

Popis dokumentů a přístup k nim – nové výzvy

Juha Hakala

Jedním z úkolů knihoven je katalogizace. Od rukou psaných katalogů přes kartičky psané na stroji až po on-line veřejně přístupné katalogy (OPAC) byl katalog vždy srdcem knihovny.

OPAC se skládá z metadat. Řada pokusů termín „metadata“ definovala od prostého „data o datech“ přes „klasifikace obsahu publikací na webu“ až po „vztah, který někdo připisuje dvěma entitám“. Existují i tvrzení, že na internetu není zřejmé rozlišení mezi obsahem a meta-obsahem. Z technického hlediska to může být pravda, ale pro knihovníky a jejich uživatele je právě toto rozlišení podstatné. V tomto textu se spokojíme s metadaty prostě jako s daty o datech a budeme se zabývat zejména tím, jak nakládat s uměleckými a intelektuálními výtvoři – díly – a jejich fyzickými provedeními.

Knihovny mají dobře zavedené nástroje a postupy, jak zpracovat tištěné publikace. Zkatalogizovat tištěnou knihu, zařadit ji na polici a později ji zde nalézt a půjčit čtenáři nepředstavuje technický problém. Udělat totéž s elektronickou knihou nebo článkem však už není tak prosté. I když cíl je týž – učinit informaci dostupnou uživateli – příslušné úkony jsou docela odlišné. A nástroje, které používáme – katalogizační pravidla, identifikátory, integrované knihovní systémy – postrádají vlastnosti a funkce potřebné pro práci v tomto novém prostředí.

V této přednášce se budu zabývat dvěma oblastmi, kde si internet vynucuje změny dosavadní infrastruktury:

- Identifikátory, zejména ISBN, ISSN a SICI (Serial Item and Contribution Identifier), které se užívají kupř. k identifikaci článků.
- Vznikající identifikátory děl, zejména ISTC (International Standard Textual Work Code).
- Směrovací služby, to jest systémy, které propojují popis zdroje a zdroj samotný (jako je signatura či přírůstkové číslo u tradičních knihovních materiálů a URL v http protokolu pro zdroje na internetu).

Na několika místech budu odkazovat na některé formáty metadat, jako je formát MARC, protože metadata jsou podstatná pro identifikaci zdrojů a pro fungování směrovacích služeb. Nebudu se pokoušet o porovnání soudobých modelů, neboť by to zabralo příliš mnoho času. Přehled z poslední doby poskytuje Snijder, 2001.

ISBN, ISSN a USMARC byly vyvinuty asi před třiceti lety. Po dlouhou dobu byly považovány za samozřejmost; nebylo třeba se zabývat základními vlastnostmi a omezeními těchto systémů. Technický pokrok a zejména vznik webu to však změnily: dnes je zřejmé, že systémy navržené začátkem 70. let pro tištěné publikace nemusí být vhodné pro použití na internetu, aniž budou modifikovány systémy samotné, nebo i pravidla jejich užívání. A abychom byli schopni zacházet s novými druhy zdrojů publikovanými na webu, budeme potřebovat zcela nové identifikátory a formáty.

Propojování zdrojů je dobrý příklad nového druhu služby, jakou nelze poskytnout v tradiční knihovně – kromě primitivního uvedení signatury v katalogovém záznamu. V elektronickém světě poskytuje podobnou službu URL. Bohužel to, co po staletí dobře fungovalo pro tištěné materiály, se ukázalo nedostatečné pro knihovnu digitální. URL mají velmi omezenou životnost; potřebujeme trvalejší směrovací služby kombinované s účinnými metodami digitální archivace. Jednotná jména zdrojů (URN) a identifikátory digitálních objektů (DOI) jsou nejdůležitějšími příklady takových služeb. Ale je zřejmé, že v internetu identifikátory nestačí: potřebujeme ještě metadata, která umožní vytvořit propojení vázaná na kontext; propojení, která berou kupříkladu ohled na oprávnění individuálního uživatele a jeho mateřské organizace.

IDENTIFIKACE ELEKTRONICKÝCH ZDROJŮ

Knihovny, vydavatelé a knihkupci užívají ISBN, ISSN a jiné identifikátory ve svých systémech od 70. let. Pokud objednávka knihy nebo žádanka MVS obsahuje ISBN, je daleko snazší vyřadit požadavek správně. Identifikátory také usnadňují údržbu, například odstraňování duplicit ze souborného katalogu. A jsou i podstatou směrovacích služeb.

Před internetem bylo třeba identifikovat konkrétní tištěné jednotky, jinými slovy provedení děl. Naproti tomu elektronické publikování si žádá vícevrstvou identifikaci, počínaje autorem samotným až po nejmenší jednotky v internetu individuálně dosažitelné, jako jsou články nebo obrázky publikované v seriálech.

V síťovém prostředí jsou k dispozici přinejmenším tyto kategorie identifikátorů:

- Identifikátory autorů. Mezinárodní standardní číslo autority (ISADN) jednoznačně identifikuje každého autora. To je nesmírně důležité tam, kde existuje mnoho „správných“ podob jména autora, daných např. transliterací. Hledání podle autora ve virtuálních katalozích by se podstatně zjednodušilo, pokud by bylo možno použít ISADN pro sloučení různých podob jména. Také ochranné autorské organizace potřebují ISADN, aby mohly předat vybrané poplatky z autorských práv správně osobě.
- Identifikátory děl. ISO nyní vyvíjí mezinárodní standardní číslo rodiny mezinárodních standardních kódů děl. Pro začátek vzniknou tři standardy: mezinárodní standardní číslo audiovizuálního díla (ISAN) pro audiovizuální materiály, mezinárodní standardní kód hudebního díla (ISWC) pro díla hudební a mezinárodní standardní kód textového díla (ISTC) pro textové materiály. Vznikne také mezinárodní standardní kód pro nepohyblivé obrázky, jeho vývoj však ještě nezačal.
- Identifikátory provedení děl, jako jsou jednotlivá vydání knihy. Tahle kategorie je nám důvěrně známá: všechny tradiční identifikátory jako ISBN či ISSN patří do této třídy identifikátorů.

- Identifikátory příspěvků či součástí provedení díla. V této kategorii vznikají dva identifikační systémy: identifikátor položek a příspěvků seriálů (SICI) pro články v časopisech a identifikátor položek a součástí knih (BICI) pro kupř. kapitoly knihy.

Řízení všech těchto systémů si vyžádá spoustu práce. Aby fungovaly dobře, musí být všechny systémy identifikátorů doplněny souvisejícími metadaty, např. ISSN by nebylo k ničemu bez metadat, která seriál, jemuž byl tento identifikátor přidělen, popisují. Skutečně: standard ISSN vyžaduje, aby seriál, kterému má být přiděleno ISSN, byl zkatologizován. Systém ISBN však (bohužel) takový požadavek nemá.

Také identifikátory děl a složek potřebují metadata, i když pouze pro identifikátory děl budou metadata povinná. Ale kdo by si troufal zkatologizovat například všechna textová díla, jako jsou knihy a články dostupné na internetu? Tato otázka nemá jasné řešení, ale spolupráce autorů, vydavatelů a knihoven pomůže pokrýt nejpodstatnější materiály. Nové technologie nám pomohou: kupříkladu bude-li XML používán jako produkční formát, je možné vložit do dokumentu dostatečná metadata a později tyto informace z textu vytěžit.

Pro identifikaci elektronických zdrojů jsou nyní využívány tradiční identifikátory. Už tisíce elektronických časopisů mají ISSN a tisíce e-knih a CD-ROMů mají ISBN. Ale to neznamená, že ISSN, ISBN a další identifikátory jsou schopné se s elektronickými zdroji vypořádat.

O omezeních ISSN, ISBN a SICI a BICI bude pojednáno později v této kapitole, především se budu zabývat všeobecnými problémy ovlivňujícími jakýkoliv systém.

- Ne každý identifikační systém lze rozšířit tak, aby mohl pojednat každý dokument patřící do tohoto systému.
- Nelze-li přidělování identifikátoru automatizovat, rozšíření katalogizace pouze z tištěných materiálů také na elektronické zdroje si vyžádá mnoho dalšího personálu.
- Často není jasné, jaký identifikátor použít pro elektronický zdroj, protože pravidla pro současné identifikační systémy původně počítala s tištěnými zdroji. Na rozšíření pravidel tak, aby pokryla také elektronické materiály, se pracuje, není to však snadné, mimo jiné proto, že elektronické zdroje se stále vyvíjejí. Zásáhnout pohyblivý cíl je bohužel vždy obtížné.

Po kritickém pohledu na tradiční identifikátory následuje krátký přehled identifikátorů děl, zejména mezinárodního standardního kódu textového díla (ISTC).

ISBN

Mezinárodní standardní číslo knihy bylo vyvinuto koncem 60. let. Systém byl velice úspěšný, v roce 2000 užívalo ISBN 152 zemí. Agentury ISBN zpravidla fungují dobře, i když některé trpí nedostatkem personálu. Čísla ISBN jsou zpravidla poskytována zdarma, i když jsou země, kde vydavatelé za ISBN musí platit.

ISBN má čtyři části: identifikátor země, identifikátor vydavatele, identifikátor publikace a kontrolní číslici. Identifikátor země může označovat jednotlivou zemi (951 = Finsko) nebo jazykovou oblast (3 = Německo, Rakousko a německy mluvící část Švýcarska). S touto strukturou ISBN poskytuje dobré vodítko k určení zeměpisné polohy vydavatele a databáze (národní bibliografie), která by mohla obsahovat informace o knize (neexistuje však závazek katalogizovat každou knihu, která ISBN dostane).

Z hlediska ISBN není fyzická podoba publikace rozhodující. Jakákoliv kniha, tištěná či elektronická, má dostat ISBN. Používání ISBN v internetu by tedy mělo být snadné. Bohužel tu jsou přinejmenším dva zásadní problémy.

V internetu může být vydavatelem v zásadě kdokoliv. Protože ISBN identifikuje také vydavatele, potřeba identifikátorů vydavatelů roste exponenciálně. S tím se ISBN do jisté míry dokáže vyrovnat, neboť zpravidla každá země vyhradila nějaký identifikátor vydavatele (ve Finsku 952-91) pro knihy vydané individuálními vydavateli. Ale tento mechanismus není dostatečně pružný pro webovské vydavatele. Je tu tedy potřeba rozšířit ISBN tak, aby obsáhlo velmi vysoký počet vydavatelských identifikátorů.

Jak už jsem se zmínil, internet zvýšil počet vydaných knih. Lze předpokládat, že v budoucnu bude velký podíl knih publikován rovněž v elektronické podobě. Je možné, že bude vznikat více souběžných elektronických verzí, a každá kniha se může skládat z mnoha složek, například každá kapitola může být samostatným souborem. Z pohledu vydavatele může být každá složka prodávána samostatně po síti, a tedy vyžaduje identifikaci.

Novým aspektem internetu je, že dvě úlohy: zdroj najít a zakoupit – například na základě jeho metadat – mohou následovat bezprostředně po sobě, jako by celý proces byla jediná akce. V minulosti, kdy nebylo možné informace získávat přímo ze sítě, bylo vyhledání informací o publikacích a jejich fyzické zakoupení oddělené: knihovny se staraly o popis zdrojů a vydavatelé a ochranné autorské organizace se staraly o vypořádání autorských práv a o obchodní transakce. Tato hranice mezi knihovnami a vydavateli začíná být nezřetelná, a tak se „knížní řetězec“ – postup knihy od vydavatele přes knihkupce a knihovnu ke klientovi – může citelně zkrátit.

Systém ISBN vznikl v 60. letech byl navržen pro tradiční vydávání knih. Dlouhou dobu byl pro tištěné knihy ISBN dostatek. Příliv elektronických publikací pravděpodobně způsobí vyčerpání čísel ISBN kolem roku 2010. Aby se tak nestalo, musí být systém ISBN rozšířen tak, aby bylo dost čísel pro budoucí roky. A o tomto rozšíření je třeba rozhodnout rychle, protože jinak nestihnou dodavatelé knihovních [ani vydavatelských – pozn. překl.] systémů své aplikace včas modernizovat. Údaje ISBN jsou v knihovních systémech uloženy na mnoha místech, proto je třeba mít na paměti, že jakákoliv jejich změna může být velmi obtížně proveditelná.

Národní centrum ISBN Spojených států navrhlo rozšíření ISBN z 10 na 13 číslic. Toho by se dosáhlo zařazením kódu knih EAN „978“ na začátek ISBN. Tím by se prakticky zdvojnásobila kapacita systému ISBN. Pro budoucnost by bylo možné rezervovat i EAN kód pro hudebniny „979“ také pro knihy.

Hlavním přínosem tohoto nového ISBN by bylo – kromě zvýšené kapacity – sladění se systémem EAN, který připouští nanejvýš třináctiznakový identifikátor. Uvážíme-li klíčovou roli ISBN v elektronickém obchodu, je nepravděpodobné, že by nové ISBN mohlo být na EAN nezávislé. Jinak by bylo možno zvolit rovnou ISBN dlouhé například 16 znaků, což by byla zcela jistě postačující délka.

Aby se do ISBN kompatibilního s EAN dalo „vmáchnout“ ještě víc identifikátorů, padl také návrh, aby ISBN bylo – podobně jako ISSN – němé, tedy nestrukturované číslo. To by samozřejmě výrazně zvýšilo kapacitu systému, ale řada center ISBN s touto myšlenkou nesouhlasí. Skutečnost, že ISBN je „mluvící“ identifikátor udávající zemi nebo region vydání a vydavatele, je považována za významný rys systému. Tato vlastnost vskutku není důležitá jenom z „ideologických“ důvodů, ale i proto, že směřování založené na ISBN – to jest nalezení bibliografických informací a posléze i zdroje samotného vycházející od identifikátoru – je s dnešním ISBN a s využitím národních bibliografií možné. Bohužel, němé ISBN by nebylo schopno směřovacímu procesu naznačit, kde hledat správnou směřovací službu [národní bibliografii – pozn. překl.]. Směřování podle ISSN v rámci systému URN je možné jen proto, že existuje databáze ISSN spravovaná centrem ISSN v Paříži. ISBN žádnou takovou globální databázi nemá a vytvořit ji by bylo nesmírně obtížné.

Ti, kdo navrhují nové ISBN, poukazují na to, že někteří mezinárodní vydavatelé občas užívají špatné kódy zemí, například vydavatel se sídlem v Německu může přidělit ISBN s kódem země „3“ i knize vydané v USA. Nicméně i taková kniha by měla být uložena v Německu a katalogizována v německé národní bibliografii, což je právě místo, kde by směrovací proces data hledal.

V této chvíli nikdo ještě neví, zda nové ISBN bude mít 13 nebo více číslic a zda bude němé, nebo ne. Jisté je, že obec ISBN tlačí čas, nový identifikátor by měl být uveden do používání od 1. ledna 2005. To bude možné jen tehdy, pokud se podaří dohodnout nové ISBN s předstihem jednoho či dvou let, aby je dodavatelé knihovních systémů mohli včas upravit. Naneštěstí nikdo neví, kdy bude nové ISBN schváleno ISO; dokud se obec ISBN nedohodne, je obtížné pokročit dále.

ISSN

Mezinárodní standardní čísla seriálů jsou široce využívána k identifikaci seriálů, jako jsou časopisy, noviny, periodika atd. Na rozdíl od ISBN jsou čísla ISSN naprosto němá, neposkytují žádný náznak, kde byl časopis vydán. Naneštěstí však mezinárodní centrum ISSN, které koordinuje užívání ISSN, udržuje globální databázi ISSN. Ta na jaře 2001 obsahovala asi milion záznamů. Každé přidělené ISSN musí být doprovázeno metadaty, která musí být předána mezinárodnímu centru ISSN, jež je zařadí do globální databáze ISSN.

Syntaxe ISSN, osm číslic, z nichž poslední je kontrolní, poskytuje 10 milionů ISSN. Poněvadž doposud byl přidělen pouze jeden milion ISSN, identifikátor může růst tak, aby dostatečně pokryl i elektronické časopisy. Dosud jsou ISSN přidělována tempem asi 50 000 ročně. I když pravidla ISSN požadují samostatná ISSN pro elektronickou a tištěnou verzi téhož časopisu, spotřebu čísel ISSN to citelně neovlivnilo – zatím.

Ale i obec ISSN má problém s elektronickými seriály. Elektronický časopis nemusí být vydáván v ročnících a číslech; fakticky kterákoliv webová stránka, na níž se shromažďují texty, by mohla teoreticky být považována za seriál či přinejmenším za určitý druh pokračující publikace. Kvůli této změně ve vydávání už byla modernizována katalogizační pravidla: je k dispozici nové ISBD pro pokračující zdroje. Rovněž Anglo-americká katalogizační pravidla byla upravena; sladění ISBD a AACR nebylo snadné, nicméně v době, kdy píšete tuto přednášku, obojí pravidla víceméně souhlasí v tom, jak zacházet s takovými seriály.

Naneštěstí existence revidovaných pravidel a pokynů pro uživatele ještě nutně neřeší celý problém. S internetovými časopisy se obtížně zachází pro jejich tendenci mizet nebo se stěhovat (měnit URL). A protože tištěné časopisy nehodlají vymizet, musí národní a regionální centra ISSN dodávat či upravovat daleko více záznamů než dříve. Je zřejmé, že se to neobejde bez dalšího personálu.

Časté změny elektronických seriálů se netýkají jen nekomerčních vydavatelů, zasahují i seriály vydávané „opravdovými“ vydavateli. Finské univerzitní knihovny mají licenci na přístup k 3 500 elektronických seriálů, udržet informace o nich v korektním stavu vyžaduje spoustu práce. Samozřejmě, zacházet s takovým počtem seriálů samostatně v každé z 22 finských univerzit by bylo ještě daleko náročnější. Některé úkoly se zjednodušily nebo zcela vymizely: u elektronických časopisů není třeba sledovat došlá čísla a směřovat je k subskribentům. Ale zatímco tyto úlohy zanikly, nové se objevily: nějaká organizace musí sjednávat licence s vydavateli, a pokud vydavatel ukončí činnost, musí být vyřešeno zachování obsahu pro budoucí čtenáře.

V prostředí internetu lze nyní řadu úkolů vykonávat centrálně, místo aby se o ně nezávisle starala každá knihovna sama. Centralizace je často nejlevnější způsob, jak věci zařídit: tím se odstraní

zbytečná duplicita práce. Knihovny tuto situaci reflektují vytvářením konsorcií: kupříkladu finská univerzitní síť Linnea nejenže sdílí společný knihovní systém Voyager, ale také jediný databázový server Sun E10000, na němž zpočátku sídlilo 26 bází dat. Naprosto zásadní je, aby centralizované služby byly budovány ve spolupráci, nikoliv diktovány shora.

Vytváření nových centralizovaných služeb zlepšuje výkonnost a vedení knihoven má možnost přesunout pracovníky na nové úkoly – pokud nejsou snižovány stavy, což by eliminovalo přínosy nových systémů. Takové přesuny zdrojů jsou podstatné, opět kvůli internetu.

Ve srovnání s tradičním vydáváním publikací je internetové vydávání daleko členitější. Místo získání či koupě celé knihy nebo čísla časopisu může uživatel získat přístup nebo si i koupit kapitolu z knihy či článek z časopisu – pokud existují metadata, která mu pomohou relevantní informace najít. Je třeba připustit, že knihovny zpravidla nebyly schopné zvládat články příliš dobře, hlavně pro příliš velký objem dat. Nyní, když internet přímý přístup k článkům nabízí, je tento problém ještě palčivější než dříve.

A co hůře, dokument s tímž intelektuálním obsahem – například tento text – se může na webu objevit na mnoha místech v několika různých podobách.

Vydavatelská obec reaguje na tyto změny různě. Co se týká identifikátorů, vývoj identifikátorů pro složky (jako jsou články a kapitoly knih) a díla (jako je Shakespearův Hamlet) však považuje za věc prvohradé důležitosti.

SICI a BICI

Součásti a příspěvky v seriálu mohou být identifikovány pomocí Serial Item and Contribution Identifier neboli SICI. Naneštěstí tak činí jen velmi málo vydavatelů seriálů; jedním z důvodů tohoto nezájmu může být i to, že SICI je doposud jen americká norma (ANSI Z 39.56), přestože první verze SICI byla publikována již v roce 1991. V době, kdy toto píšete, nemá ISO v plánu SICI standardizovat.

Dalším důvodem nezájmu vydavatelů je složitost SICI. Skládá se z ISSN seriálu, údajů o čísle (segment sešitu), údajů o článku/příspěvku (segment příspěvku) a řídicího segmentu. Článek Marka Needlemana „Computing resources for an online catalogue – 10 years later“, zveřejněný v *Information Technology and Libraries*, ročník 11, číslo 2 (červen 1992), strana 168 a další by dostal SICI: 0730-9295(199206)11:2<168:CRFAOC>2.0.TX;2- #

Uvedený řetěz znaků laikovi mnoho smyslu nedává. Knihovnicko dokáže vyluštit s výjimkou řídicího segmentu, který říká, že tento SICI odpovídá verzi 2.0 standardu a že identifikovaný článek je tištěný text. Pokud by to byl dokument na síti, byl by kód TX nahrazen OC.

Vytváření identifikátorů SICI by nicméně byl problém, jen pokud by nemohly být generovány automaticky z článků resp. z metadat článků. Automatizované vytváření SICI však bylo součástí jeho návrhu a ve světě existuje několik pilotních systémů, které to dělají.

Na opačném konci řetězce využití SICI by použití SICI bylo svízelné, pokud by měly být užívány jinak než jako klikací hypertextové odkazy. Jen málo lidí by se dokázalo naučit tvořit SICI a psát je například do dotazu na katalog rovněž není pohodlné.

Nicméně v dnešním síťovém prostředí není důvod, proč by se SICI nemohl stát velmi populárním nástrojem pro vytváření vazeb mezi odkazem na článek a odkazovaným článkem samotným jako součást rozsáhlejší směrovací služby. Vzhledem k jeho pružné struktuře je rozsah použití SICI ohromný, může identifikovat miliony, ne-li biliony článků. V tomto ohledu jediným problémem současného SICI je, že v některých případech mohou různé elektronické

verze téhož článku dostat stejná čísla SICI. V době, kdy píše tento text, uživatelská příručka ISSN říká, že originál tištěného periodika a jeho digitalizovaná kopie dostane stejné číslo ISSN. Máme-li článek, který byl skenován jednou s rozlišením 600 dpi pro tisk a podruhé 75 dpi pro zobrazení, dostanou obě verze totéž SICI. Věcný obsah obou dokumentů je sice v obou případech stejný (pokud byly všechny detaily zachyceny i při nižším rozlišení), ale užití dokumentu je zcela jiné, a proto by bylo důležité tyto dvě varianty odlišit.

Přesto nevěřím, že hlavním důvodem pomalého přijímání SICI jsou technické problémy. Na rozdíl od ISBN a ISSN zde nejsou žádná mezinárodní centra, která by koordinovala rozvoj a užití identifikátorů SICI, ani žádná národní a regionální centra pečující o rozvoj systému ve svých oblastech. Předpokládá se, že vydavatelé budou vytvářet SICI sami na základě čísel ISSN. Protože neexistuje žádná podpůrná organizace, o existenci SICI ví jen málo vydavatelů. A těm, kteří o něm vědí, se nedostane žádné podpory při vývoji služeb na SICI založených.

Národní centra ISSN by měla vydavatele informovat o SICI a vyhledávat možnosti jejich využívání. Výhledově bude značná část vydávání vědeckých a novinových článků probíhat na webu a národní knihovny budou dostávat články jako dobrovolný nebo povinný výtisk. Pokud budou články obsahovat SICI nebo pokud struktura dokumentu – například XML DTD vyvinuté pro novinové články – umožní automatické generování SICI, bude mít národní knihovna dobré východisko pro vedení elektronického archivu těchto článků. Panuje všeobecná shoda, že identifikátor je jednou z nejdůležitějších částí metadat pro uchování. Pokud tedy články SICI neobsahují a identifikátory nelze zkonstruovat z prvků v článku, musí národní knihovna vytvořit SICI ručně – což je pravděpodobně nemožné – nebo použít jiný identifikátor, jako je číslo národní bibliografie.

Na rozdíl od standardů ISO jsou všechny ANSI/NISO standardy na internetu k dispozici zdarma. Adresa je <http://www.techstreet.com/nisogate.html>. Snadná dosažitelnost textu SICI poněkud usnadňuje úkol informovat vydavatele. Mimo to přidělovat ISSN článkům nemá žádný smysl. To, že by jediné ISSN vyhledalo všechny články kdykoliv publikované v elektronické verzi časopisu Time, by bylo, jak říkají Američané, kontraproduktivní. Vydavatelé nakonec budou elektronické články identifikovat buď SICI, nebo interními identifikátory začleněnými například do směrovací služby DOI. Knihovny by je měly pomoci přesvědčit, že volba založená na SICI je ta správná.

Identifikátor položky a součásti knihy Book Item and Component Identifier neboli BICI se velmi podobá SICI. Neméně BICI je ohrožen ještě více než SICI. Především proto, že BICI zatím není dokončený ANSI/NISO standard, ale až do ledna 2002 pouze návrh. Co se stane poté, až období přezkoumávání skončí, nikdo neví. Experimenty založené na BICI mezi vývojáři software nebyly a nejsou populární.

Nebude-li BICI přijat v nejbližší budoucnosti, povede elektronické obchodování k radikálnímu růstu požadavků na čísla ISBN, protože vydavatelé začnou brzy prodávat části knih jako kapitoly samostatně. Zdá se, že pro identifikaci těchto částí budou pravděpodobně používána čísla ISBN. Zavedení BICI by si vyžádalo zásadní revizi informačních systémů používaných v knižním obchodě, zatímco rozšířené využívání ISBN je možné i se současnými aplikacemi. Američtí vydavatelé se již připravují na elektronický obchod ovládaný ISBN tak, že si rezervují množství čísel ISBN.

Možnosti knihoven jak podpořit využívání BICI jsou v současnosti omezené. Minimum, které můžeme udělat, je prostudovat standard, abychom zjistili, co může udělat pro nás a co pro vydavatele.

Je také důležité poskytnout ANSI/NISO zpětnou vazbu o BICI a zdůraznit, že i když není ještě užíván, existuje velká potřeba možnosti identifikovat obsah knih detailněji, než může nebo by mělo umožnit ISBN.

IDENTIFIKACE TEXTOVÝCH DĚL

ISTC, neboli mezinárodní standardní kód textového díla, je ISO standard v procesu vývoje. Předpokládá se, že návrh komise – první zveřejněná verze textu – bude uvolněna na podzim 2001. Následující popis standardu je založen na verzi textu z února 2001.

Cílem ISTC je umožnit účinnou identifikaci textových děl. ISTC by neměly označovat provedení textových děl; k tomu již existují jiné identifikátory (ISBN, BICI, ISSN a SICI). Tak například původní verze finského národního eposu Kalevala dostane jediný kód ISTC, ale její tištěné verze budou mít mnohá ISBN.

ISTC se skládá ze 16 hexadecimálních číslic označených číslicemi 0–9 a písmeny A–F. Kód má mít následující složky:

- prvek registrační agentury;
- prvek roku;
- prvek díla;
- kontrolní číslice.

Příklad: ISTC 0A9-2002-12B4A105-6

První prvek poskytne prostor pro 4 096 registračních agentur. Registrační Úřad ISTC bude dohlížet na tyto agentury, bude propagovat a koordinovat systém ISTC.

Každá agentura bude moci přidělit miliardu kódů ISTC až do r. 9999. Členové komise věří, že tento počet postačí dokonce i pro internet vzdálené budoucnosti.

Komise ISTC předběžně diskutovala o kritériích kvality pro registrační agentury. Všichni souhlasili, že důležitým kritériem je průkazná schopnost vytvářet metadata pro díla (nebo provedení). To dělá z národních knihoven nejsilnější kandidáty na tuto práci. Ale jak obtížné to bude?

Kalevala dobře ilustruje hlavní problém při vytváření identifikátoru pro díla: jak dílo definovat? Je Kalevala v angličtině totéž dílo jako finský originál? Je Kalevala ilustrovaná Akseli Gallen-Kallelou jiné dílo než první vydání, které nemá obrázky? A co zkrácená verze Kalevaly určená dětem, kterou napsal Elias Lönnrot v padesátých letech 19. století?

Kdyby byla pracovní skupina ISTC složená jen z knihovnických expertů, mohli bychom použít terminologie a principů definovaných ve Funkčních požadavcích na bibliografické záznamy (FRBR) IFLA a v Anglo-amerických katalogizačních pravidlech (jejich dobrý přehled dává Tillett, 2001).

Studie FRBR IFLA definovala entity dílo, vyjádření, provedení a jednotka. Dílo, jako je Kalevala, může být realizováno jedním nebo více než jedním vyjádřením, která mohou být ztělesněna v jednom či více provedeních, od nichž může existovat jediná položka či více položek.

Takže Kalevala je dílo, které je vyjádřeno například ve finštině a v českém překladu; ten měl jedno nebo více provedení (první vydání, další vydání) a české knihovny a milovníci knih mají jednotky těchto vydání na svých poličkách.

Termíny dílo, provedení a jednotka jsou známé a do jisté míry intuitivní. Ale „vyjádření“ bylo zavedeno studií IFLA. Hraniční čára mezi dílem a vyjádřením není vždy jasná, protože jak dílo, tak vyjádření jsou intelektuální anebo umělecké výtvoři myslí. Například překlady románu nebo provedení skladby jsou vyjádření, ale to je zřejmé pouze čtenáři s knihovnickým vzděláním. A pro neknihovníky bude obtížné přijmout detailní analýzu, s jejíž pomocí naše katalogizační pravidla určují, kde vést hranici mezi díly.

Například věrný překlad je pouze vyjádření, ale volný překlad je dílo. To znamená, že například každý překlad Joyceových Plaček nad Finneganem (Finnegan's Wake) je nové dílo, protože jinak než volně se toto pozdní Joyceovo mistrovské dílo přeložit nedá.

Není divu, že ostatní partneři na knižním trhu koncepci vyjádření neschválili. Například model vyvinutý vlivným projektem Interoperability dat pro systémy elektronického obchodu (INDECS) nerozlišuje mezi tak abstraktními pojmy, jako je Smetanova Vltava a její různá provedení. Ve skutečnosti ještě hůře: v modelu INDECS se vyskytuje termín vyjádření, ale označuje událost, která je sama výtvořem (dílem), jako jsou právě jednotlivá provedení Smetanovy Vltavy.

Běžný čtenář by si mohl myslet, že tento konflikt v použití termínu vyjádření zůstane abstraktní a nebude komplikovat denní práci knihovníků. Bohužel, tak tomu asi není.

V komisi vyvíjející standard ISTC jsou lidé jak ze světa INDECS, tak z knihovnického. S takto rozdílným zázemím nelze hovořit o vyjádřeních, protože tento termín má naprosto odlišný význam podle toho, kdo hovoří. Je také nemožné přijmout kterýkoliv z existujících modelů pro definici díla tak, jak je. Nicméně by mělo být možné vytvořit dokument, který by nebyl v rozporu ani s FRBR IFLA, ani s dalšími běžnými a široce užívanými modely.

Všechny strany v komisi ISTC souhlasí, že ISTC by měl být přidělen všem odlišným výtvořům myslí, to jest dílům a jejich vyjádřením (ve významu jak tato slova užívá IFLA). Český překlad Kalevaly tedy dostane někdy v budoucnu svůj ISTC. K němu a ke každému jinému ISTC se váží metadata; nejdůležitějším prvkem metadata jsou propojení k dalším vyjádřením téhož díla a případně i k jiným dílům, která patří do téže rodiny děl (jako jsou všechna provedení Vltavy nebo všechny překlady Kalevaly). Jsou tu i problematické oblasti: například kompilace jsou někdy svěbytnými díly, jindy mohou být pouhými sbírkami existujících děl, a jako takové si vlastní ISTC nezaslouží.

Podrobnosti týkající se definice díla nebudou uvedeny ve standardu ani v jeho přílohách, ale v uživatelské příručce, kterou lze později změnit snáze než standard samotný. To usnadní úpravy textu podle potřeb uživatelů. Cílem je vyhnout se příliš striktní specifikaci, jak má být ISTC užíván, protože to by mohlo bránit užitím, která budou dávat smysl až za několik let.

Z hlediska knihoven je ISTC velkou výzvou. Každý ISTC musí být doprovázen metadatami, a protože ISTC bude jedním ze základních stavebních bloků budoucího systému elektronického obchodu, po ISTC s odpovídajícími metadatami bude velká poptávka.

Naše současné systémy však podporují pouze popisy provedení (a položek). Jen několik dodavatelů začalo s vývojem integrovaných knihovních systémů podporujících katalogizaci děl. Zatím nikdo neví, zda by bylo možné dostatečně přesně odvodit popisy děl z existujících databází. Pokud ne, retrospektivní katalogizace děl v širokém měřítku by byla pro nedostatek pracovníků nemožná; celý proces by probíhal postupně a trval by desetiletí.

A aby se věci ještě zkomplikovaly, vznikají další standardy ISO pro identifikaci děl. Jsou to ISAN (Mezinárodní standardní audiovizuální číslo) a ISWC (Mezinárodní standardní kód hudebního díla). Žádný z těchto systémů není zatím dokončen, ale ISAN je již ve fázi finálního návrhu mezinárodního standardu (FDIS), zatímco ISWC dosáhl pouze statutu návrhu mezinárodního standardu DIS. Přestože jsou ISAN, ISWC a ISTC připravovány více-méně ve stejné době, v komisích zasedají většinou různí lidé s různým zázemím. To pak v praxi způsobilo, že standardy – alespoň v přípravné fázi – nebyly vzájemně sladěny. Budou-li mít tyto problémy i konečné verze, může výsledný systém trpět nesourodstím.

ISO si to uvědomila a nedávno zřídila skupinu, která bude analyzovat všechny identifikační standardy ISO, aby se ukázalo, jaká míra souladu mezi těmito standardy je žádoucí a možná. Při pomalém vývoji ISO standardů bude trvat nějakou dobu, než se výsledky této práce projeví v praxi. Tento druh koordinace je však velmi důležitý v situaci, kdy identifikačních systémů rychle přibývá. A může se docela dobře stát, že jednotlivá organizace, jako je národní knihovna anebo velký vydavatel, se budou muset vyrovnat s mnoha systémy souběžně.

Uvážíme-li ohromné množství metadata, které nové identifikační systémy vyžadují, je zřejmé, že knihovny, vydavatelé i autoři musí spojit své síly, aby toho dosáhli. Články nepůjde podrobně katalogizovat, pokud lidé, kteří je píšou, knihovníkům tuto práci neusnadní. Ustavení těchto kooperací napříč organizacemi bude zajímavou výzvou jak technickou, tak ideovou. Dlouhou dobu považovali knihovníci katalogizaci za svou doménu. Dovolit „amatérům“, aby dělali tuto práci, nebude pro profesionály lehké. Na druhé straně z uskutečněných pokusů víme, že autoři se na to, až budou moci své články vybavovat metadatami, příliš neteší.

SMĚROVACÍ SLUŽBY

V tradiční knihovně bylo nalezení zdroje snadný úkol, pokud byla kniha nebo číslo časopisu správně zařazena. Klient si našel signaturu v lístkovém katalogu nebo v záznamu OPAC, došel k udanému místu na polici a knihu si vyzvedl.

U elektronických zdrojů je situace daleko složitější. Systémy, které propojují zdroj s jeho popisem nebo zdroj přímo s jiným zdrojem, nazýváme směrovací služby [v originále „resolution services“ – pozn. překl.]. V této chvíli je vývoj těchto služeb překotně se rozvíjející oblastí automatizace knihoven i internetu samotného.

Směrovací služby lze zhruba rozdělit na statické a dynamické. URL a protokol http jsou dokonalem příkladem statické služby, která pracuje dobře, dokud dokument zůstává na stejném místě a uživatel má právo zdroj získat. Pokud se umístění změní, neexistuje žádný snadný způsob, jak zdroj na webu opět nalézt, pokud o změně nebyla informována doménová služba jmen (DNS). Na druhé straně, každý uživatel obvykle odkaz vidí, a to i tehdy, když nemá právo dokument získat.

Na dynamickou – a účinnou – směrovací službu pro elektronické zdroje jsou tedy dva požadavky. Musí být schopna přizpůsobit se změnám umístění (URL) a musí být schopna reflektovat uživatelskou oprávnění daná jeho identitou a/nebo jeho místem na síti. Dobrý systém by měl být také schopen poskytovat individuální informační služby založené na profilu zájmu uživatele. A konečně by systém měl být dobře integrován do infrastruktury internetu.

Budeme zkoumat následující směrovací služby:

- Identifikátory digitálních objektů (DOI) a Ovládací systém;
- Jednotná jména zdrojů (URN);
- OpenURL a SFX.

DOI a URN jsou všeobecně známé jako identifikátory elektronických zdrojů. Tím ale nejsou tyto systémy zajímavé, zajímavé jsou službami, které mohou poskytovat.

DOI a Odkazový systém

Iniciativa DOI (<http://www.doi.org>) začala v roce 1996. Hlavní síly stojící za tímto projektem byly Corporation of National Research Initiatives (CNRI, <http://www.cnri.reston.va.us/>) a americký vydavatel R. R. Bowker, který řídí přidělování ISBN v USA. V současnosti je tato iniciativa řízena Mezinárodní nadací DOI (IDF), která je ovládána velkými vydavateli.

Syntaxe DOI specifikuje ANSI/NISO Z39.84-2000. Podle tohoto standardu se DOI skládá z předpony a přípony oddělených lomítkem.

Příklad: 10.153/34571

Předpona je rozdělena do kódu adresáře (jediná v současnosti povolená hodnota je 10) a kódu přihlašovatele, které jsou odděleny tečkou. Kód přihlašovatele obvykle, ale ne vždy, odkazuje na vydavatele. Na podzim r. 2001 mělo kód přihlašovatele, který stojí 1 000 USD, více než 150 vydavatelů.

Přípona obsahuje identifikátor digitálního objektu. Protože předpona zaručuje, že DOI jsou unikátní, pokud sám vydavatel nepřidělí týž identifikátor dvakrát, nejsou nutná žádná pravidla určující, jaké identifikační systémy jsou přípustné; každý přihlašovatel, který má předponu DOI, může použít jakýkoliv systém, třeba vlastní vnitřní.

Směrování identifikátorů DOI je založeno na ovládacím systému (Handle System; <http://www.handle.net>), který vyvinula CNRI. Současná verze ovládacího systému je založena na protokolu http. To v praxi znamená, že každý DOI je zabudován do dokumentu jako hyperlink odkazující na URL <http://dx.doi.org>. Je-li DOI 10.1006/rwei.1999.0001, je skutečná podoba odkazu <http://dx.doi.org/10.1006/rwei.1999.0001>.

Výhodou tohoto řešení je, že funguje i v současném internetu. První služby DOI začaly fungovat v r. 1997 a rozsah systému se stále zvětšuje. Prostřednictvím směrovací služby DOI je přístupný velký počet vědeckých článků.

Hlavní slabinou Odkazového systému je, že pokusy o jeho standardizaci v internetové obci neuspěly z technických důvodů. To by mohlo být kvůli nějaké zásadní technické slabině (z hlediska Skupiny pro návrh internetu IETF) odkazové technologie. To však považují za nepravděpodobné. Daleko prostší vysvětlení je, že IETF chce bránit přemnožení směrovacích služeb tohoto typu, a proto podporuje svoji vlastní iniciativu URN, která zahrnuje obojí, jak identifikátory typu DOI, tak i směrovací služby podobné Odkazovému systému.

Současná verze DOI zaručuje pouze trvanlivost odkazů. Existují však ambiciózní plány dalšího rozvoje systému. Podle Priscilly Caplan (Caplan, 1998):

Záměrem iniciativy DOI je samozřejmě více než pouhé zajištění trvanlivosti. Pokud by to bylo všechno, oč jim šlo, byli by mohli vydavatelé s daleko menšími obtížemi použít PURL server. Mezinárodní nadace DOI hodlá vybudovat kompaktní systém pro řízení oprávnění a má pracovní skupiny aktivně se zabývající některými z těchto aspektů, včetně politiky, aplikací, deskriptivních metadat a metadat pro řízení oprávnění.

Iniciativa CrossRef (<http://www.crossref.org>) je nástrojem, jímž konsorcium DOI směřuje k vytvoření uvedeného systému.

V květnu 2001 bylo 70 vydavatelů – asi polovina z těch vydavatelů, kteří získali Registrační kód DOI – zapojeno do CrossRef. V databázi CrossRef bylo přes 3 miliony záznamů článků z 3 875 časopisů; předpokládá se, že systém poroste o přibližně 500 000 až 1 000 000 záznamů ročně (Brand).¹⁾ Záznamy a vlastní články jsou samozřejmě propojeny identifikátory DOI, které jsou součástí metadat dodaných vydavatelí.

Ve své současné podobě neposkytuje CrossRef kromě trvanlivého propojení odkazů a článků mnoho dalších speciálních funkcí. Ke stejnému účelu mohlo konsorcium CrossRef využít PURL. A pokud jde o pokročilé služby, jako je dynamické propojování (viz níže), je DOI systém dnes technicky na stejné úrovni, jako byl v r. 1998. CNRI zkoumá rozšíření Odkazového systému s cílem umožnit dynamické propojování, ale tato iniciativa nevyšla z práce DOI, ale z projektu OpenURL, jak uvidíme dále.

Jak úspěšně bude Odkazový projekt integrovat OpenURL a další případné vnější vlivy a jak iniciativa DOI využije tyto technologie k rozšíření služeb CrossRef, to teprve ukáže čas. Dokument CrossRef Frequently Asked Questions (Otázky často kladené CrossRef – <http://www.crossref.org/faqs.html>) ale příliš neosvětluje, jaké služby by měly být vyvíjeny.

Samozřejmě je možné, že takové rozšíření odkazového systému vůbec není nutné. Vedle sebe může existovat řada směrovacích služeb; zdroj může být dosažen jak prostřednictvím Odkazového serveru, tak SFX serveru (viz ukázkou od Herberta van de Sompela a jiných na <http://www.sfxit.com/crossref/prototype1.html>). Ale pokud SFX server dosáhne na všechno, co může obstarat Odkazový server, ba i dále, je ještě vůbec třeba Odkazový systém udržovat?

Z hlediska budování rozsáhlých směrovacích služeb je DOI dobrým identifikačním systémem, neboť předpona DOI říká, kde nalézt příslušnou směrovací službu. Tak by bylo možné použít identifikátorů DOI jako URN a vytvořit směrovací služby pro prostor jmen DOI v rámci systému URN. V současné době toho není zapotřebí: CNRI udržuje infrastrukturu (server nebo skupina serverů na adrese <http://dx.doi.org>), která identifikátory DOI směřuje na URL, ať už v rámci služby CrossRef nebo mimo ni.

Bohužel, navzdory svým technickým zásluhám, se DOI nikdy nestane skutečně typickým řešením pro všechny druhy internetových dokumentů. To má jak technické, tak ekonomické důvody.

Generovat DOI retrospektivně z dnešních identifikátorů je nemožné ze dvou důvodů. Především přidělení DOI není zadarmo. Přidělit identifikátory DOI retrospektivně k např. celé české a slovenské národní bibliografii by bylo dosti drahé, a užitečnost takového počínání by byla omezená, protože jen zlomek materiálů je komerčně zajímavý. Totéž platí pro kteroukoliv národní bibliografii. Na druhé straně, generovat DOI retrospektivně ani není možné, pokud nejsou určeny kódy přihlašovatelů. A pokud bychom nezvolili triviální řešení, že přihlašovatelem bude národní knihovna, prostě nemáme kódy přihlašovatele pro všechny vydavatele. Vytvoření těchto kódů pro všechny vydavatele by byla jednak ohromná práce, a pak ne všichni vydavatelé by byli ochotni za přihlašovací kód DOI platit.

Při komerční povaze systému DOI je pravděpodobné, že se systém DOI bude používat pouze pro doručování komerčního obsahu, jako jsou vědecké články. Zda systém přežije, či nikoliv, bude záviset na vytvoření služeb s přidanou hodnotou, o nichž se zmiňuje Caplanová. Dnes se zdá, že Odkazový systém zaostává za novými konkurenty, jako je SFX, ale jeho revize už probíhá. Zda CNRI dokáže vylepšit systém dostatečně rychle, ukáže blízká budoucnost.

Jednotná zdrojová jména (Uniform Resource Names)

Internetová inženýrská skupina (<http://www.ietf.org>) zahájila projekt URN krátce potom, kdy Mozaika, grafický prohlížeč založený na Windows, zpopularizoval web. V r. 1994 projekt publikoval všeobecné požadavky pro URN v RFC 1737:

- globální rozsah: jméno nesouvisí s umístěním, jeho význam je všude stejný;
- jedinečnost: dvěma různým zdrojům nesmí být dáno stejné jméno;
- trvalost: jméno musí zůstat stejné navždy;
- dostupnost: jméno může být přiděleno jakémukoliv možnému zdroji;
- podpora minulých forem: jméno musí být schopno podporovat bývalá pravidla přidělování jmen, pokud lze na tato pravidla aplikovat ostatní požadavky;
- rozšířitelnost: schéma přidělování jmen musí být možné rozšiřovat;

- nezávislost: organizace odpovídající za jména musí být zcela nezávislá.

Tyto směrnice také přijala iniciativa DOI. Ukázalo se, že jejich dodržení není triviální úkol; první pracovní skupina IETF pro URN se rozešla, aniž dosáhla svých cílů.

Práce byla znovu zahájena v r. 1996 a druhá pracovní skupina byla daleko úspěšnější: na jaře r. 2001 byla práce prakticky ukončena. Vyžádalo si to mnohem více času, než se očekávalo, ale tento odklad měl velice dobré důvody, jak ukážeme později.

Pracovní skupina pro URN (<http://www.ietf.org/html.charters/OLD/urn-charter.html>) definovala první syntaxi URN. Podle RFC 2141 (<http://www.ietf.org/rfc/rfc2141.txt>) se URN skládá ze tří částí:

- řetěz znaků „urn“. Tato předpona je nutná pro to, aby bylo možné nalézt a indexovat jména URN z nestrukturovaných internetových zdrojů, v nichž není čím přítomnost URN indikovat (jako např. užitím prvku Dublin Core Identifier);
- identifikátoru prostoru jmen (NID), který jednoznačně určuje použitý identifikační systém;
- řetězec specifický pro prostor jmen (NSS), který obsahuje vlastní identifikátor.

Každý prostor jmen daný NID musí být registrován internetovým standardem. Tento proces a povinné údaje žadatelů definuje RFC 2611 (<http://www.ietf.org/rfc/rfc2611.txt>). Mezinárodní centrum ISSN zaregistrovalo NID „ISSN“ pro Mezinárodní standardní číslo seriálů (<http://www.ietf.org/rfc/rfc3044.txt>). Registrace NID pro Mezinárodní standardní číslo knihy (NID „ISBN“) a pro čísla národních bibliografií (NID „NBN“) budou schváleny Agenturou pro přidělování internetových jmen (IANA) během léta 2001. O registraci NID pro SICI bude IETF požádána v létě 2001.

Proces registrace těchto prostorů jmen (namespaces) potvrdil předběžné analýzy provedené v RFC 2288 (<http://www.ietf.org/rfc/rfc2280.txt>): nejdůležitější identifikátory používané knihovnami mohou být použity jako jednotná zdrojová jména.

Podle RFC 3044 má URN založené na ISSN velmi jednoduchou syntaxi:

Urn:issn:<číslo issn> například urn:issn:1560-1560

Jednoduchá syntaxe jmen URN (spolu s tím, že jsou zdarma) umožňuje generovat URN automaticky z existujících identifikátorů. Dokonce ani není nutné ukládat jména URN do databáze. Pokud požadavek směřovat URN přijde na OPAC, systém prostě odstraní předponu urn:, zjistí NID a dotazem na příslušný index zjistí, zda „zná“ identifikátor obsažený v NSS části URN. Pokud je odpověď kladná, může knihovní systém požadovanou informaci – bibliografická data nebo samotný dokument – poskytnout.

Standardizací registrace prostorů jmen může internetová obec zajistit, aby byly plněny požadavky RFC 1737. Tím se výrazně liší od iniciativy DOI, kde v zásadě každý, kdo má tisíc dolarů, může získat registrační kód DOI a použít identifikační systém, jaký chce. V systému URN musí žadatelé popsat sami sebe, své identifikátory a techniku, kterou bude vytvořen směrovací mechanismus. Současná realizace systému DOI je velmi centralizovaná, kdežto systém URN je volnější: jeden prostor jmen – například prostor jmen ISSN – může být založen na centralizované službě, kdežto jiné, jako prostor jmen ISBN, jsou decentralizované.

Služby systému URN definuje RFC 2483 (<http://www.ietf.org/rfc/rfc2483.txt>). Uživatel může požadovat bibliografické informace o zdroji, o jeho umístění (URL) nebo zdroj samotný. Aby nebyl uživatel zmaten, měl by být jeho interface natolik inteligentní, aby nedostupné služby skryl. V kontrastu k tomu současná verze ovládacího systému DOI poskytuje pouze jedinou službu na

identifikátor, takže pro dodání dokumentu a jeho popisu je třeba dvou identifikátorů DOI.

V současném internetu bude směrování podle jmen založeno na doménové službě jmen (DNS). Představa je taková, aby uživatel mohl napsat URN do okna pro adresu svého prohlížeče nebo jiného internetového klienta, stejně jako tam dnes píše adresu URL. Jakmile uživatel stiskne klávesu enter, DNS služba najde zdroj podle metadat uložených v DNS systému.

Na adrese urn.net bude služba určení směrovačů URN (Resolver Discovery Service) a server doménové služby jmen (DNS), který – na základě údajů při registraci a z dodatečných informací poskytnutých provozovatelem [příslušné směrovací služby – pozn. překl.] – zná umístění všech směrovacích služeb URN v internetu. Například urn.net server bude znát adresu databáze ISSN v doméně issn.org. Informace o směrovacích službách URN se pak šíří do dalších služeb jmen podle běžných pravidel DNS. Takto koncipovaný systém nemá žádnou slabinu a nejsou žádná technická omezení pro počet možných směrovacích služeb URN.

Metadata potřebná pro směrování podle URN jsou uložena v záznamech DNS, které normálně propojují internetové adresy s IP adresami (jako je www.helsinki.fi a 128.214.4.1), mohou ale poskytovat i složitější informace. Nejběžnější DNS aplikace BIND směrování podle URN podporuje již dnes. Budoucí verze služeb URN, které mohou potřebovat složitější metadata pro dynamické služby, si možná vyžádají úpravu systémů DNS. Prozatím neexistují plány na revizi služeb URN, ale technické principy systému to snadno umožní.

Přestože je infrastruktura URN hotová, zatím příliš mnoho služeb DNS neexistuje. V knihovnické oblasti byla první směrovací služba vytvořena Mezinárodním centrem ISSN (viz. <http://www.issn.org>), v době, kdy píše tuto přednášku, však dodavatelé knihovních systémů zatím nevytvořili služby založené na URN. To má dva hlavní důvody.

Jednak je pro konečné uživatele podstatné, aby podpora směrování podle URN byla vestavěná přímo do webovských prohlížečů. V současné době to dovede pouze Microsoft Internet Explorer, bohužel směrování podle URN v provedení Internet Exploreru neodpovídá tak docela internetovým standardům.

Za druhé, přestože jména URN samotná jsou zdarma, zřízení a údržba směrovací služby už zadarmo není. Někdo musí účet zaplatit, ale dobrovolníci se nehrnou. Nicméně národní knihovny samy se zavázaly, že budou iniciativu URN podporovat, což posléze dá vzniknout službám založeným na URN. Tyto služby mohou nebo nemusí být zdarma; podstatné je, že pro propojování bibliografických záznamů s jimi popsány elektronickými zdroji musíme používat lepší nástroj, než je URL.

To si vyžádá specifikaci nových prvků metadat nejméně ve dvou oblastech: v oblasti dlouhodobého uchovávání elektronických dokumentů a v oblasti dynamického propojování. V obou oblastech je specifikace těchto prvků dobře rozjetá v některých komunitách, dá ale ještě hodně práce dohodnout se, které prvky jsou opravdu nezbytné. A až se to stane, bude třeba tyto prvky metadat integrovat do katalogizačních pravidel, do formátů užívaných knihovnami a do těch integrovaných knihovnických systémů, které přežijí do éry digitálních knihoven.

OpenURL a SFX

URN a DOI jsou příklady směrovacích služeb, které ve své současné podobě poskytují trvanlivá propojení – pokud jsou trvanlivé propojené elektronické zdroje samy – ale technický princip, na němž jsou postaveny (jako je Odkazový systém), není zatím schopen poskytnout dynamické propojení mezi zdroji. Oba identifikátory

mohou nieméně sloužit jako základ pro poskytování dynamického propojování. Pro DOI to prokázal Van de Sompel se spolupracovníky; pro URN to zatím nikdo nepředvedl.

Van de Sompel (1999) poukazuje na to, že vytvoření účinných propojovacích služeb je velmi důležitá služba s přidanou hodnotou na webu a že vydavatelé, agentury pro předplatné, výzkumníci i knihovny, ti všichni vyvíjejí nástroje pro propojování souvisejících informačních entit. Většina propojovacích systémů je zatím statická: to jest zachycují vztahy mezi entitami přístupnými v kontrolovaném prostředí. Například záznamy v národní bibliografii mohou obsahovat katalogizátorem vytvořené odkazy URL na elektronické zdroje uložené v depozitu. Při pohledu do národní bibliografické databáze uvidí uživatel vždy stejný odkaz URL, který byl „navrtdo zapsán“ do pole 856.

Informační zdroj, který by chtěl poskytovat dynamické propojování v rámci podobným jako SFX – podle popisu syntaxe OpenURL (<http://www.sfxit.com/openurl/openurl.html>) – musí:

- zavést techniku, která zdroji umožní rozlišit mezi uživatelem, který smí využít servisní složku poskytující kontextově vázanou službu, a uživatelem, který toto právo nemá;
- uživateli oprávněnému využít servisní složku poskytnout OpenURL pro každý objekt metadat.

Narozdíl od statického je dynamické propojování založeno na předpokladu, že ne všechna data, potřebná pro vytvoření vzájemně propojeného informačního prostředí, mohou být v moci orgánu, který toto prostředí vytváří, jako je katalogizátor. V dynamickém systému se přinejmenším některá propojení vytvářejí za pochodu, podle toho, jak uživatel klade dotazy, na základě jeho požadavků a podle informací obsažených v systému.

Van de Sompel velice zajímavě poznamenává, že dynamické systémy budou spíše vytvářet knihovny a jiné nekomerční organizace, protože vydavatelé nemají tradici spolupráce potřebné pro budování takových vzájemně výhodných služeb. Zda budou knihovny skutečně úspěšné, se teprve uvidí, ale současné výsledky jsou slibné.

OpenURL je syntaxe, která umožňuje převod metadat z informační služby do servisní složky poskytující kontextově vázané služby pro převedení metadatu. Jednou takovou servisní složkou je SFX, v budoucnu mohou vzniknout další. Jejich společným jmenovatelem je schopnost zpracovat OpenURL. Existuje již řada systémů OpenURL akceptujících. Je pravděpodobné, že tuto schopnost bude mít čím dále tím více systémů, protože v lednu 2001 ANSI/NISO zahájila standardizaci OpenURL.

Standard Z39.50 ANSI/NISO pro vyhledávání informací také umožňuje převod metadat mezi dvěma systémy. Mezi těmito dvěma protokoly jsou však významné rozdíly. OpenURL využívá protokol http užívaný na webu, kdežto Z39.50 definuje vlastní protokol. Většina existujících systémů Z39.50 dodává záznamy MARC v syntaxi ISO 2709 (tak zvaný výměnný formát), zatímco OpenURL je vložen do GET požadavku http. Hypotetický příklad OpenURL: <http://sfxserver.uni.edu/sfxmenu?issn=1234-5678&date=1998&volume=12&issue=2&page=134>

Z39.50 má velký výběr vyhledávacích termínů, definovaných v sadě atributů Bib-1 a v dalších sadách. Současný OpenURL má podstatně menší sadu vyhledávacích termínů, jsou to zejména takové, které jsou nutné pro nalezení určitého čísla časopisu a článku v něm nebo kapitoly v knize.

Uvážíme-li rozdíly mezi Z39.50 a OpenURL, je jasné, že snazší je implementace OpenURL. Ale oba protokoly jsou stále relevantní; OpenURL neposkytuje účinnou službu vyhledávání informací nebo přenosu záznamů MARC. To také nikdy nebylo záměrem vývoje tohoto systému. OpenURL a aplikace, které ho používají, jako je

SFX, umožňují propojování souvisejících zdrojů novými způsoby. V příštích letech uvidíme, jak úspěšné toto úsilí bude.

CESTA VPŘED

Zájem o OpenURL je dobrý začátek, ale integrované knihovní systémy budou ještě potřebovat řadu dalších funkcí, aby se staly částí sémantického webu, jehož principy zveřejnili Berners-Lee, Hendler a Lassila. Dynamická propojení a spolupráce mezi různými webovskými aplikacemi předvídané těmito vizionáři možná budou v knihovních systémech založeny na pokročilých verzích protokolu typu OpenURL a na důmyslných metadatech. Ale je zřejmé, že mezera mezi dynamickým propojováním aplikací poskytovaných tandemem OpenURL/SFX a Sémantickým webem je ještě velká.

Ačkoliv úplně základní stavební bloky Sémantického webu – XML a RDF²⁾ – jsou už nějakou dobu k dispozici, jiné podstatné věci stále chybí. Neexistují ontologie, jejichž pomocí by počítače mohly správně porozumět významu sémantických dat na webové stránce. Neexistují účinní agenti – aplikace, které shromažďují, filtrují a zpracovávají informace nalezené na webu bez přímé lidské kontroly nebo dohledu.

Možná se podaří vyvinout skutečně funkční ontologie – soustavy tvrzení v jazyce srozumitelném strojům – které definují vztahy mezi pojmy a specifikují logická pravidla pro jejich dokazování a odvozování. A možná se podaří vyvinout účinné agenty, kteří člověku pomohou při vyhledávání informací. Ale ani přes tento pokrok nebudou knihovny zbytečné. Ve skutečnosti si myslím, že knihovny by mohly být mezi prvními a nejsikovnějšími uživateli těchto nástrojů, jimiž mohou své sbírky snáze zpřístupnit.

Poskytovat informace uživatelům je součástí našeho poslání. Technické prostředky pro to se v průběhu staletí měnily, ale důležitost popisu zdroje zůstává stejná. Metadata budou mít životně důležitou roli v sémantickém webu, kde musí být data srozumitelná jak lidem, tak jim sloužícím aplikacím. Autoři, vydavatelé a knihovníci musí spojit síly, aby zajistili správný popis užitečného obsahu webu. Díky našim dlouhodobým zkušenostem s popisem zdrojů budou knihovny hrát v tomto procesu klíčovou roli. Ačkoliv automatické indexování poskytuje stále lepší výsledky, lidé jsou a ještě dlouhou dobu zůstanou lepší než počítač v obsahové analýze dokumentu a v popisu zdrojů pro ostatní lidi.

ODKAZY

- Berners-Lee, Tim & Hendler, James & Lassila, Ora: The semantic Web. *Scientific American*, sv. 284 č. 5 (květen 2001) str. 28–37.
- Brand, Amy: CrossRef turns one. *D-Lib Magazine*, sv. 7 č. 8 (květen 2001). Elektronický zdroj dostupný z <http://www.dlib.org/dlib/may01/brand/05brand.html>
- Caplan, Priscilla: DOI or don't we? *The Public Access Computer Systems Review*, sv. 9, č. 1 (1998). Elektronický zdroj dostupný z <http://info.lib.uh.edu/pr/v9/n1/capl9n1.html>
- Snijder, Ronald: Metadata standards and information analysis: a survey of current metadata standards and the underlying models. Verze z 1. května 2001. Elektronický zdroj dostupný z <http://www.geocities.com/ronaldsnijder/>. Referred at 2. 5. 2001.
- Sollins, Karen & Masinter, Larry: Functional requirements for Uniform resource names. RFC 1737. Elektronický zdroj dostupný z <http://www.ietf.org/rfc/rfc1737.txt>.

- Tillett, Barbara: Bibliographic relationships. In: C. A. Bean & R. Green: Relationships in the Organization of Knowledge, 19–35. 2001
- Van de Sompel, Herbert & Hochstenbach, Patrick: Reference linking in a hybrid library environment. Part 1: Frameworks for linking. D-Lib Magazine, sv. 5, č. 4 (duben 1999). Elektronický zdroj dostupný z http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html

PŘEHLED POUŽÍVANÝCH ZKRATEK:

ANSI	American National Standards Institute Americký národní úřad pro normalizaci
AACR	Anglo-American Cataloguing Rules Anglo-americká katalogizační pravidla
BICI	Book Item and Component Identifier Identifikátor položek a součástí knih (např. pro knižní kapitoly)
CNRI	Corporation of National Research Initiatives Korporace národních výzkumných iniciativ
DIS	Draft International Standard Návrh mezinárodního standardu
DNS	Domain Name Service Doménová služba jmen
DOI	Digital Object Identifiers Identifikátor digitálního objektu
EAN	European Article Numbering Evropské číslování zboží
FDIS	Final Draft International Standard Finální návrh mezinárodního standardu
FRBR	Functional Requirements for Bibliographic Records Funkční požadavky na bibliografické záznamy
IANA	Internet Assigned Names Authority Agentura pro přidělování internetových jmen
IDF	International DOI Foundation Mezinárodní nadace DOI
IETF	Internet Engineering Task Force Skupina pro návrh Internetu

IFLA	International Federation of Library Associations Mezinárodní federace knihovnických asociací
INDECS	Interoperability of Data for Electronic Commerce Systems Interoperabilita dat pro systémy elektronického obchodu
IP address	Internet Protocol address IP adresa – daná čtveřicí čísel 0-255
ISADN	International Standard Authority Data Number Mezinárodní standardní číslo autority
ISAN	International Standard Audiovisual Number Mezinárodní standardní audiovizuální číslo
ISBN	International Standard Book Number Mezinárodní standardní číslo knihy
ISO	International Organization for Standardization Mezinárodní organizace pro standardizaci
ISSN	International Standard Serial Number Mezinárodní standardní číslo seriálu
ISTC	International Standard Textual Work Code Mezinárodní standardní kód textového díla
ISWC	International Standard Musical Work Code Mezinárodní standardní kód hudebního díla
MARC	Machine-Readable Cataloguing Strojem čitelná katalogizace
NID	Namespace Identifier Identifikátor prostoru jmen
NSS	Namespace Specific String Řetězec specifický pro prostor jmen
OPAC	Online Public Access Catalogue On-line veřejně přístupný katalog
RDF	Resource Description Framework Rámec popisu zdroje
	Registrant Code Kód přihlašovatele
SICI	Serial Item and Contribution Identifier Identifikátor součástí a příspěvků seriálů
URL	Uniform Resource Locator Jednotná adresa zdroje
URN	Uniform Resource Names Jednotné jméno zdroje

Juha Hakala, Knihovna Univerzity Helsinky (Helsinki University Library)

e-mail: juha.hakala@helsinki.fi

POZNÁMKY:

- ¹⁾ V říjnu 2003 bylo v databázi přes 9 milionů záznamů z 8 563 časopisů 250 vydavatelů.
- ²⁾ Resource Description Framework – rámec popisu zdroje.

Příspěvek byl přednesen na konferenci CASLIN 2001, Beroun, Česká republika, květen 2001 a je publikován s laskavým souhlasem autora. Anglická verze příspěvku je dostupná na <http://www.caslin.cz:7777/caslin01/sbornik/hakala.html>.