

## ZBIRANJE IN ARHIVIRANJE SLOVENSКИH ELEKTRONSKIH PUBLIKACIJ NA MEDMREŽJU

*Eva Kodrič - Dačić \**

**UDK: 004(497.4):930.25**

*Eva Kodrič - Dačić: Zbiranje in arhiviranje slovenskih elektronskih publikacij na medmrežju. Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja, Zbornik referatov z dopolnilnega izobraževanja, Maribor 4/2005, št. 1, str. 501-511.*

*Izvirnik v slovenščini, izvleček v slovenščini in angleščini, povzetek v angleščini.*

Z razvojem tehnologije se spreminjajo tudi nosilci zapisov informacij, s tem pa tudi načini zbiranja in hranjenja publikacij. Elektronske publikacije na tradicionalnih nosilcih zapisov in tiste, ki so dosegljive na spletu, so, tako kot dokumenti, izdelani na papirju, del naše kulturne dediščine, in naloga nacionalne knjižnice je, da jih zbira in hrani. V prispevku je predstavljen projekt Zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju, ki sta ga v okviru Ciljnih raziskovalnih programov leta 2003 začela izvajati Narodna in univerzitetna knjižnica in Institut Jožef Stefan. Narodna in univerzitetna knjižnica je izdelala teoretične podlage, v katerih je opredelila vrste elektronskih publikacij, ki so dostopne na spletu, kriterije, po katerih naj bi jih zbirala, določila standarde za bibliografski popis publikacij ter avtorsko-pravne aspekte zbiranja, hranjenja in reproduciranja. Del projektne skupine iz Instituta Jožef Stefan pa je izdelal robota (WebBird), s katerim bo publikacije mogoče zajemati.

**UDC: 004(497.4):930.25**

*Eva Kodrič - Dačić: Collecting and Archiving Slovenian Online Publications. Technical and Field Related Problems of Traditional and Electronic Archiving. Conference Proceedings, Maribor 4/2005, No. 1, pp. 501-511.*

*Original in Slovenian, abstract in Slovenian and English, summary in English.*

Due to the technical development the information carriers are changing rapidly. Therefore, the ways of acquiring and archiving publications have to change, too. Electronic publications on portable carriers as well as those which are distributed online via World Wide Web, are part of the national heritage. So the national library has the task to collect and preserve them just like printed publications. The paper presents the project Methodology of Collecting and Archiving Slovenian Online Publications, which was launched in 2003 by the National and University Library of Slovenia and the Jožef Stefan Institute. The task of the national library research group was to define the online electronic resources, which should be collected, to provide the selection criteria, to define metadata and to analyse copyright issues. The researchers from the Jožef Stefan Institute programmed the harvester called WebBird.

### UVOD

Z razvojem tehnologije se spreminjajo tudi nosilci zapisov informacij in z njimi načini zbiranja, obdelave in hranjenja. Elektronske publikacije na statičnih nosilcih zapisov in tiste, ki so dosegljive na medmrežju, so tako kot knjižno gradivo del naše kulturne dediščine in naloga nacionalnih knjižnic je, da jih zbirajo in hranijo. Slovenska nacionalna knjižnica - Narodna in univerzitetna knjižnica (v nadaljevanju NUK) je na to opozorila že v strokovnih predlogih osnutka novega zakona o obveznem izvodu publikacij, ki je nastal v prvi polovici devetdesetih let, kasneje pa je z objavo

---

\* Dr. Eva Kodrič - Dačić, bibliotekarski svetnik, vodja Centra za razvoj knjižnic, Narodna in univerzitetna knjižnica, Turjaška 1, 1000 Ljubljana, Slovenija.

dveh brošur<sup>1</sup> poskušala o njihovem pomenu ozavestiti tudi javnost. Prvi korak k dejanskemu zbiranju elektronskih publikacij je knjižnica naredila s prijavo projekta Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju, ki ga je prijavila na razpis v okviru Ciljnih raziskovalnih programov leta 2002. Predlog projekta sta izdelali skupini raziskovalcev iz NUK-a in Instituta Jožef Stefan. NUK je bil v okviru projekta zadolžen za izdelavo teoretičnih podlag, v katerih je določil vrste elektronskih publikacij, dostopnih na medmrežju, ki bodo predmet zbiranja, kriterije, po katerih naj bi jih zbirali, določil standarde za bibliografski popis publikacij ter pripravil predloge rešitev za avtorsko-pravne probleme, ki se lahko pojavljajo v zvezi z zajemanjem, hranjenjem in reproduciranjem tega gradiva. Raziskovalci iz Instituta Jožef Stefan pa so izdelali robota (WebBird), računalniški program za zajemanje publikacij z medmrežja.

Rok za zaključek projekta je bil september 2004, vendar NUK še vedno testira zmogljivosti in karakteristike robota.

### **ČEMU SPLOH ZBIRATI IN ARHIVIRATI ELEKTRONSKE PUBLIKACIJE, DOSTOPNE NA MEDMREŽJU?**

Odgovor na vprašanje, čemu zbirati in arhivirati elektronske publikacije na medmrežju, fluidno gradivo, ki je pogosto dvomljive vrednosti, izhaja iz samih temeljev naše kulture. Naše spoštovanje do pisane besede izhaja iz časov, ko so prvi zapisani teksti prihajali iz božjih ust ter iz ust njihovih odposlancev in namestnikov na zemlji. Beseda, zapisana na pergament ali papir, vezan v obliko kodeksa, je definirala našo kulturo do te mere, da še danes, ne glede na vse zlorabe zapisane besede v preteklih stoletjih, dojemamo njeno uničenje ali prepoved kot enega izmed najhujših etičnih prekrškov. Knjiga, to je publikacija v obliki kodeksa, ima namreč v naši kulturi poseben pomen in vrednost, ki sta do neke mere neodvisna od pomena njene vsebine oziroma, kakor bi se danes izrazili v strokovnem jeziku, informacije, ki jo vsebuje. Tehnična definicija publikacije, iz katere smo izločili njen kulturni presežek namreč opisuje publikacijo kot zapis informacije na kateremkoli mediju, ki je razmnožen in namenjen javnosti. Ta sicer korekten opis določa le konstitutivne dele publikacije, ki pa so že sami po sebi kompleksni predmeti, okoli katerih so se razvile cele znanstvene discipline. Definicija obide semantičen nivo zapisa, njegov kodikološki nivo (jezik in pisava, v katerem je vsebina zapisana), na materialnem nivoju pa abstrahira tehnike zapisa, materiale, uporabljene za zapis (tinte, tuše, tiskarske barve, s katerimi je informacija zapisana), medij, na katerem je nastal zapis (papir, pergament, magnetni trak ...), način reproduciranja publikacije, tehnike vezave in podobno.

Knjiga nam kot kompleksni kulturni predmet torej ne prenaša le sporočila njenega avtorja, temveč s svojo materialno pojavo priča o kulturnem in tehnološkem razvoju časa, v katerem je nastala. Papir in vodni znak na njem nam ne sporočata le, kdaj in kje je knjiga po vsej verjetnosti nastala, temveč nam govorita tudi o tehnološki razvitosti obdobja in območja. Enako je s tehnikami vezave in tiska. Dojemati publikacijo (knjigo) le kot zapis informacije, čeprav je ta ključnega pomena, pomeni sprejemati le del njenega sporočila.

---

<sup>1</sup> *Elektronske publikacije: kako jih bomo ohranili za prihodnost? Ljubljana, 1998 in Elektronske publikacije: Kodeks prakse prostovoljnega depozita. Ljubljana, 2001.*

Z razvojem novih načinov zapisov in medijev, nosilcev informacij (magnetni trak, nosilci digitalnih zapisov) se je sicer povečala hitrost zapisovanja in razmnoževanja, obenem pa se je skrajšala njihova življenjska doba. Relativno kratka ni le življenjska doba njihovih materialnih nosilcev, temveč tudi vsebine publikacije. Kakor da bi lahkotnost njihovega nastanka pogojevala tudi obstojnost njihove vsebine: elektronske publikacije na medmrežju so nam danes sicer lahko dostopne, vendar lahko naslednji dan že izginejo, se dopolnijo, skratka, lahko se spremenijo v nekaj drugega. Ta fluidnost publikacij na modernih nosilcih zapisov in cenenost izdelave publikacije vpliva na naše dojemanje njene vrednosti. Videokasete, CD-ROM-i in publikacije, ki so dosegljive na medmrežju, so izgubile vrednostni piedestal knjige. Če je danes še vedno etično sporno uničiti knjigo pa zavrženega CD-ROM-a, ki je lahko vseboval celotno knjižnico, sploh ne opazimo.

Elektronske publikacije, tako tiste na fizičnih nosilcih zapisa (CD-ROM, DVD in podobno), kot tudi tiste, ki so dostopne na medmrežju, s svojo avtentično pojavnostjo in s svojo vsebino pričajo o našem kulturnem razvoju. Tako kot se danes navdušujemo nad starimi vezavami, kvaliteto ročno izdelanega papirja in mojstrstvom starih tiskarjev, se bodo čez stoletja navduševali nad formati zapisa, specifiko programske opreme in tehnikami reproduciranja, seveda, če nam bo uspelo sodobne zapise ohraniti za prihodnost.

Nacionalne knjižnice, katerih naloga je zbiranje pisne kulturne dediščine naroda imajo med nacionalnimi kulturnimi institucijami posebno mesto. Če druge kulturne institucije zbirajo, hranijo in prezentirajo le značilne, ekskluzivne ali selekcionirane primerke oziroma zbirke (muzeji, arhivi, galerije) oziroma dokazujejo svoj nacionalni pomen s kvaliteto svojih storitev (filharmonija, opera, gledališče), so nacionalne knjižnice dolžne graditi zbirko patriotike. Ta zajema brez izjeme vse publikacije, ki so nastale na ozemlju nekega naroda, publikacije, katerih avtorji so pripadniki tega naroda in so napisane v njegovem jeziku ali pa se po svoji vsebini nanašajo na narod in njegovo nacionalno ozemlje. Zbirka patriotike je osnova za izdelavo popisa celotne pisne produkcije nekega naroda, to je nacionalne bibliografije.

Izgradnjo zbirke patriotike omogoča mehanizem obveznega izvoda tiskov. Obvezni izvod tiska oziroma publikacije je privilegij, na osnovi katerega nacionalna knjižnica po zakonu brezplačno prejema vse tiske oziroma publikacije, ki so jih v državi založili založniki, natisnili tiskarji ali izdali avtorji. Obvezni izvod praviloma temelji na zakonskih predpisih, zato je omejen na območje države, v kateri je izdan. Prvi zakonski predpis te vrste je bil uveljavljen v Franciji v prvi polovici 16. stoletja, torej manj kot 100 let po iznajdbi tiska. Razlogi zanj so bili cenzurne narave, saj naj bi natisnjeni izvod dokazoval, da so bile spremembe cenzorjev upoštevane oziroma, da bi v nasprotnem primeru preprečil distribucijo spornega dela. Vendar nadzor tiska ni bil edini razlog za uvedbo mehanizma obveznega izvoda. ZDA ali Nizozemska na primer so obvezni izvod tiska uvedle kot dokazilo o avtorstvu. Nekatere države pa so ta privilegij preprosto priznale svojim kraljevim ali cesarskim knjižnicam, ali pa so ga uvedle, tako kot na Slovenskem, za podporo študiju. Licejska knjižnica v Ljubljani je tako kot prva knjižnica na Kranjskem dobila privilegij obveznega izvoda tiska že leta 1807 in ta privilegij, seveda vezan na različna geografsko-administrativna območja, je ohranila vse do danes.

Obvezni izvod tiska je bil sprva omejen le na tiskane publikacije: knjige, časnike, časopise, zemljevide in bakroreze. V drugi polovici devetnajstega stoletja najdemo v zakonu, ki je določal predmet obveznega izvoda že široko definicijo tiska,

ki je obsegala vse izdelke s področja umetnosti in literature, ki so razmnoženi z mehničnimi ali kemičnimi sredstvi. Izvzete so bile le interne publikacije in nekateri drobni tiski, saj je zakon iz obveznosti oddaje izključeval publikacije, ki so bile namenjene potrebam trgovine in prometa ali domači uporabi, kot so na primer obrazci, računi ali posetnice. Definicija iz zakona je predmet obveznega izvoda že približala današnjemu razumevanju publikacije. To danes opisujemo kot zapis informacije na mediju (dokument), ki je razmnožen in namenjen za javno uporabo. S pojmom »publicirati« označujemo dejanje, s katerim so postale informacije oziroma intelektualna stvaritev dostopne javnosti, s pojmom »razmnoževanje« pa vsako multipliciranje informacije oziroma intelektualne storitve, pa naj si bo ta izvedena na klasičen način (tisk) ali v elektronski obliki, kjer za izvod šteje vsakokratno sprejemanje radijskih in televizijskih programov preko radijskih in televizijskih sprejemnikov oziroma dostopanje uporabnika do domače strani.

Posebna značilnost obveznega izvoda je dejstvo, da razen omenjenih kriterijev, ki so vezani na definicijo publikacije, zavezanca in administrativno območje, kjer se zakonske določbe uveljavljajo, ne vsebuje selekcijskih kriterijev ne v formalnem ne v vsebinskem pogledu. Obvezni izvod tiska namreč ni pomemben zaradi posamičnih izjemnih ali dragocenih del, ki jih je nacionalna knjižnica na osnovi predpisa pridobila. Vrednost obveznega izvoda je v izčrpnih zbirkah, ki kaže na vsakokratno kulturno, gospodarsko stanje naroda in je materialni dokaz o moči njegove kulture. Kot tak je obvezni izvod nenadomestljiv in relevanten vir za zgodovinopisje, sociološke, kulturološke in druge študije. Gradivo, pridobljeno na osnovi predpisov o obveznem izvodu tiska, ima zato v Sloveniji po veljavnem Zakonu o knjižničarstvu<sup>2</sup> status kulturnega spomenika.

Elektronske publikacije so v razvitih državah postale predmet zakonov o obveznem izvodu že v začetku devetdesetih let prejšnjega stoletja. To velja tako za elektronske publikacije na fizičnih nosilcih kot za tiste, ki so dosegljive le na medmrežju. Prva država, ki je z zakonom predpisala zbiranje elektronskih publikacij na medmrežju, je bila Norveška leta 1989. Francija, Danska, Kanada in Avstralija so zakonodajo o obveznem izvodu elektronskih publikacij gradile postopoma. Najprej so predpisale zbiranje obveznih izvodov na fizičnih nosilcih, kasneje pa so jo dopolnile še z določbami o obveznosti oddaje oziroma zbiranja publikacij, dostopnih na medmrežju. Leta 2002 in 2003 sta take določbe v svojo zakonodajo vključili tudi Nemčija in Velika Britanija.

Vendar sodobna slovenska zakonodaja, ki ureja oddajo obveznih izvodov publikacij, ni sledila razvoju in razmahu novih medijev (radio, televizija, medmrežje). Uzakonjeno ni niti obvezno zbiranje in hranjenje slovenskih radijskih in televizijskih programov, vse številčnejše slovenske publikacije, ki so nam danes dostopne preko medmrežja, pa izginjajo dnevno in brez opozorila, na enak način, kot se pojavljajo.

## IZZIVI ELEKTRONSKIH PUBLIKACIJ KOT OBJEKTOV ZBIRANJA IN HRANJENJA

S pojavom elektronskih publikacij na medmrežju knjižničarji na materialni in na konceptualni ravni izgubljammo objekt svojega dela. Na materialni zato, ker objekt v otipljivem predmetnem smislu ne obstaja več, v konceptualnem pa zato, ker je publikacija v elektronski obliki izgubila klasične attribute kulturnega predmeta. Za knjižničarje neknjižno gradivo seveda ni nobena novost. Prvi dokumentirani problemi

<sup>2</sup> Zakon o knjižničarstvu. Uradni list RS, 2001, l. 11, št. 87, str. 8685-8693.

so nastali z razmahom fotografije v drugi polovici devetnajstega stoletja, ko so se pojavili predlogi, naj bi obvezni izvod tiska razširili tudi na fotografije. Vendar fotografija ni izpolnjevala enega izmed ključnih kriterijev: ni bila namenjena javnosti. Temeljni premik pri zbiranju in hranjenju publikacij je nastal že s publikacijami v strojno berljivi obliki. Tehnološki razvoj, ki je v dvajsetem stoletju prinašal vedno nove in nove izume na področju nosilcev zapisov, je pred knjižničarje postavljala domala nerešljive probleme, povezane z dolgoročnim ohranjanjem in reproduciranjem neknižnega gradiva.

V nasprotju s klasičnim knjižnim gradivom sama spretnost branja in poznavanje jezika naenkrat ni bila več dovolj. Za razumevanje zapisane informacije je bil potreben vmesnik, ki nam omogoča dekodiranje zapisa v čutno sprejemljivo obliko. S tem se je pojavil nov faktor, ki je pomembno vplival na dolgoročno hranjenje publikacij: dostop do zapisanih informacij ni bil omejen le z obstojnostjo medija, temveč tudi z ohranjenostjo aparature za dekodiranje. V dvajsetem stoletju so se že dolgo pred izumom osebnega računalnika, digitalnih zapisov in elektronskih publikacij, pojavili strojno berljivi nosilci zapisov. Vinilske plošče in magnetni avdio in video trakovi so postali del obveznega izvoda publikacij in s tem tudi del nacionalne zbirke. Celo več: dvajseto stoletje je med obvezne izvode publikacij vključilo tudi radijske in televizijske programe, ki jih načrtno in sistematično zbirajo in hranijo razvite evropske države.

Kljub tej pomembni novosti pa so bile tudi te nove publikacije, prav tako kot knjižno gradivo, statične. Statični, to je materialni nosilci zapisa namreč ohranjajo vsebinsko in oblikovno nespremenjene dokumente oziroma publikacije, pa najsi imamo v mislih rokopise, videokasete, knjige ali radijske oddaje. Vsaka sprememba zapisa ali forme pomeni novo izdajo in s tem tudi novo publikacijo. To ni sicer nič novega, saj v analognem okolju poznamo fenomen različnih izdaj predvsem pri enciklopedijah, leksikonih in slovarjih. V to skupino sodijo tudi elektronske publikacije na statičnih nosilcih zapisov. Pri elektronskih publikacijah na medmrežju pa je značilna prav njihova fluidnost: podatkovne zbirke se neprestano dopolnjujejo, domače strani spreminjajo grafično podobo in vsebino ter celo svojo lokacijo. Raziskave o obstojnosti domačih strani so pokazale, da jih le 56% preživi časovno obdobje enega leta in le 35% dve leti.<sup>3</sup> Zaradi teh značilnosti so postale elektronske publikacije povsem nova kategorija knjižničnega gradiva, ki se nam izmika iz klasične tipologije in bibliografske obdelave publikacij in je problematična tako glede zbiranja, dolgoročnega hranjenja, reproduciranja in možnosti javne uporabe.

## ZBIRANJE ELEKTRONSKIH PUBLIKACIJ NA MEDMREŽJU

Elektronskih publikacij na medmrežju ni mogoče preprosto in brez dodatnih določb vključiti med gradivo, ki so ga zavezanci dolžni dostavljati nacionalni knjižnici na osnovi zakona o obveznem izvodu publikacij. Njihovo zbiranje, hranjenje in dajanje v uporabo je potrebno opremiti z varovali, da ne bi pri tem oškodovali nosilcev avtorskih pravic na eni ali ogrozili zbiranje kulturne dediščine in pravico državljanov do dostopa do publikacij in informacij na drugi strani. Publikacija na papirju je nespremenljiva in do iznajdbe kopirnega stroja pred nekaj desetletji je praktično ni bilo mogoče enostavno in poceni kopirati. Elektronske publikacije pa je mogoče brez večjih stroškov ne le kopirati in razmnoževati, temveč tudi poljubno

---

<sup>3</sup> Več o tem glej Kavčič - Čolič, A. *Arhiviranje elektronskih publikacij: nov izziv za nacionalne knjižnice*. Knjižnica, 2004, l. 48, št. 4, str. 97-119.

spreminjati. Predvsem zaradi hitrosti in lahkote, s katero je mogoče te postopke izvesti, avtorji in založniki teh publikacij upravičeno zahtevajo od knjižnic s privilegijem obveznega izvoda posebno zaščito oziroma omejitev dostopa do publikacij.

Da bi evropskim deponitarnim knjižnicam olajšali zbiranje elektronskih publikacij na medmrežju v prehodnem obdobju, dokler ne bo deponit elektronskih publikacij ustrezno rešen, sta Združenje direktorjev evropskih nacionalnih knjižnic (CENL) in Združenje evropskih založnikov (FEP) pripravila »Kodeks prakse prostovoljnega deponita«<sup>4</sup>. Načela kodeksa naj bi nacionalne knjižnice upoštevale do zakonske rešitve oddaje obveznega izvoda. Kodeks opredeljuje temeljna načela deponita, ki izhajajo iz klasičnih določil, uveljavljenih v analognem okolju in že omenjenih posebnosti elektronskih publikacij na medmrežju. Tako naj bi bile elektronske publikacije deponirane v obliki, v kateri so dostopne javnosti ter skupaj s potrebno programsko opremo, priročniki in spremnim gradivom. Če na primer publikacija izide v več formatih, naj bi imel pri deponitu prednost format, ki je kompatibilen z IBM PC. Prav tako naj bi založnik pri deponitnih kopijah odstranil vgrajeno zaščito, kot je na primer datum poteka uporabnosti.

Posebna navodila se nanašajo na zagotavljanje dolgoročne dostopnosti in veljavnosti elektronskih publikacij. Fizični nosilci elektronskih zapisov (na primer CD-ROM-i) nimajo dokazanega zadostnega roka trajanja. Zato morajo imeti deponitarne knjižnice možnost, da elektronske publikacije kopirajo na druge medije ali izvedejo migriranje na nove tehnološke platforme.

Tudi pogoji dostopa in uporabe elektronskih publikacij naj bi bili v zakonu posebej opredeljeni. Dostop do tega gradiva naj bi bil enouporabniški, za kar bi morala nacionalna knjižnica zagotoviti ustrezne tehnične ukrepe. To pomeni, da knjižnica omogoča dostop do elektronske publikacije le enemu uporabniku naenkrat, in sicer prek varovanega intranetnega omrežja na posebej določenem računalniku, v območju, ki je dostopno le pooblaščenim uporabnikom določene deponitne knjižnice. Založnik naj bi imel pravico v času do sprejema zakona v okviru pogodbe z nacionalno knjižnico določati pogoje, obliko in frekvenco dostopanja uporabnikov do elektronske publikacije. Glede izdelave kopij na papirju naj bi veljala enaka načela kot pri kopiranju tiskanega gradiva. Uporabnik knjižnice torej praviloma ne bi smel natisniti elektronskih publikacij v celotnem obsegu. Podobno naj bi se tudi pri elektronskem presnemavanju elektronskih publikacij zgedovali po omejitvah, ki so uveljavljene v tiskanem okolju, sicer pa naj bi bilo elektronsko presnemavanje, shranjevanje in nadaljnja uporaba teh del predmet pogodbenih licenc.

## PODROČJE TERMINOLOGIJE IN TIPOLOGIJE

Uveljavljena tipologija publikacij je do sedaj temeljila na štirih kriterijih: načinu izhajanja publikacije (monografske in serijske publikacije ...), nosilcu zapisa (knjižno in neknjižno gradivo, avdiovizualno gradivo ...), načinu zapisa (kartografsko gradivo, knjige, brošure, notno gradivo ...), po vsebini (uradne publikacije, patenti ...). Elektronske publikacije na medmrežju le deloma sodijo v to shemo. Ker so elektronske publikacije na medmrežju relativno mlade, prihaja do zadreg že pri njihovem poimenovanju. Ne le različni avtorji, celo mednarodni standardi uporabljajo na tem področju različne terminološke aparate. Srečujemo se z izrazi,

<sup>4</sup> *Elektronske publikacije: Kodeks prakse prostovoljnega deponita. Ljubljana, 2001.*

kot so digitalni objekt, digitalni vir, digitalni dokument, elektronski vir, e-besedila, ki so, odvisno od konteksta oziroma stroke obravnave včasih uporabljeni kot sinonimi, drugič spet pa so med njimi jasne pomenske razlike.

Ključni problem pri razvrščanju elektronskih publikacij na medmrežju izhaja iz njihove relativne neobstoynosti, saj se predvsem posamezni tipi publikacij na medmrežju neprestano spreminjajo. Mednarodni standardi za bibliografsko obdelavo knjižničnega gradiva in knjižnično statistiko, ki so bili izdani v zadnjih petih letih, prinašajo različne termine in tipologijo. Tako na primer mednarodni standardni bibliografski opis elektronskih virov ISBD (ER) iz leta 2000 razlikuje med elektronskimi viri na fizičnem nosilcu in med gradivom, ki je dostopno na daljavo (on-line). Leta 2002 objavljeni standard iz iste skupine, mednarodni standard za bibliografsko obdelavo virov ISBD (CR), je uvedel delitev na zaključene in nezaključene vire. Problematiko razvrščanja elektronskih publikacij je razrešil s pojmom integrirnega vira. Standard ločuje med zaključenimi nekompletnimi integrirnimi viri (revidirana e-besedila, nekatere nevezane strani z zamenljivo vsebino) in nezaključenimi (kontinuiranimi) integrirnimi viri (večina nevezanih strani z zamenljivo vsebino, baze podatkov, spletne strani). Vendar pa bo preteklo še nekaj časa, da se bo ta terminologija uveljavila v strokovnih krogih. Standard za knjižnično statistiko ISO 2789 : 2003, ki je izšel le leto kasneje, omenjenega izrazoslovja ne pozna.

## **PODROČJE BIBLIOGRAFSKE OBDELAVE IN IDENTIFIKACIJE ELEKTRONSKIH PUBLIKACIJ**

Bibliografska obdelava publikacij na formalnem nivoju je določena s standardi ISBD, na vsebinskem nivoju pa s klasifikacijskimi sistemi (univerzalna decimalna klasifikacija - UDK; Dewey). Vendar običajni bibliografski podatki, ki obravnavajo območje naslova in navedbe odgovornosti, izdaje, zvrsti in dosega vira, založništva in distribucije, fizičnega opisa, zbirke, opomb ter območje standardne številke in pogojev dostopnosti, ne zadoščajo več za identifikacijo elektronskih publikacij na medmrežju. Potrebni je še vrsta podatkov, ki izhajajo iz same specifičnosti nosilca informacij in iz načina zapisa. To so tako imenovani metapodatki (podatki o podatkih), ki so se pojavili prav z razvojem elektronskih virov. Omogočajo nam njihovo identifikacijo in lokalizacijo ter obenem vključujejo podatke o strukturi, delovanju, okolju in o pogojih za dostop do njih. Glede na njihovo vsebino se uveljavlja delitev na deskriptivne, strukturne, tehnične in administrativne metapodatke. Ker je naš namen ne le zbirati, temveč tudi trajno hraniti elektronske publikacije in omogočati dostop do njih, potem predvsem ne smemo pozabiti na metapodatke, ki vsebujejo informacije o formatu oziroma o strukturi bitnega toka elektronske publikacije in podatke o aplikaciji, s katero lahko dostopamo do dokumenta<sup>5</sup>.

## **PROJEKT METODOLOGIJA ZBIRANJA IN HRANJENJA ELEKTRONSKIH PUBLIKACIJ NA MEDMREŽJU**

Potem ko je NUK že predlagal razširitev oddaje obveznih izvodov publikacij tudi na elektronske publikacije na medmrežju in po poskusih informiranja javnosti in založnikov o nujnosti hranjenja le-teh, je prešel na reševanje praktičnih problemov

---

<sup>5</sup> O tem izčrpno piše Alenka Kavčič - Čolič v svojem članku *Metapodatki za trajno ohranjanje elektronskih virov* (Knjižnica, 2004, l. 48, št. 4, str. 97-119).

njihovega zbiranja, obdelave in hranjenja. Leta 2002 je na razpis Ciljnega raziskovalnega programa »Konkurenčnost Slovenije 2001-2006« prijavil projekt Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju. Projekt je bil sprejet in financerji (Ministrstvo za šolstvo, znanost in šport, Ministrstvo za kulturo, Ministrstvo za informacijsko družbo) so za njegovo realizacijo namenili 11.998.000,00 SIT. Projektna skupina je združevala raziskovalce NUK-a (Eva Kodrič - Dačić kot nosilka projekta, Alenka Kavčič - Čolić, Vilenka Jakac - Bizjak, Irena Kavčič, Renata Šolar, Dunja Kalčič) in Instituta Jožef Stefan (Dunja Mladenič, Marko Grobelnik, Maja Škrjanc, Polona Domadenik - Raziskovalni center Ekonomske fakultete v Ljubljani). Projekt naj bi bil realiziran v času od 1. 10. 2002 do 30. 9. 2004.

Cilj tega projekta je bil izdelati model za sistematično zbiranje in arhiviranje slovenskih elektronskih publikacij na medmrežju. Naloga raziskovalcev NUK-a je bila določiti objekt zbiranja, to je vrste elektronskih publikacij na medmrežju, ki jih bomo zbirali, kriterije, po katerih jih bomo zbirali, formate in normative za zbiranje in hranjenje, formate za vnos bibliografskih podatkov o gradivu ter analizirati avtorskopravno problematiko, povezano z zbiranjem, hranjenjem in dajanjem v uporabo. Izdelava programske aplikacije za zbiranje elektronskih publikacij na medmrežju pa je bila naloga raziskovalcev Instituta Jožef Stefan.

Zbiranje in arhiviranje elektronskih publikacij na medmrežju je v zadnjem desetletju ena izmed prioritetenih nalog nacionalnih knjižnic. Zato smo se lahko zgledovali po tujih mednarodnih in nacionalnih projektih. V prvi vrsti so nas zanimali modeli sistemov za dolgoročno arhiviranje elektronskih publikacij (na primer Open Archival Information System: OAIS) in projekti, povezani z zajemanjem vseh vrst elektronskih publikacij (na primer ZDA: Internet Archive, Mapping the Internet: the Electronic Resources Virtual Archive-MINERVA, Avstralija: Preserving and Accessing Networked Documentary Resources of Australia-PANDORA, Evropa: Nordic Web Index-NORDINFO, KULTURARW3 - poskus zbiranja švedskih elektronskih dokumentov na Internetu, Nordic Web Archive-NWA, NEDLIB, skupni projekt evropskih nacionalnih knjižnic, francoski poskus zbiranja domačih strani in tako dalje).

Na podlagi tujih izkušenj, domače prakse in razpoložljivih virov smo se odločili, da bomo začeli zbirati oziroma arhivirati štiri vrste elektronskih publikacij, ki so dostopne na medmrežju: domače strani, serijske elektronske publikacije, monografske publikacije in kartografsko gradivo.

Osnovni kriteriji za zbiranje so bili povzeti po klasičnih vsebinskih kriterijih za zbiranje slovenike (navedeni so v enem prejšnjih poglavij). Skušali smo jih prilagoditi okolju interneta. Hipotetično naj bi zbirali vse elektronske publikacije na medmrežju, ki izpolnjujejo enega izmed naslednjih treh pogojev:

- domena \*.si,
- strežnik naj bi bil na ozemlju Republike Slovenije,
- slovenski jezik.

Za potrebe projekta smo definirali domačo stran kot vrsto spletne strani, ki je vhodna stran oziroma glavni dokument podjetja, ustanove ali posameznika na svetovnem spletu, preko katere imamo dostop do drugih njihovih oziroma njegovih spletnih strani. To je informacijska enota, ki je sestavljena iz ene ali več dostopnih http povezav in jo lahko obiščemo v celoti preko enega samega URL-ja. Domače strani naj bi zajemali neselektivno. Z domeno \*.si je mogoče zajeti vse vstopne strani



organizacij, uredništev, agencij, podjetij, društev, zvez in podobno. Po podatkih ARNES-a je bilo v juniju 2004 registriranih 17.381 takih domen. Domače strani naj bi zajemali do 4. nivoja. Selektivno naj bi zajemali slovenske domače strani z domenami \*.com, \*.net in \*.org. Ker je slovenski splet zelo majhen, ocena je približno 0,4 Tb, bi bilo možno zajemanje 2 do 3-krat letno.

Zdi se, da je od vseh publikacij na medmrežju še najbolj obvladljiv segment slovenskih serijskih publikacij. V septembru leta 2003, ko je bila v okviru projekta izvedena raziskava teh publikacij, je bilo v Sloveniji pri ISSN centru Slovenija<sup>6</sup> registriranih 184 elektronskih založnikov, in sicer: posamezniki (53), podjetja (46), športna društva in zveze (36) in razna društva (25), ki so izdajali okrog 200 naslovov elektronskih serijskih publikacij. Glede na vrsto publikacij je bilo največ, in sicer 82, naslovov elektronskih časopisov, ki so se dnevno ali tedensko prenavljali z dodajanjem vsebin oziroma besedil (Dnevnik, Delo, 24ur.com), ali pa so po zgledu papirnatih verzij predstavljali zaključene celote v obliki številčk ali letnikov (Gospodarski vestnik, Manager)<sup>7</sup>. V okviru projekta je bil izdelan tudi vzorčni seznam tipičnih serijskih elektronskih publikacij za testno zajemanje.

Bolj problematično bo zbiranje in hranjenje monografskih publikacij in kartografskega gradiva. V skupino monografskih publikacij spadajo predvsem normativni akti, pravilniki, letna in druga poročila, priložniki, navodila, programi, članki, literarna dela in podobno. Predvsem fakultete objavljajo vedno več študijskega gradiva v elektronski obliki, vendar so monografske publikacije tako razpršene po spletu, da bo potrebno pripraviti posebno analizo in načrt njihovega postopnega zajemanja.

Digitalno kartografsko gradivo na medmrežju lahko ločimo na dva dela: na digitalizirano kartografsko gradivo in izvorno digitalno kartografsko gradivo. Največja izdelovalca digitalnega kartografskega gradiva ter topografskih kartografskih baz sta Geodetski zavod Slovenije in Geodetski inštitut Slovenije, glavni založniki pa Geodetska uprava RS, Geodetski zavod Slovenije, Ministrstvo za obrambo RS, Ministrstvo za okolje in prostor RS. Kartografsko gradivo pa digitalizirajo tudi knjižnice same, (na primer NUK) in je prosto dosegljivo preko medmrežja. Pri izvornem digitalnem kartografskem gradivu gre za obsežne podatkovne zbirke, zato bi bila smiselna rešitev v dogovoru o dolgoročnem hranjenju med nacionalno knjižnico in institucijo, ki to gradivo proizvaja.

Za potrebe projekta smo se tudi pri definiranju metapodatkov naslonili na model OAIS (Open Archival Information System), ki je bil leta 2003 registriran kot standard ISO 14721 za izdelavo arhivskih sistemov elektronskih virov<sup>8</sup>. V okviru projekta je bil izdelan predlog nabora metapodatkov za trajno hranjenje slovenskih elektronskih publikacij. Nabor se deli na dva dela, in sicer na metapodatke, ki vsebujejo informacije o načinu prikazovanja vsebine (sem sodita opis načina prikazovanja elektronskega vira ter opis tehničnih značilnosti in okolja delovanja elektronskega vira) in na metapodatke, ki vsebujejo informacije, pomembne za trajno ohranjanje (sem sodijo referenčne informacije, informacije o kontekstu,

---

<sup>6</sup> Številka ISSN (International Standard Serial Number) je standardizirana mednarodna oznaka, ki omogoča identifikacijo publikacij. Slovenskim serijskim publikacijam jo dodeljuje ISSN center Slovenija, ki deluje v okviru NUK-a.

<sup>7</sup> Kodrič, N. *Analiza slovenskih elektronskih kontinuiranih virov in njihovih založnikov*. Ljubljana: NUK, september 2003 (raziskovalna naloga).

<sup>8</sup> Kavčič - Čolić, A. *Metapodatki za trajno ohranjanje elektronskih virov*. Knjižnica, 2004, l. 48, št. 4, str. 99.

informacije o provenienci in informacije o pristnosti)<sup>9</sup>. V projektu smo predvideli tudi možnost generiranja popisov elektronskih publikacij v obstoječi lokalni bazi COBIB, kar bi bilo mogoče izvesti z uvedbo dodatnih polj v bloku IX formata COMARC.

Zaradi spletnih strani, ki se neprestano spreminjajo, je bil v okviru projekta razvit prototipni spletni servis, ki zbira domače strani in jih sprotno ažurira oziroma dodaja vse spremembe, ki se zgodijo na posamezni spletni strani. V tem okviru so bili razviti naslednji moduli: robot za sprehajanje po spletu (WebBird), modul za shranjevanje celotnega stanja spleta, indeksiranje shranjenega arhiva, iskanje informacij na spletu (po vzoru standardnih iskalnikov) in modula za statistične analize spleta, spletnih strani, topološke strukture in evolucije spleta skozi čas. Aplikacijo so izdelali raziskovalci Instituta Jožef Stefan. Trenutno je v testni fazi druga verzija robota za zbiranje publikacij na spletu.

## ZAKLJUČEK

Projektna skupina je z realizacijo projekta Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju izpolnila zastavljene cilje. Vendar pa smo še daleč od utečenega in celovitega zajemanja publikacij z medmrežja. V analizah testnih zajemanj spleta smo ugotovili, da bomo predvsem pri domačih straneh težko dosegali optimalne cilje. Čeprav smo se odločili za neselektiven pristop k zbiranju, dokumenti zaradi svojih strukturnih posebnosti ne bodo zajeti v celoti. Nedostopne nam bodo ostale domače strani, ki imajo v ozadju interaktivne baze ali pa jih podpirajo posebne aplikacije (Java script) oziroma publikacije, ki imajo zaščiten dostop.

S projektom smo naredili šele prvi korak k vključevanju elektronskih publikacij na medmrežju v knjižnične zbirke. Te problematike pa seveda ni mogoče razrešiti v okviru enega samega projekta. Zbiranje, dolgoročno hranjenje in omogočanje dostopa do elektronskih publikacij na medmrežju so v tem času najpomembnejši izzivi bibliotekarske stroke. Čeprav elektronske publikacije nikoli ne bodo nadomestile knjige, so postale, neobstoje, fluidne in izmuzljive, del našega vsakdana. Njihove značilnosti odražajo specifično kulturo na prelomu tisočletja, zato jih moramo ohraniti za prihodnje generacije.

## LITERATURA

- *Guidelines for the preservation of digital heritage (draft). National Library of Australia; UNESCO, 2002.*
- Šolar, Renata. *Digitalno kartografsko gradivo, nov izziv kartografskih zbirk. Knjižnica, 2003, l. 47, št. 4, str. 7-22.*
- Kavčič - Čolić, Alenka. *Teoretični model digitalnih arhivov. Knjižnica, 2004, l. 48, št. 4, str. 63-75.*
- Kavčič - Čolić, Alenka. *Metapodatki za trajno ohranjanje elektronskih virov. Knjižnica, 2004, l. 48, št. 4, str. 97-119.*
- Kalčič, Dunja. *Novi viri na internetu in njihova taksonomija. Knjižnica, 2004, l. 48, št. 4, str. 77-95.*
- Kodrič, Nataša. *Analiza slovenskih elektronskih kontinuiranih virov in njihovih založnikov. Ljubljana: NUK, september 2003 (raziskovalna naloga).*
- *Elektronske publikacije: kako jih bomo ohranili za prihodnost? Ljubljana, 1998.*
- *Elektronske publikacije: Kodeks prakse prostovoljnega depozita. Ljubljana, 2001.*
- *ISBD (ER): mednarodni standardni bibliografski opis elektronskih virov. Ljubljana, 2000.*
- *ISO 2789 : 2003, Information and documentation - International library statistics.*

<sup>9</sup> Celoten predlog nabora metapodatkov je predstavljen v Kavčič - Čolić, A. *Metapodatki za trajno ohranjanje elektronskih virov. Knjižnica, 2004, l. 48, št. 4, str. 114-117.*

## SUMMARY

### COLLECTING AND ARCHIVING SLOVENIAN ONLINE PUBLICATIONS

The main purpose of a national library is to collect, preserve, record and distribute publications which are published in a country, written in the language of the nation, by authors born or associated with the country or nation, and which subject is related to the nation, its country and the culture. The so-called collection of "patriotica" is being acquired regardless to the quality of publications or their physical form. The major part of this national collection is being acquired on a mandatory basis, due to the Legal Deposit Act. Each editor, printer or even author in a certain country is obliged to send a copy of the published work to the library or libraries which have the privilege of a legal deposit copy. According to this regulation the national library is a compulsory recipient of a legal deposit copy. According to the Librarianship Act the collection of Slovenian legal deposit copies has the status of cultural heritage.

The object of the legal deposit is a publication. The publication is defined as an object: information written on a media, which is multiplied and made available for a wide readership. Since the invention of printing press and letter-press printing, the publications have changed considerably. In the past century publications in machine readable forms (vinyl gramophone records, videocassettes, CDs, CD-ROMs, DVDs) were introduced. They became an object of legal deposit as well as broadcasts, telecasts and electronic publications accessible online.

Electronic publications available online have, compared with the traditional book material, special features which make them difficult to be collected and preserved. Their complexity and fluidity make them the prime challenge to contemporary library and information science.

The Slovenian National and University Library has recognized the importance of electronic publications (those on portable carriers as well as those which are distributed online via World Wide Web) as national heritage and highlighted the importance of their preservation. Since the beginning of the 19<sup>th</sup> century it has been suggesting that they should become the object of legal deposit act. In 2002, the library made the first step towards the archiving of Slovenian electronic publications distributed via World Wide Web by launching the project Methodology of Collecting and Archiving Slovenian Online Publications. The project was supported by The Ministry of Culture, Ministry of Education, Science and Sport and by The Ministry of Information Society. Together with the national library another major Slovenian research institution, the Jožef Stefan Institute, was involved into the project. The task of the national library research group was to define the online electronic resources, which should be collected, to provide the selection criteria and metadata. The researchers from the Jožef Stefan Institute were engaged with the programming of the harvester.

The project was carried out in 2003 and 2004 and its main goals were reached. The national library started to collect Slovenian websites and Slovenian online electronic serial publications. The additional criteria have been defined as electronic resources available online via URL \*.si, servers located in Slovenia using the Slovenian language. The metadata have been defined according to the OAIS model. The harvester called WebBird has been programmed and tested.