funkce a cílové určení a forma uspořádání informací*.

Nejrozsáhlejší dělení bylo definováno pro typ „**text**“, v jehož rámci se vyskytovaly jak podtypy známé z tradiční typologie textových dokumentů (monografie, seriál, technická zpráva, korespondence aj.), tak podtypy novodobé (formulář, domovská stránka aj.). Výčet podtypů nebyl úplný, pro nové verze se předpokládala další úprava a doplňování. Totéž platilo i pro podtypy dalších speciálních typů. Komplikovaná situace byla ve vymezení typů „**data**“ (soubory dat) a „**interaktivní zdroje**“, které byly doplňovány i v souvislosti s rozvojem informačních technologií.

Další návrh typologie informačních zdrojů DC pochází z roku 1998⁶⁵. Nyní si uvedeme přehled základních typů se stručnou charakteristikou.

Jedná se o následujících 8 typů informačních zdrojů:

⁶⁵ Připraveny 4 verze, poslední úprava schválena 23. 10. 1998.

1. **text** (text) - zdroj, jehož obsah je určen především pro čtení (např. knihy, dopisy, básně, noviny aj.); k tomuto typu zařazovali tvůrci typologie i faksimile nebo obrázky textů
2. **obraz** (image) - zdroj, jehož obsah představuje symbolickou vizuální (zrakovou), resp. obrazovou reprezentaci, ale jinou než u textu (např. obrázky, fotografie fyzických objektů, obrazy, kresby, animace, filmy, videozáznamy, diagramy, mapy, hudební notace)
3. **zvuk** (sound) - zdroj, jehož obsah je určen především k poslechu (např. hudba, projev, záznam zvuku)
4. **soubor dat** (dataset) - zdroj, jímž se rozumí strukturované informace kódované v seznamech, tabulkách, databázích apod., které se normálně vyskytují ve formátu pro přímé strojové zpracování (např. tabulkové procesory, databáze, data geografických systémů aj.); nestrukturované číselné nebo slovní údaje jsou považovány za typ „text“
5. **program** (software) - počítačový program ve zdrojové nebo kompilované formě, který je dostupný trvale pro instalaci na jiných počítačích; programy, které vytvářejí interaktivní prostředí při komunikaci, je nutné zařadit k typu „interaktivní zdroj“
6. **interaktivní [zdroj]** (interactive) - zdroj vyžadující interakci uživatele, aby informace byly pochopeny, provedeny nebo využity (např. formuláře na webovských stránkách, aplety, multimediální výukové předměty, konverzační služby, virtuální realita)
7. **událost** (event) - zdroj s časově omezeným výskytem (např. výstava, konference, představení aj.); metainformace o události (akci) nemusí identifikovat vyhledatelný zdroj, pokud její čas již vypršel a více se nevyskytuje
8. **fyzický objekt** (physical object) - trojrozměrné objekty nebo látky, které nelze považovat za texty, obrázky nebo jiné typy výše uvedené (např. osoba, počítač, velká pyramida, socha); digitální reprezentace nebo zástupce uvedených objektů musí tedy figurovat jako typ „obraz“, „text“ apod.

Novinkou této verze typologie bylo zařazení typu „**událost**“ a „**fyzický objekt**“. Byly posuzovány i pojmy „**složený (kompozitní) zdroj**“, popř. „**smíšený zdroj**“ a pojem

„*sbírka*“ (collection). Pro základní jednoduchou variantu „DC.Typy“ byly však tehdy odmítnuty, a to především z důvodů jejich vyhledávání. Složené zdroje bylo výhodnější zařazovat k více dílčím typům (např. multimediálnímu programu s jedním URL je možné přiřadit 4 typy – zvuk, text, obraz a interaktivní zdroj). Sbírký (kolekce) zdrojů měly být přiřazeny spíše k jednomu základnímu dílčímu typu. Na místě bylo také uplatňování vazeb (relací), zvažovalo se i jejich případné uplatnění jako podtypů (např. Text.Sbírka).

Nejrozsáhlejší verze typologie DC všech informačních zdrojů (včetně tradičních) byla přijata v srpnu 1999. Zahrnuty byly (kurzívou jsou označeny tehdy nově navržené typy):

1. *Sbírka (Collection)*
2. Soubor dat (Dataset)
3. Událost (Event)
4. Obraz (Image)
5. Interaktivní zdroj (Interactive resource)
6. *Model (Model) jako abstrakce*
7. *Strana (Party) jako osoba nebo korporace*
8. Fyzický objekt (Physical object)
9. *Místo (Place) jako geografická entita*
10. *Služba (Service)*
11. Program (Software)
12. Zvuk (Sound)
13. Text (Text)

Novými položkami byl typ „*Sbírka*“, který byl specifikován jako agregát (soubor) dokumentů, dále typ „*Model*“, považovaný za symbolickou reprezentaci (abstrakci reálné věci), typ „*Strana*“, jíž mohla být fyzická osoba, organizace nebo instituce, a typ „*Místo*“ ve smyslu geografické oblasti. Poslední tři jmenované typy, které více specifikovaly původně navržený typ „Fyzický objekt“, nebyly přijaty jednoznačně řadou odborníků, a proto byly z dalších návrhů vyřazeny. Posledním novým typem, zařazeným do sestavy v roce 1999, byl typ „*Služba*“, jíž je míněn systém zajišťující jednu nebo více funkcí koncovým uživatelům (například MVS, autentikační služba, bankovní služba aj.).

Typologie informačních zdrojů DC z července 2000 znamenala návrat k menšímu počtu typů. Zařazeny byly:

1. Sbírka (Collection)
2. Soubor dat (Dataset)
3. Událost (Event)
4. Obraz (Image)
5. Interaktivní zdroj (Interactive resource)
6. Služba (Service)
7. Program (Software)
8. Zvuk (Sound)
9. Text (Text)

Své místo si uhájily typy „**Sbírka**“ a „**Služba**“, po dlouhých debatách však byl (ale jenom na určitý čas) vyňat problémový a velmi diskutovaný typ „**Fyzický objekt**“. Do typologie zdrojů byl tento typ opět zařazen až v červenci 2002, takže jejich celkový počet dosáhl počtu deseti typů zdrojů. Tento stav v podstatě potvrdila i další verze z února 2003. Nový směr v rozvoji typologie všech informačních zdrojů znamenala až verze z 19. listopadu 2003. Její novinkou bylo (kromě formální úpravy celého seznamu) zařazení dvou nových typů, které znamenaly sémantické zjemnění typu „**Obraz**“. Šlo o typy:

➔ **Stálý obraz (Still Image)**, který je definován jako statická vizuální reprezentace (malba, plán, mapa, kresba, grafický design aj.) a

➔ **Pohyblivý obraz (Moving Image)**, který je definován jako série vizuálních reprezentací, které, když jsou ukazovány, vyvolávají dojem pohybu (animace, film, televizní program, video aj.)

Zařazením těchto dvou nových typů se typologie DC opět, pokud jde o obrazové informace, se přiblížila typologii standardu internetové normy MIME. Celkový počet položek této verze typologie všech informačních zdrojů DC zahrnoval tedy dvanáct typů.

Zatím poslední verze typologie informačních zdrojů DC pochází z června 2004. Tyto verze potvrdila a s ní i 12 typů informačních zdrojů, změnou bylo pouze jejich abecední

uspořádání (dle angličtiny) v seznamu:

1. Sbíрка (Collection) - agregát jednotek, popisovaných jako skupina (popis dílčích jednotek je možný)
2. Soubor dat (Dataset) - informace kódované v definované struktuře, určené pro přímé strojové zpracování
3. Událost (Event) - zdroj s netrvalým a časově omezeným výskytem
4. Obraz (Image) - symbolická vizuální reprezentace jiná než text (obraz v elektronické či fyzické formě)
5. Interaktivní zdroj (Interactive resource) - zdroj, který vyžaduje interakci s uživatelem
6. Pohyblivý obraz (Moving Image) - série vizuálních reprezentací, které, když jsou ukazovány, vyvolávají dojem pohybu
7. Fyzický objekt (Physical object) - neanimovaný trojrozměrný objekt (předmět nebo látka)
8. Služba (Service) - systém zajišťující jednu či více funkcí pro koncové uživatele
9. Program (Software) - počítačový program ve zdrojové či kompilované formě
10. Zvuk (Sound) - zdroj, jehož obsah je určen především k poslechu
11. Stálý obraz (Still Image) - statická vizuální reprezentace
12. Text (Text) - zdroj, jehož obsahem jsou především slova ke čtení

V systému formátu DC se uplatňuje i další typologie síťových elektronických zdrojů, která má důležitý doplňkový význam pro jejich vyhledávání. Formálním hlediskem se v tomto případě stává konkrétní formát zdroje (uspořádání dat) a je jen logické a rozumné, že tvůrci DC sáhli v tomto případě po známém internetovém standardu typologie elektronických zpráv MIME v jeho poslední verzi RFC 2045 - RFC 2049. Celkový seznam schválených typů médií, který udržuje a neustále doplňuje úřad IANA (Internet Assigned Numbers Authority), je k dnešnímu dni již poměrně dlouhý, v rámci již existujících provozů se jich zatím však užívá méně.

Vybrané typy DC formátů elektronických zdrojů uplatňované dle typologie MIME:

text/plain

text/richtext

text/html

text/sgml

text/xml

image/jpeg

image/gif

image/tiff

audio/basic

video/mpeg

application/postscript

application/rtf

application/wordperfect5.1

application/pdf

application/powepoint

application/msword

application/pgp-encrypted

application/pgp-signature

application/marc

model/vrml

3.4. Závěr typologií

Rozbor vybraných třídění a typologií internetových a webových informačních zdrojů podává v celkovém průniku zajímavý, ale velmi komplikovaný obraz rozsahu a obsahu těchto dvou na sebe navazujících základních pojmů. Objektivně jde o jednu, resp. dvě velké množiny všech informačních zdrojů, avšak při pokusu poskládat pestrou mozaiku jednotlivých zjištěných skupin, kategorií, druhů a typů informačních zdrojů, uspořádat je za

účelem jejich porovnání a zhodnocení zjistíme, že to zdaleka není tak jednoduchou záležitostí, jak by se na první pohled mohlo zdát.

Rozsah našeho předmětu neumožňuje prezentovat podrobnější závěry a hodnocení, proto se zatím omezím pouze na následující poznámky:

- ➔ rozbor ukazuje velké a neustálé změny jak v tradičních knihovnických, tak v novodobých tříděních a typologiích systémů provozovaných přímo v prostoru Internetu; řada rozdílů, ale postupně se sblíží; nastává značný posun k podstatným znakům zdrojů – především typ a forma komunikované informace
- ➔ řeší se hlavní problém – zavedení jednotného principu pro hlavní třídění všech informačních materiálů a jeho sladění s dalšími návaznými typologiemi elektronických informačních zdrojů včetně síťových; vzájemné efektivní spolupráce
- ➔ typologie má velmi dobrou úroveň; mezi jednotlivými reprezentanty existují zatím jisté rozdíly; vzájemná spolupráce přináší své ovoce; typologii a třídění zdrojů rozvíjenou v rámci formátu „Dublin Core“
- ➔ rozbor třídění a typologií digitálních informačních zdrojů dostupných v síti Internet prokázal jejich další jasný trend: vyčleňování jednotlivých kategorií či typů je na základní úrovni realizováno podle **typu přenášené informace**; v další úrovni je velmi časté třídění podle **formy (formátu) uložené a přenášené informace**;
- ➔ v dalším stručném přehledu jsou uvedeny jednak základní typy zdrojů společné pro všechna třídění a typologie, jednak další specifické typy, které se vyskytují spíše u systémů Internetu:

- 1) **textové informační zdroje**: zahrnují je všechny systémy; kromě termínu „text“ se více méně synonymně uplatňuje i termín „dokument“ nebo jazykový/písemný materiál; tento typ koresponduje i s položkou „text“ v typologii MIME
- 2) **obrazové informační zdroje**: kromě speciální kanadské typologie je zahrnují všechny systémy, ovšem u některých se pod něj začleňovaly i videozáznamy (např. to do nedávné doby platilo pro standard Dublin Core, situace se však v roce 2003 změnila, zavedeny byly dva jemnější typy obrazu – stálý a pohyblivý obraz); vedle typologie MIME rozlišují zvláště „obraz“ a „video“ také systém ROADS a knihovnická třídění – ta

dokonce jdou ještě dále tím, že vyčleňují kartografická díla (mapy), resp. i hudebniny, popř. partitury (jako reprezentanty ideografických informací)

- 3) **zvukové informační zdroje:** kromě kanadské typologie je zahrnují všechny systémy; tento typ zahrnuje i typologie MIME
- 4) **počítačové programy (software):** kromě kanadské typologie zahrnují tento typ zdroje všechny systémy; jde o významný typ, který se uplatňuje především při práci počítačů; typologie MIME jej zahrnuje pod obecněji pojatý typ „aplikace“ s tím, že k němu patří i textové zdroje závislé na speciálních programech
- 5) **fyzické objekty/modely:** tento typ, resp.kategorie zahrnují bez rozdílu všechny systémy (pro jeho označení se užívá různých dílčích termínů (trojrozměrný objekt, fyzický objekt, model, vrml aj.); v typologii Dublin Core byl s tímto typem určitý problém, protože byla původně zamýšlena jenom jako typologie pro elektronické zdroje, v současné době jde však o typologii, které pokrývá jak tradiční, tak elektronické zdroje, takže fyzické objekty jsou již její součástí
- 6) **soubory dat:** pod různými dílčími jmény je obsažen zejména v systémech samotné sítě Internet; knihovnické systémy spíše uvádějí dílčí typy číselných, statistických, bibliografických aj. dat
- 7) **interaktivní zdroje:** jeden z nejzajímavějších, ale ve své podstatě velmi složitý typ; zahrnují jej prakticky všechny systémy, některé jdou dále v dalším vyčleňování např. interaktivních multimédií a online služeb nebo ještě více specifitějších typů, jako jsou diskusní skupiny, elektronické konference, vyhledávací nástroje Internetu, nebo dokonce internetové obchody, agendy výzkumných projektů, popř. celá webovská sídla (websites) či webovské systémy
- 8) **události:** diskutovaný typ informačního zdroje, který představuje komplex různých dílčích zdrojů; zahrnují je některé systémy Internetu

4. Architektura digitální knihovny

Digitální knihovny, archivy, atd. provozované na území USA, jsou budována v řadě případů na základě architektury (modelu), jejíž základy byly připraveny v rámci výzkumných projektů v první polovině 90. let 20. století (zejména v letech 1993-1996). Podle konkrétních tvůrců tohoto modelu (Robert Kahn a Robert Wilensky⁶⁶) se někdy označuje jako „**architektura Kahn-Wilensky**“. Významný byl především projekt Computer Science Technical Report Project (CSTR), vedený Korporací pro národní výzkumné iniciativy (Corporation for National Research Initiativec, CNRI⁶⁷) ve spolupráci s dalšími organizacemi. Byl financován z prostředků Agentury pro rozvinuté výzkumné projekty⁶⁸ ministerstva obrany USA.

Základními částmi architektury digitální knihovny jsou:

- **Uživatelské rozhraní** (User Interface) v jehož rámci koncový uživatel, knihovník nebo správce využívá systému digitální knihovny prostřednictvím prohlížeče a návazných klientských služeb
- **Repozitář** (Repository) pro ukládání **digitálních objektů**. Repozitáře (specifická skladiště pro digitální objekty) mohou být moderního typu (zahrnují digitální objekty libovolného typu), může jít o databáze neskenovaných dokumentů, nebo mohou být samotné webové servery. K transferu informací v repozitáři je využíván Přístupový protokol repozitáře (RAP⁶⁹).
- **Identifikační systém** (Handle System) pro generování či přidělování a správu jednoznačných identifikátorů digitálních objektů.
- **Vyhledávací systém** (Search System) pro vyhledávání informací v digitální knihovně. Zahrnuje indexy pro vyhledávání, tvořené z popisných metadat (informací o objektech), popř. z úplných textů, jsou-li předmětem zpracování.

Řadu principů architektury CSTR je možné částečně implementovat v rámci WWW (korespondují se zásadami vyhledávanými organizací IETF), například:

⁶⁶ Kahn, R.; Wilensky, R., nestr.

⁶⁷ Organizace úzce spolupracuje se skupinou Internet Engineering Task Force (IETF).

⁶⁸ Advanced Research Projects Agency (ARPA)

⁶⁹ Repository Access Protocol

- digitální objekty včetně meta-objektů mohou být uloženy ve webových archivech
- identifikační server (Handle Server) podporuje webová URL
- pojem URC⁷⁰ pocházející z IETF je formou meta-objektu
- identifikátor (Handle) je specifickou formou URN⁷¹

4.1. Identifikátory a systémy pro jejich správu a řízení

Pro komunikaci digitálních objektů v digitálních knihovnách hrají důležitou roli jejich **identifikátory** (amer. angl. *Handles*). Jde o jména, která by:

- měla být nezávislá na lokaci digitálního objektu
- měla být jedinečná v globálním měřítku
- měla být trvalá z hlediska času
- by ůa generována automaticky (nebo se případně používalo přidělování těchto jmen)

Systém pro řízení a správu identifikátorů (*Handle Management System*) by měl zajistit následující úlohy:

- rychlé směrování identifikátorů (v rámci vyhledávacích procesů)
- tvorbu kopií (replik) digitálních objektů a jejich ukládání ve vyrovnávací paměti (catching)
- decentralizovanou správu a řízení
- kontrolu změn digitálních objektů

Současné aktivity jsou zaměřeny na podporu identifikátorů ze strany standardních uživatelských rozhraní.

Identifikátory (*Handles*) jako unikátní jména mají následující strukturu:

Příklady:

loc.ndlp.amrlp/3a16116

cnri-1/1995.02.12.16.42.21;9

⁷⁰ Uniform Resource Characteristic

⁷¹ Uniform Resource Name

Údaj před lomítkem (prefix) představuje administrativní jednotku, která přiděluje jména, údaj za lomítkem (sufix) je lokálním jedinečným identifikačním číslem digitálního objektu. Administrativní jednotky mohou být zřizovány v hierarchické struktuře.

Systém pro správu a řízení identifikátorů by měl být veřejně dostupný a měl by mít následující části:

1. **Generátory identifikátorů** (handle generators) – které vytvářejí globálně jednoznačné identifikátory, jež mohou být spojeny s digitálními objekty
2. **Identifikační servery** (handle servers), které slouží k: **a)** k ukládání identifikátorů včetně příslušných lokálních informací (údajů); **b)** ke zpracování dotazů klienta týkajících se směrování identifikátorů; a **c)** k vrácení údajů o lokaci směrem ke klientovi
3. **Adresář identifikačních serverů** (handle server directory), který zajišťuje správu jejich seznamu.

Schéma ukazuje zcela obecný způsob využití identifikačních serverů ke směrování identifikátorů a poskytování odkazů (referencí) na digitální objekty, které identifikátory identifikují.

Hašovací tabulka (hash table) je využívána ke zjištění a označení identifikačního serveru (handle server), který vlastní informace spojené s určitým identifikátorem. Existují 3 běžné konfigurace:

- ▶ klient se může spojit se serverem pro vyrovnávací paměť (catching server), který zná našívací algoritmus (hash), má kopii našívací tabulky a má ve vyrovnávací paměti (cache) zaznamenaný naposledy využívané identifikátory (jde o konfiguraci zobrazenou ve schématu)
- ▶ klientský program může znát našívací algoritmus a vlastnit kopii našívací tabulky. Může udržovat vlastní vyrovnávací paměť. Jde o ekvivalent klienta fungujícího podobně jako jeho vlastní server pro vyrovnávací paměť.
- ▶ Klient se může spojit přímo s jakýmkoliv identifikačním serverem. Server provede vyhledání našívací tabulky a předá dotaz dále příslušnému počítačovému serveru.

Identifikační servery vracejí buď všechny uložené údaje, nebo jenom dílčí soubory údajů. Identifikátory a systémy pro jejich správu a řízení byly navrženy pro využívání

v digitálních knihovnách, ale lze je využívat i pro ukládání širší skupiny údajů, např. URL, email adres, apod.

Některé z digitálních archivů (zejména v USA zavedly přidělování URN typu „handle“, jehož model byl připraven v rámci systému The Handle System⁷²). Systém založila a spravuje národní americká organizace CNRI⁷³. Struktura tohoto trvalého identifikátorů (**HDL**) se skládá ze dvou hlavních částí, které jsou odděleny lomítkem (/). Část před lomítkem identifikuje jednoznačně korporaci přidělující identifikátory, část za lomítkem identifikuje konkrétní digitální objekty. Globální server CNRI spravuje bázi korporací (Naming Authorities), které přidělují identifikátory v rámci své působnosti, v rámci lokálního serveru jsou pak e-disertacím přidělována konkrétní HDL (URN) a zároveň je zajišťována a neustále aktualizována vazba mezi nimi a příslušnými URL. Program k zajištění uvedených procesů lze stáhnout ze sídla systému „The Handle System“ (Local Handle System Server-JAVA Version). Součástí celého programového balíku je i plug-in (Handle System Revolver Plug-In), který zajišťuje směrovací službu.

Americký systém „The Handle System“ se stal také základem mezinárodního identifikačního systému **DOI**⁷⁴, který je hojně využíván komerčními vydavateli (proxy server pro směrování DOI identifikátorů je k dispozici veřejně na URL: <http://dx.doi.org>).

Identifikační systém by měl potřebné funkce (tvorba identifikátorů a jejich směrování) zajišťovat **distribuovaným způsobem**. Schéma znázorňuje **model globálního identifikačního systému**. Tvorbu identifikátorů i jejich směrování zajišťují vyhrazení klienti (webové prohlížeče). Ve všech případech je komunikace s identifikačním systémem zabezpečována pomocí protokolů tohoto identifikačního systému (existují i specifické implementace v rámci protokolů pro komunikaci ze strany klienta zajišťujícího tvorbu identifikátorů (Administration) i klienta zajišťujícího směrování identifikátorů (Resolution)). V systému dochází k následujícím operacím:

1. V rámci směrování identifikátorů (Resolution) potká klient (proces 1) identifikátor na Internetu, popř. Intranetu v podobě hyperodkazu nebo jiného typu odkazu. Klient odesílá identifikátor ke směrování

⁷² Dostupný z WWW: <<http://www.handle.net/>>

⁷³ Corporation for National Research Initiatives; Dostupný z WWW: <<http://www.cnri.reston.va.us/>>

⁷⁴ Dostupný z WWW: <<http://www.doi.org/>>

k identifikačnímu systému (může k tomu dojít přímo na straně klienta nebo prostřednictvím proxy serveru).

2. Vlastní identifikační systém (proces 2) je tvořen skupinou identifikačních služeb. Každá služba může být fyzicky i logicky distribuována na libovolném počtu serverů a každý server může být replikován (kopie). Jedna ze služeb – Globální registrace identifikátorů – zajišťuje správu všech ostatních lokálních identifikačních služeb, jež jsou zpětně propojené s globální centrální jednotkou. Dotaz ke směrování jakéhokoli identifikátoru proto může být identifikačnímu systému zaslán z libovolné služby
3. Každý identifikátor (proces 3) může být spojen s jedním nebo více zapsanými údaji. Například identifikátor je spojen jak s URL, který zajišťuje komunikaci v rámci WWW, tak s novým protokolem RAP (Repository Access Protocol), zajišťujícím komunikaci v rámci repozitáře. Taková informace je zpět poskytnuta po směrování klientovi (proces 4). Propojit je možné i jeden identifikátor s vícenásobným digitálním objektem, který má více URL.
4. V rámci procesů administrativního klienta dochází jednak k vytváření identifikátorů a doplňování a editování informací spojených s jejich směrováním (proces 5), jednak k zasílání zpráv o úspěšných nebo chybových řešeních (procesy 6).

4.1.1. Identifikátor URN

K identifikaci digitálních objektů uložených v digitálních archivech je žádoucí uplatnit identifikátory, které by tyto objekty určovaly jednoznačně, trvale, globálně a nezávisle na jejich umístění. Jedním ze základních trvalých identifikátorů (Persistent Identifiers), který naplňuje uvedené charakteristiky je jednotné jméno zdroje **URN** (Uniform Resource Name), jehož syntax je dána standardem RFG 2141 z roku 1997. Vychází ze syntaxe jednotného identifikátoru zdroje URI (Uniform Resource Identifier, RFC 2396). URN se podle uvedeného standardu skládá z několika hierarchicky uspořádaných jmenných prostorů (Namespaces). Obecnou strukturu lze zapsat následujícím způsobem:

URN:NID:SNID:NSS

kde **NID** označuje identifikátor jmenného prostoru (Namespace Identifier)

SNID identifikátor dílčího jmenného prostoru (SubNamespace Identifier)

NSS specifický řetězec jmenného prostoru (Namespace Specific String). Dílčích jmenných prostorů může být i více.

Vytváření a přidělování konkrétních identifikátorů URN nepředstavuje v podstatě žádný problém (prvním krokem je zavedení schématu jmenného prostoru). Problémem je teprve realizace vlastního procesu nalezení digitální zdroje na základě daného URN v podmínkách současného Internetu. Tento proces se označuje v angličtině odborným termínem „*resolution*“, což v češtině používám jako „*směrování*“. K efektivnímu využívání URN je nezbytné síťově dostupného směrovacího systému (Resolution Systém), který by byl schopen uživateli okamžitě předložit minimálně jedno URL zdroje s daným URN. V principu by však takový systém měl být schopen uživatele nasměrovat na jakýkoliv zdroj, službu, atd. dostupnou v síti Internet.

Současné prohlížeče Internetu však zatím nejsou připraveny pro práci s URN. Tento identifikátor nelze zapisovat přímo do řádky adresy a požadovat nalezení zdroje. *Globální směrovací mechanismus* dosud není zaveden. Některé směrovací systémy již ale existují, k jejich využití je však nutné nainstalovat speciální rozšíření ke klientskému prohlížeči (plugin) nabízené příslušným systémem (službou).

Nadějný experiment s přidělováním a směrováním trvalého identifikátoru **URN e-disertací** je realizován v Německu v rámci národního projektu DissOnline, které získává Německá knihovna ve Frankfurtu nad Mohanem v rámci PV, od roku 2001.

K jiným trvalým identifikátorům patří ještě také identifikátor **PURL** (Persistent Uniform Resource Locator), jehož systém byl zaveden a je spravován OCLC. Jeho aplikaci zavedla např. Národní knihovna Portugalska v rámci projektu národního systému e-disertací.

4.2. Digitální objekt

Digitální objekt (Digital Object) zahrnuje:

- jedinečný identifikátor (handle)
- obsah (kontent) jako sekvenci bitů nebo soubor sekvencí bitů
- příslušné informace o vlastnostech (properties) k jejich zápisu (metadata)
- příslušné informace o transakcích (transaction log), tj. o užití objektu
- digitální signaturu (volitelná) garantující, že objekt nebyl změněn

Digitální objekty mohou být proměnlivé, pokud je dovoleno, aby byl jeho obsah měněn, nebo neproměnlivé.

Digitální objekty mohou být jednotuché, nebo může jít o soubor či skupinu (agregát) několika digitálních objektů.

4.3. Skupina (agregát) digitálních objektů

V rámci digitálních knihoven je v řadě případů nutné, aby jejich tvůrci seskupovali příbuzné digitální objekty. V repozitáři např. mohou být uloženy výzkumné zprávy ve více formátech (Postskript, PDF, aj.), různé interpretace hudebního díla, apod. Skupiny digitálních objektů mohou být reprezentovány buď jako **složené objekty** (composite objects), jejichž obsah zahrnuje soubor několika objektů, nebo jako **meta-objekty** (meta-objects), jejichž obsah zahrnuje identifikátory pro soubor objektů.

4.4. Digitální skladiště a repozitáře

Informace v digitální knihovně může být uložena v „digitálních skladištích“, která mohou mít *formu*:

- FTP archivu
- WWW serveru
- online databáze
- repozitáře

Repozitář je specifickou kategorií digitálního skladiště. Je vhodný pro materiály obsahující hodnotné informace. K základním charakteristickým repozitářů patří:

- v repozitáři jsou ukládány jak digitální objekty, tak příslušné vlastnosti (metadata) i informace o jejich užití
- jeden objekt smí být uložen ve více repozitářích (replikace-kopie objektů)
- v repozitářích musí být zajištěna bezpečnost uložených objektů. Individuální objekty mohou být (nepovinně) opatřeny signaturou, která může například garantovat platnost a neporušitelnost objektu, apod.

- ke každému repozitáři náleží server s vyhledávacím systémem, který obsahuje informace o objektech v repozitáři. Takový server nemusí být součástí repozitáře.

4.5. Identifikátory informačních entit

4.5.1. Identifikátory tvůrců děl

- identifikátory ISADN (International Standard Authority Data Numer) a INSAN (International Standard Author Numer)
- identifikátor IPI (Interested Parties Information)

4.5.2. Identifikátory tvůrčích děl

- ISWC (International Standard Musical Work Code)
- ISAN (International Standard Audiovisual Number)
- ISTC (International Standard Text Code)

4.5.3. Identifikátory vyjádření děl

- ISRC (International Standard Recording Code)
- DOI (Digital Object Identifier)

4.5.4. Identifikátory zhmotnění děl

- ISBN (International Standard Book Numer)
- ISMN (International Standard Music Numer)
- ISSN (International Standard Seriál Numer)
- ISRN (International Standard Technical Report Numer)
- EAN (European Artikle Numer)

5. 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 metadat 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. Problematika metadat 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 10-12 let.

Knihovníky a informační pracovníky, zejména pak katalogizátory problematika metadat zajímá proto, že síťově dostupné elektronické dokumenty byly a jsou také, i když zatím jenom v malé míře, předmětem jejich zpracovatelské činnosti. Základním odborným termínem, který doposud pro označení zpracování dokumentů knihovníci užívali a stále užívají, je katalogizační popis či záznam, resp. bibliografický popis či záznam. Proto mnohé překvapilo, že se začalo hovořit a psát o metadatach, jakožto nástrojích, či prostředcích popisu elektronických dokumentů, především pak síťově dostupných. Těmi, kdo se danými problémy prioritně zabývá, nejsou však knihovníci, nýbrž počítačová odborníci, kteří rozvíjejí další technologie i celkovou novou architekturu komunikace informací na WWW pro 21. století.

Zhruba v polovině 90. let se začaly mezi knihovníky objevovat první definice a výklady metadat, které byly velmi jednoduché a nepostihovaly zcela podstatu, funkce a význam metadat 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 metadat objevilo zařazení tradičních bibliografických formátů typu MARC (USMARC, UKMARC, UNIMARC, PICA aj.) jako jednoho z typů metadat⁷⁵. Někteří odborníci přesněji navrhli i základní kategorie (skupiny) metadat, například z hlediska typů organizací, které metadata používají. 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 metadat, 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.

Významný autor Stefan Gradmann⁷⁶ se pokusil porovnat a odhalit podstatné rozdíly mezi oběma pojmy. Poukazuje na fakt, že tvůrci dnes nejznámější specifikace (sémantiky) metadat „Dublinského jádra“ (Dublin Core, dále též DC) při jeho navrhování jednoduše nepřevzali a neupravili existující formát MARC, ale navrhli zcela nový soubor údajů k popisu digitálních dokumentů. Pravdou je, že pro současnou 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. Kdo by chtěl považovat proces tvorby metadat za nějaký typ zjednodušené katalogizace, bude se hluboce mýlit. Podstatné rozdíly jsou funkční i strukturální povahy, vězí v **celém kontextu produkce a užití metadat** v rámci síťové digitální komunikace informací⁷⁷. Konkrétněji autor poukazuje na:

⁷⁵ Dempsey, L.; Heera, R., nestr.

⁷⁶ Grandmann, S., nestr

⁷⁷ Who does it, and How is it done?

- 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 metadat se podtrhuje účel zjišťování a vyhledávání zdrojů v síťovém prostředí, u bibliografických či katalogizačních záznamů hraje podstatnou roli stránka deskripce; 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 metadat na WWW se předpokládá primární tvorba ze strany autorů, editorů či vydavatelů)
- *nezávislost či samostatnost jednotlivých metadat* a jejich účelného shromažďování, přeskupování, propojování a zejména zpracování a nové užití pro nejrůznější formy výstupů a služeb včetně služeb v oblasti znalostních systémů
- otázku *užití metadat*, která je klíčová: v rámci WWW jsou primárními uživateli metadat 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 povolane - 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 metadat“⁷⁸ podal několik základních pregnantně vyjádřených předpokladů a definic s komentáři, které se týkají metadat jakožto podstatné součásti celého prostoru WWW. Základní charakteristikou webu 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 webové 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 počítačovní odborníci běžně nazývají „**metadata**“.

⁷⁸ Berners-Lee, T., nestr.

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 „métadonnées“), řada dalších národních jazyků však preferuje výraz pocházející z angličtiny. Již zmiňovaný T. Berners-Lee ve svém výkladu rovněž nečiní podstatnější rozlišování obou pojmů. 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 také uvádí základní definici pro metadata: jde o „**stroji srozumitelné informace o webovských zdrojích nebo dalších věcech**“. 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. Zdůrazněna je klíčová charakteristika metadat, 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 metadat musí být naprosto formálně logicky správný a jednoznačný. V budoucnu, až se podaří zdokonalit informační jazyky metadat 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 **systemu pro informace o informacích** (informačních zdrojích).

Pro metadata platí podle T. Berners-Leea 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. **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 vyskytnou 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.

5.1. Metadata pro digitální zdroje v zahraničních systémech a službách

Formát metadat označovaný výrazem „Dublinské jádro“, jehož základní charakteristikou se budeme dále zabývat, 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.

5.1.1. 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)⁷⁹

5.1.1.1. 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-

⁷⁹ Bernard, L.; Light, R., nestr; část 1.

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. V příloze č. 1 je připraven podle dostupné dokumentace 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).

5.1.1.2. 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. 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).

5.1.1.3. CIMI (Computer Interchange of Museum Information)

Jde o formát metadat velmi významného a dynamicky se rozvíjejícího systému mezinárodního charakteru (USA a Velká Británie), který je řízen Konsorciem CIMI <<http://www.cni.org/pub/CIMI/framework.html>>. Formát vznikl v roce 1988 a byl rozvíjen zejména v letech 1990-1993. 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.

5.1.2. Formát GILS

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 WWW⁸⁰.

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.

V příloze č. 2 je zobrazen ilustrativní záznam stejného webového dokumentu jako u obrázku 1 ve struktuře vybraných údajů metadat GILS. Záznam byl připraven na základě

⁸⁰ Dempsey, L.; Heery, R., nestr.

konverze z původně připraveného formátu DC do formátu GILS pomocí britského konvertoru „**DC-dot**“, který je k dispozici zdarma na WWW⁸¹.

5.1.3. 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). 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 webové 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.

5.1.4. Internetová norma RFC 1807

V krátkém přehledu nejvýznamnějších formátů metadat uveďme 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).

⁸¹ DC-dot, nestr.



5.2. Formát Dublinské jádro (DC)

Nejnámějším formátem metadat, který byl navržen pro popis webových informačních zdrojů. Jde o formát, jehož obsah se týká podstatných vlastností informačních zdrojů z hlediska potřeby jejich zjišťování a vyhledávání v rámci rozvíjení nové architektury celého systému WWW. Formát 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. Má značný význam a perspektivu pro komunikaci informací v 21. století.

5.2.1. 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 informací o informacích (metadat) na WWW. Podstatným důvodem je i ohled na budoucí tvůrce záznamů, jimiž mohou být i samotní **autoři (tvůrci)** nebo **vydavatelé (zpřístupňovatelé)** 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 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 komunikace, 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 jedněm z nejsložitějších⁸², **forma** (Form), představující reprezentaci (uspořádání) dat na počítačovém médiu, specifický **identifikátor** (Identifier), kterým je pro webové 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, tezurů, 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 třídníků Deweyho desetinné klasifikace pro webovské dokumenty včetně tvorby struktury RDF na Univerzitě ve Wolverhamptonu, Velká Británie⁸³.

Akteři 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).

5.2.2. Další rozvoj 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 jádra⁸⁴ a také experimentální konvertor pro

⁸²Bratková, E. nestr.

⁸³Automatic RDF, nestr.

⁸⁴DC Metadata Templáře, nestr.

převod metadat do formátů typu MARC⁸⁵. 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 autorit⁸⁶.

V průběhu příprav na realizaci prvních projektů využívajících specifikace údajů DC vystala 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. 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 realizacích 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

⁸⁵ d2m, nestr.

⁸⁶ Dempsey, L.; Weibel, S., nestr.

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 naléhavou, protože je měli vytvářet autoři dokumentů nebo vydavatelé. Příkladů 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 předpokládalo 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="MetadataArchitecture">
```

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="MetadataArchitecture">
```

```
<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 webové 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 webového dokumentu, který obsahuje metadata v hlavičce <HEAD>, do interaktivního formuláře konvertoru. Příloha č. 3 a příloha č. 4.

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 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 provedeny dílčí změny (soubor byl aktualizován). Poslední aktualizace byla provedena 2. října 1997⁸⁷. 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 normu⁸⁸.

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 systémů⁸⁹.

Nicméně, canberrský seminář projednal návrh na podrobnější specifikaci DC pomocí tzv. kvalifikátorů (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, jak je vidět i na příloze č. 8).

⁸⁷ DCMI, nestr.

⁸⁸ Weibel, S., nestr.

⁸⁹ Weibel, S., et al., nestr.

- **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 přenášených 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 přílohách č. 3 a 4. Za povšimnutí stojí v příl. č. 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“).
- **ŽTyp (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**⁹⁰. Problematika dat jako časových údajů (např. jejich typologie,

⁹⁰ Weibel, S.; Hakala, J., nestr.

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 vazbách 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 webových stránek sdružených v rámci webových 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 DC⁹¹ 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 RDF⁹².

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

⁹¹ DCMI, nestr.

⁹² DCMI, nestr.



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).

5.2.3. 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**)⁹³. 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. 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

⁹³ DCMI, nestr.

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. 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:

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. Přispěvatel (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“ přispěvatele 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 článku Dr. Bratkové v časopise Národní knihovna).

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, které 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 rozproutila 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 zdroje“.

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.

5.3. 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 digitálních 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)**⁹⁴, 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.

⁹⁴ W3C, nestr.

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** a **specifikací schématu** 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:

1. **zdroje** (webová stránka, část stránky, kolekce stránek nebo celé webové sídlo) identifikované jednoznačným identifikátorem zdroje URI (URL, URN aj.); zdrojem podle RDF ale může být i tradiční dokument. Příloha č. 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 příloha 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 příloha 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 příloze 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 příloze 7. Takový zdroj může mít sám o sobě své vlastní vlastnosti (údaje) s příslušnými hodnotami. Příloha č. 6 a příloha č. 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. 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 webové 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 identifikujícím 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 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>). Příloha č. 8.

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. 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. 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 neznamena, že by model DC musel být vyjádřen jenom prostřednictvím něho samotného.

V příloze č. 8 je soubor metadat původní 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 „xmlns“). 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 sólový 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 - hlavičky <HEAD>, ovšem musí být užito pouze tzv. „zkrácené syntaxe RDF“, 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. 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 příloze č. 9.
- V první řádce záznamu je uvedena povinná deklarace verze jazyka XML (verze1.0) 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. 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, která kvalifikátory zatím jen navrhuje.

5.4. Shrnutí

Několikaletá komunikace informací na WWW vstupuje v současné době do své další významné vývojové fáze. Na scénu vstupuje, kromě řady nových technologií, nový značkový 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.

6. Interoperabilita

Obecná představa DK vychází z toho, že digitální knihovna není nějaký monolitický produkt, ale naopak systém dynamicky propojovaných spolupracujících komponent, které samy o sobě mohou být autonomní a nezávisle spravované. Termínem *interoperabilita* bývá označována schopnost spolupráce mezi technicky různorodými a organizačně nezávislými komponentami při řešení určitého úkolu. Někdy se s mírnou nadsázkou tvrdí, že všechny technické problémy a výzvy digitálních knihoven nejsou nic jiného než jen různé aspekty interoperability.

6.1. Úvod do problematiky a stručný přehled

Existuje velice široké spektrum pohledů na interoperabilitu: na jedné straně lze pohlížet na interoperabilitu jen jako na použití společných nástrojů a rozhraní pro vytvoření *povrchní jednoty* pro přístup a navigaci, na opačné straně je pak vysoce ambiciózní *hluboká sémantická interoperabilita*, kdy inteligentní technologie dokáží poskytnout koherentní pohled na různorodý informační obsah a služby digitálních knihoven (zatím jde o hudbu budoucnosti). Někde mezi těmito dvěma extrémy je primárně *syntaktická interoperabilita*, kdy výměna metadat a použití protokolů pro přenos digitálních objektů a formátů založených na těchto metadatach umožňují poskytnout omezenou koherenci obsahu, která pak musí být ještě doplněna lidskou interpretací.

Při zkoumání interoperability se ukazuje závislost mezi funkcionalitou a cenou. Většina v současnosti používaných metod pro interoperabilitu (např. webové standardy HTTP, HTML, URL) dosahují jen průměrné funkcionality, ale za nízkou cenu a s velmi širokým uplatněním (příklad webových vyhledávačů). Naopak většina kvalitních koncových služeb (založených např. na využití standardů Z39.50 či SGML) dosahuje vysoké funkcionality, ale za vysokou cenu, která často brání jejich širšímu využití. Většina výzkumu v oblasti digitálních knihoven je pak vedena snahou najít ten správný „zlatý střed“.

Systematický pohled na interoperabilitu a přístupy k jejímu dosažení shrnuje⁹⁵. Uvádí, že problém interoperability se bezprostředně dotýká všech pěti základních funkcí

⁹⁵ Paepcke, A.; Chang, C.K.; García-Molina, H., et al., S. 33-43.

digitálních knihoven – správy informací (ukládání, organizace a získávání informace), prezentace informací uživatelům, komunikace mezi částmi systému, řízení systému a ochrany informačních zdrojů a uživatelů včetně jejich práv. Ačkoliv porovnávání úspěšnosti jednotlivých řešení je v oblasti interoperability velmi obtížné (různé přístupy vycházejí z různých předpokladů a mají různé, často protikladné cíle), navrhuje šest základních kritérií, které přece jen poskytují určité vodítko:

- vysoký stupeň autonomie komponent;
- nízká cena infrastruktury;
- snadnost přidání nové komponenty;
- snadnost používání komponenty;
- širší celkové složitosti;
- škálovatelnost v počtu komponent.

V některých případech se může stát, že rozhodnutí optimalizující jedno z těchto kritérií mohou negativně ovlivnit jiné (např. systém, který minimalizuje cenu infrastruktury, může být použitelný jen pro jednoduché úkoly, nebo ho může být obtížné používat vůbec).

Existuje velmi rozdílných přístupů k dosažení požadovaného stupně interoperability; práce uvedených autorů popisuje pět základních tříd přístupů:

1. *silné standardy* – nejstarší přístup založený na tom, že heterogenní komponenty se shodnou na standardu, který zajistí určitou omezenou míru homogenity mezi nimi. Příkladem jsou standardy Z39.50, HTML/HTTP.

2. *rodiny standardů* – v tomto případě má implementátor komponenty k dispozici nikoliv jediný standard, ale celou rodinu standardů, z nichž může volně vybírat a dosáhnout tak vyššího stupně autonomie než v předchozím případě. Příkladem je elektronický obchod, kdy při implementaci platebního modulu může uživatel volit mezi řadou standardizovaných platebních schémat (DigiCash, First Virtual, některá z mnoha platebních karet).

3. *vnější zprostředkování* – cesta k dosažení velmi vysokého stupně autonomie komponent. Mechanismus pro zajištění interoperability je umístěn mimo spolupracující komponenty v podobě samostatných zprostředkujících modulů nazývaných „wrappers“ nebo „proxies“, které překládají datové formáty a komunikační protokoly komponent do/z interního

standardu systému (provádí mapování mezi globálními a lokálními schémata na úrovni komponent). Příkladem z oblasti propojování sítí jsou gateways. Nevýhodou tohoto řešení může být vyšší cena přidání nové komponenty zahrnující i vytvoření příslušné zprostředkující mezikomponenty.

4. *interakce založená na specifikacích* – cílem je umožnit použití komponenty bez pomoci speciálních předběžných opatření a prostředníků. Pro každou komponentu existuje přesný formální popis sémantiky a struktury všech jejích dat a operací; komponenty pak mohou mezi sebou interagovat díky tomu, že jsou schopny zjistit specifikace jiných systémů a zohlednit je při vzájemné komunikaci a spolupráci. Příkladem nástrojů pro implementaci tohoto typu přístupů je nástroj pro sdílení znalostí (knowledge-sharing) pro softwarové agenty (jazyk Agent Communication Language a jeho Knowledge Interchange Format – KIF) nebo jazyky SETL a PAISley pro opakovatelné využití komponent (software-reuse), které umožňují popsat sémantiku funkcionality dané komponenty čistě deklarativním způsobem. Tento přístup přináší vysokou míru autonomie, avšak současně vysokou náročnost přidání nové komponenty (popsat dostatečně podrobně komponentu může být velmi složité a někdy v praxi i nemožné).

5. *mobilní funkcionalita* – je založena na mobilních softwarových agentech, kteří cestují sítí na místa, kde zpřístupní potřebné služby. Z novějších technologií umožňují např. javovské applety přístupy pro doručení nových funkcionalit klientským komponentám až v době běhu (takovou novou funkcionalitou může být např. schopnost komunikovat s jinou komponentou systému). Tento přístup je velmi lákavý a efektivní zejména z pohledu snadnosti přidání nové komponenty. Na druhou stranu jeho implementace může být nákladná nejen z hlediska komunikačního (pokud na straně klienta neexistuje dlouhodobá vyrovnávací paměť typu *cache*, může identický programový kód cestovat po síti opakovaně stále dokola), ale i z hlediska bezpečnostního (kontroly autenticity a bezpečnosti kódu na každém přijímacím místě sítě). Tento přístup je také silně závislý na existenci silného standardu (v daném případě širokém rozšíření prohlížečů podporujících příslušný javovský standard).

Jiné obecné pohledy na interoperabilitu přináší⁹⁶. Rozlišuje několik abstraktních úrovní interoperability, od obecné transportní vrstvy a na aplikační oblasti nezávislého

⁹⁶ EU-NSF Digital Library ..., nestr.

middleware (Z39.50, distribuované objekty např. s technologií CORBA) přes úroveň specifické pro digitální knihovny (vrstva informačního modelu, správy informací/dokumentů, správy vlastnických a autorských práv) až po nejvyšší vrstvu týkající se sociálních souvislostí. Pokorný zmiňuje čtyři úrovně praktické kooperace digitálních knihoven⁹⁷:

úroveň	způsob kooperace
federace	striktní použití standardů (syntaktické, sémantické, obchodní) příklad: MARC, Z39.50
sklizení metadat	digitální knihovny nabízejí základní metadata; jednoduchý protokol a registrace příklad: otevřené archivy, např. iniciativa OAI
shromažďování dat	digitální knihovny nekooperují; informace se musí hledat explicitně pomocí služeb příklad: internetové vyhledávače
vyhledávací middleware	zdroje vybavené metadaty jsou volně zapojovány do kooperace příklad: Z39.50, XML, RDF, SDLIP.

Z obrovského množství nástrojů, přístupů a projektů z oblasti interoperability uvedme jen několik málo vybraných zástupců z těch všeobecně nejznámějších, které charakterizují typické nebo nové perspektivní přístupy: vyhledávací protokol Z39.50, přístup přes sklizení metadat v Open Archives Initiative (OAI), Stanfordský projekt InfoBUS a technologii OpenURL a SFX pro otevřené kontextově citlivé propojování zdrojů (reference linking).

6.2. Protokol Z39.50

Z39.50 je mezinárodním standardem pro komunikaci mezi počítači, který umožňuje jednomu počítači (klient, *origin*) vyhledávat a získávat informace na jiném počítači

⁹⁷ Pokorný, J., nestr.

(databázový server, *target*), a to v heterogenním prostředí, nezávisle na operačních systémech, databázích a dotazovacích jazycích. I když koncepčně není vázán na žádný konkrétní druh informací ani typ databází, největší jeho současnou aplikační oblastí jsou bibliografická data a knihovní katalogy. Základ standardu vznikl v roce 1984 jako výsledek projektu předních amerických knihoven *Linked Systems Project* a od té doby prošel několika ANSI/NISO verzemi: 1988 (v1), 1992 (v2), 1995 (v3) a 2002 (Z39.50-2003). Verze 1 není s ostatními kompatibilní; verze 3 zahrnuje verzi 2 a byla přijata jako mezinárodní standard ISO 23950. Podrobný přehled historie a motivací ve vývoji Z39.50 nalezneme v Lynchově práci⁹⁸.

Z39.50 je založen na abstrakci databázového vyhledávání, která je obecnější než např. u SQL. Server provozuje jednu či více databází obsahujících záznamy; s každou databází je spojena množina přístupových bodů (indexů), které mohou být použity pro vyhledávání. Protokol je stavový (na rozdíl od bezstavového HTTP) a relačně orientovaný, interakce mezi klientem a serverem je založena na koncepci seance (*session*): klient otevře spojení se serverem, provede sekvenci interakcí a uzavře spojení. Během sezení si klient i server pamatují stav jejich interakce. Zdůrazněme, že Z39.50 je protokol mezi dvěma počítači, nijak nespécifikuje uživatelské rozhraní, kterým bude ke klientskému počítači přistupovat uživatel.

Typická seance začíná tím, že klient naváže spojení se serverem a vyvolá inicializační službu *init*, během které si obě strany vyjednájí podrobnější detaily spolupráce (kterou verzi protokolu podporují, používanou množinu znaků a jazyk, maximální délku záznamu předávaného ze serveru, požadavek na autentikaci uživatele apod.). Poté může klient pomocí služby *explain* zjistit detaily o serveru a jím nabízených službách: databáze dostupné pro prohledávání a jejich přístupové body (indexy), podporovaná syntaktická schémata a datové formáty, třídící možnosti, ale také obecné charakteristiky, jako popis serveru, provozní doba, případná omezení a cena za použití. Po těchto úvodních operacích může klient vyslat vyhledávací dotaz pomocí služby *search*; standard specifikuje šest typů vyhledávání od booleovského přes standard ISO 8777 Commands for Interactive Text Searching, ANSI standard Common Command Language (CCL) až po SQL – běžně však bývá plně implementováno jen booleovské vyhledávání. Dotaz tedy může mít následující význam:

⁹⁸ Lunch, C., nestr.

Najdi v databázi 'Knihy' všechny záznamy, pro které přístupový bod 'title' obsahuje hodnotu 'sen' a přístupový bod 'author' obsahuje hodnotu 'shakespeare'.

Server provede hledání, vytvoří výsledkovou množinu, tzv. *result set*, a uloží si ji, takže klient se na ni může následně v dalších příkazech odvolávat – zmenšit velkou výsledkovou množinu upřesňujícím hledáním, setřídít ji, vymazat apod. V závislosti na parametrech příkazu hledání vrátí server klientovi buď jen počet vyhledaných záznamů, nebo přímo jeden či více záznamů z výsledkové množiny. Jakmile je hledání dokončeno, vyšle klient službu *present*, v níž serveru specifikuje, které záznamy z výsledkové množiny a v jakém formátu mu mají být zaslány (standardně se používá textový formát, nebo formát MARC, ale možné jsou i jiné varianty).

Kromě dosud popsaných služeb nabízí protokol ještě řadu dalších – pro procházení indexů, řízení přístupu (možnost serveru vyslat žádost o autentizaci uživatele, informovat o postupu dlouhotrvajícího vyhledávání), možnost účtování, ukončení seance a také tzv. *rozšířené služby*, což je v zásadě mechanismus pro asynchronní vzdálené volání procedur, pomocí nichž lze realizovat např. další operace nad výsledkovou množinou – jako její uchování mezi seancemi, zařazení do fronty pro zaslání e-mailem nebo tisk, pro zaznamenání dotazů, které mohou být na serveru prováděny opakovaně v určenou dobu (SDI) a další. Ve verzi Z39.50-1995 je možné provádět ze strany klienta také aktualizaci záznamů v databázi na serveru.

Protokol Z39.50 je na jednu stranu velmi mocný a flexibilní, na druhou stranu hodně rozsáhlý (jeho úplná specifikace má kolem 160 stran) a náročný na implementaci i správné nastavení pro bezchybnou funkci v dané doméně (potřeba společného *profilu*, který specifikuje vlastnosti a nastavení protokolu pro komunikaci). Jak již bylo řečeno, hlavní oblastí nasazení protokolu Z39.50 jsou bibliografické knihovní databáze, ale existují i profily pro využití standardu v oblasti vládních informačních systémů, vědecko-technických databází, geografických informačních systémů, muzeí a digitálních knihoven.

Má-li knihovnický systém zabudován Z-klienta, lze použít protokol Z39.50 jako meziplatformní standard pro interoperabilitu při vyhledávání následujícím způsobem: uživatel zformuluje dotaz v jazyce svého knihovního systému a vybere pro vyhledávání cizí vzdálený katalog se Z-serverem. Dotaz je přeformulován do Z39.50 a zaslán Z-serveru cizího

katalogu; ten přeloží dotaz do vyhledávacího jazyka cílové databáze a přijme výsledek vyhledávání. Výsledek pošle Z-klientovi, který ho předá knihovnímu systému pro zobrazení v jeho standardním uživatelském rozhraní. Z-klient může být implementován také tak, aby vyhledávací dotaz rozeslal paralelně více specifikovaným Z-serverům, což např. umožňuje realizovat virtuální souborné katalogy.

Existuje několik volně dostupných samostatných Z-klientů, které lze instalovat a využívat pro prohledávání informačních zdrojů podporujících protokol Z39.50, např. BookWhere, Z-navigator a další. Další rozvoj protokolu Z39.50 řídí mezinárodní skupina Z39.50 Implementors Group (ZIG) pod patronací Kongresové knihovny, která zodpovídá za Z39.50 v roli Agentury pro jeho údržbu a rozvoj⁹⁹. Vedle nedávno dokončené verze Z39.50-2003 probíhají paralelně i diskuse o možnostech přiblížení protokolu směrem k webovým technologiím a snížení náročnosti jeho implementace.

6.3. Open Archives Initiative (OAI)

Za vznikem Open Archives Initiative¹⁰⁰ koncem roku 1999 je rostoucí nespokojenost vědců s tradičním modelem vědeckého publikování (dlouhá doba od nabídnutí příspěvku k jeho zveřejnění a stále rostoucí cena předplatného časopisů) spolu s pozitivními zkušenostmi s novými modely publikování v podobě online repozitářů typu e-print (viz arXiv.org, NCSTRL, atd.). OAI je zaměřena na podporu rozvoje tohoto typu publikování tím, že nabízí technický mechanismus a organizační struktury pro podporu interoperability mezi *otevřenými archivy* (pojem „otevřený“ je zde ve smyslu architektury systému, nikoliv nutně ve smyslu bezplatného či neomezeného přístupu; pojem „archív“ je chápán volně jako jakýkoliv repozitář pro ukládání informací na webu). Jako metoda pro dosažení potřebné interoperability s nízkými náklady bylo zvoleno tzv. *sklizení metadat*, kdy *poskytovatelé dat* (archivy) mají k dispozici relativně snadno implementovatelný mechanismus pro externí zviditelnění informací (metadat) o obsahu archivu, což umožňuje třetí straně – *poskytovatelům služeb* – tyto informace z mnoha archivů automatizovaným způsobem shromažďovat a budovat nad nimi různé nadstavbové služby. Technický aspekt tohoto řešení zahrnuje tři komponenty:

⁹⁹ Z39.50 Maintenance Agency

- *společný metadatový standard* – Open Archives Metadata Set (OAMS) – povinnou součástí metadat je nekvalifikovaný Dublin Core, různé odborné komunity mohou volitelně doplnit další metadata v jejich specifickém schématu. Metadata jsou zabalena do záznamu XML, který obsahuje záhlaví (jednoznačný identifikátor, datum vytvoření či změny záznamu), metadata a popis metadat. Záznamy jsou uloženy u poskytovatele dat v repozitáři, který musí podporovat OAI sklízecí protokol, a mohou obsahovat odkaz na vlastní dokument, který může nebo nemusí být volně dostupný;
- *jednotné identifikační schéma* – musí být jednoznačné a má následující tvar: oai:arXiv:hep-th01. První část tvoří konstantní řetězec „oai“, za ním je jednoznačný identifikátor repozitáře (archiv ho obdrží při registraci u OAI), poslední částí je libovolný identifikátor jednoznačný uvnitř daného repozitáře. Směrování identifikátorů bude probíhat přes centrální OAI směrovací službu s podporou OpenURL (viz níže);
- *protokol pro sklízení metadat* – původní návrh počítal s využitím protokolu Dienst, ale pro zjednodušení implementace byl nakonec vytvořen samostatný OAI protokol na bázi HTTP obsahující šest jednoduchých příkazů.

V současné době patří OAI mezi dobře zavedené standardy a jeho využívání stále roste. Jeho obrovskou předností je jednoduchost a velmi snadná implementace prakticky do jakýchkoliv informačních systémů požadujících sdílení metadat.

6.4. Stanfordský InfoBus

Jedním z nejobsáhlejších prakticky realizovaných řešení interoperability byl projekt *The Stanford Integrated Digital Library Project* realizovaný na Stanfordské univerzitě v 2. polovině 90. let v rámci amerického programu DLI-1. Projekt byl zaměřen na vývoj technologií pro integraci širokého spektra existujících i budoucích heterogenních sbírek a informačních zdrojů do virtuální digitální knihovny s jednotným přístupem ke všem jejím

¹⁰⁰ Dostupný z WWW: <<http://www.openarchives.org>>

komponentám. Výsledky výzkumu byly realizovány v systému *InfoBus*¹⁰¹ (název vychází z analogie s hardwarovou sběrnici propojující různé hardwarové komponenty do jednoho funkčního celku) využívajícího technologii distribuovaných objektů na bázi systému CORBA (Common Object Request Broker Architecture).

Namísto pokusu adaptovat existující informační systémy je InfoBus ponechává v původním stavu. Pro každý z nich je zkonstruován zprostředkující 'wrapper', což je objekt systému CORBA reprezentující příslušnou online službu. Tyto zprostředkující objekty (*proxies*) komunikují s existujícími systémy v jejich „mateřském“ komunikačním jazyku a transformují zprávy do/z interního standardního rozhraní, kterým je protokol DLIOp (Digital Library InterOperability Protocol) podporující distribuované objekty. Např. určitý klient s vyhledávacím rozhraním Z39.50 chce vyhledávat v nějaké online informační službě, kterou může být např. systém Dialog. K tomu je zapotřebí dvou zprostředkujících objektů, jeden pro překlad mezi Z39.50 a DLIOp, druhý pro překlad mezi Dialogem a DLIOp. Ve Stanfordu vyvinuli řadu takových zprostředkujících objektů umožňujících prostřednictvím InfoBusu komunikovat libovolným Z39.50 klientem s velkou škálou informačních služeb, které Z39.50 nepodporují (souběžně byla na Michiganské univerzitě implementována proxy služba, která zprostředkovává pomocí systému InfoBus zdroje s protokolem Z39.50). Dále byly vyvinuty proxy služby pro HTTP, webové vyhledávače a řadu dalších služeb.

Architektura InfoBusu obsahuje řadu dalších komponent potřebných pro realizaci komplexního systému:

- SMA – standardní metadatová architektura pro unifikovaný popis informačních služeb a jejich zdrojů pro podporu vyhledávání;
- STARTS (STANford protocol proposal for internet ReTrieval and Search) – vrstva sloužící k organizaci metavyhledávání (*metasearching*), včetně výběru zdroje, vyhodnocení dotazů a slučování výsledků hledání;
- UPAI (Universal Payment Application Interface) – řeší mechanismus placení za poskytnuté služby;

¹⁰¹ Stanford University Digital Libraries Project

- FIRM (Framework for Interoperable Rights Management) – řada propracovaných technik pro řízený přístup ke zdrojům s ohledem na dodržování konkrétních podmínek vlastnických práv.

Jako drobnou zajímavost k výzkumu digitálních knihoven na Stanfordské univerzitě lze uvést skutečnost, že vedl mj. i k technologiím, které stály u zrodu dnes neznámějšího a nejpoužívanějšího internetového vyhledávače Google (hlavní tvůrci Googlu se jako postgraduální studenti Stanfordské univerzity podíleli na výzkumu vyhledávání v oblasti digitálních knihoven).

6.5. Open URL a SFX

Problematika otevřeného, kontextově citlivého propojování zdrojů (*open and context-sensitive linking*) patří v posledních dvou letech k jedné z nejživějších oblastí digitálních knihoven. Představme si následující situaci: dnešní typická digitální knihovna určité instituce se skládá z řady heterogenních informačních zdrojů, ať již vlastních (knihovní katalog, digitalizované sbírky), nebo cizích (licencované plnotextové časopisy v elektronické podobě, informační databáze, abstraktové a citační služby, zdroje volně přístupné na internetu), které jsou dostupné buď externě v repozitářích příslušných producentů či zprostředkovatelů, nebo lokálně v podobě zrcadlených zdrojů či dle místních potřeb upravených systémů. Provozovatel a uživatelé takové digitální knihovny mají zájem na tom, aby informace z jednotlivých zdrojů byly co nejvíce integrovány, např. *provázány* pomocí hypertextových vazeb jdoucích *napříč těmito zdroji*: z citace v komerční citační databázi na záznam publikace v lokálním katalogu, ze záznamu v katalogu nebo z citace v seznamu referencí nějakého článku na plný text článku v elektronickém časopise příslušného nakladatele, ze slov v názvu článku nebo předmětových hesel na relevantní informace v příslušném internetovém vyhledávači apod. Navíc by tyto vazby měly být „inteligentní“ v tom smyslu, aby zohledňovaly konkrétního uživatele a odkázaly ho vždy na zdroj odpovídající jeho statusu (např. na plný text licencovaného článku v případě zaměstnance instituce, na volně dostupný abstrakt, pokud je uživatelem cizí osoba). Standardní „linkovací“ řešení nabízená v posledních letech komerčními producenty informačních zdrojů jsou omezená (mají dosah jen v rámci informačního prostoru daného producenta),

kontextově necitlivá (odkazují vždy na stejný cíl bez ohledu na to, který uživatel a s jakými právy je používá) a uzavřená (nedovolují třetí straně – např. knihovně – nastavovat tyto vazby podle svých vlastních potřeb).

Řešení, které umožňuje překonat omezení dřívějších přístupů a realizovat představy z úvodu tohoto odstavce, nabízí standard OpenURL a nad ním postavený aplikační rámec SFX (Special Effects), které vycházejí z výsledků výzkumu na konci 90. let na univerzitě v belgickém Gentu¹⁰². Podstatou řešení je, že na rozdíl od klasických vazebních referencí, kdy *výchozí zdroj* (např. citace článku) odkazuje hypertextovou vazbou přímo na *cílový zdroj* (plný text článku), se oddělí popis zdroje (citace s odkazem) od poskytování vazeb, takže obecné vazební schéma pak vypadá následovně:

výchozí zdroj odkazuje na *servisní službu* (service component), která teprve odkazuje na *správný cílový zdroj*.

Implementace tohoto schématu v kontextu SFX je založena na několika principech:

1. servisních služeb existuje více, uživatel je i se svými právy registrován u některé z nich (servisní službu může implementovat např. jeho knihovna nebo nějaká třetí strana);
2. aby servisní služba mohla (dynamicky) určit správné cílové zdroje (nemusí být jeden) pro daný výchozí zdroj a daného uživatele, potřebuje znát podrobnosti o výchozím zdroji – jeho metadata;
3. tato *metadata nese v sobě přímo URL výchozího zdroje*, na který uživatel klikl, a to zakódována v podobě OpenURL. Např. výchozím zdrojem nechť je citace článku v databázi Medline nakladatele Ebsco Publishing: *Moll, JR. Attractive electrostatic interactions. J Biol Chem. 2000 Nov 3, 275(44):34826-32. doi:10.1074/jbc.M004545200.*

Nakladatel doplní k této citaci OpenURL, které může mít následující tvar:
<http://sfx1.exlibris.com/demo?sid=ebsco:medline&aulast=Moll&auinit=JR&date=2000-11-03&stitle=J%20Biol%20Chem&volume=275&issue=44&spage=34826>

První částí OpenURL je adresa servisní služby, za ní následuje identifikátor zdroje,

¹⁰² Van De Sompel, H.; Hochstenbach, P., nestr.

v němž uživatel klikl na OpenURL, a poslední částí jsou metadata a identifikátory výchozího zdroje zakódovaná dle specifikace OpenURL [61] (NISO již zahájilo tzv. zrychlené řízení pro přijetí OpenURL jako ANSI standardu);

4. protože OpenURL jsou ve výchozím zdroji vytvářena dynamicky, je možné a potřebné v nich adresu servisní služby měnit tak, aby odpovídala správné servisní službě příslušného uživatele. K propojení uživatele a jeho servisní služby nenabízí současná infrastruktura webu žádný solidní a dostatečně obecný mechanismus. Nicméně existuje několik pragmatických řešení (např. mechanismus CookiePusher);

5. koncepce předpokládá spolupráci producentů informačních zdrojů ve smyslu doplnění odkazů OpenURL do jejich zdrojů a kooperaci třetích stran při implementaci servisních služeb. Překvapivě během velmi krátké doby od zveřejnění specifikace OpenURL ohlásila řada světově významných producentů informací dostupnost svých „OpenURL enabled“ zdrojů a izraelská firma Exlibris (producent knihovního systému Aleph používaného v Národní knihovně ČR a v mnoha dalších velkých knihovnách u nás) získala licenci na SFX [70] a uvedla na trh první komerční implementaci servisní služby SFX-server a komplexní řešení pro integraci heterogenních digitálních zdrojů (zahrnujících i SFX server) pod názvem Metalib.

Souhrnný scénář práce v prostředí SFX vypadá následovně:

- uživatel přes standardní webový prohlížeč vyhledá v informačním zdroji (např. v citační databázi Web of Science) výchozí zdroj (citaci článku) a klikne na jeho OpenURL;
- OpenURL zdroje odkazuje na servisní službu uživatele; ta je aktivována a z obdrženého OpenURL si vyzvedne metadata výchozího zdroje;
- servisní služba vyhodnotí metadata výchozího zdroje (např. provede vyhledání informací o výchozím zdroji v různých databázích, k nimž má uživatel oprávnění);
- vrátí uživateli hypertextový seznam příslušných cílových zdrojů (*appropriate extended service links*), který může zahrnovat např. plný text zdroje, odkaz na

záznam v lokálním online katalogu s uvedením lokace dokumentu, odkazy na další práce autora výchozího zdroje vyhledané na internetu atd.

V literatuře je uveden příklad jednoho z dalších možných využití technologie OpenURL a SFX v kombinaci se systémem DOI, který umožňuje aplikovat výše uvedený scénář i na informační zdroje, které nepodporují OpenURL.

Technologie OpenURL a SFX otevírá nové možnosti pro širokou integraci (interoperabilitu) heterogenních informačních zdrojů v současných digitálních a heterogenních knihovnách.

7. Globální vyhledávání zdrojů

Podobně jako jsou navzájem provázány oblasti metadat a vzájemného propojování digitálních knihoven (metadata pro interoperabilitu a interoperabilita metadat), i oblast globálního vyhledávání zdrojů v distribuovaném prostředí digitálních knihoven souvisí velmi těsně s metadaty i s interoperabilitou – a naopak.

7.1. Úvodní poznámky

Detailní rozbor všech aspektů této problematiky lze nalézt v práci¹⁰³; stručně je lze shrnout do pěti podoblastí: organizace, systémy, digitální obsah, rozhraní a metriky.

Organizace – v oblasti distribuovaného vyhledávání má každé řešení svůj organizační aspekt. Mezi heterogenními, distribuovanými, nezávisle spravovanými systémy musí vždy existovat určitá forma koordinace, má-li být vyhledávání zdrojů dostatečně efektivní. Jak již bylo naznačeno u interoperability, tato koordinace může mít velmi rozdílné formy – od rozsáhlého rozšíření silných standardů a komunikačních protokolů až po velmi volnou kooperaci založenou jen na použití stejných základních technologií (shromažďování dat webových serverů internetovými vyhledávači). Strategie pro organizaci distribuovaných komponent digitální knihovny musí brát v úvahu různorodost zainteresovaných institucí, jejich rozdílné priority, potřeby, cíle – ale také např. bezpečnostní a cenové otázky.

Systémy – existuje silná potřeba vyvinout systémovou infrastrukturu podporující vyhledávání, navigaci, zprostředkovávání a získávání informací v záplavě různorodých dat dostupných online. Součástí této infrastruktury musí být nástroje pro výběr informačních bází na systémové úrovni (přesměrování (*routing*) dotazů ke správným fyzickým serverům), interakci informačních bází s překonáním jejich heterogenity (mezirepozitářové protokoly, distribuované vyhledávací protokoly, mechanismy pro zajištění bezpečnosti, soukromí, kooperativní autentifikace, placení) a zajištění konzistence ve složitém distribuovaném systému.

Obsah – množství a variabilita forem digitálního obsahu vyžaduje schopnost řešit efektivně problémy, jako je optimální výběr informačních bází na logické úrovni (za použití

metadat pro popis celých informačních bází zahrnujících na jedné straně obsah a jeho kvalitu, ale na druhé straně též výkonnostní, cenové a další přístupové parametry), dotazovací jazyky pro netextové informační zdroje (multimediální a dynamické dokumenty), nástroje pro ohodnocování vyhledaných informačních zdrojů (*ratings*) a efektivní filtraci informací a konečně také mechanismy pro překonání sémantické heterogenity mezi informačními bázemi umožňující přechod od vyhledávání explicitních informací k získávání implicitních poznatků (*knowledge discovery*).

Rozhraní – oblast komunikace člověk – počítač (Human-Computer Interaction – HCI) lze z pohledu digitálních knihoven rozdělit zhruba do čtyř rovin; první dvě se tradičně týkají vstupu a výstupu (mechanismy konstrukce a zadávání dotazů na vstupu, prezentace či vizualizace výsledků při výstupu), další dvě se týkají pokusů o strojové porozumění tomu, co uživatel zamýšlí provádět (*task understanding*), a naopak pochopení procesů realizovaných systémem ze strany uživatele (*process exposure*) – zatímco někteří uživatelé jsou mnohem produktivnější, když rozumí tomu, jak jejich nástroj pracuje, jiní mohou být větším množstvím detailů zmateni a preferují přístup „černé skříňky“. Řešením může být podpora pro široký individualizovaný přístup.

Metriky – pro vyhodnocování efektivity různých řešení a přístupů jsou vytvářeny nejrůznější taxonomie pro různé třídy uživatelů a vzorce jejich chování, dotazovací mechanismy, prezentaci výsledků apod., které je nutné testovat na reálných datech a reálných uživateli. Silně je pocíťována potřeba odpovídajících rozsáhlých ověřovacích prototypových řešení (*testbeds*), které by zahrnovaly velké množství distribuovaných informačních bází, široké spektrum médií a formátů a diverzifikovanou informaci z pohledu kvality, časových charakteristik a cílových tříd uživatelů – to vše spolu s distribuovanou sdílenou kolekcí služeb a vyhledávacích a navigačních nástrojů.

Dosavadní praxe ve sféře globálního distribuovaného vyhledávání zdrojů potvrzuje řadu poznatků z historie v tom smyslu, že hrubá výpočetní síla zatím vítězí nad přístupy založenými na umělé (a někdy i přirozené) inteligenci. Arms popisuje oblasti, v nichž využití hrubé síly přineslo v posledních letech překvapivě dobré výsledky¹⁰⁴: vyhledávání informací (webové vyhledávače), rozhodování, nakolik vyhledaný dokument odpovídá zadanému

¹⁰³ Ressource Discovery in a Globálky-Distributed Digital Library, nestr.

dotazu (přístupy z oblasti sémantiky dokumentů, viz např. projekt digitální knihovny Illinoiské univerzity¹⁰⁵), vyhodnocování důležitosti dokumentů (řadicí algoritmus systému Google), archivace digitálního dědictví (automatizovaný přístup v Internet Archive nebo švédském programu Kulturarw3), citační analýza (ResearchIndex), kontextové propojování informačních zdrojů (SFX), automatická extrakce metadat z multimediálních digitálních objektů (Infermedia Digital Video Library na univerzitě Carnegie Mellon¹⁰⁶) nebo pokusy o vytvoření automatického referenčního knihovníka (projekt na univerzitě ve Washingtonu¹⁰⁷).

7.2. DK a internetové vyhledávací systémy

Informační exploze na internetu vyvolala potřebu okamžitého pragmatického řešení problému, jak v chaotickém moři informací vyhledávat a zprostředkovávat přístup k požadovaným informacím. Odpovědí byly internetové vyhledávací služby – vyhledávače (*search engines*) a adresáře (*directories*). Při srovnání vyhledávačů s přístupy klasických knihoven jsou rozdíly markantní; stručně a výstižně to charakterizuje citát: „*Almost everything that is best about a library catalog is done badly by a web search service. ... On the other hand, web search services are strong in ways that catalogs are weak.*“¹⁰⁸. V tomto duchu byly až donedávna i digitální knihovny a internetové vyhledávače považovány obecně za dvě naprosto nezávislá paradigmatu využívající webového prostředí k vytváření informačních repozitářů. Práce autorů U. Hanani aj., F. Ariel ukazuje, že ve skutečnosti mají obě hodně společného a je třeba je chápat nikoliv jako konkurenční, nýbrž alternativní, doplňující se přístupy (vyhledávače pro rychlou první odpověď, digitální knihovna pro vysoce kvalitní cílenou informaci). Digitální knihovny jsou teoreticky dobře podložené, perspektivní, nabízejí či slibují širší a v mnoha aspektech lepší služby; prakticky jsou však zatím stále ještě nedostatečně zvládnuté a v globálním měřítku nerealizované. Webové vyhledávače jsou naopak prakticky realizované a široce dostupné, avšak jejich vyhledávání je obecně málo přesné, zaměřené pouze na oblast volně dostupných zdrojů na tzv. povrchovém webu (pro

¹⁰⁴ Arms, W.Y., nestr.

¹⁰⁵ UIUC Digital Library Testbed

¹⁰⁶ Infermedia Project

¹⁰⁷ Automatic Reference Librarian Project

¹⁰⁸ Arms, W.Y., nestr.

vyhledávače nedostupný „hluboký“ web je údajně až 500krát rozsáhlejší¹⁰⁹) a řadu dalších služeb nad rámec vyhledávání nerealizují vůbec.

Ve své krátké historii prošly oba přístupy třemi etapami s mnoha podobnými charakteristikami:

Vyhledávače – první generace (základní vyhledávače) je představována relativně jednoduchými přístupy založenými na jednoduchých metadatových strukturách a plnotextových indexech. Existují v podobě vyhledávačů buď univerzálních (např. AltaVista, Lycos), nebo specializovaných (např. MedHunt, TravelFinder). Druhá generace (metavyhledávače, multivyhledávače) klade důraz na snazší metody pro lokalizaci zdrojů, redukci nasbíraných výsledků, jednoduché metody jejich ohodnocování a kombinaci více různých základních vyhledávačů (např. MetaCrawler, SavvySearch). Třetí generace (paralelní vyhledávače, portálové vyhledávače) spojuje vyhledávače a adresářové služby a nabízí pokročilejší techniky pro vyšší kvalitu služeb (lepší ohodnocování, kontextové techniky pro identifikaci relevantních vazeb), zohlednění uživatelských potřeb (uživatelská zpětná vazba a individualizace) a rychlejší vyhledávání (např. Google, FAST, DirectHit, FizziLab).

Digitální knihovny – první generaci (samostatné (*stand-alone*) digitální knihovny) představovaly víceméně klasické, plně digitalizované a izolované digitální knihovny s lokálně ohraničeným a centralizovaným digitálním materiálem. Existovaly buď jako univerzálněji zaměřené (např. digitální knihovna Kongresové knihovny, projekt Alexandria), nebo specializované (např. Making of America na Michiganské univerzitě, digitální knihovna ACM). Druhá generace (federalizované digitální knihovny) byla nejčastěji organizována jako federace několika nezávislých samostatných digitálních knihoven organizovaných na základě společného tématu a nabízející jednotné uživatelské rozhraní pro transparentní přístup k heterogenním komponentám (viz např. Networked CompSci Technical Reference Library). Třetí generace (sklizené digitální knihovny) je představována virtuálními digitálními knihovnami poskytujícími sumarizovaný přístup k relevantnímu materiálu rozmístěnému po globální síti. Obsahem takové knihovny bývají pouze metadata získávaná s využitím automatizovaných technik sklizení (*harvesting*) na základě definic informačního prostoru

¹⁰⁹ Bergman, M.K., nestr.

vytvářených informačními specialisty a při kontrole potřebné kvality (např. SourceBank, ArticleCentral.com).

Autoři také předpovídají postupně konvergující vývoj obou přístupů, který bude postupovat přes inteligentní vyhledávače a inteligentní digitální knihovny více využívající technik umělé inteligence a správy znalostí až po společný megaportál/metaportál poskytující unifikovaný přístup a deklarativní vyhledávání ve všech datových repozitářích vytvořených libovolnými technikami obou přístupů.

8. Související oblasti s problematikou DK

Na závěr se velmi stručně zmíníme o dvou dalších oblastech, které sice nesouvisí přímo s technologiemi digitálních knihoven, přesto však hrají klíčovou roli v tom, zda digitální knihovny jako takové budou úspěšné a naplní očekávání svých tvůrců a uživatelů.

8.1. Copyright, intelektuální vlastnictví

Digitální knihovny nejsou zdaleka jen problémem technologickým; technický rámec digitálních knihoven vždy působí v určitém legislativním, ekonomickém a společenském kontextu. Přizpůsobení tohoto společenského kontextu tak, aby systém digitálních knihoven mohl efektivně a v globálním měřítku fungovat, je přitom záležitost mnohem složitější a časově náročnější než realizace vlastního technického řešení. Nejdůležitější komponenty, ekonomika a legislativa, se přitom úzce vzájemně ovlivňují a podmiňují. Mezi základní otázky teorie a praxe digitálních knihoven patří zejména: (1) *funkce autorského práva v digitálním prostředí*, tedy otázka, jak vyvážit veřejné právo na přístup k informacím s oprávněnými ekonomickými zájmy autorů a vydavatelů; (2) *jak pokrýt nákladové položky* v procesu vzniku, organizace, zpřístupňování a uchovávání digitální informace tak, aby byly zajištěny ekonomické podmínky dlouhodobé provozuschopnosti digitální knihovny. V ekonomické oblasti jsou zkoumány různé obchodní modely jak pro oblast *otevřeného přístupu*, kdy informační zdroje jsou z pohledu uživatelů k dispozici bezplatně (náklady ovšem hradí někdo jiný), tak i oblast přístupu placeného (využívajícího různá platební schémata – od předplatného, přes poplatky typu pay-by-use až po mikroplatby). *Právní aspekty* digitálních knihoven zahrnují nejen proces vytváření nové legislativy či zkoumání dopadu šíření a využívání zdrojů v digitální formě na oblast ochrany duševního vlastnictví, ale celý provázaný komplex otázek, kam patří také ochrana osobních údajů a ochrana soukromí, odpovědnost za obsah poskytovaných informací, otázky zodpovědnosti provozovatele systému za nelegální jednání jeho uživatelů a mnohé další. Intenzívně jsou zkoumány a rozvíjeny také *technologie ochrany* digitálních informací před neoprávněným přístupem a kopírováním (hardwarové zámky, steganografie, šifrování digitálního obsahu, flickering aj.).

8.2. Dlouhodobé uchování digitálních informací

Po četných negativních zkušenostech s rozpadajícími se tisky a ztracenými či zničenými fondy ve světě klasických knihoven se nástup digitálních technologií jevil jako dlouho očekávané „definitivní“ řešení problému efektivní ochrany a trvalého uchování informací. Bity nestárnou, neznechodocují se používáním či rozmnožováním, lze je snadno kopírovat v nezměněné kvalitě. Rychle se však ukázalo, že přes nesporné přínosy a výhody přináší přechod na digitální informační zdroje vážné problémy paradoxně právě z hlediska dlouhodobého uchování. Na vině je jednak relativně krátká životnost nosičů digitální informace, ale zejména velmi krátký a stále se zrychlující inovační cyklus digitálních technologií (v průměru zhruba 5 let). Situaci vystihuje citát „*digital information lasts forever – or five years, whichever comes first*“. Navzdory aktivnímu výzkumu v oblasti *strategií pro dlouhodobé uchování* (replikace, ožívování, technické muzeum, migrace, emulace, konverze do analogové formy aj.) není obecně současná situace z hlediska dlouholetého uchování digitální informace zatím nijak příznivá. Systematický koncepční přístup k problematice uchování zahrnující technické i organizační a systémové přístupy představuje referenční model OAIS – Open Archival Information System. Zajímavé projekty hraničící s intelektuálními cvičeními z oblasti opravdu „dlouhodobého“ uchování (po dobu tisíciletí) shromažďuje nadace The Long Now – zmiňme alespoň projekt „Rosettská deska“, usilující vytvořit trvalý lingvistický archiv a překladatelský nástroj pro obnovu tisícovky soudobých jazyků ztracených v *hluboké budoucnosti* metodou konverze do analogové formy, a to zaznamenáním obrazů až stovky tisíc stran textů vyrytím do niklového disku (s životností několika tisíc let) prostřednictvím optické nanolitografie.

9. Programy a projekty DK

V současnosti existují tisíce dokončených nebo probíhajících projektů digitálních knihoven po celém světě. Popsat stručně jen malou část z nich by vyžadovalo samostatný rozsáhlý článek. V tomto textu již byly zmíněny či odkázány některé projekty, v rámci kterých se vyvíjí vybrané klíčové komponenty současné infrastruktury digitálních knihoven. Doplňme proto vybrané, celosvětově nejdůležitější programy, které podporují výzkum i praktický vývoj a budování konkrétních digitálních knihoven a přinášejí nejvýznamnější podněty pro celou oblast.

9.1. *Digital Library Initiative – Phase 1*

Od počátku 90. let 20. století probíhala v odborných kruzích ve Spojených státech široká diskuse o potřebě zásadní pomoci výzkumu na podporu vlny nově vznikajících projektů z oblasti digitálních knihoven a jeho začlenění do programu národní informační infrastruktury. Pod koordinací National Science Foundation (NSF) a za spolupráce s agenturou DARPA (Defense Advanced Research Project Agency) a kosmickou agenturou NASA vznikl pětiletý program Digital Library Initiative, Phase 1 (DLI-1) pro období 1994-1998, jehož cílem bylo *dosáhnout zásadního technologického pokroku při sběru, ukládání a organizaci digitálních informací a jejich uživatelsky přívětivého zpřístupnění v globální síti*". Jako prostředek k dosažení tohoto cíle byla zvolena masivní finanční podpora jen velmi omezenému počtu špičkových výzkumných projektů z různých oblastí digitálních knihoven, které měly šanci na dosažení zásadního průlomu v poznání nových technologií a jejich ověření prostřednictvím rozsáhlých prototypových řešení (*testbeds*). Celkem bylo vybráno šest projektů předních amerických univerzit (každá z nich vytvořila k řešení projektu výzkumné konsorcium zahrnující řadu dalších subjektů, včetně významných komerčních firem), z nichž každý dostal podpůrný grant ve výši 4 milionů USD (včetně dalších zdrojů dosáhly celkové náklady na řešení těchto projektů 75 milionů USD). Šlo o tyto projekty:

University of Michigan DL Project – projekt zaměřený na vytváření rozsáhlé multimediální digitální knihovny z oblasti věd o zemi a výzkumu vesmíru, která byla tvořena velkým množstvím informačních repozitářů a systematickým způsobem zpřístupňovala velké množství informací z mnoha různých tematických oblastí na internetu.

University of Illinois – Building the Interspace: DL Infrastructure for a University Engineering Community – projekt zaměřený na integraci přístupu k textovým dokumentům ve formě (různě označovaných) elektronických verzí článků v SGML z odborných technicky zaměřených časopisů od různých producentů. Součástí bylo i zkoumání algoritmů využívajících statistických technik pro analýzu sémantiky dokumentů.

University of California (Berkeley) – The Environmental Electronic Library: A Prototype of a Scalable, Intelligent, Distributed Electronic Library – projekt zaměřený na vývoj technologií pro inteligentní přístup k obrovským distribuovaným databázím obsahujícím fotografie, satelitní snímky, mapy, videozáznamy, plné texty a další typy dokumentů s cílem zpřístupnit rozsáhlé množství veřejně přístupných dat z oblasti životního prostředí.

Carnegie Mellon University – Informedia: Integrated Speech, Image and Language Understanding for Creation and Exploration of Digital Video Libraries – využití integrovaných technologií z oblastí rozpoznávání řeči, porozumění přirozenému jazyku a zpracování obrazu/videosekvencí pro obsahově založené vyhledávání v terabytové digitální videoknihovně.

Stanford University Integrated Digital Library Project – vývoj technologií pro integraci širokého spektra existujících i budoucích heterogenních sbírek a informačních zdrojů do virtuální digitální knihovny s jednotným přístupem ke všem jejím komponentám. Vyvíjené technologie byly ověřovány na prototypu InfoBus.

University of California (Santa Barbara) – The Alexandria Project: Towards a Distributed DL with Comprehensive Services for Images and Spatially Referenced Information – digitální knihovna pro snadný přístup k rozsáhlým a různorodým sbírkám map, obrázků, leteckých snímků z kalifornské oblasti s využitím nástrojů z geografických informačních systémů.

9.2. Digital Library Initiative – Phase 2

Na program DLI-1 bezprostředně navázal v období 1998–2002 jeho následník DLI-2. Nebyl již zaměřen jen na výzkum, ale také na budování digitálních sbírek a rozšíření sféry působnosti i do nových oblastí, především do lékařských a humanitních oborů. Jeho motem

bylo „zajistit vedoucí roli ve výzkumu klíčovém pro vývoj nové generace digitálních knihoven, zvýšit využívání a použitelnost globálních distribuovaných síťových informačních zdrojů a povzbudit stávající i nové komunity v zaměření na nové inovativní aplikační oblasti digitální knihoven“. Stručně by se dal program charakterizovat podle následujících hesel: lépe využívat to, co již existuje a zjistit, co ještě chybí; komunikovat a spolupracovat; učinit technologii (pro uživatele) neviditelnou.

Ke třem vyhlášovatelům DLI-1 se přidaly další instituce (Kongresová knihovna, Národní lékařská knihovna a další), zvýšil se objem finanční podpory na 15 milionů USD ročně po dobu pěti let a program se stal otevřeným (průběžně vypisovaná nová kola výběrového řízení, širší zaměření projektů a různá délka řešení, možná účast zahraničních partnerů). Grantovou podporu získalo celkem více než 50 projektů, z toho několik mezinárodních, pokrývajících velmi široké spektrum výzkumných a aplikačních oblastí; mezi nimi byly i projekty navazující na oněch šest původních projektů z DLI-1. Podrobnější informace o programu a jednotlivých projektech lze získat v dalších webových zdrojích.

9.3. *Electronic Library Programme (eLIB)*

Britský program eLIB, the Electronic Library Programme, probíhal ve třech etapách v letech 1994-2000 a na rozdíl od převážně výzkumně zaměřeného programu DLI, byl orientován ryze prakticky s cílem pokrýt co nejširší oblast středoškolského a vysokoškolského sektoru – při řešení celkem 80 projektů se do něj zapojila více než stovka vzdělávacích institucí. Mezi podporované oblasti v prvních fázích programu patřily elektronické publikování a digitalizace, přístup k elektronickým zdrojům a elektronické dodávání dokumentů, vzdělávání a výuka; v závěrečné fázi se podpora soustředila na hybridní knihovny, dlouhodobé uchovávání digitálního materiálu, realizace souborných virtuálních katalogů s využitím technologie Z39.50 a zejména transformaci řešení a služeb vytvořených v prvních fázích projektu do podoby trvale provozovatelných služeb. Program se stal katalyzátorem pro široký rozvoj elektronických informačních služeb a digitálních knihoven a získávání teoretických i praktických zkušeností z oblasti digitálních knihoven na britských vzdělávacích institucích.

9.4. National Digital Library Program (NDLP)

Kongresová knihovna získala první rozsáhlejší zkušenosti s velkoplošnou digitalizací a zpřístupňováním digitálního obsahu v pilotním projektu *American Memory* (1990-1995). V návaznosti na něj pak vyhlásila pětiletý program National Digital Library Program (NDLP), který James O'Donnell označil za knihovní „projekt Apollo“. Cílem tohoto programu bylo ve velmi krátké době zdigitalizovat a zpřístupnit na síti 5 milionů artefaktů ze sbírek Kongresové knihovny týkajících se americké historie (jedinečné fotografie, rukopisy, vzácné knihy, mapy, zvukové nahrávky, filmy), zejména pro potřeby výuky na všech typech škol, od mateřských až po univerzity (hlavní cílovou skupinou jsou žáci základních a středních škol). Výsledky programu, na kterém spolupracuje s Kongresovou knihovnou řada dalších významných knihoven, škol i komerčních organizací, jsou soustředěny do více než stovky digitálních multimediálních sbírek sdružených pod *American Memory Historical Collections*; v době psaní tohoto příspěvku obsahovaly sbírky přes 9 milionů digitálních položek. Jen pro představu: v polovině roku 1999 zaměstnával program NDLP více než 100 osob a měl roční rozpočet 12 milionů USD. Program vyvinul vlastní standardy, digitalizační postupy a doporučení, metody integrace heterogenních digitálních sbírek a prezentační metody; velká pozornost je věnována problematice dlouhodobého uchování digitální informace.

9.5. Další projekty

Evropská unie zatím nemá žádný samostatný program zaměřený výlučně na digitální knihovny. Nicméně v rámci tematického programu IST 5. rámcového programu existuje ve skupině *Multimedia Content and Tools* jako jedna z hlavních oblastí *Digital Heritage and Cultural Content*, ve které je každoročně vyhlášováno několik témat s problematikou digitálních knihoven úzce souvisejících (např. téma *Next Generation Digital Collections* vyhlášené pro rok 2001). Kromě výzkumně zaměřeného programu IST existoval paralelní aplikační program *eEurope Initiative* na léta 2001-2004, v jehož rámci byla vyhlášena aktivita *eContent* zaměřená na vytváření a zpřístupňování evropských digitálních sbírek.

Z dalších zemí, které jsou na poli digitálních knihoven velmi aktivní, zmiňme především Německo, Francii (a její projekt *Bibliotheca Universalis*, který se postupně rozvinul v mezinárodní kooperativní digitalizační program *Paměť světa* pod záštitou

UNESCO) a z mimoevropských zemí především Austrálii a Kanadu. Široce pojaté a štědře dotované programy na podporu rozvoje digitálních knihoven nejsou však jen doménou státních rozpočtů, jak o tom svědčí např. ambiciózní program Library Digital Initiative Harvardovy univerzity.

10. Závěrem

Digitální knihovny představují fascinující a dynamicky se rozvíjející směr v oblasti pořádání a zpřístupňování digitálních informací. Ačkoliv jde o oblast velmi širokou po stránce výzkumných témat, používaných technologií i způsobů realizace, objevuje se stále více standardizovaných dílčích přístupů prověřených praxí, umožňujících vytvářet již dnes velmi rozsáhlé systémy spolupracujících digitálních knihoven. Ty vytvářejí předobraz inteligentních, vzájemně propojených digitálních knihoven budoucnosti, které budou poskytovat rychlé, spolehlivé, kvalitní a vyčerpávající informace a všestranné služby – přesně dle požadavků a potřeb uživatelů – ve kteroukoliv dobu, na kterémkoliv místě a na jakémkoliv téma.

Použitá literatura:

1. ARMS, W.Y. *Digital libraries*. Cambridge : MIT Press, 2000.
2. ARMS, W. Y. *Open Access to Digital Libraries : Must Research Libraries Be Expensive?* [online]. Invited Talk to European Conference on DL 2000, Lisbon, 2000. Dostupný z WWW: <<http://www.cs.cornell.edu/wya/papers/Lisbon-2000.doc>>.
3. *Automatic RDF Metadata Generátor* [online]. De-signed and implem. By Ch. Jenkins. Wolverhampton : Wolverhampton Univ., [cit. 2006-05-11]. Dostupný z WWW: <<http://www.scit.wlv.ac.uk>>.
4. *Automatic Reference Librarian Project* [online]. University of Washington. Dostupný z WWW: <<http://www.cs.washington.edu/research/diglib/>>.
5. BERGMAN, M. K. *Deep Web : White Paper* [online]. BrightPlanet, c2004. Dostupný z WWW: <<http://brightplanet.com/technology/deepweb.asp>>.
6. 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>>.
7. BORGMAN, Christine L. What are digital libraries, who is building them, and why? In *Digital libraries : interdisciplinary concepts, challenges and opportunities*. Zagreb : Benja, 1999, S. 29.
8. 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>.
9. BURNARD, L; LIGHT, R. *Three SGML metadata formats : TEI, 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/>>.
10. *Česká terminologická databáze knihovnictví a informační vědy : (TDKIV)* [online]. Praha : Národní knihovna ČR, c2004 [cit. 2006-02-16]. Dostupný z WWW: <<http://sigma.nkp.cz/F>>.
11. DAY, Michael. *The ROADS metadata registry* [online]. Bath : Metadata Group, UKOLN, University of Bath, last upd. 2000-05-16 [cit. 2006-04-03]. Dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/roads/templates/>>.

12. 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/>>.
13. 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>.
14. 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>.
15. DCMI. *Dublin Core Metadata Initiative : Home Page* [online]. Dublin : DCMI, c1999 [cit. 1999-07-27]. Přístup z: <<http://purl.org/dc/>>.
16. 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>>.
17. 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>.
18. 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/>>.
19. 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>>.
20. *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>.

21. DRABENSTOTT, Karen M. *Analytical review of the library of the future*. Washington : Council Library Resources, 1994.
22. *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/>>.
23. *EU-NSF Digital Library Working Group on Interoperability between Digital Libraries : Position Paper* [online]. ERCIM-DELOS, 1999. Dostupný z WWW: <<http://www.iei.pi.cnr.it/DELOS/NSF/interop.htm>>.
24. 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>>.
25. HANANI, U., ARIEL, J. F. The Parallel Evolution of Search Engines and Digital Libraries : Their Convergence to the Mega-Portal. In *Proceedings of Kyoto International Conference on Digital Libraries : Research and Practice*. IEEE Computer Society Press, 2000, s. 269–276.
26. HEERY, Rachel. *ROADS Templates : how they are used* [online]. Bath, last upd. 1998-10-05 [cit. 2006-04-03]. Dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/templates>>.
27. *Informedia Project* [online]. Carnegie Mellon University, c2004. Dostupný z WWW: <<http://www.informedia.cs.cmu.edu/>>.
28. KAHN, Robert; WILENSKY, Robert. *A Framework for Distributed Digital Object Services* [online]. 1995-05-13 [cit. 2006-04-19]. Dostupný z WWW: <<http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>>.
29. KIRRIEMUIR, John. *Chat is ROADS?* [online]. Bath : University of Bath, UKOLN, Metadata Group [cit. 2006-04-03]. Dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/roads/what/>>.
30. KNIGHT, John; HAMILTON, Martin. *Dublin Core standards resource types*. Loughborough : Loughborough University, 1997.
31. LAGOZE, C.; SHAW, E.; DAVIS, J.R.; KRAFFT, D.B. *Dienst : implementation reference manual*. Ithaca : Cornell University, 1995.

32. LYNCH, C. The Z39.50 Information Retrieval Standard. Part 1: A Strategic View of Its Past, Present and Future. *D-Lib Magazine* [online], April 1997. Dostupný z WWW: <http://www.dlib.org/dlib/april97/04lynch.html>.
33. MACCALL, S.L.; CLEVELAND, A.D.; GIBSON, I.E. Outline and preliminary evaluation of the classical digital library model. In *Proceedings of the fall 1999 annual meeting of the American society for information science*. 1999.
34. PAEPCKE, A., CHANG, C. K., GARCÍA-MOLINA, H., WINOGRAD, T. Interoperability for Digital Libraries Worldwide. *Communication of the ACM*, 1998, roč. 41, č. 4, s. 33–43.
35. POKORNÝ, J. Digitální knihovny : principy a problémy. In RAMAJZLOVÁ, B. (ed.). *Automatizace knihovnických procesů 8*. Praha : ČVUT, 2001, s. 27–38. Dostupný též z WWW: <http://knihovny.cvut.cz/akp/clanky/03.pdf>.
36. POWELL, Andy. *Metadata : resource organisation and discovery in subject-based services* [online]. Bath : UKOLN, Metadata [cit. 2006-04-03]. Dostupný z WWW: <http://www.ukoln.ac.uk/metadata/roads/>.
37. *Open Archives Initiative* [online]. Dostupný z WWW: <http://www.openarchives.org/>.
38. *Resource Discovery in a Globally-Distributed Digital Library : Working Group Report* [online]. ERCIM-DELOS, 1999. Dostupný z WWW: <http://www.iei.pi.cnr.it/DELOS/NSF/resourcediscovery.htm>.
39. SCHAMBER, Linda. What is a dokument? : rethinking the koncept in uneasy times. *Journal of the American Society for Information Science*. 1996, vol. 47, no. 9, s. 669-671.
40. SCHWARZ, Candy. LIS 462 – definitions [online]. 1999 [cit. 2006-02-16]. Dostupný z WWW: <http://web.simmons.edu/~schwartz/462-defs.html>.
41. *Stanford University Digital Libraries Project* [online]. Dostupný z WWW: <http://www-diglib.stanford.edu/diglib/pub/userinfo.html>.
42. *UIUC Digital Library Testbed* [online]. University of Illinois. Dostupný z WWW: <http://dli.grainger.uiuc.edu/idli/idli.htm>.
43. VAN DE SOMPEL, H., HOCHSTENBACH, P. Reference Linking in a Hybrid Library Environment. Part 1, Frameworks for Linking. *D-Lib Magazine* [online], 1999, roč. 5, č. 4. Dostupný z WWW: http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html.
44. VAN DE SOMPEL, H., HOCHSTENBACH, P. Reference Linking in a Hybrid Library Environment. Part 2, SFX, a Generic Linking Solution. *D-Lib Magazine* [online], 1999, roč.

- 5, č. 4. Dostupný z WWW:
<http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html>.
45. WATERS, D.J. What are digital libraries? *CLIR Issues* [online], 1998, č. 4, [cit. 2006-02-16]. Dostupný z WWW: <<http://www.clir.org/pubs/issues/issues04.html#dlf>>.
46. 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>>.
47. 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>>.
48. 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>>.
49. WHALEY, Tom. *Definitions of Digital Library* [online]. Lexington : Washington and Lee University, [2004]. [cit. 2006-02-16]. Dostupný z WWW:
<<http://home.wlu.edu/~whaley/classes/DigiLib/Whaley/Definition.html>>.
50. 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>>.
51. *Z39.50 Maintenance Agency* [online]. Washington : Library of Congress, 2004. Dostupný z WWW: <<http://lcweb.loc.gov/z3950/agency/>>.