

Národní knihovna České republiky

**Registrace, ochrana a zpřístupnění domácích
elektronických zdrojů v síti Internet**

Souhrnná zpráva za rok 2000

Předkládá: PhDr. Vojtěch Balík, ředitel NK ČR

Zpracovala: Mgr. Ludmila Celbová, řešitelka

Praha, listopad 2000

OBSAH

OBSAH.....	2
A KONSTATAČNÍ ČÁST	3
A.1 Rešerše	3
A.2 Současný stav ve světě a v ČR.....	4
A.3 Cíl, vstupní data	5
B ANALYTICKÁ ČÁST	6
B.1 Vlastní řešení.....	6
B.2 Přínos řešitele	7
B.2.1 Oblast problematiky vztahů knihoven, vydavatelů a legislativy	7
B.2.2 Oblast problematiky informačních technologií	9
B.3 Posun znalostí.....	10
C NÁVRHOVÁ ČÁST	11
C.1 Výsledky řešení.....	11
C.1.1 Výběr vydavatelů, resp. dokumentů pro testování	11
C.1.2 Legislativa	11
C.1.3 Dublin Core	12
C.1.4 Harvester.....	13
C.1.5 Ověřovač URL.....	13
C.2 Závěr	14
C.3 Návrhy opatření.....	15
D RESUMÉ A KLÍČOVÁ SLOVA	16
D.1 Resumé	16
D.2 Klíčová slova	16
E PŘÍLOHY	17

A KONSTATAČNÍ ČÁST

A.1 Rešerše

1. *Biblink* [online]. Bath (Anglie) : UKOLN, last updated 12-Jul-2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://hosted.ukoln.ac.uk/biblink/>>.
2. *Cobra+* [online]. Boston Spa : British Library [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://portico.bl.uk/gabriel/en/projects/cobra.html>>.
3. OCLC. *Preservation resources* [online]. Dublin (Ohio, USA) : OCLC, c1999, 21-Sep-2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.oclc.org/oclc/man/catproj/catcall.htm>>.
4. OCLC. *Internet Cataloging Project* [online]. Dublin (Ohio, USA) : OCLC [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.oclc.org/oclc/man/catproj/catcall.htm>>.
5. *Dublin Core Czech* [online]. Brno : Masarykova univerzita, c1999, posl. aktual. 31. října 2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <http://www.ics.muni.cz/dublin_core/>.
6. *Dublin Core Metadata Initiative* [online]. Dublin (Ohio, USA) : DCMI, c2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://purl.org/dc/>>.
7. *EVA : the acquisition and archiving of electronic network publications* [online]. Helsinki (Finsko) : Helsinki University Library, last updated 2-Oct-2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.lib.helsinki.fi/eva/english.html/>>.
8. *Guidelines for the selection of online Australian publications intended for preservation by the National Library of Australia* [online]. Canberra (Austrálie) : NLA, last updated 21-Dec-1999 [cit. 10. listopadu 2000]. Dostupné na World Wide Web: <<http://www.nla.gov.au/scoap/guidelines.html>>.
9. *INDOREG : Internet Document Registration Project report* [online]. Ballerup (Dánsko) : Dansk Bibliotheks Center, 16-Sep-1997 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.purl.dk/rapport/html.uk>>.
10. International DOI Foundation. *The Digital Object Identifier* [online]. Kidlington, Oxford (Anglie) : IDF, updated 8-Nov-2000. [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.doi.org/>>.
11. Internet Engineering Task Force. *Uniform Resource Names (urn)* [online]. Preston (Virgin., USA); Leeds : IETF, last modified 05-Sep-2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.ietf.org/html.charters/urn-charter.html>>.
12. *Kulturarw3 Heritage Project* [online]. Stockholm (Švédsko) : Royal Library, 1998, updated 06-Sep-1999 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://kulturarw.kb.se/html/projectdescription.html>>.
13. *Metadata for preservation : the Cedars project outline specification : draft for public consultation* [online]. Leeds : Consortium of University Research Libraries, March 2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>>.
14. *National Library of Canada Electronic Collection* [online]. Ottawa (Kanada) : NLC, revised 2000-06-13 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://collection.nlc-bnc.ca/e-coll-e/index-e.htm>>.

15. *National strategy for provision of access to Australian electronic publications : a National Library of Australia position paper* [online]. Canberra (Austrálie) : NLA, last updated 07-Dec-1999 [cit. 10. listopadu 2000]. Dostupné na World Wide Web: <<http://www.nla.gov.au/policy/paep.html>>.
16. *Networked European Deposit Library* [online]. Hague (Nizozemí) : Koninklijke Bibliotheek, c1998 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.kb.nl/nedlib/>>.
17. *The Nordic Metadata project* [online]. Helsinki (Finsko) : Helsinki University, 1996, last updated 21-Feb-2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.lib.helsinki.fi/meta/>>.
18. OLSON, Nancy B. *Cataloging Internet Resources : a manual and practical guide* [online]. Second edition. Dublin (Ohio, USA) : OCLC, c1997 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.purl.org/oclc/cataloging-internet>>.
19. *PANDORA* [online]. Canberra (Austrálie) : NLA, last updated 10-Oct-2000 [cit. 10. listopadu 2000]. Dostupné na World Wide Web: <<http://www.nla.gov.au/pandora/>>.
20. MARTIN, Elizabeth. *Management of networked electronic publications* [online]. Ottawa (Kanada) : NLC, 1999 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.nlc-bnc.ca/consult/consult4-e.pdf>>.
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. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.vutbr.cz/konference/rufis2000/sbornik/25-zabicka.pdf>>.

A.2 Současný stav ve světě a v ČR

Exploze v elektronickém publikování, tj. zejména v tvorbě elektronických zdrojů přístupných v síti Internet, vyžaduje nový přístup ke zpracování, ochraně a zpřístupňování těchto informací. Dálkově 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 zpracování je třeba vycházet ze skutečnosti, že v průběhu procesu manipulace s těmito zdroji není možné aplikovat tytéž metody jako v případě tradičních (tištěných) dokumentů a v neposlední řadě také elektronických dokumentů uložených na fyzických nosičích, což je dáno skutečností, že jejich vlastnosti se zásadně, resp. částečně liší.

Ve světě (zejména v USA, Kanadě, Austrálii a v evropských severských 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 společné mezinárodní projekty (CoBRA+, BIBLINK a NEDLIB), 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. Ze zemí bývalého východního bloku zatím žádná s řešením obdobného projektu nezačala.

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 České republiky. V řešení problematiky informačních technologií spolupracuje Národní knihovna ČR s Ústavem výpočetní techniky Masarykovy univerzity v Brně, na řešení okruhu problémů knihovnických a legislativních se podílejí externí spolupracovníci Národní knihovny ČR.

A.3 Cíl, vstupní data

Cílem projektu je připravit podmínky pro zpracování české národní bibliografie elektronických zdrojů, se zaměřením zejména na zdroje dálkově přístupné. S bibliografickým zpracováním souvisí zajištění trvalého uchování domácích elektronických dokumentů monografických i seriálových publikovaných v síti Internet a jejich zpřístupnění, respektující autorské právo vydavatelů.

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.

V zásadě je třeba řešit několik okruhů problémů:

- Výběr elektronických zdrojů pro registraci a archivaci v souladu s politikou budování fondů depozitních knihoven (součást národní bibliografie - konzervační fond)
- Získávání elektronických zdrojů pro konzervační fond (legalizace povinného výtisku)
- Organizace a řízení virtuálních elektronických publikací, bibliografická kontrola (standards a nástroje pro ukládání a vyhledávání dat o elektronických publikacích i pro ukládání a vyhledávání primárních zdrojů)
- Zajištění trvalého přístupu k vybraným publikacím uloženým v elektronickém archivu (metadata, URN, údržba prostředí)
- Stanovení podmínek přístupu k archivovaným elektronickým zdrojům (autorské právo)

B ANALYTICKÁ ČÁST

B.1 Vlastní řešení

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 v prvním roce práce na projektu značnou pozornost informačním průzkumům, získání dostupných informačních materiálů publikovaných v tištěné a zejména v elektronické formě, jejich analýze a navázání kontaktů s vytypovanými zahraničními pracovišti, od nichž lze získat cenné informace, zkušenosti i softwarové nástroje jako výsledky řešených národních i mezinárodních projektů. Tento postup je racionální a ekonomicky únosný v rámci finančních, personálních a organizačních možností českého knihovnictví a současně umožní zařadit se poměrně rychle mezi pokrokové země v oblasti získávání, ochrany a zpřístupňování elektronických dokumentů publikovaných v síti Internet.

Jedním z nejrozsáhlejších a nejvýznamnějších zahraničních projektů na tomto poli je projekt NEDLIB, na jehož realizaci se podílí osm národních knihovny západoevropských zemí a tři instituce zajišťující technickou stránku řešení. Tento projekt, který navazuje na řadu podobně zaměřených národních i nadnárodních projektů (v oblasti našeho zájmu jmenujme zejména projekty Nordic Metadata I, II a Nordic Web Archive severovýchodních zemí) a jehož předmětem je budování depozitní knihovny elektronických dokumentů, se zabývá všemi elektronickými dokumenty, včetně méně problematických off-line zdrojů. Díky jeho širokému záběru je vhodné a možné převzít mnohé z toho, čeho v něm bylo a ještě bude dosaženo.

Možnost získání zkušeností z řešení klíčových projektů při zahraniční cestě do Finska (návštěva Helsinki University Library - národní projekty i účast v mezinárodních projektech) 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 bylo třeba změnit původně plánovanou výpočetní techniku v rámci rozpočtových investičních prostředků, tj. místo původně plánovaných šesti terminálových stanic zakoupit dva terminály s odpovídajícími současnými technickými parametry a jeden vysoce výkonný PC s velkými kapacitami operační i diskové paměti pro připojení na síť, pracující v OS Linux. Tento stroj bude ve fázi testování suplovat unixový server a bude sloužit k instalování nástrojů pro stahování a archivaci dokumentů, pro ukládání údajů pro popis zdrojů aj. a pro ukládání zdrojů do webového 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 součástí národní publikační produkce podle stanovených kritérií výběru pro účely České národní bibliografie; tato činnost klade značné nároky na intelektuální práci zpracovatelů;
- shromažďování a archivace domácích zdrojů z Internetu v relativní úplnosti (automatizovaný proces).

B.2 Přínos řešitele

V této části je vhodné z hlediska přehlednosti rozdělit informace o řešení do dvou částí – za prvé na oblast problematiky knihovnické, resp. vydavatelské a právní a za druhé na oblast problematiky informačních technologií.

B.2.1 Oblast problematiky vztahů knihoven, vydavatelů a legislativy

Pro účely pilotní fáze tohoto projektu, jejímž smyslem je testovat stanovené postupy při zpracování online přístupných 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 součástí České národní bibliografie od května 1999).

Výběr časopisů byl proveden na základě spolupráce s vysokoškolskými a dalšími odbornými knihovnami v říjnu 2000 po předchozím tříměsíčním sledování jednotlivých vytypovaných titulů. Jejich přehled je uveden v příloze F.1. Základním kritériem výběru byla neomezená dostupnost na webu, dalšími kritérii bylo vydávání minimálně po dobu jednoho roku a nereklamní charakter časopisu (není pouze prostředkem prezentace vydavatele - soukromé osoby či 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 (např. EkoList po drátě), 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, kterou je však třeba brát rovněž v úvahu, tvoří průběžně aktualizované systémy (někdy označované jako zpravodajské servery – např. Česká škola), 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.

Záznamy v metadatovém schématu Dublin Core vygenerované pomocí lokalizového formuláře (viz odst. C.1.3) a programu Metabrowser jsou pokusně zařazovány do zdrojových kódů článků elektronického časopisu Ikaros (viz přílohy F.4, F.5, F.7 a F.8).

Předběžně byla dohodnuta spolupráce při testování využití metadatového schématu Dublin Core s několika informačními a dalšími institucemi (např. Parlamentní knihovna, Knihovna Evangelické teologické fakulty UK, Národní vzdělávací fond, Vědecká lékařská knihovna IKEM), které působí současně jako vydavatelé elektronických zdrojů, neboť v rámci svých webových prezentací mj. publikují dokumenty, které nejsou jiným způsobem dostupné, avšak z hlediska cílových uživatelských skupin jsou považovány za významné.

Spolupráce s vytypovanými vydavateli bude nutná hlavně z právních 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 v této souvislosti je 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 (pojmu „dokument“ a „vydavatel“), otázky postupů a řízení celého procesu. Implementace by měla být průběžně monitorována a na základě zkušeností by se měla navrhnout účinná a oběma stranám vyhovující legislativa.

Ustanovení CENL/FEP vycházejí z předchozí rozsáhlé práce provedené v rámci projektu CoBRA+, 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. vytvářet 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 současnou i budoucí společnost, vydavatelům v uchování jejich produkce elektronických dokumentů a zpřístupnění informací o jejich existenci pro širší veřejnost prostřednictvím soupisů národních bibliografií.

B.2.2 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 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

Byly realizovány první pokusy o zpřístupnění webového archivu s využitím standardních uživatelských přístupových technologií (webového prohlížeče) – viz švédský projekt Kulturarw3. 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), případně z něj odvozené standardy. Za uplynulý rok bylo dosaženo jistého pokroku v rozvoji kvalifikovaného DC a nastartovány významné iniciativy k širšímu uznání standardu DC (ANSI/NISO standardizace).

Byla vytvořena česká verze standardu DC jako základ 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) – 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 Nordic Metadata Project I a II a NEDLIB.

B.3 Posun znalostí

K výraznému posunu znalostí došlo zejména v těchto oblastech:

- aplikace kritérií výběru zdrojů pro Českou národní bibliografii, výběr vzorku zdrojů pro testování,
- analýza právních aspektů - přehled situace v uzákonění povinného výtisku pro elektronické online dokumenty v nejvýznamnějších zemích a související problematika autorskoprávní,
- analýza technických nástrojů pro zabezpečení testování postupů sloužících k získávání, registraci, ochraně a zpřístupňování domácích elektronických zdrojů publikovaných na Internetu,
- možnost přístupu k potřebným nástrojům pro testování jako výsledkům zahraničních a mezinárodních projektů.

C NÁVRHOVÁ ČÁST

C.1 Výsledky řešení

O celkové koncepci řešení projektu byla odborná veřejnost informována v přednášce na konferenci Inforum 2000 [1] a prostřednictvím článku v časopise Ikaros [2]. Kromě toho byly publikovány informace k dílčím řešeným problémům - viz odst. C.1.3.

C.1.1 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 přístupných 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 či osob nebo instituce)

Jako typ dokumentů nejvhodnějších pro testování byly vybrány elektronické časopisy. Jde o kategorii dokumentů poměrně stabilních. Časopisy mají navíc 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 dohodnuta spolupráce při využití metadatového schématu Dublin Core.

Problematika výběru vzorku pro testování a souvisejících problémů je blíže popsána v odst. B.2.1 této zprávy.

C.1.2 Legislativa

Byla provedena důkladná analýza legislativního zabezpečení získávání elektronických online přístupných publikací a souvisejících otázek ve vybraných zemích. Platný zákon o povinném výtisku zahrnující i dálkově přístupné elektronické zdroje je

[1] CELBOVÁ, Ludmila. Registrace a zpřístupňování elektronických zdrojů publikovaných v síti Internet. In *Inforum 2000 : 6. ročník konference o profesionálních informačních zdrojích* [online]. Praha: Albertina icome Praha, 2000 [cit. 2000-08-31]. Dostupné na World Wide Web: <<http://www.inforum.cz/inforum2000/prednasky/registraceazp.htm>>.

[2] CELBOVÁ, Ludmila. Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie. *Ikaros* [online elektronický časopis]. 2000, roč. 4, č. 6 [cit. 2000-08-31]. Dostupné na Internetu: <<http://ikaros.ff.cuni.cz/ikaros/2000/c06/elzdroje.htm>>. ISSN 1212-5075.

zatím pouze v Dánsku, Norsku a nově (květen 2000) po úpravách původního zákona i na Slovensku; připraven ke schválení je v Austrálii, Finsku, Švédsku; v Holandsku funguje bez problémů spolupráce s vydavateli na základě dohod. Výsledky analýzy jsou uvedeny v příloze F.2.

C.1.3 Dublin Core

Byl vytvořen český překlad nejnovější verze standardu Dublin Core Metadata Element Set, Version 1.1 – viz [1]. Národní česká verze DC byla zaregistrována v rámci DCMI (Dublin Core Metadata Initiative) – viz [2]. Pro iniciativu Dublin Core byly vytvořeny české webové stránky – viz [3]; pracuje se na rozšíření českých stránek iniciativy DC o další dokumenty. Byly zpracovány přehledové analýzy z oblasti vývoje a využití standardu Dublin Core – viz [4], [5].

Byla vytvořena beta-verze lokalizovaného DC-metadatového formuláře (převzatého od Helsinské univerzitní knihovny z projektu Nordic Metadata). Tento nástroj podporuje kvalifikovaný DC podle nejnovější specifikace a zároveň umožňuje propojení na nástroj pro automatické přidělování URN (není zatím lokalizováno). 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, na českých stránkách iniciativy DC.

Standard Dublin Core byl již v letošním roce propagován na významných odborných konferencích (RUFIS 2000, Datasem 2000) a z ohlasu lze usuzovat, že tento standard najde mezi autory dokumentů a vydavateli své příznivce. Ukázkové webové zdroje s vloženými DC-metadaty jsou prezentovány na webu – např. [6].

[1] *Dublin Core Czech : soubor metadatových prvků Dublin Core, verze 1.1 : referenční popis* [online]. Brno : Masarykova univerzita, 12-06-2000 [cit. 11. listopadu 2000]. Dostupné na World Wide Web: <http://www.ics.muni.cz/dublin_core/DC-czech-1.1.html>.

[2] *DCMI Multiple Languages Special Interest Group* [online]. DCMI, c2000 [cit. 11. listopadu 2000]. Dostupné na World Wide Web: <<http://purl.org/dc/groups/languages.htm>>.

[3] *Dublin Core Czech* [online]. Brno : Masarykova univerzita, posl. aktualizace 31. října 2000 [cit. 11. listopadu 2000]. Dostupné na World Wide Web: <http://www.ics.muni.cz/dublin_core/>.

[4] 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 a Masarykova univerzita, 2000 [cit. 8. listopadu 2000]. Dostupné na World Wide Web: <<http://www.vutbr.cz/konference/rufis2000/sbornik/25-zabicka.pdf>>.

[5] ŽABIČKA, Petr. *Dublin Core – metadata pro popis elektronických dokumentů*. Předneseno na konferenci DATASEM 2000, konané 21. až 24. října 2000 v Brně. Nepubl. 6 s.

[6] URL: <http://www.mzk.cz/DC/rufis2000.html> a http://www.ics.muni.cz/dublin_core/DC-czech-1.1.html.

C.1.4 Harvester

Je analyzován aktuální stav, možnosti a podmínky lokalizace nástroje z projektu NEDLIB na Linux-serveru, který bude instalován v závěru roku z prostředků projektu.

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 napsaný 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.

Tento nástroj má relativně vysoké nároky na hardwarové vybavení pro svůj provoz. Vzhledem k jeho specifikům 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 a má být hotova až na konci roku. Také specifikace množiny dokumentů pro archivaci (specifikace „českého webu“) bude relativně náročná.

C.1.5 Ověřovač URL

Vzhledem k chybějícímu zadání bylo provedeno jen velmi 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.

webxref: jednoduchý program, dokáže kontrolovat interní a volitelně i externí HTML odkazy. Kromě upozornění na neexistující odkazy může volitelně vygenerovat souhrnnou zprávu o všech odkazech různých typů (interní, externí, obrázky, cgi skripty, mailto, ftp). Pomocí regulárních výrazů je možno specifikovat množiny odkazů, které se mají z kontroly vyjmout. Program je volně k dispozici a byl řešiteli testován. Je použitelný pro menší systémy odkazů, u rozsáhlejších systémů představuje nasazení tohoto systému příliš velký objem ruční práce pro ověřování podezřelých vazeb.

Linklint2.1: vychází z výše uvedeného programu webxref, nabízí však mnohem širší nabídku voleb a možností konfigurace. Je možno do větších detailů specifikovat, jaké druhy zpráv má program generovat, je možné vytvářet konfigurační soubory pro

opakované dávkové kontroly celých webových systémů. Jedná se o shareware (cena pro jednotlivce 20 USD, organizace 300 USD), testování plně funkční verze je možné zdarma.

Další perspektivní možností je využití existujících volně dostupných perlovských modulů k vytvoření aplikace přesně podle potřeb projektu. V celosvětově dostupném archivu CPAN jsou k dispozici moduly pro HTML parsing, komunikaci pomocí HTTP protokolu a mnoho dalších. Samotná aplikace tak může být poměrně jednoduchá, záležitosti týkající se formátů a protokolů mohou být v kompetenci hotových perlovských modulů.

C.2 Závěr

Byla analyzována řada zahraničních i mezinárodních projektů zabývajících se problematikou získávání, registrace, ochrany a zpřístupňování elektronických on-line 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 v projektu 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 opominout, je záležitost dlouhodobá, k níž bude nutné nejprve 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)*)

C.3 Návrhy opatření

1. Pokračovat v řešení pilotního projektu v roce 2001 s cílem připravit podmínky pro realizaci bibliografické kontroly v oblasti elektronických online dokumentů.
2. V rámci řešení usilovat o co nejvyšší efektivitu aplikací výsledků řešení podobných zahraničních či mezinárodních projektů, včetně aplikace volně dostupných nástrojů i nástrojů poskytovaných za úhradu.
3. V pokračování řešení zohlednit související projekty a činnosti v ČR
 - zpřístupňování národní článkové bibliografie a plných textů elektronických periodik
 - integrace automatizovaných systémů pro systematické a věcné pořádkání (Mezinárodní desetinné třídění)
 - digitalizace klasických knihovnických fondů
 - navrhovaný programový projekt Ministerstva kultury ČR „Jednotná informační brána pro hybridní knihovny“
4. Včasné přidělení finančních prostředků na pokračování projektu, aby nedocházelo k silnému časovému stresu v řešení a ke kumulaci prací na projektu pouze do druhé poloviny roku.

D RESUMÉ A KLÍČOVÁ SLOVA

D.1 Resumé

Projekt se zabývá problematikou, která je v České republice i v zemích bývalého východního bloku zcela nová. Jeho cílem je připravit podmínky pro zpracování české národní bibliografie elektronických zdrojů, se zaměřením zejména na zdroje dálkově přístupné. S bibliografickým zpracováním souvisí zajištění trvalého uchování domácích elektronických dokumentů monografických i seriálových, publikovaných v síti Internet, a jejich zpřístupnění, respektující autorská práva.

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.

Při řešení je do značné míry využíváno zkušeností a výsledků zahraničních i mezinárodních projektů, zejména z evropských severských zemí (Nordic Metadata I, II, Nordic Web Archiv, NEDLIB).

D.2 Klíčová slova

*národní bibliografie * ochrana knihovních fondů * elektronické zdroje * Internet * elektronický archiv * legislativní dokumenty * povinný výtisk * autorské právo * akvizice * zpřístupňování dokumentů * indexace * mezinárodní standardy * metadata * Dublin Core*

E PŘÍLOHY

- E.1 Domácí odborně zaměřené elektronické časopisy vybrané k testování metadatového schématu Dublin Core**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e1.pdf>

- E.2 Přehled situace ve zpracování síťových elektronických publikací ve vybraných zemích**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e2.pdf>

- E.3 Překlad Dublin Core Element Set, Version 1.1**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/dc.htm>

- E.4 Generátor záznamu metadat ve schématu Dublin Core vyvinutý v rámci Nordic Metadata Project**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e4.pdf>

- E.5 Vygenerovaný záznam metadat ve schématu Dublin Core pomocí aplikace vyvinuté v rámci Nordic Metadata Project**

http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e5_e6_e8_e11.pdf

- E.6 Záznam metadat ve schématu Dublin Core ve struktuře Resource Description Framework**

http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e5_e6_e8_e11.pdf

- E.7 Zpracování metadat článku z elektronického časopisu Ikaros pomocí programu Metabrowser**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e7.pdf>

- E.8 Záznam metadat ve schématu Dublin Core vložený pomocí programu Metabrowser do zdrojového kódu článku v elektronickém časopisu Ikaros**

http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e5_e6_e8_e11.pdf

- E.9 Generátor jednoznačného identifikátoru URN vyvinutý v rámci Nordic Metadata Project**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e9.pdf>

- E.10 Konvertor metadat ve schématu Dublin Core do formátů typu MARC vyvinutý v rámci Nordic Metadata Project**

<http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e10.pdf>

- E.11 Záznam metadat ve schématu Dublin Core zkonvertovaný do formátu USMARC**

http://www.webarchiv.cz/files/dokumenty/zpravy/zprava2000/e5_e6_e8_e11.pdf