

DIGITÁLNÍ KNIHOVNY – TEORIE A PRAXE

Miroslav Bartošek
Ústav výpočetní techniky,
Masarykova univerzita v Brně
bartosek@ics.muni.cz

Obsah

1. Úvod
 - 1.1 Co je to digitální knihovna?
 - 1.2 Krátce z historie
 - 1.3 Proč digitální knihovny
 - 1.4 Aktuální stav, hlavní současné aktivity a zdroje informací
2. Klíčové oblasti výzkumu a praxe digitálních knihoven
3. Obecný rámec a architektura digitální knihovny
 - 3.1 Kahn-Wilenského architektura
 - 3.2 Začlenění do právního a sociálního prostředí
 - 3.3 Hierarchická abstrakce intelektuálního díla (model IFLA)
4. Jména a identifikátory
 - 4.1 Koncept URN
 - 4.2 Klasické identifikátory: ISBN, ISSN, SICI/BICI, ISTC
 - 4.3 PURL – perzistentní URL
 - 4.4 Systém „handles“
 - 4.4 DOI – Digital Object Identifier
5. Metadata
 - 5.1 Úvod a stručný přehled
 - 5.2 Dublin Core
 - 5.3 Metadata Kongresové knihovny
 - 5.4 Standardy vycházející ze struktury MARC
 - 5.5 XML a RDF
6. Interoperabilita
 - 6.1 Úvod a stručný přehled
 - 6.2 Protokol Z39.50
 - 6.3 Open Archives Initiative (OAI)
 - 6.4 Stanfordský InfoBus
 - 6.5 OpenURL a SFX
7. Globální vyhledávání zdrojů
 - 7.1 Úvod a stručný přehled
 - 7.2 Digitální knihovny a internetové vyhledávače
8. Stručné poznámky k dalším oblastem digitálních knihoven
 - 8.1 Intelektuální vlastnictví a ekonomika
 - 8.2 Dlouhodobé uchovávání digitální informace
9. Vybrané programy a projekty
 - 9.1 Digital Library Initiative – Phase 1
 - 9.2 Digital Library Initiative – Phase 2
 - 9.3 Electronic Library Programme (eLIB)
 - 9.4 National Digital Library Program (NDLP)
 - 9.5 Ostatní

Literatura

1 Úvod

1.1 Co je to digitální knihovna?

Přestože pojem *digitální knihovna* (angl. digital library) patří v posledních letech k těm nejfrekventovanějším, panuje řada nejasností, co vlastně tento termín obnáší. Jednou z příčin tohoto stavu je skutečnost, že obsah pojmu digitální knihovna se průběžně vyvíjí – tak, jak se vyvíjí jeho technologická základna, výpočetní technika. Jiný důvod souvisí s tím, že problematikou digitálních knihoven se zabývá mnoho různorodých odborných komunit, z nichž každá si vytváří vlastní náplň tohoto pojmu v souladu se svým zaměřením: z pohledu databázového specialisty představuje digitální knihovna informační systém využívající architekturu federativních databází, pro odborníky zabývající se hypertextem a šířením informací je to jen jedna z nadstavbových aplikací webu, knihovník vidí v digitální knihovně další krok v automatizaci na cestě od knihovny analogové (papírové) přes automatizovanou či hybridní (fyzické sbírky s automatizovaným katalogem) až po digitální (většina či veškeré informace a služby knihovny jsou elektronické). V neposlední řadě přispívá ke zmatení pojmů i skutečnost, že pojmem digitální knihovna jsou někdy označovány systémy, které – přinejmenším z pohledu informačního specialisty – představují úplně jiné entity (např. soubory aplikací a procedur, systémy na správu dokumentů apod.).

S masovým rozšířením internetu po nástupu webových technologií se objevily i představy, že celý internet, resp. web jsou vlastně jednou digitální knihovnou. S tím však odborníci z oblasti informační vědy nesouhlasí. Clyford Lynch, jeden z předních amerických informačních specialistů, připomíná, že web nebyl pro podporu organizovaného publikování a vyhledávání informací vůbec navržen. Výstižně to charakterizoval Carl Lagoze: „*Although the internet provides access to an enormous amount of information, the current state-of-the-art falls far short of what is commonly viewed as a library service – that is, relatively easy navigation of and access to a set of documents that are part of a collection. The notion of a collection is important in that it implies that the set of documents was not selected haphazardly, but by some trusted intermediary. Current users of the internet confront an information space where the quality of documents is far from reliable, facilities for locating documents are primitive, and access to a specific document frequently means wading through a Tower of Babel of architecture dependencies and file formats.*”^[1] [47]

Přestože od zveřejnění tohoto názoru uplynulo již několik let a vývoj např. v oblasti tzv. sémantického webu dosáhl od té doby pozoruhodných výsledků, má výše uvedená charakteristika stále svou platnost.

Co tedy jsou ony digitální knihovny? Z mnoha desítek existujících „definic“ uveďme alespoň dvě. 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.* [4]

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 družicí na Zemi není knihovnou. Avšak tatáž data, jakmile jsou systematicky uspořádána, stávají se sbírkou v digitální knihovně. Podobně málokdo bude považovat za digitální knihovnu databázi obsahující finanční záznamy jedné společnosti; ale soubor takových záznamů z mnoha společností 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, jako je výběr, zpřístupnění a uchovávání materiálu, a zdůrazňuje, že digitální knihovny budou vždy budovány tak, aby sloužily konkrétní komunitě uživatelů (představa všeobjímající univerzální digitální knihovny 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 vše s ohledem na snadné a ekonomické využití určitou komunitou nebo množinou komunit uživatelů.* [80]

Z mnoha definic a projektů vyplývají určité společné základní 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 (pro zdůraznění tohoto aspektu mnozí autoři používají zásadně a výhradně 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);
- 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.

Na vývoji a nasazení digitálních knihoven v praxi se podílí zejména dvě skupiny odborníků. První z nich jsou informační profesionálové (včetně knihovníků, nakladatelů a široké skupiny poskytovatelů informací, jako jsou např. indexační a abstraktní služby). Druhou skupinu tvoří počítačoví specialisté a vývojáři internetu.

1.2 Krátce z historie

Vize digitálních knihoven provází v různých podobách větší část historie výpočetní techniky. Podstatný pokrok však v této oblasti nastal až počátkem 90. let minulého století, kdy prudký rozvoj informačních a komunikačních technologií umožnil začít v praxi realizovat představy teoretiků a efektivně zpřístupňovat první slibné výsledky širokému okruhu uživatelů.

V literatuře jsou nejčastěji uváděni dva průkopníci, kteří nejvíce inspirovali generace výzkumníků a propagátorů digitálních knihoven. Prvním z nich je Vannevar Bush, profesor MIT a ředitel amerického Národního úřadu pro vědecký výzkum a vývoj v období 2. světové války. Ve svém vizionářském článku *As We May Think* publikovaném v roce 1945 [13] se zabýval problémem efektivnějšího „automatizovaného“ zpracování odborných informací („*our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose*“²⁾). Analyzoval potenciální možnosti, které pro získávání, ukládání a vyhledávání informací nabízelo využití soudobých (analogových, ještě nikoliv digitálních) technologií a nastínil vizi systému využívajícího fotografické postupy a kompresi dat pomocí mikrofilmů. Bushem navržený systém *Memex* koncepčně odpovídá dnešnímu osobnímu počítači, v němž jsou informace provázány asociativními vazbami, a je tak předchůdcem hypertextu a koncepce dnešního webu.

Druhou často citovanou osobností je J. C. R. Licklider, který v 60. letech minulého století studoval na MIT možnosti transformace knihoven s využitím digitálních počítačů (na rozdíl od Bushe, který – ačkoliv již číslicové počítače znal – vycházel ještě z analogových technologií). V roce 1965 publikoval knihu *Libraries of the Future*, v níž identifikoval výzkum a vývoj potřebný k realizaci skutečně použitelné digitální knihovny a nastínil vizi digitální knihovny po 30 letech – tedy v roce 1994. V obecné rovině jsou jeho předpovědi pozoruhodně přesné a mnohé z nich se vyplnily, i když ne vždy v jím očekávané podobě; Licklider celkově výrazně podcenil výsledky, kterých se dá dosáhnout využitím hrubé výpočetní síly, a naopak přecenil pokroky založené na rozvoji umělé inteligence a počítačových metod zpracování přirozeného jazyka.

V 60. letech se také objevují první významné praktické výsledky v nasazení výpočetní techniky pro zpracování informací v knihovnách, mezi které bezesporu patřil jednak vývoj formátu MARC (Machine Readable Cataloguing) v americké Kongresové knihovně (Library of Congress), který standardizuje strukturu bibliografického záznamu v elektronické podobě, a využití tohoto formátu pro sdílenou katalogizaci knihoven v systému OCLC, jednak rozvoj online knihovních katalogů (knihovny označovaných termínem OPAC, Online Public Access Catalogue). Navzdory všem překážkám vyplývajícím z tehdejších technických omezení podnítily tyto první výsledky řadu optimistických předpovědí. Jeden příklad za všechny: A. L. Samuel předpovídal v roce 1964, že papírové knihovny do 20 let zaniknou [69]. Důvody, proč se většina předpovědí ze 60. let nenaplnila, byly samozřejmě různé; často však mezi ty hlavní patřily důvody finanční. Pro vyplnění Samuelovy vize by bylo třeba zdigitalizovat zhruba 100 milionů titulů knih, přičemž údaje z amerického prostředí [14] uvádějí cenu digitalizace v rozmezí 2–6 USD za stránku³⁾; ještě mnohem větší náklady by ovšem byly třeba na kompenzaci autorských práv.

Počátkem 90. let začíná v oblasti digitálních knihoven skutečný boom. Zásahu na tom měla skutečnost, že technologický pokrok ve všech třech oblastech, které jsou pro digitální knihovny kritické a které zahrnují

- *computing* (výpočetní a krátkodobá i dlouhodobá paměťová kapacita),
- *communications* (globální síť a přenosová kapacita) a
- *content* (množství informace v digitální podobě),

dosáhl dostatečně vysokého stupně při rozumně nízké jednotkové ceně a široké všeobecné dostupnosti, což umožnilo začít realizovat projekty reagující na skutečné potřeby uživatelů. To vše odstartovalo prudký rozvoj v oblasti digitalizace, elektronického publikování a šíření informací, což přineslo i nový impuls pro výzkum a vývoj v oblasti digitálních knihoven (dalšími výraznými podněty bylo celosvětové masové rozšíření webových technologií a všeobecná potřeba efektivnějšího sdílení vědeckých poznatků). Vyspělé země podpořily tento trend zřízením štědrě dotovaných programů na podporu výzkumu a vývoje (nejvýznamnějším z nich byl americký program DLI-1 (Digital Library Initiative Phase 1) a na něj v současnosti navazující DLI-2), ale i prakticky orientovaných projektů (např. britský program eLIB). Podrobněji se o nich zmíníme v závěru příspěvku.

1.3 Proč digitální knihovny

Počáteční představa digitální knihovny vycházela z koncepce klasické knihovny a byla orientována především na digitalizaci existujících sbírek jako nástroje pro zlepšení klasických knihovních služeb, a to zejména v následujících oblastech:

- vzdálený a nepřetržitý přístup k informacím;
- efektivnější metody vyhledávání (např. plnotextové);
- lepší využití fondu (souběžný přístup k jednomu a témuž dokumentu);
- sdílení informací mezi různými knihovnami;
- dokonalejší ochrana fondu (nahrazení zranitelných fyzických objektů digitálními).

Záhy se však ukázalo, že potenciální možnosti digitálních knihoven jdou nad rámec možností klasických knihoven s fyzickými dokumenty a projevují se např. možnostmi neomezené globální integrace digitálních repozitářů v celosvětovém měřítku, novými formami a formáty informací, možností permanentní aktualizace informace uložené v digitální knihovně nebo zcela novými typy služeb (přeformátovávání dokumentů on-fly do různých formátů či dokonce jazykových verzí, vytvářením složených děl, vyjednáváním autorských a přístupových práv aj.). 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 uchovávání materiálu.

Metody a postupy klasických knihoven jsou za mnoho staletí svého vývoje dobře propracovány a tvoří ucelený, efektivně fungující systém. Digitální knihovny však přinášejí nové výzvy a problémy, pro jejichž řešení nelze často klasické postupy použít vůbec nebo jen ve velmi omezené míře. Po počátečním optimismu z první poloviny 90. let se ukázalo, že problém budování funkčních digitálních knihoven je mnohem složitější, než se zdálo. Principiálním problémem a základem všech obtíží je nedostatečně propracovaná technologie na straně jedné a nepřipravené společenské prostředí zahrnující složité komplex navzájem provázaných problémů z oblasti ekonomické, právní, sociální a etické na straně druhé. To, na co měly klasické knihovny dlouhá staletí, musí digitální knihovny řešit za pochodu a během několika málo let.

1.4 Aktuální stav, hlavní současné aktivity a zdroje informací

V oblasti digitálních knihoven probíhá v současnosti velké množství aktivit jak v *základním a aplikovaném výzkumu* (vůdčí roli v tomto směru hrají zejména Spojené státy americké s množstvím nejrůznějších odborných aktivit především na univerzitách a ve velkých výzkumných knihovnách v čele s Kongresovou knihovnou), tak i v *praxi*, kde existují stovky velmi rozsáhlých a ambiciózních projektů zaměřených na digitalizaci či budování konkrétních digitálních knihoven poskytujících cenné informace a služby příslušným komunitám, nebo na implementaci nových prototypů ověřujících v praxi nové přístupy, potřeby a chování uživatelů. Oproti situaci z konce minulého století je u těchto projektů znát posun od „experimentování“ k budování obecné infrastruktury. Ačkoliv dosud neexistuje žádné univerzální a všeobecně přijaté řešení digitální knihovny a v mnoha směrech chybí potřebná globální infrastruktura, která by umožnila škálovat a propojovat znalostní sítě reprezentované jednotlivými digitálními knihovnami obdobně, jako je tomu dnes u komunikačních sítí reprezentovaných internetem a webem, je již k dispozici řada základních technologických kamenů v podobě standardů (Z39.50, OAI, Dublin Core, Handles, DOI) a volně dostupných nástrojů pro implementaci základních funkcí digitálních knihoven (za všechny uveďme systém Greenstone z University of Waikato na Novém Zélandu [33] a systém DSpace vyvinutý ve spolupráci MIT a HP Labs [83]). Velmi často se výzkum v oblasti digitálních knihoven překrývá s jinými oblastmi, kupříkladu elektronickým obchodováním (metadata, interoperabilita, bezpečnost, autorská a vlastnická práva).

Dění v oblasti digitálních knihoven mapuje řada časopisů, konferencí, specializovaných workshopů a také courseware – kurzů na vysokých školách [31].

Mezi nejvýznamnější časopisy patří:

- *D-Lib Magazine* [16] – elektronický časopis zaměřený hlavně na prakticky orientovaný výzkum v oblasti digitálních knihoven; vychází měsíčně od roku 1995 v CNRI s podporou DARPA. Je volně dostupný na webu a v současnosti je patrně nejprestižnějším zdrojem odborných informací o dění v oblasti digitálních knihoven.

- *International Journal on Digital Libraries* – klasický tištěný časopis z nakladatelství Springer-Verlag. Specializuje se spíše na teoretický výzkum, vychází od roku 1997, bohužel však s nepravidelnou periodicitou.
- *Ariadne* [2] – elektronický časopis pro informační specialisty zejména z Velké Británie. Informuje o aktivitách z oblasti digitálních knihoven, vydává ho čtvrtletně UKOLN (UK Office for Library and Information Networking); volně dostupný na webu.
- *RLG DigiNews* [68] – elektronický časopis zaměřený na oblast digitalizace a uchovávání digitální informace. Je vydáván Cornellovou univerzitou ve spolupráci s organizací Research Libraries Group (RLG) sdružující kolem 160 výzkumně zaměřených knihoven, archivů a dalších paměťových institucí převážně z USA. Vychází dvakrát měsíčně, je volně dostupný na webu.

Problematicke digitálních knihoven byla věnována také některá speciální čísla časopisů přehledově zaměřených na informační technologie, jako *Communications of the ACM* (vrací se k digitálním knihovnám vždy pravidelně po 3 letech, viz čísla z dubna 1995 a 1998 a května 2001) nebo *IEEE Computer*.

Z nejvýznamnějších odborných konferencí je třeba uvést od roku 1996 pravidelně každoročně pořádané konference ADL – Advances in Digital Libraries (IEEE) a ACM Conference on Digital Libraries (od roku 2001 pořádané společně pod názvem JCDL – Joint Conference on Digital Libraries) a evropskou konferenci ECDL – European Conference on Research and Advanced Technology for Digital Libraries.

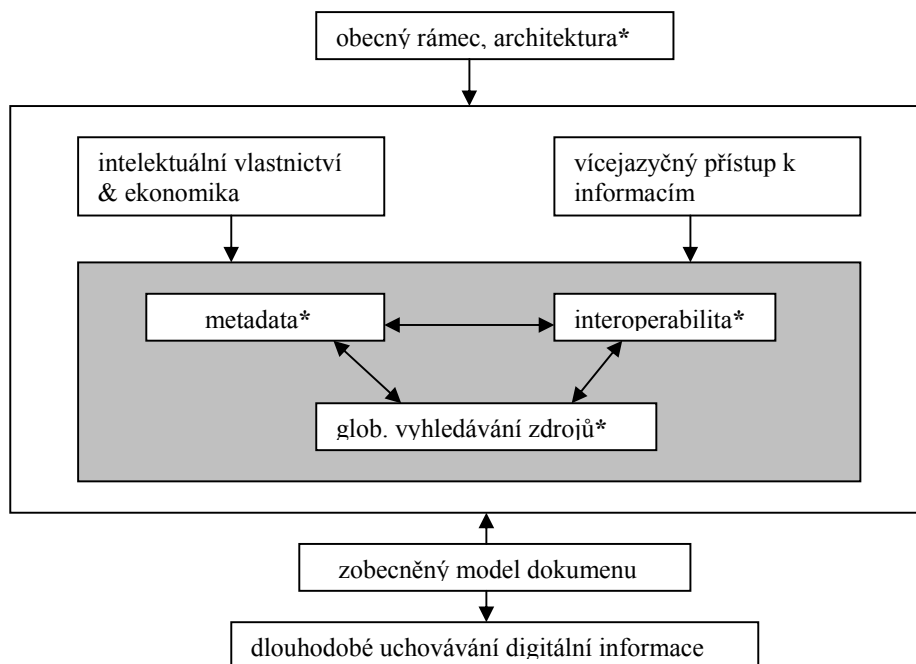
Z metazdrojů zahrnujících různé projekty digitálních knihoven vyberme např.:

- IFLA Electronic Collections [39]
- Berkeley Digital Library Sunsite [11]
- ARL Digital Initiatives Database [3]

Oblasti digitálních knihoven se věnuje i řada monografií. Práce [48] je prakticky zaměřená přehledová publikace a popisuje spíše technické aspekty digitálních knihoven, kniha [4] má povahu obecné přehledové encyklopedie celé oblasti, [84] přináší rozbor jednotlivých oblastí při návrhu a realizaci digitální knihovny a způsob jejich řešení v systému Greenstone. Řada dalších monografií se věnuje již konkrétním dílčím aspektům digitálních knihoven, jako jeden příklad za mnohé uvedme [45] se zaměřením na problematiku digitalizace obrazové informace.

2 Klíčové oblasti výzkumu a praxe digitálních knihoven

Termín „digitální knihovny“ je typicky zastřešující pojem. Problematika digitálních knihoven a aspekty jejich realizace jsou totiž natolik široké, že se s trochou nadsázky dá říci, že pod tento pojem lze zahrnout „téměř cokoliv“ z mnoha oblastí počítačové vědy (databáze, informační systémy, umělá inteligence, počítačové sítě, bezpečnost), ale



navíc i mnoho aspektů z řady společenských věd (z knihovní a informační vědy, práva, ekonomie, sociologie, psychologie, lingvistiky).

Takové bezbřehé pojetí nám však příliš nepomůže. Zaměříme-li se na oblasti, které jsou pro digitální knihovny skutečně klíčové, dostaneme výše uvedený obrázek (adaptováno dle [71]); oblasti popisované v další části našeho příspěvku jsou označeny hvězdičkou.

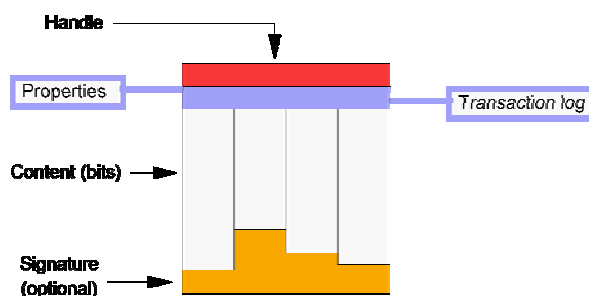
3 Obecný rámec a architektura digitální knihovny

K tomu, aby mohl vzniknout globální systém kooperujících digitálních knihoven, je třeba vytvořit a uvést do života potřebnou globální informační infrastrukturu. Obecná architektura digitální knihovny navržená na dostatečně vysoké úrovni abstrakce umožňuje formalizovat představy o funkcích a fungování digitálních knihoven a současně identifikovat „middleware“ internetu potřebný pro realizaci distribuovaných digitálních informačních služeb (všechny dnešní systémy digitálních knihoven jsou realizovány ve webovém prostředí internetu).

3.1 Kahn-Wilenského architektura

Nejpropracovanější obecnou architekturu digitálních knihoven podali Kahn a Wilensky [44]; experimentální systém vycházející z této architektury byl pak realizován např. v rámci projektu National Digital Library Project v Kongresové knihovně [6].

Základním prvkem architektury je *digitální objekt*, datová struktura pro základní samostatně použitelnou informační jednotku tvořená dvěma základními částmi: obsahem (content) a klíčovými metadaty (tvořenými globálním jednoznačným identifikátorem digitálního objektu, označovaným jako *handle*, a dalšími blíže nespecifikovanými neměnnými metaúdaji, např. „autor“). Obsahem digitálního objektu může být sekvence bitů reprezentující konkrétní digitální materiál (může být zahrnut i ve vícero formách), množina jiných datových objektů (složený objekt), množina identifikátorů objektů (metaobjekt), případně jiné datové typy – model tak poskytuje dostatečnou flexibilitu pro reprezentaci libovolně složitých informačních objektů a vztahů mezi nimi. Digitální objekty mohou být buď proměnlivé (obsah objektu lze měnit i po jeho uložení do repozitáře – ať již jde o jednorázové změny nebo přímo dynamické informační objekty), nebo fixní. Schéma jednoduchého digitálního objektu ukazuje následující obrázek:



Podle typu materiálu mohou být digitální objekty rozděleny do *kategorí* (např. text formátovaný pomocí SGML, počítačový program, digitalizovaný zvuk) a pro každou kategorii mohou být stanovena pravidla pro převod materiálu do jednotlivých typů digitálních objektů, struktura metadat apod. Tak je tomu např. v realizovaném systému [6]; obecná architektura ovšem úmyslně s žádnými specifickými typy materiálů nepracuje, aby udržela co nejvyšší míru obecnosti, neomezovala či nepředjímala budoucí technologický vývoj a ponechávala dostatečnou míru flexibility pro konkrétní implementace.

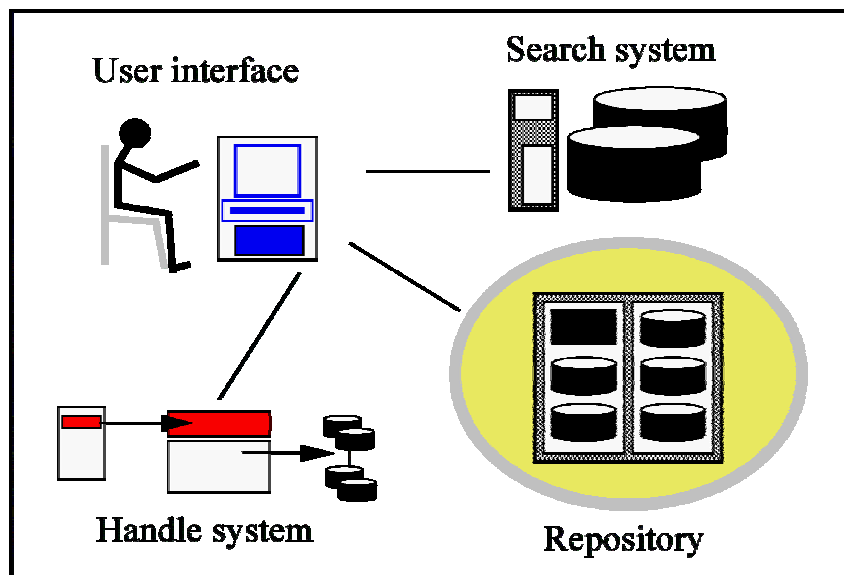
Digitální objekty jsou uloženy v *repozitářích*, které mají přiřazeno jednoznačné globální jméno. Repozitář umožňuje řízený přístup k digitálním objektům v něm uloženým. Pro každý digitální objekt udržuje dva typy metadatových záznamů. Prvním z nich je *záznam vlastností objektu* (properties record) obsahující údaje např. o autorských právech a podmínkách pro zpřístupnění digitálního objektu, technické vlastnosti jako formáty a přístupové protokoly, bibliografické údaje nebo administrativní data (např. datum/čas uložení objektu do repozitáře). Druhým je *transakční záznam* (transaction log) zaznamenávající veškeré transakce repozitáře týkající se daného digitálního objektu. Spolu s neměnnými a na repozitáři nezávislými klíčovými metadaty tvoří tyto dva záznamy základní metadatový popis digitálního objektu.

Každý repozitář komunikuje s okolím prostřednictvím jednoduchého *repozitářového přístupového protokolu* RAP (Repository Access Protocol) umožňujícího ukládání a zpřístupnění digitálních objektů, případně i další operace – to vše při zajištění odpovídajícího zabezpečení. Digitální knihovna může sestávat z mnoha repozitářů různých typů.

Další komponentou architektury je tzv. *handle-system* sloužící jako globální distribuovaný směrovací

mechanismus, který pro digitální objekt identifikovaný svým identifikátorem vrátí seznam repozitářů, které tento objekt udržují. Handle-system byl v praxi úspěšně realizován v Corporation for National Research Initiatives (CNRI) a patří dnes mezi jedny z nejpropracovanějších a v projektech digitálních knihoven nejužívanějších systémů pro globální identifikaci a směrování informačních objektů na internetu [36]. Podrobnější popis je uveden níže v kapitole o identifikátorech.

Schéma kooperace jednotlivých komponent digitální knihovny je naznačeno na obrázku:



1. **search:** uživatel specifikuje svůj požadavek přes inteligentní koncové uživatelské rozhraní, které jej přeformuluje na dotaz pro globální vyhledávací systém, jenž vrátí seznam informačních zdrojů splňujících požadavek uživatele;
2. **select:** uživatel vybere ze seznamu informační zdroj, který chce získat;
3. **retrieve1:** uživatelské rozhraní předá identifikaci digitálního objektu odpovídajícího zvolenému informačnímu zdroji globálnímu směrovacímu systému a získá identifikaci příslušného repozitáře;
4. **retrieve2:** přes příslušný protokol RAP si vyžádá potřebný digitální objekt z repozitáře;
5. **display:** získaný objekt zobrazí uživateli.

Z diskusí o výše naznačené obecné architektuře vyplynula řada podmínek klíčových pro její realizaci v praxi [7]. V dalším textu rozebereme alespoň tři z nich:

- (a) technický návrh musí být začleněn do konkrétního právního a sociálního prostředí;
- (b) uživatelé požadují intelektuální díla a nikoliv digitální objekty;
- (c) klíčovými stavebními bloky informační infrastruktury jsou jména a identifikátory.

3.2 Začlenění do právního a sociálního prostředí

První informační systémy na internetu byly vytvořeny konkrétními odbornými a vědeckými komunitami s ohledem na jejich vlastní potřeby; primárním cílem byla rychlá, efektivní a zejména *bezplatná* výměna informací. Mnohé z těchto systémů jsou velmi úspěšné dodnes a dále se rozvíjejí (jako dva příklady z oblasti elektronického publikování připomeňme historický systém RFC používaný IETF a novější preprintový systém ArXiv.org [8] sloužící celosvětové komunitě fyziků). Situace u obecných digitálních knihoven je ale mnohem složitější, neboť aby byly v praxi použitelné, musí respektovat mnohem širší ekonomický, sociální a právní kontext. Typickým příkladem jsou hudební díla, jež představují živobytí skladatelů i hudebníků, kteří vyžadují poplatky za jejich používání (stejně tak i nahrávací studia). Taková díla se mohou stát součástí digitální knihovny pouze a jen tehdy, pokud digitální knihovna nebude ohrožovat ekonomické zájmy jejich původců a zprostředkovatelů.

Legislativní rámec upravující způsob klasického vytváření, publikování a využívání intelektuálních výtvořů, který zahrnuje množství složitě provázaných konceptů od autorského práva, provedení díla, volného užití díla, soukromí, ochrany osobnosti až po komunikační zákony, daně, národní bezpečnost aj., se utvářel velmi dlouhou dobu tak, aby vyvážil zájmy všech subjektů (zájem autorů tvořit, vydavatelů vydávat, uživatelů díla využívat, společnosti ochraňovat produktivní prostředí a zajišťovat bezpečnost). Digitální publikování a šíření informací, nebude-li právně vhodně

ošetřeno, představuje pro tento složitě vyvážený systém hrozbu s mnoha potenciálně obrovskými dopady (dle [4] představuje informační a zábavní průmysl asi 5 % ekonomiky USA; radikální změna technologie může znamenat obrovské ekonomické změny vedoucí až k zániku velkých firem a celých odvětví, a to i s příslušnými sociálními dopady v komerční i nekomerční sféře). Právní situace týkající se digitálních knihoven je o to složitější, že vzhledem k jejich globálnímu charakteru nestačí upravit a aplikovat národní legislativu, ale je třeba dojednat, vytvořit a uvést do života odpovídající legislativu na mezinárodní úrovni.

Hlavní problém spočívá v tom, že vytvoření potřebného legislativního, společenského a ekonomického prostředí není již problém technologický, ale společenský, a jako takový je mnohem složitější a časově mnohem náročnější. Zkušenost také ukazuje, že nelze předvídat všechny aspekty vývoje, že nejprve musí ve společnosti vzniknout určité vzorce chování, které lze později právně kodifikovat.

3.3 Hierarchická abstrakce intelektuálního díla (model IFLA)

Digitální objekty jsou sice základními stavebními kameny obecné architektury digitální knihovny, uživatelé ale obvykle potřebují odkazovat na informační zdroje na vyšší úrovni abstrakce. Navržená architektura umožňuje reprezentovat libovolně složité objekty a vztahy mezi nimi, pro její naplnění je však třeba aplikovat nějakou všeobecně přijatelnou a dostatečně obecnou abstrakci informačních objektů. Nejrozšířenější a nejvyužívanější kategorizace informačních objektů pochází ze studie *Functional Requirements for Bibliographic Records* (FRBR) [41] organizace IFLA (International Federation of Library Associations and Institutions) z roku 1998 a rozlišuje čtyři úrovně:

- *dílo (work)* – intelektuální či umělecký výtvar jako abstraktní koncept (Homérova Iliada, Beethovena 5. symfonie, operační systém Unix apod.);
- *vyjádření (expression)* – konkrétní realizace, časoprostorová fixace daného díla (např. Iliada byla nejprve realizována ústním podáním, poté formou písemného zápisu; symfonie může být realizována jako zápis partitury nebo jako některé z mnoha jejích hudebních provedení);
- *projev (manifestation)* – fyzické „ztělesnění“ nějakého vyjádření daného díla (např. text Iliady může být „projeven“ v několika různých rukopisech nebo v různých knižních vydáních; určité provedení symfonie může být zaznamenáno na hudební CD nebo na videokazetě jako záznam televizního přenosu);
- *jednotka (item)* – jeden (z potenciálně mnoha) exemplářů daného projevu, kopie (např. jeden výtisk daného vydání knihy, konkrétní výtisk podepsaný autorem, kopie souboru apod.).

Stručně vyjádřeno: *dílo* je realizováno prostřednictvím jednoho či více *vyjádření*, to je převedeno do fyzické podoby v jednom či několika různých *projevech*, ty jsou pak rozmnoženy v jedné či mnoha *jednotkách*. Praktické zkušenosti ukazují, že tento čtyřúrovňový model je schopen plně postihnout všechny aspekty uživatelských zájmů o informační objekty nejen v oblasti digitálních knihoven, ale i v ostatních oblastech (např. elektronickém obchodu). Navíc, na rozdíl od jiných modelů v jiných komunitách, nechybí knihovnám jasně propracovaný systém jeho uplatňování (např. přesný překlad díla je považován za jeho nové vyjádření, zatímco volný překlad může být novým dílem; podobně je za nové dílo považována změna žánru, např. dramatizace určité novely).

Přesné rozlišování různých abstrakcí informačních objektů je užitečné a v některých případech dokonce zcela zásadní nejen v oblasti digitálních knihoven (pro vyhledávání, odkazování, správu vlastnických a autorských práv); oblast elektronického obchodování je toho typickým příkladem (viz např. [38]).

4 Jména a identifikátory

Čím více se při výměně informací (resp. obchodu) snižuje potřeba fyzického kontaktu mezi uživatelem informace a jejím poskytovatelem (resp. mezi kupujícím a prodávajícím), tím víc roste potřeba schopnosti věci jednoznačně pojmenovávat a identifikovat. Schopnost jednoznačně globálně identifikovat informační objekty (a tuto identifikaci dynamicky jednoznačně propojit s informačním objektem nacházejícím se kdekoli v globální síti) je *zcela zásadní* pro nasazení jakéhokoliv distribuovaného globálního informačního systému.

4.1 Koncept URN

Stávající internet zatím nenabízí dostatečně univerzální, všeobecně podporovaný a rozšířený identifikační systém pro informační objekty, který by splňoval základní požadavky zformulované již počátkem 90. let pro koncepci URN (Uniform Resource Name) [76], a to:

- globální rozsah:* dané jméno je celosvětově jednoznačné a nezávislé na lokaci;
perzistence: přidělené jméno trvá „na věky“, i po zániknutí popisované entity;

škálovatelnost: jméno musí být použitelné pro jakýkoliv představitelný typ entity;
legacy support: systém musí podporovat existující identifikační systémy;
rozšiřitelnost: musí být umožněno budoucí rozšíření identifikačního schématu

Stávající URL tyto požadavky nesplňuje, neboť identifikuje lokaci, nikoliv entitu (intelektuální obsah). Smyslem návrhu URN je naopak identifikovat entitu bez ohledu na její momentální umístění; musí ovšem existovat tzv. *směrovací mechanismus*, který pro zadané URN zjistí aktuální umístění entity tímto URN identifikované. Syntaxe URN je následující:

URN:nid:nss

kde *nid* (Namespace Identifier) je identifikátor určitého identifikačního systému (např. DOI, viz níže) a *nss* (Namespace-specific String) je konkrétní identifikátor v daném systému. Jak je vidět, URN není založeno na jediném identifikačním systému, ale naopak poskytuje zastřešení pro neomezený počet schémat splňujících stanovené podmínky (zahrnujících i popis technik pro realizaci směrovacího mechanismu – viz RFC-2611).

Přestože obecná idea URN je jasná a také návrh jeho infrastruktury byl nedávno již dokončen, implementace globálních řešení na internetu jsou zatím omezené. Příčinou jsou tyto skutečnosti:

- globální směrovací systém pro URN (na bázi DNS) ještě není rozšířen v celém internetu;
- stávající webové prohlížeče zatím nepodporují URN tak, jak podporují URL;
- panují nejasnosti, kdy a kterým entitám URN přidělovat (problém verzí, různých formátů, úrovní);
- problém jednoduchého (single-point) nebo násobného (multiple) směrování: směrovací mechanismy vycházející ze současné webové technologie vracejí jediné URL, zatímco služby založené na URN mohou obecně vyžadovat identifikaci více instancí entity či více jí odpovídajících služeb;
- v neposlední řadě jde o otázku finanční: identifikátory URN jsou sice zdarma, ale budování a udržování směrovacích služeb nikoliv – náklady jsou obrovské a někdo je musí uhradit.

Knihovny se problémem identifikace fyzických informačních zdrojů zabývají již dlouho; popíšme současný stav a nástroje, které jsou k dispozici i digitálním knihovnám [34].

4.2 Klasické identifikátory: ISBN, ISSN, SICI/BICI, ISTC

Již zhruba 30 let používají knihovny a nakladatelé identifikaci ISBN (International Standard Book Number), ISSN (International Standard Serial Number) a další identifikátory k identifikaci tištěných publikací neboli projevu díla (v terminologii modelu FRBR). Digitální knihovny a elektronické publikování však vyžadují komplexnější vícevrstvou identifikaci, počínaje samotnými autory a konče těmi nejmenšími jednotkami informací, s nimiž lze v internetu samostatně manipulovat a prodávat je, jako jsou např. články v časopisech. Současný stav v oblasti standardů je následující:

a) autoři:

ISADN (International Standard Authority Data Number) – umožňuje jednoznačně identifikovat každého autora. Zatím v praxi nerealizováno (již několik let se diskutuje o otázce skutečné potřeby a technické realizovatelnosti tohoto identifikátoru).

b) dílo:

ISTC (International Standard Textual Work Code) – pro textová díla.

ISAN (International Standard Audiovisual Number) – pro audiovizuální díla.

ISWC (International Standard Musical Work Code) – pro hudební díla.

Standardy ISAN a ISWC jsou již několik let platnými standardy ISO; práce na ISTC postoupily v září 2004 do závěrečné schvalovací fáze.

c) projev:

ISBN, ISSN, ISMN a další identifikátory, které se využívají již dlouhou dobu.

d) komponenta:

SICI (Serial Item and Contribution Identifier) – pro články.

BICI (Book Item and Component Identifier) – pro kapitoly v knize, obrázky, poznámky apod.

Norma SICI existuje jako standard ANSI/NISO již od roku 1996, ale zatím se v praxi moc nevyužívá; má však budoucnost. BICI byl v pozici pracovního návrhu a jeho další osud je nejasný.

Popis správy vybraných standardů dobře charakterizuje rozdílné přístupy ke koncepci identifikátorů:

ISBN – International Standard Book Number

Číslo ISBN přiděluje nakladatel. Podle normy ISO 2108-1978 začíná číslo ISBN vždy zkratkou ISBN následovanou 10místným číslem rozděleným do čtyř bloků proměnlivé délky oddělených spojovníkem, např. ISBN 80-00-01978-9. První blok přiděluje mezinárodní agentura ISBN a identifikuje zemi, v níž nakladatel působí (0 a 1 anglická oblast, 80 ČR a SR). Druhý blok přiděluje národní agentura ISBN a identifikuje nakladatele. Může mít délku 2–6 číslic; čím větší nakladatel, tím kratší je jeho číslo. Třetí blok přiděluje nakladatel a určuje konkrétní vydání knihy či její formy; délka je volena tak, aby celková délka čísla ISBN byla deset znaků. Poslední blok tvoří kontrolní číslice, která se vypočítává z předchozích devíti cifer podle modulu 11 s využitím váhových koeficientů. Národní agentura shromažďuje informace o všech přidělených ISBN v dané zemi.

ISBN je příkladem tzv. *inteligentního* či složeného identifikátoru, který kromě vlastní identifikace nese ještě další explicitní informaci (země, nakladatel). Systém je u klasických fyzických informačních zdrojů velmi úspěšný, ale v digitálním světě má problémy:

1. na internetu může být nakladatelem kdokoliv, což vede k exponenciálnímu nárůstu požadavků na nakladatelská čísla; částečně se to dá řešit vyčleněním vyhrazených identifikátorů pro knihy publikované jednotlivci, ale obecně není tento systém pro potřeby webu dostatečně flexibilní;
2. dramatický nárůst publikací po vzniku elektronického publikování a také to, že ISBN je často přidělováno i menším informačním jednotkám než kniha, vede k tomu, že ve velmi krátké době (odhaduje se do roku 2010) se prostor ISBN čísel vyčerpá.

Z výše uvedených důvodů vyžaduje systém ISBN důkladnou revizi, a to nejpozději do roku 2006. Přitom jakákoliv změna dosavadního systému bude mít obrovské dopady s velkými náklady na knihovní a informační sektor, srovnatelnými s náklady na řešení problému Y2K. Přijaté řešení (které není sice koncepční, ale umožní získat čas) je rozšíření ISBN na 13 číslic tím, že na jeho začátek bude přidán kód EAN „978“ (knihy), používaný celosvětově v obchodu pro označování zboží čárovými kódy. Po dohodě s EAN bude využit pro ISBN i druhý prefix „979“ (hudebniny) a číselný prostor ISBN se tak téměř zdvojnásobí.

ISSN – International Standard Serial Number

Na rozdíl od ISBN je číslo ISSN (ISO norma 3297-1998) tzv. *hloupý* či jednoduchý identifikátor, který v sobě nenese žádnou sémantiku. Má tvar 8 cifer rozdělených do dvou bloků po čtyřech cifrách oddělených spojovníkem, např. ISSN 0885-2308. Posledním znakem je kontrolní znak (obdobně jako u ISBN). Všechna čísla ISSN jsou přidělována a centrálně spravována Mezinárodním centrem pro ISSN; v současnosti obsahuje jeho databáze 1,1 milionu záznamů (celková kapacita je 10 milionů). Spolu s každým přiděleným ISSN je v databázi uložen metadatový záznam o příslušném periodiku či seriálové publikaci.

Elektronické časopisy zatím kapacitu ISSN vážněji neohrožují (ročně se zatím přiděluje kolem 50 000 čísel), problém je však s velmi krátkým poločasem jejich rozpadu. Navíc elektronické časopisy nemusí být vydávány v ročnících, svazcích a jednotlivých číslech, takže podle posledních aktualizací katalogizačních pravidel může být za kandidáta na přidělení čísla ISSN považována každá webová stránka, v jejímž rámci jsou shromažďovány nové dokumenty.

SICI – Serial Item and Contribution Identifier

I když je SICI americkou normou již od poloviny 90. let (ANSI/NISO Z39.56, viz [59]), zatím se příliš nevyužívá. Důvodem může být jak neexistence jeho podoby v mezinárodní ISO normě, tak skutečnost, že mezi nakladateli je zatím o něm poměrně malé povědomí a může jim připadat poměrně složitý. Chybí také mezinárodní centrum, které by využívání tohoto standardu dostatečně propagovalo.

Příklad: článek Marka Needlemana „Computing resources for an online catalog – 10 years later“ publikovaný v časopise *Information technology and libraries*, svazek 11, číslo 2 (červen 1992), str. 168–, bude mít SICI:

0730-9295(199206)11:2<168:CRFAOC>2.0.TX;2-#

Identifikátor je tvořen ISSN číslem časopisu následovaným údaji o čísle, údaji o článku (první písmena slov z názvu) a kontrolní částí (verze standardu 2.0, typ zdroje je tištěný text TX). SICI je příkladem identifikátoru, který může být plně „vypočítán“, tj. automaticky vygenerován přímo z článku nebo jeho metadat. Identifikátor BICI vypadá podobně, jeho standardizace ale ještě není dokončena.

ISTC – International Standard Textual Work Code

Podle pracovní verze návrhu standardu tvoří číslo ISTC šestnáct hexadecimálních cifer rozdělených do čtyř bloků, např.:

První blok představuje identifikátor některé registrační agentury (kterých může být až 4096). Každá z nich může přidělit až miliardu čísel každým rokem až do roku 9999. Jedním z požadavků na agenturu je schopnost vytvářet a udržovat metadata pro díla; přirozenými kandidáty se tak stávají např. národní knihovny (množství všech existujících textových intelektuálních výtvorů zahrnujících mj. i články, básně, eseje – či jejich komponenty – je obrovské a jejich kompletní „katalogizace“ je nepředstavitelně náročný úkol). Druhý blok představuje rok, třetím je identifikátor díla a posledním kontrolní číslice. Fungování a rozvoj celého systému bude koordinovat mezinárodní registrační autorita; výběrové řízení na ni vyhrálo v létě 2004 konsorcium CISAC / Nielsen BookData / R.R. Bowker.

Kromě dosud uvedených identifikačních schémat existují mnohá další, ať již ve formě oficiálních, nebo faktických standardů (jakým je také PII – Publisher Item Identifier), viz např. [64]. Pro všechny výše uvedené příklady identifikátorů lze na internetu implementovat příslušný globální směrovací systém s využitím koncepce URN. U „hloupých“ identifikátorů typu ISSN (identifikátor nenese sám o sobě žádnou informaci o tom, kde hledat informaci o informačním objektu) je k tomu zapotřebí globální centrální databáze; naproti tomu směrování ISBN lze postavit na distribuovaném systému např. národních bibliografií (databází bibliografických záznamů při národních knihovnách mapujících knižní produkci daného národa či země).

Vedle výše uvedených identifikátorů vycházejících primárně ze světa klasických dokumentů existuje několik systémů vytvořených již přímo pro internetové zdroje. Zmíníme stručně tři z nich: PURL, Handle a DOI.

4.3 PURL – perzistentní URL

Tento systém [66] realizovaný organizací OCLC byl jedním z prvních pragmatických řešení vyvinutých pro knihovnicko-informační komunitu s cílem využít to, co již současný internet nabízí (http a URL), a přitom co nejjednodušším způsobem odstranit základní problém s identifikací pomocí URL – závislost identifikace zdroje na jeho umístění. PURL je URL poskytující *nepřímou adresaci* zdroje. Princip je velmi jednoduchý: informační zdroj na internetu dostane přidělený identifikátor PURL např. ve tvaru <http://purl.oclc.org/catalog/item1> a teprve na této „adrese“ je uloženo skutečné URL zdroje. Funkčně je tedy PURL normálním URL, které však neodkazuje přímo na umístění zdroje, nýbrž na zprostředkující směrovací službu. Ta propojí identifikátor PURL se skutečným URL a vrátí ho klientovi. Klient pak dokončí URL transakci standardním způsobem (přes http příkaz redirect). Pokud se změní umístění zdroje, změní jeho správce hodnotu uloženou na adrese <http://purl.oclc.org/catalog/item1>, ale samotné PURL (externí jméno) se nikdy nemění. Uživatelé se mohou volně zaregistrovat na PURL serveru a poté si vytvářet vlastní identifikátory PURL, mohou dokonce volně stahovat příslušný software a instalovat vlastní směrovací PURL server.

4.4 Systém „handles“

Technologii známou pod názvem „handles“ [36] vyvinula CNRI jako součást obecné architektury digitální knihovny pro jednoznačnou identifikaci digitálních objektů [44]. Ačkoliv byl tento systém vyvinut nezávisle na konceptu URN, je s ním kompatibilní a lze ho považovat za vůbec první systém URN použitý v oblasti digitálních knihoven. Současná verze systému je založena na protokolu HTTP s identifikátorem vloženým do dokumentu ve formě hypertextové vazby odkazující na směrovací server systému handle. Identifikátor handle má následující tvar:

`hdl:cnri.dlib/magazine`

kde první část (prefix `cnri.dlib`) je tzv. pojmenovávací autorita, která je přidělována hierarchicky (nejvyšší úroveň *cnri* je přidělována centrální autoritou, zbylá část již lokálně). Část za lomítkem je libovolný řetěz znaků jedinečný v rámci dané pojmenovávací autority. Architektura systému handle je dvojúrovňová – jeden globální registr a neomezený počet lokálních serverů; z důvodů lepší výkonnosti a lepší dostupnosti služeb je implementována jako distribuovaný systém s decentralizovanou administrací (globální registr identifikátorů tak není centralizován fyzicky, nýbrž virtuálně). Každá z jeho komponent může být rozprostřena mezi více počítačů a data mohou být automaticky replikovány, k dispozici je řada služeb *cache*. Pro maximální využití a přímé směrování identifikátorů (včetně násobného směrování) je třeba instalovat do uživatelského webového prohlížeče příslušný software ve formě plug-in (jsou volně k dispozici); prostřednictvím proxy serverů lze systém používat i s neadaptovanými prohlížeči, avšak s neúplnou funkcí.

Systém je velmi propracovaný; jeho slabá stránka však spočívá v tom, že se jej nepodařilo prosadit jako internetový standard (patrně hlavně proto, že IETF nechtělo připustit rozmnožování různých koncepcí směrovacích služeb a podporuje pouze vlastní koncept v podobě URN). Nicméně řada současných, velmi úspěšných systémů je na handles založena; jmenujme alespoň program Kongresové knihovny NDLP [51], NCSTRL – distribuovanou digitální knihovnu technických zpráv z oblasti počítačové vědy [57] a iniciativu amerických nakladatelů DOI.

4.5 DOI – Digital Object Identifier

V roce 1996 vznikla z popudu Asociace amerických nakladatelů iniciativa DOI [23], jejímž cílem bylo vytvořit systém pro identifikaci digitálních objektů (prací chráněných autorským zákonem) pro potřeby komerčních vydavatelů. Vznikl systém, který je od roku 1998 dále rozvíjen mezinárodní nadací International DOI Foundation (IDF). Jako směrovací mechanismus identifikátorů DOI je využíván systém handle popsany výše. Syntaxe DOI byla specifikována normou ANSI/NISO Z39.84-2000. Příklad:

doi:10.1006/123456

Prefix 10.1006 sestává z konstanty 10 (slouží k odlišení DOI od ostatních implementací systému handle), za níž po tečce následuje numerický identifikační kód registrující organizace (1006 je např. kód Academic Press). Sufix za lomítkem obsahuje identifikátor digitálního objektu a může jím být cokoliv za předpokladu, že v rámci dané registrující organizace bude jednoznačný. To dává registrující organizaci možnost použít volně libovolné identifikační systémy – jak globální, např. doi:10.1000/ISBN1-900512-44-0, tak i lokální (to je zásadní rozdíl oproti koncepci URN, která použití každého identifikačního systému umožňuje pouze tehdy, pokud bylo stanoveným postupem standardizováno v internetové komunitě). Číslo DOI identifikuje dílo, nikoliv projev díla (viz model FRBR výše), takže tištěná verze článku a jeho digitální kopie mají totéž číslo. Systém DOI je silně centralizovaný; každá registrující organizace musí všechna jí vydaná DOI čísla registrovat u (zatím jediné) registrační agentury, směrování probíhá přes tuto centrální databázi (<http://dx.doi.org/10.1000/ISBN-1-900512-44-0>). Zajímavým rysem je, že povinnou součástí registrace čísla DOI (kromě stavových dat specifikujících umístění objektu) je také předání metadat DOI popisujících daný objekt; ta pak mohou být vrácena jako výsledek procesu směrování, když není možno zpřístupnit objekt samotný, např. z licenčních důvodů.

V současné verzi poskytuje DOI pouze perzistentní identifikátory (čili v zásadě totéž, co mnohem jednodušší PURL), ale IDF má ambice rozvinout ho do komplexního systému na podporu řízení správy vlastnických a autorských práv. Všeobecně se zatím nepředpokládá, že by se DOI měl rozvinout ve všeobecně použitelné řešení pro identifikaci jakýchkoliv typů dokumentů na internetu, a to zejména z důvodů ekonomických (systém není dostupný bezplatně, platí se jak za registraci registrující organizace, tak i za každé zaregistrované číslo DOI). Nicméně k dnešnímu datu se jedná o systém velmi úspěšný a v praxi opravdu fungující. V polovině roku 2004 bylo v systému zaregistrováno přes 250 registrujících organizací (nakladatelů). V systému CrossRef [15], který využívá DOI pro vytváření citačních vazeb v oblasti vědeckých publikací (citation-linking) a zpřístupňování plných textů, bylo v té době aktivních 700 subjektů (přes 450 knihoven) a přiděleno více než 12 milionů čísel DOI článkům z 10 000 titulů periodik. Od června 2001 jsou k dispozici nástroje umožňující realizovat vícenásobné směrování čísel DOI.

5 Metadata

Obecně jsou metadata *informace o informacích*; v kontextu digitálních knihoven je lze charakterizovat jako počítačově zpracovatelné strukturované informační objekty popisující vlastnosti jiných informačních objektů. Protože metadata jsou klíčovou komponentou pro obrovskou škálu velice různorodých služeb (vyhledávání informačních zdrojů a jejich výběr, autentizaci, interoperabilitu, správu vlastnických práv, dlouhodobou archivaci a řadu dalších), existuje velmi mnoho různých metadatových schémat.

5.1 Úvod a stručný přehled

Klasické knihovny jsou od počátku své existence postaveny na vytváření a využívání metadat (bibliografických záznamů) a totéž platí i pro knihovny digitální. Avšak mezi bibliografickými metadaty v klasických knihovnách a síťovými metadaty pro digitální síťové prostředí je jeden základní koncepční rozdíl: zatímco bibliografický záznam usiluje o kompletní popis zdroje, síťová metadata jsou *specializovaná* – pokrývají vždy jen určitou část, jen některé aspekty zdroje. Tento rozdíl je dán dvěma faktory: za prvé organizačním modelem používaným při tvorbě metadat (u klasických knihoven je to jedna centrální autorita, např. národní knihovna, zatímco u digitálních knihoven jde o řadu různých komunit pracujících nezávisle na sobě a podle svých specifických potřeb). Za druhé je zde jiný model přístupu k samotnému zdroji: protože v klasické knihovně neměli uživatelé metadat obvykle přímý přístup k informačnímu zdroji, museli být schopni učinit své rozhodnutí o užitečnosti zdroje výhradně na základě znalosti jeho metadat. V digitálním síťovém prostředí jsou naproti tomu často zdroje dostupné přímo, uživatel je může bezprostředně prohlížet, což eliminuje potřebu komplexního popisu. Stejně tak je možné snadno zpřístupňovat různá metadata daného zdroje a nové technologie na bázi RDF nabízí možnost je vzájemně propojovat, kombinovat a dynamicky tak vytvářet složitější popisy podle aktuální potřeby.

Metadata lze klasifikovat podle různých hledisek. Z hlediska jejich obecného použití se obvykle dělí na metadata *popisná* (slouží k obecnému popisu zdroje za účelem jeho vyhledání, identifikace a selekce), *strukturální* (zachycují formát a strukturu zdroje za účelem jeho správného uložení a zobrazování) a *administrativní* (slouží ke správě zdroje,

včetně řízeného přístupu a archivace). Jiná typologie může rozčleňovat metadatová schémata podle bohatosti jejich struktury a míry detailnosti popisu: od jednoduchých často proprietárních schémat s jednoduchými nepřiliš strukturovanými formáty (např. automaticky generovaná metadata internetových vyhledávačů) až po velké propracované mezinárodní standardy typu MARC nebo značkovací systémy typu TEI [74]. Zatímco popisná metadata bývají často uložena v katalozích a indexech udržovaných vně repozitářů s digitálními objekty, strukturální a administrativní data bývají naopak často vložena přímo do digitálního objektu.

V tomto příspěvku se zmíníme o některých přístupech z oblasti popisných metadat. Zájemce o problematiku metadat pro účely dlouhodobé archivace odkazujeme na [22], příkladem z oblasti správy vlastnických a autorských práv je DOI [24], domovská stránka iniciativy INDECS – INteroperability of Data in E-Commerce Systems [38] zase poskytuje dobrý vstupní bod pro studium problematiky metadat v oblasti elektronického obchodování. Ze zástupců metadat pro netextové dokumenty zmiňme alespoň standard MPEG-7, Multimedia Content Description Interface [56]. Podrobný přehled popisných metadat lze nalézt v [18] a [72], rozsáhlý seznam internetových zdrojů na téma metadat s odkazy na různá metadatová schémata je udržován na [40] a analýzu výzkumných témat v oblasti metadat podává [28].

Hlavním účelem síťových popisných metadat je zlepšit přesnost vyhledávání a výběru digitálních informačních zdrojů ve srovnání s tím, co dnes nabízí internetové vyhledávače (velká úplnost (*recall*), malá přesnost (*precision*)). Prvním pokusem v tomto směru byl návrh RFC-1807 (Bibliographical Format for Technical Reports) využitý např. v DIENST [19], protokolu a implementaci systému distribuovaných serverů digitálních knihoven použitým v řadě projektů digitálních knihoven. Nejznámějším a patrně nejperspektivnějším formátem v této oblasti je však standard Dublin Core.

5.2 Dublin Core

Klasická bibliografická metadata vycházející např. ze standardu MARC jsou příliš složitá a pravidla pro jejich použití (nejčastěji Angloamerická katalogizační pravidla – AACR2) příliš komplikovaná na to, aby je byl schopen používat i někdo jiný než jen profesionální katalogizátoři. Naproti tomu metadata generovaná automaticky internetovými vyhledávači či ad hoc doplňovaná do záhlaví webových dokumentů jsou příliš primitivní na to, aby mohla podstatněji ovlivnit přesnost vyhledávání. Dublin Core (DC) [25] je pokusem o kompromis: vytvořit metadatový standard rozumně jednoduchý na to, aby ho mohli využívat i nezaškolení autoři publikací na webu, na druhou stranu však dostatečně propracovaný, univerzální a flexibilní i pro netriviální aplikace v co nejširším spektru oborů a oblastí. DC se snaží vytvářet minimální jádro atributů společných všem typům intelektuálních výtvorů (knihy, články, obrazy, fotografie, hudební díla, programy, webová stránka apod.) bez ohledu na jejich formu a způsob realizace. Avšak ambice DC jdou ještě dál: stát se všeobecně rozšířeným a používaným standardem, který může díky své jednoduchosti posloužit i jako základna pro sémantickou interoperabilitu mezi jinými složitějšími formáty. Podrobnější popis motivace a historie DC lze nalézt v [10].

Dublin Core lze používat dvěma způsoby; prvním je tzv. *nekvalifikovaný Dublin Core* (tj. DC bez kvalifikátorů), kdy uživatel má k dispozici 15 základních metadatových prvků popisujících obsah (název, předmět, popis, pokrytí, typ, zdroj, vztah), intelektuální vlastnictví (tvůrce, přispěvatel, vydavatel, práva) a instanci síťového zdroje (identifikátor, datum, jazyk, formát). Každý prvek je nepovinný, opakovatelný a na pořadí prvků nezáleží. Implementátor konkrétní aplikace může dokonce přidávat své vlastní specifické prvky, ty však samozřejmě nebudou globálními aplikacemi využitelné. Druhý způsob představuje tzv. *kvalifikovaný Dublin Core*, kdy pro zpřesnění popisu zdroje lze jednotlivé prvky DC specifikovat pomocí dvou typů kvalifikátorů: kvalifikátor prvku (zužuje sémantiku prvku – např. Autor, Ilustrátor) a kvalifikátor hodnoty (specifikuje způsob interpretace zadané hodnoty prvku – např. Datum = „1999-04-12“ : ISO8601). Existuje seznam standardizovaných kvalifikátorů pro jednotlivé prvky DC, implementátoři mohou opět doplňovat vlastní kvalifikátory. Všechny kvalifikátory však musí splňovat tzv. princip „dumb-down“, což znamená, že kvalifikovaný záznam DC musí být korektně zpracovatelný i aplikací navrženou pro nekvalifikovaný DC (aplikace jednoduše ignoruje kvalifikátory a hodnoty všech prvků musí i po této restrikci odpovídat sémantice základních prvků nekvalifikovaného DC).

Uvedme příklad kvalifikovaného záznamu Dublin Core:

```
IDENTIFIER=http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf : URL
TITLE = Guidance on expressing the Dublin Core within the RDF
TITLE = Dublin Core in RDF: Eine Anleitung
CREATOR = Eric Miller
CREATOR = Paul Miller
CREATOR.Illustrator = Dan Brickley
DESCRIPTION.Abstract = This work describes work carried out by ...
SUBJECT.Keywords = Dublin Core; DC; Resource Description Framework; RDF; XML
PUBLISHER = Dublin Core Metadata Initiative
CONTRIBUTOR = Dublin Core Data Model Working Group
DATE.Created = 1999-07-01 : ISO8601
```

DATE.Revised = 1999-11-10 : ISO8601
LANGUAGE = en : RFC1766
TYPE = Working Draft
FORMAT.Medium = text/html : IMT
MYELEMENT.Checksum = 123456 : XYZ

Tvůrci Dublin Core (mezinárodní komunita informačních specialistů, knihovníků a vydavatelů koordinovaná organizací OCLC) se soustředili pouze na precizní specifikaci sémantiky a syntaxi záměrně ponechávali volnou (nepředjímalo se, kde a jak budou metadata Dublin Core ukládána). Časem se však ukázalo, že je vhodné vypracovat základní návody na způsoby zápisu alespoň v základních formátech, jakými jsou HTML, XML a RDF. Dublin Core se stal internetovým doporučením již ve druhé polovině 90. let (RFC-2413) a koncem minulého století byl již v pozici faktického standardu přeloženého do desítek jazyků (česká verze byla vytvořena na Masarykově univerzitě v Brně, kde je i spravována [26]). V roce 2001 byl DC schválen jako americký ANSI/NISO standard Z39.85.

5.3 Metadata Kongresové knihovny

Tento „firemní standard“ metadat [50] byl vypracován pro potřeby projektů digitálních knihoven zastřešených velmi ambiciózním programem *National Digital Library Program* [51] vedeným Kongresovou knihovnou. Na rozdíl od Dublin Core se tedy nejedná o univerzální standard, nýbrž o příklad schématu navrženého (byť dostatečně obecně) pro potřeby jednoho konkrétního praktického programu. Toto metadatové schéma zahrnuje současně metadata strukturální, administrativní a částečně i popisná (plná popisná data jsou uložena v katalozích), a to pro podporu všech funkcí digitální knihovny: řízení přístupu, vyhledávání, prezentaci, administraci, perzistentní identifikaci a dlouhodobé uchovávání digitálních objektů. Metadatové záznamy mají pět úrovní odpovídajících pětiúrovňové hierarchické struktuře objektů. Nejvyšší úroveň, tzv. *seť*, je jedna digitální sbírka. Sběrka je tvořena jedním či více *agregáty*, tj. skupinami digitálních objektů stejného typu (text, video) či podléhajících stejné správě. Agregát sdružuje *primární objekty*, koherentní jednotky odpovídající fyzické jednotce (knize, nahrávce). Primární objekt může sestávat z několika komponent, respektive pohledů – *meziobjektů* (např. kniha může mít dva meziobjekty: první z nich je obrazová prezentace knihy, sestávající z obrázků vzniklých naskenováním jejích stran; druhý meziobjekt zastupuje textovou komponentu knihy, soubor textů vzniklých např. převodem z obrázků pomocí OCR a sloužící pro plnotextové vyhledávání). Konečně poslední úroveň tvoří *terminální objekty*, což jsou jednotlivé soubory s digitální informací.

Metadatové schéma má celkem 77 prvků, všechny však nemusí být v metadatovém záznamu vyplněny; některé jsou specifické pro konkrétní typ média (jiné pro digitalizovanou zvukovou nahrávku, jiné pro digitalizovaný obrázek), další zase pro úroveň objektu (např. metadata pro terminální objekty jsou především strukturální).

5.4 Standardy vycházející ze struktury MARC

Předchozí dva příklady popisovaly metadatová schémata vytvořená přímo pro digitální, resp. síťové zdroje. Aby byl náš obrázek úplnější, musíme zmínit i opačný pól: národní a mezinárodní bibliografické standardy založené na struktuře MARC (Machine Readable Cataloguing), které byly vytvořeny původně pro popis fyzických informačních objektů v knihovnách a až posléze byly upravovány i pro potřeby digitálního světa. Na základě společné koncepce vyvíjené od poloviny 60. let v Kongresové knihovně vznikla postupně celá rodina více či méně kompatibilních standardů zohledňujících národní specifika, např. USMARC (USA), UKMARC (UK), CANMARC (Kanada) a v roce 1977 pod záštitou IFLA i mezinárodní standard UNIMARC, který pak mnohé další země převzaly za svůj národní formát. Pro tyto standardy je charakteristická velmi bohatá struktura polí a podpolí, velice jemná granularita popisu a detailní propracovanost návazných pravidel pro tvorbu záznamů. Pro představu uveďme příklad zjednodušeného záznamu ve formátu UNIMARC v tzv. řádkové struktuře:

```
001    CASLIN0000001
005    19960312
010    $a80-7050-237-1
100    $a19960305d1996####k##y0czy0103####ba
101 0#  $acze
102    $aCZ
200 1#  $aZáznam pro souborný katalog
        $eUNIMARC$Itištěnémonografie$fPracovní skupina CASLIN pro standardizaci a jmenné zpracování ... [et
        al.]
205    $a1. vyd.
210    $aPraha$cNárodní knihovna České republiky$d1996
215    $a 31 s.
225 1#  $aStandardizace$vc. 4
675    $a025.3$9v
```

711 02 \$aCASLIN\$bPracovní skupina pro standardizaci a ...
 801 #0 \$aCZ\$bABA001\$c19960312\$gAACR2\$91
 801 #3 \$aCZ\$bABA001\$c19960515
 910 \$a ABA001

Aniž bychom zabíhali do podrobností, uveďme alespoň základní vysvětlení: řádek začíná vždy třípísmenným názvem pole, za kterým mohou (ale nemusí) následovat až dva poziční indikátory upřesňující obecné charakteristiky pole. Hodnota pole je strukturována do podpolí označených znakem „\$“ následovaným jednoznakovým názvem podpole (má-li pole jen základní dále nestrukturovanou hodnotu, je uložena v podpoli \$a). Např. pole 200 obsahuje názvové údaje, které jsou v našem příkladě tvořeny hlavním názvem *Záznam pro souborný katalog* v podpoli \$a, další názvové informace *UNIMARC* v podpoli \$e, název části *Tištěné monografie* v podpoli \$i a údaj o odpovědnosti v podpoli \$f. Standard definuje pro každé pole seznam všech jeho možných podpolí, jejich charakteristiky, vzájemné vazby a závislosti.

Formáty typu MARC umožňují popsat (fyzický) informační zdroj do těch nejjemnějších nuancí způsobem vhodným pro počítačové zpracování. Na druhé straně jsou však příliš složité pro použití mimo oblast knihoven vybavených specializovanými informačními pracovníky.

5.5 XML a RDF

Definovat strukturu a obsah metadat je jedna věc, další otázkou je, jak metadata zapsat a kam je uložit. Protože v současnosti realizované digitální knihovny jsou pevně svázány s webem, je přirozené využít k zápisu metadat standardních webových technologií. Jednoduchá síťová metadata jsou často vkládána přímo do digitálního informačního zdroje, např. pomocí metaznaček do záhlaví HTML dokumentů. Jedním z klíčových důvodů tak úspěšného prosazení webu byla i jednoduchost jazyka HTML. V současnosti je však tato jednoduchost i jeho největší slabinou; HTML je orientován na prezentaci dokumentu a nikoliv na zachycení jeho struktury. Značkovací jazyk XML [81], jako nástupce HTML, je pokusem o kompromis mezi jednoduchostí HTML a silou SGML (který je naopak pro široké použití příliš flexibilní a komplikovaný). XML dokáže dobře zapsat potřebné struktury a navíc byl navržen s ohledem jednak na snadné vytváření programů pro manipulaci s XML dokumenty, jednak na relativně snadný přechod z HTML. V současnosti je XML nejčastější forma zápisu síťových metadat všeho typu a proniká i do takových oblastí, jako je standard MARC (viz projekt XMLMARC [82]).

Každé metadatové schéma má tři aspekty – *sémantiku, syntaxi a strukturu*. Sémantika definuje interpretaci jednotlivých prvků, jejich význam (CO chceme říci o informačním zdroji). Dublin Core se týká sémantiky – přesně specifikuje význam každého prvku formátu. Syntaxe je naopak způsob, jak metadata formálním a přesným způsobem zapsat (čili JAK svá tvrzení o informačním zdroji vyjádříme). Nástrojem pro zápis Dublin Core je např. jazyk XML. Struktura definuje vztahy mezi metadatovými prvky, v ideálním případě i mezi prvky různých metadatových schémat. Nástrojem pro vyjádření a zápis struktury metadat je *Resource Description Framework* (RDF) [67], který jako svůj vyjadřovací jazyk používá XML. Strukturální model RDF sestává ze zdrojů, atributů a hodnot a reprezentuje vlastně způsob zápisu orientovaných grafů představujících vztahy mezi zdroji. Uvažujme např. jednoduchou větu „Shakespeare je autorem hry Hamlet“.

V Dublin Core je zdrojem popisovaný dokument a atributem (vlastností) příslušný prvek DC; zachytit v něm lze pouze jednoduchá tvrzení typu „dokument – prvek – hodnota“, takže výše uvedenou větu zaznamenejme jako:

zdroj		atribut		hodnota
Hamlet	→	CREATOR	→	Shakespeare
Hamlet	→	TYPE	→	hra

Jiná metadatová schémata mohou také obsahovat prvek pro autora, ale třeba jinak pojmenovaný (nikoliv CREATOR, ale např. AUTHOR) a naopak prvek pojmenovaný TYPE mohou používat s úplně jinou sémantikou. Proto musí RDF explicitně vyjádřit, že prvky „creator“ a „type“ mají v tomto případě takový význam, jaký jim dává standard Dublin Core – k tomu slouží mechanismus jmenného prostoru xmlns, viz příklad níže. Předpokládejme, že Hamlet je reprezentován webovým zdrojem <http://hamlet.org>; pak naši větu můžeme zapsat v RDF následovně:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:DC="http://purl.org/dc/elements/1.1/">
  <rdf:description rdf:about="http://hamlet.org">
    <DC:creator>Shakespeare</DC:creator>
    <DC:type>hra</DC:type>
  </rdf:description>
</rdf:RDF>
```

Ze zápisu je zřejmé, že atributy „description“ a „about“ jsou definovány ve schématu RDF, zatímco „creator“ a „type“

ve schématu Dublin Core. Pokud bychom chtěli doplnit popis Hamleta o atribut z nějakého jiného metadatového schématu než Dublin Core, stačilo by doplnit do záznamu jmenný prostor s odkazem na definici tohoto schématu `<xmlns:NM="http://www.abc.cz/new-metadata/">` a přidat do popisu nový atribut s příslušným prefixem `<NM.novy-atribut>xyz</NM.novy-atribut>`.

Strukturální model RDF umožňuje zachytit i mnohem složitější vztahy; např. atribut zdroje může odkazovat na jiný zdroj. Předpokládáme, že v určité databázi existuje záznam o Shakespearovi obsahující jeho metadata, životopis, informace o díle apod. V našem příkladě by atribut `DC.creator` odkázal na tento záznam následovně:

```
<DC.creator rdf:about = "http://people.net/WS/">Shakespeare</DC.creator>
```

Popsaným způsobem lze v RDF propojovat metadata s digitálními objekty, vyměňovat metadata z různých metadatových schémat a skládat z jednoduchých komponent libovolně složité metadatové popisy. Jednou z oblastí, v níž je velmi často potřebné zachytit i velmi komplikované vztahy, je oblast vlastnických a autorských práv.

6 Interoperabilita

Jak již bylo řečeno v úvodu, obecná představa 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 *inter-operabilita* 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 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 metadatech umožňují poskytnout omezenou koherenci obsahu, která pak musí být ještě doplněna lidskou interpretací.

Při zkoumání interoperability se ukazuje závislost mezi funkcionalitou a cenou [4]. Většina v současnosti používaných metod pro interoperabilitu (např. webové standardy HTTP, HTML, URL) dosahuje 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 shrnuje [63]. Uvádí, že problém interoperability se bezprostředně dotýká všech pěti základních funkcí 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;
- šíře 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 řada velmi rozdílných přístupů k dosažení požadovaného stupně interoperability; práce [63] 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 schématy 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 (popisat dostatečně podrobně komponentu může být velmi složité a někdy v praxi i nemožné).
5. *mobilní funkcionality* – 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 (kontrola 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 přináší [29]. Rozlišuje několik abstraktních úrovní interoperability, od obecné transportní vrstvy a na aplikační oblasti nezávislého 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 knihoven [65]:

úroveň	způsob kooperace
<i>federace</i>	striktní použití standardů (syntaktické, sémantické, obchodní) příklad: MARC, Z39.50
<i>sklizení metadat</i>	digitální knihovny nabízejí základní metadata; jednoduchý protokol a registrace příklad: otevřené archivy, např. iniciativa OAI
<i>shromažďování dat</i>	digitální knihovny nekooperují; informace se musí hledat explicitně pomocí služeb příklad: internetové vyhledávače
<i>vyhledávací middleware</i>	zdroje vybavené metadaty jsou volně zapojovány do kooperace příklad: Z39.50, XML, RDF, SDLIP [62].

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 (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 lze nalézt v [53].

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. Zdůrazňeme, že Z39.50 je protokol mezi dvěma počítači, nijak nespecifikuje 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 (podporovanou verzi protokolu, 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:

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, seřadit 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 uchovávání 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 rozvoj [52]. 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 Initiative [60] 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 e-Print archive [8], NCSTRL [57]). 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:

- *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 sklízecí protokol OAI, 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 směrovací službu OAI s podporou OpenURL (viz níže);
- *protokol pro sklizení metadat* – původní návrh počítal s využitím protokolu Dienst, ale pro zjednodušení implementace byl nakonec vytvořen samostatný protokol OAI 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 komponentám. Výsledky výzkumu byly realizovány v systému *InfoBus* [73] (název vychází z analogie s hardwarovou sběrnicí propojující různé hardwarové komponenty do jednoho funkčního celku) využívajícího technologii distribuovaných objektů na bázi systému CORBA (Common Object Request Broker Architecture).⁴⁾

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 DLIOIP (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 DLIOIP, druhý pro překlad mezi Dialogem a DLIOIP. 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;
- FIRM (Framework for Interoperable Rights Management) – řada zpracovaných technik pro řízení přístupu 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 nejznámějšího a nejpoužívanějšího internetového vyhledávače Google (hlavní tvůrci Googlu se jako postgraduální studenti Stanfordské univerzity podíleli na výzkumu vyhledávání v oblasti digitálních knihoven).

6.5 OpenURL 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 výše uvedené představy, 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 Gentu ([78], [79]). 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 jiná 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 zakódovaná v podobě OpenURL. Např. výchozím zdrojem necht' 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&title=J%20Biol%20Chem&volume=275&issue=44&page=34826>

První částí OpenURL je adresa servisní služby, za ní následuje identifikátor zdroje, 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 standardu ANSI);

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 zdroj 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 [77] 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, jež 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 předchozí dvě probírané oblasti 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 Úvod a stručný přehled

Detailní rozbor všech aspektů této problematiky lze nalézt v [30]; stručně 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 informačníchází na systémové úrovni (přesměrování (*routing*) dotazů ke správným fyzickým serverům), interakci informačníchází 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 informačníchází na logické úrovni (za použití metadat pro popis celých informačníchází 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 informačníchází (*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 pocítována 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 výsledky [5]: vyhledávání informací (webové vyhledávače), rozhodování, nakolik vyhledaný dokument odpovídá zadanému dotazu (přístupy z oblasti sémantiky dokumentů, viz např. projekt digitální knihovny Illinoiské univerzity [75]), vyhodnocování důležitosti dokumentů (řadící algoritmus systému Google [32]), archivace digitálního dědictví (automatizovaný přístup v Internet Archive [43] nebo švédském programu Kulturarw3 [46]), citační analýza (ResearchIndex [12]), kontextové propojování informačních zdrojů (SFX – viz výše), automatická extrakce metadat z multimediálních digitálních objektů (Informedia Digital Video Library na Mellonově univerzitě [42]) nebo pokusy o vytvoření automatického referenčního knihovníka (projekt na univerzitě ve Washingtonu [9]).

7.2 Digitální knihovny a internetové vyhledávače

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.*“^[5] [5]. V tomto duchu byly až donedávna digitální knihovny i internetové vyhledávače považovány obecně za dvě naprosto nezávislá paradigmatu využívající webového prostředí k vytváření informačních repozitářů. Práce [35] 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 vyhledávače nedostupný „hluboký“ web je údajně až 500krát rozsáhlejší [17]) 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 prostými 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 univerzálnější 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 [57]). 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 vytvářených informačními specialisty a při kontrole potřebné kvality (např. SourceBank, ArticleCentral.com).

V [35] se předpovídá 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 Stručné poznámky k dalším oblastem digitálních knihoven

Velmi stručně se 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 Intelektuální vlastnictví a ekonomika

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 [85]. 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 [86]; (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* [87], 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. Intenzivně 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 uchovávání digitální informace

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, neznechodnocují 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 uchovávání. 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*“⁽⁶⁾ [88]. Navzdory aktivnímu výzkumu v oblasti *strategií pro dlouhodobé uchovávání* (replikace, oživování, technické muzeum, migrace, emulace, konverze do analogové formy aj.) není obecně současná situace z hlediska dlouholetého uchovávání digitální informace zatím nijak příznivá. Systematický koncepční přístup k problematice uchovávání zahrnující technické i organizační a systémové přístupy představuje referenční model OAIIS – Open Archival Information System [89]. Zajímavé projekty hraničící s intelektuálními cvičeními z oblasti opravdu „dlouhodobého“ uchovávání (po dobu tisíciletí) shrmažduje nadace The Long Now [90] – 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

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 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 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) [20] 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 te-rabytové 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 digitálních 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 vyhlášeným 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 [21].

9.3 Electronic Library Programme (eLIB)

Britský program eLIB [27], 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“ [49]. Cílem tohoto programu [51] 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* [1]; 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 uchovávání digitální informace.

9.5 Ostatní

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 vyhlašováno 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* po záštitou UNESCO [55]), 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 [37].

10. Závěr

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.

Literatura

1. *American Memory* [online]. Washington : The Library of Congress, [2004]. Dostupný z WWW: <<http://memory.loc.gov/>>.
2. *Ariadne Magazine* [online]. UKOLN, 1996-. Dostupný z WWW: <<http://www.ariadne.ac.uk/>>.
3. *ARL Digital Initiatives Database* [online]. Washington : Association of Research Libraries, c2001. Dostupný z WWW: <<http://www.arl.org/did/>>.
4. ARMS, W. Y. *Digital Libraries*. Cambridge : MIT Press, 2000. ISBN 0-262-01880-8.
5. 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>>.
6. ARMS, W. Y., BLANCHI, C., OVERLY, E. A. An Architecture for Information in Digital Libraries. *D-Lib Magazine* [online], February 1997. Dostupný z WWW: <<http://www.dlib.org/dlib/february97/cnri/02arms1.html>>.
7. ARMS, W. Y. Key Concepts in the Architecture of the Digital Library. *D-Lib Magazine* [online], July 1995. Dostupný z WWW: <<http://www.dlib.org/dlib/July95/07arms.html>>.
8. *arXiv.org e-Print archive* [online]. Cornell University. Dostupný z WWW: <<http://www.arxiv.org/>>.
9. *Automatic Reference Librarian Project* [online]. University of Washington. Dostupný z WWW: <<http://www.cs.washington.edu/research/diglib/>>.
10. BARTOŠEK, M. Vyhledávání v Internetu a DUBLIN CORE. *Zpravodaj ÚVT MU*, 1999, roč. 9, č. 4, s. 1–4. Dostupný též z WWW: <<http://www.ics.muni.cz/bulletin/issues/vol09num04/bartosek/bartosek.html>>.
11. *Berkeley Digital Library SunSITE* [online]. Berkeley : University of California, c2003. Dostupný z WWW: <<http://sunsite.berkeley.edu/>>.
12. BRATKOVÁ, E. Citace odborné literatury jako nástroj rozvoje služeb a integrace digitálních knihoven. In RAMAJZLOVÁ, B. (ed.). *Automatizace knihovnických procesů 8*. Praha : ČVUT, 2001, s. 109–120. Dostupný též z WWW: <<http://knihovny.cvut.cz/akp/clanky/12.pdf>>.
13. BUSH, V. As We May Think. *Atlantic Monthly*, July 1945, s. 101–108. Dostupný též z WWW: <<http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>>.
14. CLEVELAND, G. *Digital Libraries : Definitions, Issues and Challenges* [online]. IFLA, 1998. Dostupný z WWW: <<http://www.ifla.org/VII/5/op/udtop8/udtop8.htm>>.
15. *CrossRef* [online]. Pila, c2003. Dostupný z WWW: <<http://www.crossref.org/>>.
16. *D-Lib Magazine* [online]. Corporation for National Research Initiatives, 1995-. Dostupný z WWW: <<http://www.dlib.org>>. ISSN 1082-9873.
17. BERGMAN, M. K. *Deep Web : White Paper* [online]. BrightPlanet, c2004. Dostupný z WWW: <<http://brightplanet.com/technology/deepweb.asp>>.
18. DEMPSEY, L., HEERY, R. *A Review of Metadata : A Survey of Current Resource Descriptive Formats. Work Package 3 of Telematics for Research project DESIRE (RE 1004)* [online]. UKOLN Metadata Group, 1999. Dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/desire/overview/>>.
19. *Dienst : Overview and Introduction* [online]. Ithaca : Cornell University, 2000. Dostupný z WWW: <<http://www.cs.cornell.edu/cdlrg/dienst/DienstOverview.htm>>.
20. *Digital Libraries Initiative [Phase 1]* [online]. [National Science Foundation], 1999. Dostupný z WWW: <<http://www.dli2.nsf.gov/dlione/>>.
21. *Digital Libraries Initiative Phase 2* [online]. [National Science Foundation], 2003. Dostupný z WWW: <<http://www.dli2.nsf.gov/dlione/>>.
22. *Digitization and Preservation* [online]. OCLC, c2004. Dostupný z WWW: <<http://www.oclc.org/digitalpreservation/>>.
23. *The Digital Object Identifier System* [online]. International DOI Foundation, 2004. Dostupný z WWW: <<http://www.doi.org/>>.
24. *The DOI Handbook* [online]. International DOI Foundation, 2004. Dostupný z WWW: <<http://dx.doi.org/10.1000/182>>.
25. *Dublin Core Metadata Initiative* [online]. DCMI, c2004. Dostupný z WWW: <<http://dublincore.org/>>.
26. *Dublin Core Czech* [online]. Brno : Masarykova univerzita, 2000. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/DC-czech-1.1.html>.
27. *eLib : the Electronic Library Programme* [online]. UKOLN, 2004. Dostupný z WWW: <<http://www.ukoln.ac.uk/services/elib/>>.
28. *Metadata for Digital Libraries : A Research Agenda* [online]. ERCIM-DELOS, 1999. Dostupný z WWW: <<http://www.iei.pi.cnr.it/DELOS/NSF/metadata.html>>.
29. *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>>.
30. *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>>.
31. FOX, E. *Digital Libraries : Virginia Tech Courseware* [online]. Dostupný z WWW: <<http://ei.cs.vt.edu/~dlib/>>.
 32. Google [online]. Google, 2004. Dostupný z WWW: <<http://www.google.com/>>.
 33. *The New Zealand Digital Library* [online]. University of Waikato, 2000. Dostupný z WWW: <<http://nzdl.org/>>.
 34. HAKALA, J. Document Description and Access : New Challenges. In *CASLIN 2001*. Praha : Knihovna Akademie věd ČR, 2001, s. 33–46.
 35. 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.
 36. *Handle System* [online]. Corporation for National Research Initiatives, 2003. Dostupný z WWW: <<http://www.handle.net/>>.
 37. *Library Digital Initiative* [online]. Harvard University Library, 2004. Dostupný z WWW: <<http://hul.harvard.edu/ldi/>>.
 38. *INDECS : Interoperability of Data in E-commerce Systems* [online]. indecs, c2001. Dostupný z WWW: <<http://www.indecs.org/>>.
 39. *Digital Libraries : Resources and Projects* [online]. IFLA, 2003. Dostupný z WWW: <<http://www.ifla.org/II/diglib.htm>>.
 40. *Digital Libraries : Metadata Resources* [online]. IFLA, 2004. Dostupný z WWW: <<http://www.ifla.org/II/metadata.htm>>.
 41. IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records : final report*. IFLA, 1998. Dostupný též z WWW: <<http://www.ifla.org/VII/s13/frbr/frbr.htm>>.
 42. *Informedia Project* [online]. Carnegie Mellon University, c2004. Dostupný z WWW: <<http://www.informedia.cs.cmu.edu/>>.
 43. *Internet Archive* [online]. San Francisco : Internet Archive, [2004]. Dostupný z WWW: <<http://www.archive.org/>>.
 44. KAHN, R., WILENSKY, R. *A Framework for Distributed Digital Object Services. Technical Report hdl:cnri.dlib/tn95-01*. CNRI, 1995. Dostupný též z WWW: <<http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>>.
 45. KENNEY, A. R., OYA, Y. R. *Moving Theory into Practice : Digital Imaging for Libraries and Archives*. RLG, 2000. ISBN 0-9700225-0-6
 46. *Kulturarw³ : long time preservation of electronic documents* [online]. The Royal Library, National Library of Sweden, 2004. Dostupný z WWW: <<http://www.kb.se/kw3/ENG/>>.
 47. LAGOZE, C., SHAW, E., DAVIS, J. R., KRAFFT, D. B. *Dienst : Implementation Reference Manual*. Cornell University, 1995.
 48. LESK, M. *Practical Digital Libraries : Books, Bytes, and Bucks*. San Francisco : Morgan Kaufmann Publishers, 1997. 297 s. ISBN 1-55860-459-6.
 49. *LC21 : A Digital Strategy for the Library of Congress*. Washington : National Academy Press, c2000. 265 s. ISBN 0-309-07144-5. Dostupný též z WWW: <<http://books.nap.edu/html/lc21/>>.
 50. *Library of Congress Digital Repository Development : Core Metadata Elements* [online]. Washington : Library of Congress, 2004. Dostupný z WWW: <<http://lcweb.loc.gov/standards/metadata.html>>.
 51. *National Digital Library Program* [online]. Washington : Library of Congress. Dostupný z WWW: <<http://memory.loc.gov/ammem/dli2/html/lcndlp.html>>.
 52. *Z39.50 Maintenance Agency* [online]. Washington : Library of Congress, 2004. Dostupný z WWW: <<http://lcweb.loc.gov/z3950/agency/>>.
 53. 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>>.
 54. LYNCH, C., GARCÍA-MOLINA, H. *Interoperability, Scaling, and the Digital Libraries Research Agenda* [online]. IITA, 1995. Dostupný z WWW: <<http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.htm>>.
 55. *Memoriae Mundi Series Bohemica : National Programme of Digital Access to Rare Documents = Národní program digitálního zpřístupnění vzácných dokumentů*. Praha, Národní knihovna ČR, 2004. Dostupný z WWW: <<http://digit.nkp.cz/>>.
 56. *MPEG.ORG* [online]. San Francisco : MpegTV, 2004. Dostupný z WWW: <<http://www.mpeg.org/>>.
 57. *Networked Computer Science Technical Reference Library* [online]. Dostupný z WWW: <<http://www.ncstrl.org/>>.
 58. NIKOLAOU, C., MARAZAKIS, M. System Infrastructure for Digital Libraries : A Survey and Outlook. In ROVAN, B. (ed.). *SOFSSEM'98 : theory and practice of informatics : 25th Conference on Current Trends in Theory and Practice of Informatics, Jasná, Slovakia, November 21-27, 1998 : proceedings*. Berlin ; New York : Springer, 1998, s. 186–203.
 59. *Techstreet.com : [NISO Standards and Technical Reports]* [online]. Techstreet, 2004. Dostupný z WWW: <<http://www.techstreet.com/nisogate.html>>.

60. *Open Archives Initiative* [online]. Dostupný z WWW: <<http://www.openarchives.org/>>.
61. *The Proposed OpenURL Framework Standard* [online]. Dostupný z WWW: <<http://library.caltech.edu/openurl/Standard.htm>>.
62. PAEPCKE, A., aj. Search Middleware and the Simple Digital Library Interoperability Protocol. *D-Lib Magazine* [online], 2000, roč. 6, č. 3. Dostupný z WWW: <<http://www.dlib.org/dlib/march00/paepcke/03paepcke.html>>.
63. 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.
64. PASKIN, N. Information Identifiers. *Learned Publishing*, 1997, roč. 10, č. 2, s. 135–156. Dostupný též z WWW: <<http://www.elsevier.nl/homepage/about/infoident/>>.
65. 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>>.
66. *Persistent URL : PURL* [online]. OCLC. Dostupný z WWW: <<http://purl.oclc.org/>>.
67. *Resource Description Framework (RDF)* [online]. W3C Consortium. Dostupný z WWW: <<http://www.w3.org/RDF/>>.
68. *RLG DigiNews* [online]. RLG, 1997-. Dostupný z WWW: <<http://www.rlg.org/preserv/diginews/>>. ISSN 1093-5371.
69. SAMUEL, A. L. The Banishment of Paperwork. *New Scientist*, 1964, roč. 21, č. 380, s. 529–530.
70. *SFX : context sensitive linking* [online]. ExLibris, c2004. Dostupný z WWW: <<http://www.sfxit.com/>>.
71. SCHAUBLE, P., SMEATON, A. F. (eds.). *An International Research Agenda for Digital Libraries : Summary Report of the Series of Joint NSF-EU Working Groups on Future Directions For Digital Libraries Research*. Brussels, 1998. Dostupný též z WWW: <<http://www.iei.pi.cnr.it/DELOS/NSF/Brussrep.htm>>.
72. SNIJDER, R. *Metadata Standards and Information Analysis : A Survey of Current Metadata Standards and the Underlying Models* [online]. 2001. Dostupný z WWW: <<http://www.geocities.com/ronaldsnijder/>>.
73. *Stanford University Digital Libraries Project* [online]. Dostupný z WWW: <<http://www-diglib.stanford.edu/diglib/pub/userinfo.html>>.
74. *Text Encoding Initiative* [online]. TEI Consortium, 2003. Dostupný z WWW: <<http://www.tei-c.org/>>.
75. *UIUC Digital Library Testbed* [online]. University of Illinois. Dostupný z WWW: <<http://dli.grainger.uiuc.edu/idli/idli.htm>>.
76. *Uniform Resource Names (urn)* [online]. IETF, 2001. Dostupný z WWW: <<http://www.ietf.org/html.charters/urn-charter.html>>.
77. VAN DE SOMPEL, H., BEIT-ARIE, O. Open Linking in the Scholarly Information Environment Using the OpenURL Framework. *D-Lib Magazine* [online], 2001, roč. 7, č. 3. Dostupný z WWW: <<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>>.
78. 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>. 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>.
79. VAN DE SOMPEL, H., HOCHSTENBACH, P. Reference Linking in a Hybrid Library Environment. Part 3, Generalizing the SFX solution in the „SFX@Ghent & SFX@LANL“ experiment. *D-Lib Magazine* [online], 1999, roč. 5, č. 10. Dostupný z WWW: <http://www.dlib.org/dlib/october99/van_de_sompel/10van_de_sompel.html>.
80. WATERS, D. J. What are digital libraries? *CLIR Issues* [online], 1998, č. 4. Dostupný z WWW: <<http://www.clir.org/pubs/issues/issues04.html#dlf>>.
81. *Extensible Markup Language (XML)*. W3C Consortium, 2004. Dostupný z WWW: <<http://www.w3.org/XML/>>.
82. *Medlane : A project of Lane Medial Library, Stanford Univesity Medical Center* [online]. Stanford University, 2003. Dostupný z WWW: <<http://xmlmarc.stanford.edu/>>.
83. *DSpace* [online]. MIT, c2003. Dostupný z WWW: <<http://www.dspace.org/>>.
84. WITTEN, I. H., BAINBRIDGE, D. *How to Build a Digital Library*. San Francisco : Morgan Kaufmann Publishers, c2003. 518 s.
85. EU-NSF Working Group on Intellectual Property and Economic Issues. *Intellectual Property and Economics Issues for Digital Libraries : A Framework for Future Research* [online]. 1998. Dostupný z WWW: <<http://www.iei.pi.cnr.it/DELOS/NSF/ipe.htm>>.
86. *ECUP : European Copyright User Platform* [online]. Hague : EBLIDA, c2000. Dostupný z WWW: <<http://www.eblida.org/ecup/>>.
87. ARMS, W. Y. Economic Models for Open Access Publishing. *iMP Magazine*, March 2000. Dostupný z WWW: <http://www.cisp.org/imp/march_2000/03_00arms.htm>.

88. ROTHENBERG, J. Ensuring the Longevity of Digital Documents. *Scientific American*, 1995, roč. 272, č. 1, s. 24–29. Rozšířená verze z roku 1998 dostupná též z WWW: <<http://www.clir.org/pubs/archives/ensuring.pdf>>.
89. *OAIS Reference Model* [online]. Dostupný z WWW: <<http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>>.
90. *The Long Now Foundation* [online]. The Long Now Foundation. Dostupný z WWW: <<http://www.longnow.org/>>.

Poznámky:

- 1) „Ačkoliv internet poskytuje přístup k velkému množství informací, současný stav má daleko k tomu, co je obvykle chápáno jako knihovní služba – to znamená relativně snadné vyhledávání a přístup k souboru dokumentů, které jsou součástí sbírky. Pojem sbírky je důležitý, protože implikuje, že soubor dokumentů není vybrán náhodně, ale nějakým důvěryhodným prostředníkem. Současní uživatelé internetu se setkávají s informačním prostorem, ve kterém má kvalita dokumentů daleko ke spolehlivosti, nástroje pro vyhledávání dokumentů jsou primitivní a získání konkrétního dokumentu často znamená bloudění babylónskou věží specifické architektury a formátů souborů.“ (redakční překlad).
- 2) „naše metody přenosu a hodnocení výsledků výzkumu jsou staré celé generace a dnes jsou již pro svůj účel zcela neadekvátní“ (redakční překlad).
- 3) Díky novým technologiím v poslední době klesají náklady na digitalizaci přímo závratným tempem – příkladem jsou technologie používané v programech *Google Print* (<http://print.google.com>) či *Search in the Book* (Amazon.com), umožňující dnes digitalizovat statisíce knižních svazků v krátkém časovém rozpětí a s velmi příznivými náklady.
- 4) Jedním ze základních komponent architektury CORBA je Object Request Broker (ORB), který po přidání k aplikačnímu programu realizuje vztah klient-server mezi distribuovanými objekty. Pomocí ORB může klient transparentně volat požadovanou operaci (metodu) nějakého serverového objektu, který se může nacházet na stejném počítači nebo kdekoli v síti. ORB najde objekt, který může realizovat požadovanou operaci, předá mu parametry, vyvolá jeho příslušnou metodu a vrátí výsledek. Klient samotný nemá žádné povědomí o tom, kde se tento objekt nalézá, jak a v jakém programovacím jazyku byl implementován, ani pod kterým operačním systémem je spouštěn. ORB tak umožňuje realizovat interoperabilitu mezi aplikacemi na různých počítačích v heterogenním distribuovaném prostředí.
- 5) „Prakticky všechno, co je nejlepší v knihovních katalozích, je mizerné u webových vyhledávačů. ... Na druhé straně, webové vyhledávače jsou výkonné tam, kde katalogy selhávají.“ (překlad autor a redakce).
- 6) „Digitální informace trvá věčně – nebo 5 let. Podle toho, co nastane dříve.“ (překlad autor).