

STANOU SE ONLINE DOSTUPNÉ ELEKTRONICKÉ ZDROJE INTEGROVANOU SOUČÁSTÍ DIGITÁLNÍCH KNIHOVEN?

Ludmila Celbová
Národní knihovna ČR

Úvod

V rámci programu Ministerstva kultury ČR *Zpřístupňování a ochrana knihovních fondů formou digitalizace, s využitím sítě Internet v souvislosti s vytvářením informační společnosti*, tedy programu zaměřeného na výzkum a vývoj získala Národní knihovna ČR grant také na řešení projektu *Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet* (řešení v letech 2000–2001). Cílem tohoto pilotního projektu je připravit podmínky pro zpracování české národní bibliografie elektronických zdrojů se zaměřením zejména na zdroje vzdáleně přístupné. S bibliografickým zpracováním souvisí zajištění trvalého uchování domácích elektronických monografických i seriálových dokumentů publikovaných v síti Internet a jejich zpřístupnění, respektující autorské právo vydavatelů.

Proč se elektronickými zdroji dostupnými online zabývat?

Exploze v elektronickém publikování, tj. zejména v tvorbě elektronických zdrojů přístupných v síti Internet, vyžaduje nový přístup ke zpracování, ochraně a zpřístupňování těchto informací. Vzdáleně přístupné elektronické zdroje se stávají nedílnou součástí národní produkce a národního kulturního dědictví, i „obrazem doby“, který je třeba zachytit a uchovat pro budoucnost. Při budování digitálních knihoven, které se stávají globálními integrovanými systémy umožňujícími uživatelům přístup k dokumentům bez ohledu na jejich typ či lokalitu, se vedle digitalizovaných dokumentů začíná věnovat pozornost rovněž dokumentům, které jsou publikovány na Internetu a existují často pouze v elektronické podobě.

Vzhledem k tomu, že v případě internetových elektronických zdrojů se jedná o nehmotné dokumenty, vůči nimž navíc je zpracovatelská instituce v externím postavení, je zajištění jejich pokud možno trvalého zpřístupnění podmíněno řešením řady problémů, z nichž některé jsou společné s tvorbou/zpřístupňováním tradičních dokumentů převáděných do digitální formy či digitálních dokumentů uložených na hmotných nosičích (např. digitální zvukové záznamy), některé jsou zcela specifické pro tuto kategorii digitálních dokumentů. Při procesu integrace internetových elektronických zdrojů do fondů knihoven vystupují do

popředí technické, knihovnické, organizační a legislativní aspekty, které lze stěží od sebe oddělit.

V zásadě je třeba řešit několik okruhů problémů:

- ♦ kritéria výběru (obsahová a formální) těchto zdrojů pro registraci a archivaci v souladu s politikou budování fondů depozitních knihoven (součást národní bibliografie – konzervační fond)
- ♦ legislativní opora získávání elektronických zdrojů pro konzervační fond (zahrnutí elektronických zdrojů do kategorie dokumentů, na něž se vztahuje povinný výtisk)
- ♦ bibliografická kontrola elektronických zdrojů (metadatové standardy)
- ♦ zajištění trvalého přístupu k vybraným publikacím uloženým v elektronickém archivu a s tím související otázka jejich autenticity (přidělování jednoznačných identifikátorů, např. URN)
- ♦ strategie archivace digitálních dokumentů jako nezbytného předpokladu pro jejich dlouhodobé zpřístupnění
- ♦ stanovení podmínek přístupu k archivovaným elektronickým zdrojům (autorské právo).

Ve světě (zejména v USA, Kanadě, Austrálii a v evropských severovýchodních zemích) existuje již několikaletá zkušenost s projekty zaměřenými na sbírky elektronických dokumentů publikovaných v síti Internet. V Evropě se za podpory Evropské komise zabývají touto problematikou společně mezinárodní projekty (**CoBRA+** [1], **BIBLINK** [2], **NEDLIB** [11]), jejichž cílem je stanovit budoucí úlohu evropských národních knihoven ve vztahu k elektronickým publikacím a vytvořit podmínky pro propojení národních bibliografických agentur a vydavatelů elektronických publikací, které by bylo užitečné pro obě strany. K nejvýznamnějším národním projektům patří **PANDORA** (Austrálie) [10], **EVA** (Finsko) [3], **INDOREG** (Dánsko) [4], **Kulturarw3** (Švédsko) [5] a **DNEP** (Nizozemí) [6]. Ze zemí bývalého východního bloku zatím žádná s řešením obdobného projektu nezačala.

Řešení problematiky v České republice

Ani v České republice se dosud komplexně problematikou registrace, ochrany a zpřístupňování elektronických publikací nikdo nezabýval. Vzhledem k tomu, že se jedná o národní bibliografii a národní konzervační fond, přísluší řešení této problematiky institucionálně především Národní knihovně ČR. Na projektu Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet, který by měl vejít ve známost pod jednoduchým názvem **WebArchiv**, spolupracuje NK ČR s Ústavem výpočetní techniky Masarykovy univerzity v Brně a s Ikaros, o. s. WebArchiv se prezentuje na webovských stránkách na adrese <http://webarchiv.nkp.cz>.

Aplikace v našich podmínkách představuje mj. stanovení kritérií výběru zdrojů pro národní bibliografii, legislativní zajištění akvizice domácích elektronických publikací, technické a programové řešení jejich indexace

i archivace, zajištění standardů pro budoucí čitelnost zdrojů a pro vyhledávání v síti; archivace a zpřístupnění primárních síťových elektronických zdrojů vyžadují řešení otázek autorského práva, vytvoření podmínek pro kooperaci centrálních, regionálních a specializovaných knihoven, resp. informačních pracovišť a propojení s vydavateli elektronických zdrojů.

Problematika řešená v tomto projektu je velmi komplexní, zahrnuje oblast knihovnictví a vydavatelství, práva i informačních technologií. Navíc vyžaduje aplikaci mezinárodních standardů a kompatibilitu řešení s jinými podobnými projekty zpracovávanými ve světě v nedávné minulosti i v současnosti, na nichž pracují velké týmy specialistů.

Řešitelé proto věnovali značnou pozornost informačním průzkumům, získání dostupných informačních materiálů publikovaných v tištěné a zejména v elektronické formě, jejich analýze a navázání kontaktů s vytipovanými zahraničními pracovišti, od nichž lze získat cenné informace, zkušenosti i softwarové nástroje jako výsledky řešených národních i mezinárodních projektů.

Jedním z nejrozsáhlejších a nejvýznamnějších mezinárodních projektů na tomto poli byl projekt NEDLIB [11] podporovaný Evropskou komisí, na jehož realizaci se podílelo osm národních knihoven západoevropských zemí a tři instituce zajišťující technickou stránku řešení. Tento projekt, který navázal na řadu podobně zaměřených národních i nadnárodních projektů (v oblasti našeho zájmu zejména projekty Nordic Metadata I, II [11] severovýchodních zemí) a jehož předmětem je budování evropské deponitní knihovny elektronických dokumentů, se zabýval všemi elektronickými dokumenty včetně méně problematických offline zdrojů. Díky jeho širokému záběru je vhodné a možné převzít mnohé z toho, čeho v něm bylo dosaženo.

Možnost získání zkušeností z řešení klíčových projektů při zahraniční cestě do Finska (návštěva Helsinki University Library – národní projekty i účast v mezinárodních projektech) [23] napomohla do značné míry k technickému zaměření řešení na využití nástrojů, které jsou výsledkem výše zmíněných projektů a které lze získat za výhodných podmínek. K účelům testování softwarových nástrojů na vybraném vzorku elektronických zdrojů slouží vysoce výkonný počítač s velkými kapacitami operační i diskové paměti pro připojení na síť, pracující v OS Linux. Tento stroj má ve fázi testování suplovat unixový server a slouží k instalování nástrojů pro stahování a archivaci dokumentů, pro ukládání údajů pro popis zdrojů aj. a pro ukládání zdrojů do webovského archivu.

Řešení pilotního projektu představuje principiálně testování dvou metod, které by v optimálním případě měly být aplikovány paralelně s cílem umožnit dlouhodobé uchování a využívání elektronických zdrojů:

- ♦ shromažďování, registrace a archivace vybraných domácích elektronických online dostupných dokumentů jako legitimní součásti národní publikační produkce podle stanovených kritérií výběru pro účely České ná-

rodní bibliografie (tato činnost klade značné nároky na intelektuální práci zpracovatelů),

- ♦ automatizovaný proces shromažďování a archivace domácích zdrojů z Internetu v relativní úplnosti (servery v národní doméně .cz a další omezení) pomocí speciálního indexačního programu.

Oblast problematiky vztahů knihoven, vydavatelů a legislativy

Pro účely pilotní fáze projektu, jejímž smyslem je testovat stanovené postupy při zpracování online dostupných elektronických dokumentů, bylo vybráno celkem 14 elektronických domácích odborně zaměřených časopisů jako vhodný vzorek publikačních aktivit v prostředí World Wide Web. Výrazem „domácí“ se v tomto kontextu míní ty dokumenty, které jsou zpřístupněny na serverech s doménou I. stupně „.cz“. Kategorie „elektronický časopis“ byla vymezena v souladu s příslušnými mezinárodními normativními předpisy – ISBD(S), ISBD(ER) a AACR2R – jako podmnožina seriálových publikací. Z těchto titulů pouze časopis Ikaros soustavněji podléhá bibliografické kontrole v celostátním měřítku (v Bibliografické databázi záznamů o dokumentech z oblasti VTEI a knihovnictví SPOJ od srpna 1999 a v databázi článků z českých periodik ANL jako součásti České národní bibliografie od května 1999).

Výběr časopisů byl proveden na základě spolupráce s vysokoškolskými a dalšími odbornými knihovnami. Jejich přehled je uveden v tabulce 1. Základním kritériem výběru byla neomezená dostupnost na webu, dalšími kritérii bylo vydávání minimálně po dobu jednoho roku a nereklamní charakter časopisu (tzn., že není pouze prostředkem prezentace vydavatele – soukromé osoby, osob nebo instituce). Základní identifikační údaje byly excerpovány jednak z vlastních primárních dokumentů, jednak ze sekundárních zdrojů (zejména ISSN Register). K tomu je nutné dodat, že jen menší část z těchto časopisů je označena ISSN. Většina z těchto časopisů vychází pouze v elektronické podobě, některé tituly však mají charakter tzv. online supplementu, který do jisté míry figuruje jako samostatný dokument, neboť se po obsahové stránce s tištěným časopisem zcela neshoduje a také má odlišnou periodicitu. Zvláštní skupinu seriálů, která se vymyká zaběhnuté klasifikaci seriálových publikací, avšak je třeba ji brát rovněž v úvahu, tvoří průběžně aktualizované systémy (někdy označované jako zpravodajské servery), které mají z technického hlediska charakter dynamické databáze, z níž jsou jednotlivé dokumenty generovány na základě uživatelského dotazu. Z této skutečnosti pak plynou ve srovnání se staticky zpřístupňovanými časopisy (tj. časopisy, jejichž obsah se mění v určitých intervalech) různá omezení při registraci dokumentů na analytické úrovni. Úmyslně byly opomenuty elektronické verze tištěných časopisů, přestože jim bylo přiděleno vlastní ISSN.

Tab. 1: Domácí odborně zaměřené elektronické časopisy vybrané k testování metadatového schématu Dublin Core

Název	URL	ISSN	Vydavatel	Obor	Periodicita	Kontakt
Česká škola	http://www.ceskaskola.cz	-	Computer Press	školství	průběžně	vit.sebor@cpress.cz
Deník veřejné správy	http://denik.obce.cz	-	Triada	veřejná správa	průběžně	denik@triada.cz
ECHOMAGAZIN	http://www.echoplus.cz/echomagazin.htm	-	L. Pachta-Impea	vodní hospodářství, ekologie	nepravidelně	info@echoplus.cz
EKO	http://www.bm.ipex.cz/eko	-	ZO Českého svazu ochránců přírody 54/34 "Bystřina" v Brně a Nadace pro výchovu k přátelské civilizaci 3000	ekologie	2 x měsíčně (vždy v pondělí)	alkor@ipex.cz
EkoList po drátě (supplement tištěného EkoListu)	http://www.ekolist.cz/	-	Brontosauří ekocentrum Zelený kruh	ekologie	denně	ekolist@ecn.cz
Elektrorevue	http://www.elektrorevue.cz	1213-1539	VUT Brno	elektrotechnika	průběžně	redakce@elektrorevue
Chirurgie	http://www.chirurgie.cz		soukromé osoby	chirurgické obory	průběžně	redakce@chirurgie.cz
Ikaros	http://www.ikaros.cz	1212-5075	soukromé osoby	informační a knihovní věda	10 x ročně	ikaros-redakce@ff.cun
Instantní astronomické noviny	http://www.ian.cz	1212-6691	občanské sdružení IAN a Hvězdárna a planetárium Mikuláše Koperníka v Brně	astronomie	jednou až dvakrát týdně	redakce@ian.cz
Strategie Online (supplement tištěné Strategie)	http://www.istrategie.cz	-	Strategie Praha	reklama, média, public relations	denně	strategie@istrategie.cz
Lidem.cz	http://www.lidem.cz	-	Lotos, o. p. s. a ICN	neziskový sektor	týdně	redakce@lidem.cz
Militaria	http://militaria.internet.cz	-	Elka Press	vojenství	občasník	militaria@internet.cz
Orbis pictus – Internetový zpravodaj	http://www.orbispictus.cz/	-	Orbis Pictus	školství	Průběžně	redakce@orbispictus.c
Veřejná správa online (supplement časopisu Obec a finance)	http://denik.obce.cz/vsol/	-	Triada	informační technologie ve veřejné správě	6 x ročně	denik@triada.cz

Předběžně byla dohodnuta spolupráce při testování využití metadatového schématu **Dublin Core** (viz dále – odstavec *Dublin Core*) s několika informačními a dalšími institucemi, které působí současně jako vydavatelé elektronických zdrojů, neboť v rámci svých webovských prezentací mj. publikují dokumenty, které nejsou jiným způsobem dostupné, avšak z hlediska cílových uživatelských skupin jsou považovány za významné.

Spolupráce s vytípanými vydavateli bude nutná hlavně z legislativních důvodů. Dohody o spolupráci by měly řešitelům projektu umožnit testování výše uvedených nástrojů se souhlasem vydavatelů testovaných zdrojů, tj. umožnit přístup do zdrojů a jejich stahování a uložení na serveru umístěném v NK ČR. Problematiku archivace a zpřístupňování elektronických online zdrojů z právního hlediska bude ovšem třeba výhledově řešit obdobně jako u ostatních druhů dokumentů, tj. uzákoněním práva povinného výtisku pro depozitní knihovny.

Otázka zákonů o povinném výtisku i otázka autorského zákona je v této souvislosti velmi živá, v současné době se jí intenzivně zabývá i konference CENL (Conference of European National Librarians) společně s federací evropských vydavatelů (FEP – Federation of European Publishers). Na této úrovni došlo prozatím k dohodě, že vydavatelé budou poskytovat elektronické online publikace depozitním knihovnám na bázi dobrovolnosti. Byla stanovena pravidla pro dobrovolné poskytování kopie elektronických online dokumentů do elektronického archivu.

Ve fázi pilotních projektů by měly knihovny s vydavateli dohodnout otázky definic pojmů *dokument* a *vydavatel*, otázky postupů a řízení celého procesu. Implementace by měla být průběžně monitorována a na základě zkušeností by měla být navržena účinná a oběma stranám vyhovující legislativa.

Ustanovení CENL/FEP [13] vycházejí z předchozí rozsáhlé práce provedené v rámci projektu CoBRA+ [1], podporovaného Evropskou komisí a zaměřeného na zlepšení vzájemné spolupráce evropských národních knihoven. Jeho cílem bylo nalézt taková řešení, která umožní uložení dokumentů v knihovních fondech, tj. vytváření (relativně) kompletní sbírky dokumentů, ale současně umožní také kontrolu přístupu k uloženým dokumentům tak, aby nedocházelo k narušení komerčních zájmů vydavatelů. Zdůrazňuje se, že implementace zásad v ustanovení musí přinášet výhody oběma stranám: knihovnám v uchování kompletní národní produkce pro současné i budoucí uživatele, vydavatelům v uchování jejich produkce elektronických dokumentů a zpřístupnění informací o jejich existenci pro širší veřejnost prostřednictvím soupisů národních bibliografií.

Podle analýzy provedené na základě dostupných informačních pramenů i dotazů adresovaných na vybrané depozitní knihovny je již v některých zemích právo povinného výtisku (depozitní kopie) online elektronických zdrojů uzákoněno, jinde probíhá schvalovací řízení takového zákona (viz dále – odstavec *Výsledky řešení. Legislativa*).

Oblast problematiky informačních technologií

Z výsledků zkoumaných zahraničních projektů a výzkumů jsou pro náš projekt důležité tyto body:

Registrace, ochrana, archivace

Průměrná doba existence elektronického dokumentu na Internetu je asi tři roky. Z hlediska institucí, jejichž zájmem je dlouhodobé uchování kulturního dědictví, je proto nutné přistoupit k aktivní ochraně těchto dokumentů formou archivace.

Z dosavadních zahraničních zkušeností, z počtu již existujících dokumentů a z pokračujícího exponenciálního růstu počtu elektronických online dostupných dokumentů vyplývá, že jediný prakticky reálný/zvládnutelný způsob vytváření **relativně úplného** konzervačního fondu (elektronický archiv) a národní bibliografie je postup plně automatizovaný. Selektivní přístup je reálný pouze u velmi omezeného výseku specifických publikací na Internetu.

Odhadovaná velikost „národního webu“ je překvapivě relativně malá (poměřováno technickými i cenovými parametry již dnes běžně dostupných archivačních technologií); na základě aproximací experimentálně zjištěných parametrů v severských zemích (Finsko a Švédsko) ji odhadujeme kolem 300 GB. Současné technologie nám dovolují realizovat automatizovaný způsob archivace za přijatelnou cenu, pokud se omezíme jen na oblast národních elektronických zdrojů.

V rámci projektu NEDLIB [11] jsou vyvíjeny nástroje pro sběr, archivaci a indexaci elektronických online dokumentů. Některé z těchto nástrojů jsou k dispozici zdarma a jejich lokalizace a nasazení je v našich podmínkách reálné. Nejvýznamnějším z této skupiny nástrojů je NEDLIB Harvester, nástroj pro stahování a archivaci elektronických dokumentů.

Zpřístupnění archivovaných dokumentů

Ve švédském projektu Kulturarw3 [5] byly realizovány první pokusy o zpřístupnění webového archivu s využitím standardních uživatelských přístupových technologií (webového prohlížeče). Nástroj pro dokonalejší zpřístupnění archivovaných dokumentů je v projektu NEDLIB sice také vyvíjen, ale už nebude k dispozici zdarma. Tento nástroj by měl umožňovat prohlížení archivovaných dokumentů nejen v rámci vzájemných odkazů, ale i vzhledem k časové ose.

V dlouhodobějším horizontu se zde otevírá pole pro uplatnění přístupů z oblasti analýzy přirozeného jazyka (překračuje rámec stávajícího projektu). Jak vlastní sběr a archivace, tak zejména zpřístupnění dokumentů vyžaduje odpovídající národní legislativní rámec.

Metadata

Pro zkvalitnění automaticky vytvářených indexů je vhodné propagovat mezi veřejností publikující v prostředí World Wide Web jednotné metadatové standardy pro popis elektronických zdrojů, použitelné přímo samotnými autory. Nejvýznamnějším z těchto standardů je Dublin Core (DC) [16], případně z něj odvozené standardy. Za uplynulý rok bylo dosaženo jistého pokroku v rozvoji kvalifikova-

ného DC a byly nastartovány významné iniciativy k širšímu uznání standardu DC (ANSI/NISO standardizace).

Došlo také k vytvoření české verze standardu DC jako základu pro širší národní využití. Pro podporu používání DC vzniklo několik zdarma online dostupných nástrojů, po širší lokalizaci použitelných i u nás.

Jednoznačná globální trvalá identifikace

Pro usnadnění identifikace elektronických dokumentů byl vytvořen koncept Uniform Resource Name (URN) [18] – jednoznačných identifikátorů dokumentu. Tyto identifikátory jsou generovány a žadatelům přidělovány automaticky. Jednou z aplikací URN mohou být registrační čísla národní bibliografie (NBN), dále ISBN a ISSN.

Další možností je vytvořit identifikátor URN na základě kontrolního součtu MD5 – tímto způsobem je možné snadno ověřit i to, zda byl dokument po přidělení tohoto identifikátoru změněn.

V této oblasti lze také převzít zkušenosti a postupy z projektů Nordic Metadata I a II [12] a projektu NEDLIB [11].

Výsledky řešení

Na Linux-serveru pořízeném pro testování vyvíjených, resp. lokalizovaných nástrojů (např. elektronický formulář pro jednoduchý zápis údajů Dublin Core s možností uložení do zdrojového dokumentu, nástroj pro automatické přidělování URN) a pro uložení vybraných elektronických zdrojů stažených z webu (webovský archiv) byla zřízena doména URL: <http://webarchiv.nkp.cz>, kde jsou prezentovány výsledky řešení projektu WebArchiv a další relevantní informace.

Výběr vydavatelů, resp. dokumentů pro testování

Na základě studia zahraničních projektů byla stanovena kritéria pro výběr typů dokumentů do vzorku, na němž je třeba testovat stanovené postupy při zpracování online dostupných elektronických dokumentů publikovaných v síti Internet. Hlavní kritéria výběru byla následující:

- ♦ původní vydání dokumentu v elektronické formě (online)
- ♦ vydávání minimálně po dobu jednoho roku
- ♦ dokument není pouze prostředkem prezentace vydavatele (soukromé osoby, osob nebo instituce).

Jako typ dokumentů nejvhodnějších pro testování byly vybrány elektronické časopisy. Jde o kategorii dokumentů poměrně stabilních. Časopisy mají navíc tu výhodu, že u nich není v současné době problém s přidělováním ISSN použitelným pro účely projektu jako identifikátor URN, takže je možné na nich ověřovat připravené postupy a nástroje.

Další uvažovanou skupinou dokumentů pro testování jsou domovské stránky vybraných informačních a podobných institucí. Předběžně byla s těmito institucemi dohod-

nuta spolupráce při využití metadatového schématu Dublin Core.

Problematika výběru vzorku pro testování a souvisejících otázek již byla blíže popsána v odstavci *Oblast problematiky vztahů knihoven, vydavatelů a legislativy*.

Legislativa

Byla provedena důkladná analýza legislativního zabezpečení získávání elektronických online dostupných publikací a souvisejících otázek ve vybraných zemích. Platný zákon o povinném výtisku zahrnující i vzdáleně přístupné elektronické zdroje je zatím pouze v Dánsku, Norsku a po úpravách původního zákona i na Slovensku; připraven ke schválení je zákon v Austrálii, Finsku, Švédsku; v Holandsku funguje bez problémů spolupráce s vydavateli na základě dohod. Výsledky analýzy byly publikovány v časopise Ikaros, č. 10/2000 [20].

Dublin Core

Do češtiny byla přeložena nejnovější verze standardu Dublin Core Metadata Element Set, Version 1.1 [14]. Národní česká verze DC byla zaregistrována v rámci DCMI (Dublin Core Metadata Initiative) [17]. Pro iniciativu Dublin Core byly vytvořeny české webovské stránky [15]. Byly zpracovány přehledové analýzy z oblasti vývoje a využití standardu Dublin Core [21], [22].

Byla vytvořena beta-verze lokalizovaného DC-metadatového formuláře převzatého od Helsinské univerzitní knihovny z projektu Nordic Metadata (*obr. 1*). Tento nástroj podporuje kvalifikovaný DC podle nejnovější specifikace a zároveň umožňuje propojení na nástroj pro automatické přidělování URN. Podporuje jak syntaxi HTML, tak XML (RDF). Hardwarové nároky pro provoz těchto nástrojů jsou minimální, protože jde o relativně jednoduché skripty v programovacím jazyce Perl. Oba tyto nástroje budou k dispozici všem zájemcům publikujícím především na českém Internetu.

GENERÁTOR METADAT DUBLIN CORE

Dokumenty
Formuláře
Odkazy
English

Tento nástroj, vyvinutý v rámci Nordic Metadata Project, slouží ke generování záznamu metadat podle schématu Dublin Core (verze 1.1).

Potřebujete-li nápovědu při užití tohoto nástroje, použijte následující odkazy:

- Otevírá pravidla (určená těm, kteří s nástrojem doposud nepracovali).
- Dublin Core Czech (česká verze standardu vytvořená na MU Brno)
- Odkazy na klasifikační systémy a slovníky
- Nápověda týkající se daného pole se zobrazí po kliknutí na jeho název.

Chcete-li přidat nebo ubrát výskyt pole, užití tlačítka '+' a '-'.

Přesvědčte se, zda okno vašeho prohlížeče je dostatečně široké pro zobrazení kvalifikátorů polí!

Smazat formulář Načíst metadata DC z této adresy

1 NÁZEV (jméno dané zdroje)
Povinný výtisk elektronických publikací, zejména vzdálen

2 TVŮRCE (Entita primárně zodpovědná za vytvoření obsahu zdroje)

3 PŘEDMĚT: Klíčová slova (vlastní klíčová slova, maximálně jedno v poli)
povinný výtisk
elektronické zdroje

3 PŘEDMĚT: Řízené slovníky (klíčová slova z řízených slovníků, maximálně jedno v poli)

Obr. 1: Lokalizovaný generátor záznamu metadat ve schématu Dublin Core vyvinutý v rámci projektu Nordic Metadata

Záznamy v metadatovém schématu Dublin Core vygenerované pomocí lokalizového formuláře jsou pokusně zařazovány do zdrojových kódů článků elektronického časopisu Ikaros (obr. 2-3).

```

Povinný výtisk elektronických publikací - Microsoft Internet Explorer
Soubor Úpravy Zobrazení Zobrazení Historie Nápověda
Adresa http://webarchiv.nkp.cz/cgi-bin/ikar/...

<META name="DC.Title" content="Povinný výtisk elektronických publikací, zejména vzdálených elektronick
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Title">

<META name="DC.Creator.personalName" content="Čelbová, Ludmila">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Creator.personalName">

<META name="DC.Creator.address" content="ludmila.celbova@nkp.cz">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Creator.address">

<META name="DC.Subject" content="povinný výtisk">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" content="elektronické zdroje">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="FNK" content="knihovnictví">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="FNK" content="legislativa">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="FNK" content="Česko">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="NED" content="02">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Subject" scheme="NED" content="347.7">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Subject">

<META name="DC.Description.abstract" content="Současný stav legislativy v oblasti povinného výtisk
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Description.abstract">

<META name="DC.Publisher" content="Ikaros, o. s.">
<LINK rel="SCHEMA.dc" href="http://purl.org/metadata/dublin_core_elements#Publisher">

```

Obr. 2: Vygenerovaný záznam metadat ve schématu Dublin Core ve formátu HTML

```

www.výtisk elektronických publikací - Microsoft Internet Explorer
Soubor Úpravy Zobrazení Zobrazení Historie Nápověda
Adresa http://webarchiv.nkp.cz/cgi-bin/ikar/...

<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/documents/rec-dcog-19990702.htm"
xmlns:dcq="http://purl.org/dc/documents/rec-dcog-qualifiers-20000711.htm">
<rdf:description rdf:about="Povinný výtisk elektronických publikací, zejména vzdálených elektronick
<dc:title>
<rdf:description>
<rdf:value>Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů</rdf:va
</rdf:description>
</dc:title>

<dc:Creator>
<rdf:description>
<dcq:CreatorType>personalName</dcq:CreatorType>
<rdf:value>Čelbová, Ludmila</rdf:value>
</rdf:description>
</dc:Creator>

<dc:Creator>
<rdf:description>
<dcq:CreatorType>address</dcq:CreatorType>
<rdf:value>ludmila.celbova@nkp.cz</rdf:value>
</rdf:description>
</dc:Creator>

<dc:Subject>
<rdf:description>
<rdf:value>povinný výtisk</rdf:value>
</rdf:description>
</dc:Subject>

<dc:Subject>
<rdf:description>
<rdf:value>elektronické zdroje</rdf:value>
</rdf:description>

```

Obr. 3: Vygenerovaný záznam metadat ve schématu Dublin Core ve formátu XML (RDF)

Harvester

Je analyzován aktuální stav, možnosti a podmínky lokalizace nástroje Harvester z projektu NEDLIB na Linux-serveru instalovaném z prostředků projektu v Národní knihovně ČR.

Na vývoji tohoto nástroje se stále pracuje, nicméně již bylo s jeho pomocí dosaženo v některých zemích pozoruhodných výsledků. Tento nástroj, využívající databáze MySQL pro ukládání dat, byl původně napsán v jazyce Perl, ale na zakázku Helsinské univerzitní knihovny byl přepsán do jazyka C, čímž došlo k jeho výraznému zrychlení a zároveň byla zlepšena i jeho funkčnost.

Harvester má relativně vysoké nároky na hardwarové vybavení pro svůj provoz. Vzhledem k jeho specifickým vyžaduje rychlou přípojku do sítě Internet, dostatečnou velikost paměti a odpovídající kapacitu paměti na pevných discích a případně i páskových jednotkách. Pro rutinní provoz je proto doporučen některý z větších značkových unixových serverů (Sun, HP, IBM, Compaq) kombinovaný s dostatečně velkým diskovým a páskovým polem. Pro akce menšího rozsahu je ale možné použít i stanici s Linuxem, což činí tento nástroj dostupným i pro náš projekt. Předpokladem je dostatečné množství paměti (nejlépe alespoň 512 MB), výkonný procesor (např. AMD Athlon) a především velké diskové pole (v našem případě 90 GB RAID 0+1). Pro uložení větších množství dat počítáme s možností využití páskové robotické knihovny, která je v majetku Národní knihovny ČR.

Přestože tento nástroj nebude nutné lokalizovat z hlediska uživatelského, bude nutné o to větší úsilí věnovat jeho přizpůsobení českým podmínkám na úrovni programové. Ani samotná instalace tohoto nástroje nebude jednoduchá vzhledem k tomu, že na jeho finální verzi se stále ještě pracuje. Také specifikace množiny dokumentů pro archivaci (specifikace „českého webu“) bude relativně náročná.

Ověřovač URL

Bylo provedeno předběžné orientační zmapování vybraných volně dostupných nástrojů či programů shareware určených pro kontrolu interních a externích HTML odkazů. Ve všech případech jde o prostředky na bázi perl skriptů. Pro serióznější analýzu by bylo třeba specifikovat předpokládaný způsob nasazení a požadované funkce v kontextu celkového řešení projektu.

Další perspektivní možností je využití existujících volně dostupných perlových modulů pro vytvoření aplikace přesně podle potřeb projektu.

Směry dalšího řešení

1. Připravit podmínky pro realizaci bibliografické kontroly v oblasti elektronických online dokumentů.
2. Usilovat o co nejvyšší efektivitu aplikací výsledků řešení podobných zahraničních či mezinárodních projektů včetně aplikace volně dostupných nástrojů i nástrojů poskytovaných za úhradu.

3. Zohlednit související projekty a činnosti v ČR:

- ♦ zpřístupňování národní článkové bibliografie a plných textů elektronických periodik
- ♦ integrace automatizovaných systémů pro systematické a věcné pořádání (Mezinárodní desetinné třídění, Předmětová hesla Národní knihovny a další)
- ♦ digitalizace klasických knihovních fondů
- ♦ programový projekt Ministerstva kultury ČR „Jednotná informační brána pro hybridní knihovny”.

Závěr

V rámci dosavadního řešení projektu WebArchiv byla analyzována řada zahraničních i mezinárodních projektů zabývajících se problematikou získávání, registrace, ochrany a zpřístupňování elektronických online dokumentů. Výzkum ukázal, že se jedná o velmi komplexní problematiku, zahrnující oblast spolupráce a propojení knihoven s vydavateli, oblast práva a oblast informačních technologií. Navíc vyžaduje aplikaci mezinárodních standardů a kompatibilitu řešení s jinými podobnými projekty.

Díky tomu, že lze využít výsledků jiných projektů (zahraničních a mezinárodních) a že značná část nástrojů potřebných pro realizaci projektu je volně k dispozici, je možné v relativně krátkém čase a za relativně nízkých finančních nákladů připravit v rámci pilotního projektu podmínky pro jeho realizaci. Na druhé straně je nutné počítat s tím, že jen samotná instalace, lokalizace a vzájemná integrace není samozřejmou záležitostí a že na dosažení požadované funkčnosti bude třeba intenzivně pracovat.

Řešení právních otázek, které nelze samozřejmě v této souvislosti opomenout, je dlouhodobou záležitostí. Nejprve bude nutné zvážit všechny aspekty (provozní, technické aj.), které se vážou k množině dokumentů podléhajících povinnosti vydavatelů poskytovat/ohlašovat vydané publikace, a teprve následně bude možné připravit podklady pro změnu zákona obsahujícího ustanovení o povinném výtisku seriálových publikací („tiskový zákon”), resp. výklad a směrnice k zákonu týkajícímu se povinného výtisku neperiodických publikací a k autorskému zákonu (*zákon č. 37/1995 Sb., o neperiodických publikacích; zákon č. 46/2000 Sb., o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů – tiskový zákon; zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů – autorský zákon*).

Poznámka:

Ke zpracování článku jsem použila podkladů externích spolupracovníků podílejících se na řešení projektu – RNDr. Miroslava Bartoška, CSc. (ÚVT MU Brno), Ing. Petra Žabičky (MZK Brno), Mgr. Ivy Celbové a Mgr. Filipa Vojtáška (oba Ikaros, o. s.) pro výroční zprávu k projektu za rok 2000 [24].

Použité informační zdroje

1. *Cobra+* [online]. Boston Spa : British Library, 1997 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://portico.bl.uk/gabriel/en/projects/cobra.html>>.
2. *Biblink* [online]. Bath (Anglie) : UKOLN, last updated 12-Jul-2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://hosted.ukoln.ac.uk/biblink/>>.
3. *EVA : the acquisition and archiving of electronic network publications* [online]. Helsinki (Finsko) : Helsinki University Library, last updated 15-Dec-1997 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.lib.helsinki.fi/eva/english.html>>.
4. *INDOREG : Internet Document Registration : project report* [online]. Ballerup (Dánsko) : Dansk Bibliotheks Center, 16-Sep-1997 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.purl.dk/rapport/html.uk>>.
5. *Kulturarw3 Heritage Project* [online]. Stockholm (Švédsko) : Royal Library, [1998] [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://kulturarw.kb.se/html/projectdescription.html>>.
6. FEIJEN, Martin. DNEP : fond holandské produkce elektronických publikací. *Národní knihovna*, 6, 1995, č. 6, s. 206-210. ISSN 0862-7487.
7. *National Library of Canada Electronic Collection* [online]. Ottawa (Kanada) : NLC, revised 2001-02-26 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://collection.nlc-bnc.ca/e-coll-e/index-e.htm>>.
8. OCLC. *Preservation resources* [online]. Dublin (Ohio, USA) : OCLC, c1999, 21-Sep-2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.oclc.org/oclc/man/catproj/catcall.htm>>.
9. OCLC. *Internet Cataloging Project* [online]. Dublin (Ohio, USA) : OCLC, 1996 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.oclc.org/oclc/man/catproj/catcall.htm>>.
10. *PANDORA* [online]. Canberra (Austrálie) : NLA, last updated 10-Oct-2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.nla.gov.au/pandora/>>.
11. *NEDLIB : Networked European Deposit Library* [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, c1998, last updated 11-Mar-2001 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.kb.nl/nedlib/>>.
12. *The Nordic Metadata projects* [online]. Helsinki (Finsko) : Helsinki University Library, 1996, last updated 21-Feb-2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.lib.helsinki.fi/meta/>>.
13. International declaration on the deposit of electronic publications : conference of European National Librarians/Federation of European Publishers (CENL/FEP). *Dialog mit Bibliotheken*, 2000, vol. 12, no. 3, s. 2-14. ISSN 0936-1138.
14. *Dublin Core Czech : soubor metadatových prvků Dublin Core, verze 1.1 : referenční popis* [online]. Brno : Masarykova univerzita, 12-06-2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <http://www.ics.muni.cz/dublin_core/DC-czech-1.1.html>.
15. *Dublin Core Czech* [online]. Brno : Masarykova univerzita, posl. aktualizace 26-2-2000 [2001?] [cit. 9. dubna 2001]. Dostupné na World Wide Web: <http://www.ics.muni.cz/dublin_core/>.
16. *Dublin Core Metadata Initiative* [online]. Dublin (Ohio, USA) : DCMI, c2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://purl.org/dc/>>.
17. *Dublin Core Metadata Initiative : Translations of DCMI Documents* [online]. Dublin (Ohio, USA) : DCMI, c2001 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://dublincore.org/resources/translations>>.
18. Internet Engineering Task Force. *Uniform Resource Names (urn)* [online]. Preston (Virgin., USA); Leeds : IETF, last modified 30-Mar-2001 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.ietf.org/html.charters/urn-charter.html>>.
19. CELBOVÁ, Ludmila. Elektronické zdroje publikované v síti Internet jako součást české národní bibliografie. *Ikaros* [online]. 2000, č. 6 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://ikaros.ff.cuni.cz/2000/c06/elzdroje.htm>>.
20. CELBOVÁ, Ludmila. Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů. *Ikaros* [online]. 2000, č. 10 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://ikaros.ff.cuni.cz/2000/c10/onlinePV.htm>>.
21. ŽABIČKA, Petr. Dublin Core jako standard pro popis elektronických síťových zdrojů. In *Česko-slovenská konference RUFIS 2000, Brno 5.-6. 9. 2000* [online]. Brno : Vysoké učení technické : Masarykova univerzita, 2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <<http://www.vutbr.cz/konference/rufis2000/sbornik/25-zabicka.pdf>>; <<http://www.mzk.cz/DC/rufis2000.html>>.
22. ŽABIČKA, Petr. *Dublin Core – metadata pro popis elektronických dokumentů*. Předneseno na konferenci DATASEM 2000, konané 21. až 24. října 2000 v Brně. Dostupné na World Wide Web: <<http://webarchiv.nkp.cz/datasem2000.pdf>>.
23. *Zpráva ze služební cesty do Finska*. Praha : Národní knihovna ČR, 2000. 2 s. Dostupné též na World Wide Web: <<http://webarchiv.nkp.cz/finsko.pdf>>.
24. *Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet : souhrnná zpráva za rok 2000* [online]. Praha : Národní knihovna ČR, 2000 [cit. 9. dubna 2001]. Dostupné na World Wide Web: <webarchiv.nkp.cz/zprava2000.pdf>.

