

ZBIRANJE OBVEZNEGA IZVODA SPLETNIH PUBLIKACIJ V NARODNI IN UNIVERZITETNI KNJIŽNICI: ZAKONSKA PODLAGA IN PRAKTIČNI VIDIKI

Janko Klasinc
Irena Sešek

Oddano: 9. 2. 2010 – Sprejeto: 17. 3. 2010

Kratki znanstveni prispevek
UDK 021.84:004.738.5(497.12)

Izvleček

Zbiranje in zajemanje slovenskih spletnih publikacij poteka v skladu z Zakonom o obveznem izvodu publikacij, Zakonom o spremembah in dopolnitvah Zakona o obveznem izvodu publikacij in Pravilnikom o vrstah in izboru elektronskih publikacij za obvezni izvod. V prispevku so predstavljene aktivnosti, ki jih je Narodna in univerzitetna knjižnica (NUK) izvajala na področju zbiranja obveznega izvoda spletnih publikacij od leta 2006, ko se je začel uporabljati Zakon o obveznem izvodu publikacij. Ta opredeljuje dva koncepta zbiranja spletnih publikacij. NUK lahko prosto dostopne spletne publikacije išče in pridobiva sam, obenem pa mora omogočiti zavezancem za obvezni izvod elektronsko oddajo njihovega gradiva. Koncepta se sicer delno prekrivata, vendar se bistveno razlikujeta glede postopkov pridobivanja, hranjenja in uporabe publikacij, saj zakon kot publikacije opredeljuje tudi celotne spletne strani, ki jih NUK zajema z avtomatiziranimi postopki.

Ključne besede: obvezni izvod, spletne publikacije, Slovenija

KLASINC, Janko; Irena SEŠEK. Collecting legal deposit of web publications in the National and University Library, Slovenia: legal provisions and practice. Knjižnica, 54(2010)1–2, p. 121–135

Short scientific article

UDC 021.84:004.738.5(497.12)

Abstract

The Slovenian web publications are collected in accordance with the Legal Deposit Act, Act amending the Legal Deposit Act, and Regulations on Type and Selection Criteria for Legal Deposit of Electronic Publications. Since the adoption of Legal Deposit Act in 2006, the National and University Library (NUL) has carried out a range of activities in the field of collecting legal deposit of web publications. Two different concepts of web publications acquisition are defined by The Legal Deposit Act. NUL can search for and collect freely accessible web publications but must also provide an electronic means of depositing web publications for the publishers and intellectual rights owners. The two concepts overlap to some extent, however, as web sites are also treated as publications, and are, therefore, harvested by NUL, there exist some fundamental differences in terms of acquisition, preservation and usage of different types of web publications.

Key words: legal deposit, web publications, Slovenia

1 Uvod

Zbiranje obveznega izvoda ima v Evropi že večstoletno tradicijo in tudi v Sloveniji smo že pred tremi leti praznovali 200-letnico tega pomembnega instituta za ohranjanje slovenske pisne kulturne dediščine. Z razvojem novih nosilcev informacij, ki niso več v fizični obliki, se je pri zbiranju obveznega izvoda izgubljal pomemben del pisne kulturne dediščine (Kavčič - Čolić, 2007; Kodrič, 2006).

Zakon o obveznem izvodu publikacij, ki se je začel uporabljati leta 2006, je v Sloveniji prvič omogočil zbiranje obveznega izvoda spletnih publikacij. Pogoji in osnove za sprejem zakonodaje na tem področju v Sloveniji in tujini so že bili predstavljeni (Kavčič - Čolić et al. 2004; Kavčič - Čolić, 2007; Kodrič - Dačić, 2006; Sešek, 2006). Obvezni izvod spletnih publikacij prinaša številne nove vidike, ki jih pri zbiranju obveznega izvoda tiskanih publikacij ni bilo, in nove poglede ter razsežnosti pri znanih vprašanjih: večjo vpletenost založnikov pri zbiranju obveznega izvoda, teritorialnost (kraj izdaje), elektronski dostop za uporabnike, upravljanje z digitalnimi pravicami, digitalno ohranjanje in zaščita podatkov (Gibby in Green, 2008).

2 Zakonske določbe

Zakon o obveznem izvodu publikacij določa, da so elektronske publikacije »elektronske knjige, elektronski časopisi in časniki, spletne strani in podobno, ki so objavljene na fizičnih nosilcih (npr. na magnetnih trakovih, kasetah, disketah, CD-romih ipd.) ali so dostopne na računalniških omrežjih ali svetovnem spletu«, in nadalje, da so spletne publikacije »elektronske knjige, elektronski časopisi in časniki, dostopni po spletu, ter spletne strani in podobno« (Zakon, 2006, 2. člen).

V primerjavi z nekaterimi drugimi evropskimi državami (npr. Velika Britanija) (Gibby in Green, 2008) je NUK že pred sprejemom *Zakona o obveznem izvodu publikacij* zbiral elektronske publikacije na fizičnih nosilcih. *Zakon o obveznem pošiljanju tiskov* iz leta 1972 je v 3. členu določal: »S tiskom so po tem zakonu mišljene vse publikacije (knjige, brošure, separati, muzikalije, revije, časniki, zemljevidi, atlanti vseh vrst, umetniške reprodukcije, katalogi, programi, razglednice ipd.), kakor tudi gramofonske plošče, glasbene razglednice, tonske kasete ter videokasete in podobno, ki so namenjene za razširjanje oziroma za prodajo.« Leta 1994 je bil sprejet *Zakon o pogojih za opravljanje reproduktivne video in audio dejavnosti*, ki je v drugem odstavku 4. člena določal: »Fizična ali pravna oseba, ki razmnožuje videogram oziroma fonogram, namenjen trženju, dajanju v najem, javnemu predvajanju ali drugi obliki posredovanja, mora v 15 dneh po končanem razmnoževanju en izvod videograma oziroma fonograma skupaj z dokumentacijo iz prvega odstavka 2. člena tega zakona poslati Narodni in univerzitetni knjižnici v Ljubljani.« Na podlagi obeh zakonov se je v Sloveniji z razvojem elektronskih publikacij na fizičnih nosilcih kontinuirano zbiral tudi obvezni izvod tega gradiva, niso pa se zbirale spletne publikacije.

Zbiranje obveznega izvoda spletnih publikacij je v *Zakonu o obveznem izvodu publikacij* (2006) posebej opredeljeno v dveh členih. Zakon določa, da morajo biti spletne publikacije predložene brez elektronske zaščite dostopa, ali pa morajo zavezanci zagotoviti geslo za prevzem. Na ta način je zagotovljeno prevzemanje in trajno ohranjanje spletnih publikacij v NUK. Zakon zagotavlja tudi varovanje avtorskih pravic založnikov, saj je dostop zaščitenih publikacij za končne uporabnike omogočen v skladu z dogovorom med NUK in založnikom. Vsekakor mora biti uporaba obveznega izvoda omogočena vsaj v prostorih NUK ali za študijske in raziskovalne namene (7. člen). Zakon je za zajemanje obveznega izvoda spletnih publikacij predpisal sprejem *Pravilnika o vrstah in izboru elektronskih publikacij za obvezni izvod*, zagotavlja pa tudi zajemanje različic ali novih izdaj (17. člen).

V posameznih členih so v nekaterih točkah spletne publikacije navedene kot izjeme: oddati je potrebno samo en izvod, kolofon je obvezni del spletne publikacije, a vsebuje samo tiste podatke, ki so značilni za medij ter podatek o spletni lokaciji. NUK »poišče obvezni izvod elektronskih publikacij s postopki iskanja

po spletu«, oziroma jih posreduje zavezanec sam. Zaščitene spletne publikacije posreduje zavezanec z geslom v 15 dneh po objavi na spletu. Pri prevzemu zaščitene publikacije NUK obvesti zavezanca o datumu prevzema posamezne spletne publikacije. NUK je odgovoren za zakonito uporabo in varovanje spletnih publikacij pred zlorabo. Če zaradi tehnološke zastarelosti NUK ne more zagotoviti dostopa v skladu z zakonom, pa zakon omogoča, da NUK trajno ohrani vsebino s prenosom na nadomestni medij. Pri prenosu zakon dovoljuje tudi tehnične spremembe z namenom zagotavljanja dostopa ali hranjenja. Če ni mogoče ohraniti prvotne informacijske, slikovne in slišne vsebine, ali če je publikacija uničena ali zastarela, jo NUK lahko izloči iz zbirke.

V treh letih od začetka uporabe *Zakona o obveznem izvodu publikacij* se je pri zbiranju publikacij izkazalo, da prvotna definicija ni najbolje določala spletnih publikacij, ki se pojavljajo v praksi, zato je v *Zakonu o spremembah in dopolnitvah Zakona o obveznem izvodu publikacij* (2009, 2. člen) nova definicija: »Spletne publikacije so vse elektronske publikacije, ki so dostopne preko spleta, in so lahko v besedilni, avdio ali video obliki ali so kombinacija naštetih oblik, vključno s spletnimi stranmi in podobno.«

Takoj po začetku uporabe zakona je NUK najprej pripravil *Pravilnik o vrstah in izboru elektronskih publikacij za obvezni izvod*, ki je bil sprejet leta 2007. Nato je razvil tehnično infrastrukturo, ki je potrebna za delovanje sistema zbiranja spletnih publikacij: portal za oddajo publikacij in program za avtomatsko zajemanje. Dejansko zbiranje obveznega izvoda spletnih publikacij se je začelo šele leta 2008.

Na podlagi tradicije zbiranja obveznega izvoda tiskanih publikacij so bile meje zbiranja obveznega izvoda elektronskih publikacij v svetu in v Sloveniji postavljene zelo na široko. Zaradi zahtevnosti zbiranja pa so potrebni drugačni in ožji kriteriji. Vendar so v elektronskem okolju meje med znanimi vrstami gradiva vedno bolj zabrisane, poleg tega pa se stalno pojavljajo tudi vedno novi formati. Zato zakon ne določa podrobneje spletnih publikacij, ki jih NUK zbira kot obvezni izvod. *Pravilnik o vrstah in izboru elektronskih publikacij* (2007) za obvezni izvod navaja tri splošne in pet posebnih kriterijev. Splošni kriteriji ostajajo enaki kot pri dosedanem zbiranju obveznega izvoda na fizičnih nosilcih: dela slovenskih avtorjev, dela v slovenskem jeziku in dela o Sloveniji. Posebni kriteriji so določeni glede na značilnosti spletnih publikacij in se nanašajo na vsebino, odgovornost, strukturo publikacije, domeno in format podatkov. Zbiranje elektronskih publikacij je v treh letih potekalo selektivno zaradi kadrovskih omejitev (za zbiranje in zajemanje je v NUK zaposlen en delavec, ki skrbi za administriranje portala Svarog in upravljanje programa za avtomatsko zajemanje spletnih strani) ter zaradi ogromne količine publikacij v spletnem okolju, ki jih nobena nacionalna knjižnica ne more shraniti v celoti.

Z vidika prostega dostopa do vseh zajetih vsebin, so nekateri avtorji opozorili, da se na spletu pojavljajo tudi sporne, zelo občutljive, osebne (vprašanje varo-

vanja osebnih podatkov) in celo nezakonite vsebine (teroristične vsebine, nespo-
dobne ali opolzke publikacije). S povsem avtomatiziranim zajemom, kjer so
možnosti pregledovanja pred zajemom zelo majhne, bi morda nehote zajemali
tudi vse takšne publikacije. Z omejitvijo dostopa bi lahko ublažili tveganje glede
hranjenja in omogočanja prostega dostopa do takšnih vsebin, ki bi bilo lahko za
depozitno knjižnico zakonsko sporno. Na to sta opozorili tudi dve študiji. Prvo
je leta 2003 izdelal Andrew Charlesworth za Joint Information Systems Com-
mittee and the Wellcome Trust, drugo pa sta opravila Beunen and Schiphof leta
2006 za Koninklijke Bibliotheek v Hagu (Gibby in Green, 2008). V Sloveniji tak-
šnih vsebin še nismo zajemali in tudi nismo določili posebnih kriterijev, s kate-
rimi bi jih iz zajema izločili. Zanimivo pa je, da pri reševanju različnih pravnih
vprašanj s področja zajemanja spletnih publikacij opozarjajo tudi na ta vidik.

3 Sodelovanje med založniki in knjižnicami

Uspešna oddaja obveznega izvoda je bila vedno odvisna tudi od dobrega
sodelovanja med depozitnimi knjižnicami in zavezanci (založniki ali tiskarji),
čeprav je oddaja določena z zakonom. Pri zbiranju obveznega izvoda spletnih
publikacij pa so založniki še posebej zaskrbljeni in nezaupljivi. Že leta 1997 je
bila ustanovljena skupna delovna skupina, sestavljena iz članov nekaterih ev-
ropskih nacionalnih knjižnic in predstavnikov evropskih založnikov (Sešek,
2006).

Kot sta zapisala Gibby in Green (2008), imajo tiskane publikacije zaradi značil-
nosti medija že po svoji naravi določene omejitve pri uporabi. Spletne publika-
cije pa omogočajo na eni strani zelo hitro in enostavno distribucijo, na drugi strani
pa je dosti težje preprečevanje nedovoljenega kopiranja. Zato založniki po vsem
svetu zahtevajo vsaj upoštevanje omejitev in izjem v 10. členu Pogodbe svetovne
organizacije za intelektualno lastnino o avtorski pravici iz leta 1996:

»(1) Pogodbenice lahko v svoji domači zakonodaji predvidijo omejitve ali izjeme
glede pravic, ki jih avtorjem književnih in umetniških del daje ta pogodba, v
nekaterih posebnih primerih, ki niso v nasprotju z običajno uporabo dela in pre-
tirano ne posegajo v zakonite interese avtorja. (2) Pogodbenice morajo pri upora-
bi Bernske konvencije skrbeti vse omejitve ali izjeme od pravic, ki so določene v
njej, na nekatere posebne primere, ki niso v nasprotju z običajno uporabo dela in
pretirano ne posegajo v zakonite interese avtorja.« (Zakon, 1999).

Leta 2009 je mednarodna bibliotekarska organizacija (International Federation
of Library Associations and Institutions – IFLA) v sodelovanju z mednarodnimi
organizacijami avtorjev (European Writers' Council – EWC), založnikov (Inter-

national Publishers Association – IPA), organizacijo za varstvo intelektualne lastnine (World Intellectual Property Organization – WIPO) in zvezo za reproduktivne pravice (International Federation of Reproduction Rights Organisations – IFRRO) pred knjižnim sejmom v Frankfurtu organizirala enodnevno konferenco z naslovom *The Culture of Reading and Books in the Digital Age*. Prvič so se skupaj srečali vsi ključni predstavniki navedenih organizacij in prav vsi so izrazili veliko zadovoljstvo, da je do srečanja prišlo. Kljub različnim pogledom zaradi posameznih interesov je bila izmenjava mnenj za vse zelo koristna. Vsi so pripravljani na nadaljnje sodelovanje in iskanje rešitev za trenutno najpomembnejša vprašanja na področju digitalizacije gradiva in prostega dostopa. Organizatorji so obljubili, da bodo predstavitve in nadaljnje sodelovanje predstavljene na spletni strani IFLA, a do danes žal razen najave konference še niso bile objavljene nobene druge informacije (Enhancing, 2009).

S strani knjižnic so predavatelji predvsem opozarjali, da so nacionalne knjižnice zadnji braniki nacionalne kulturne dediščine, vendar lahko trenutno brez zakonskih omejitev digitalizirajo le gradivo do leta 1890. Knjižnice so bile predstavljene kot tržišče, ustvarjalke bralcev in varuhi intelektualne lastnine. Zato je nujno zagotoviti več pravic uporabnikov in omogočiti dostop do znanja. Za knjižnice morajo tudi v digitalnem okolju veljati izjeme in omejitve pri ohranjanju gradiva, pri raziskovalnem in študijskem delu, za splošno in brezplačno uporabo v skladu s pošteno poslovno prakso ter za zagotavljanje dosegljivih formatov. Še posebej je takšna izjema potrebna pri zagotavljanju konverzije v posebne formate za potrebe uporabnikov s posebnimi potrebami. Za izboljšanje bralne kulture je potrebno prepoznavanje kompleksnosti knjižnic in pomembnejša vloga knjižnic v nacionalnih informacijskih politikah.

Sodelovanje med NUK in slovenskimi založniki spletnih publikacij poteka že od začetka zajemanja spleta. Kljub sprejetemu Kodeksu prakse prostovoljnega depozita (Sešek, 2006), se je zbiranje v praksi začelo šele, ko so bili zagotovljeni vsi tehnični pogoji in zakonske osnove. Zavezanci za obvezni izvod slovenskih spletnih publikacij so vključeni v proces oddaje gradiva in samostojno odločajo o možnostih dostopa.

4 Zajemanje in arhiviranje vsebin na spletu

Svetovni splet je prostor, na katerem se že od samega začetka pojavljajo najrazličnejše publikacije iz vseh področij človekovega ustvarjanja, tako znanstvenoraziskovalnega kot kulturno-umetniškega. Splet pa je še veliko več, je prostor komunikacije, izmenjave informacij in znanja, povezovanja ter vzpostavljanja najrazličnejših odnosov med posamezniki ali skupinami posameznikov. Kot tak

predstavlja virtualizacijo človeške družbe v vseh njenih odtenkih in subtilnostih.

Ko govorimo o nacionalni ali svetovni kulturni dediščini ne moremo več izključiti tega obsežnega in kompleksnega prostora, ki je pri produkciji in pretoku znanja danes že v precejšnji meri nadomestil klasične mehanizme in poti. S stališča bibliotekarske stroke se kulturne dediščine lotevamo predvsem z vidika trajnega hranjenja in trajnega omogočanja uporabe, zato smo se v knjižnicah morali soočiti tudi z izzivi ohranjanja spletnih strani, tako njihove vsebine kot tudi specifične strukture, ki je že sama po sebi vredna ohranitve za prihodnje rodove. Zajemanje in arhiviranje vsebin na spletu sta dejavnosti, ki ju na podlagi teh temeljev izvajajo predvsem nacionalne ali regionalne knjižnice in sta v skladu z njihovim poslanstvom logična nadgradnja zbiranja, obdelave, hranjenja in dajanja v uporabo klasičnih publikacij na fizičnih nosilcih.

Pri zajemanju in arhiviranju spletnih vsebin gre za sistematično in delno avtomatizirano shranjevanje ter arhiviranje spletnih mest. Postopek poteka s pomočjo robotskih programov za samodejno zajemanje (ang. web harvesters), ki so podobni spletnim pajkom, ki jih uporabljajo spletni iskalniki za preiskovanje in indeksacijo spleta. Bistvena razlika med obema postopkoma je, da programi, ki zajemajo spletne strani, te tudi shranjujejo. Zajete spletne strani so tako na voljo za kasnejše pregledovanje.

Spletišča so praviloma zajeta v celoti, z vsemi pripadajočimi datotekami, prav tako se ohrani njihova struktura. Vsak zajem je možno omejiti glede na globino. Lahko zajamemo le posnetek prve strani spletišča ali pa posežemo globlje do druge, tretje ali do najgloblje ravni. Prav tako lahko omejimo ali raztegnemo zajem glede na širino. Zajamemo lahko le strani in pripadajoče datoteke na eni domeni ali pa razširimo zajem na nekatere zunanje povezave, ki so vsebinsko povezane z osnovno stranjo.

Na splošno ločimo dva pristopa k zajemanju – selektivnega in celostnega. Pri prvem zajemamo posamezna spletna mesta, ki smo jih izbrali na podlagi nekih kriterijev. Na ta način skušamo zajeti spletna mesta čim bolj globoko in jih ohraniti čim bolj podobna originalnim, živim izvornikom. Selektivno zajemanje nam omogoča preverjanje kakovosti posameznih zajemov in popravljanje morebitnih napak (nedelujoče povezave, nepopolni zajemi itd). Ta način nam omogoča tudi zbiranje večje količine metapodatkov o posamezni strani. Vsako stran zajemamo z neko frekvenco, ki je odvisna od ocene pogostosti spreminjanja strani, velikosti strani in od kapacitete strojne opreme, ki jo uporabljamo. Slabost pristopa je predvsem časovna zamudnost in potratna uporaba programskih ter pomnilniških kapacitet. Poleg tega so tako zajete spletne strani izvzete iz konteksta širšega spleta, saj zunanje povezave ne delujejo. Ker je bistvena značilnost svetovnega spleta ravno medsebojna povezanost spletnih strani, tako zajeta stran do neke mere izgubi svojo izvorno funkcionalnost.

Pri celostnih zajemih gre običajno za zajemanje celotnih nacionalnih domen ali podobno velikih kosov spleta. Pri tem načinu običajno zajemamo na bolj površinski ravni in tako dobimo posnetek spleta v nekem trenutku. Prednost tovrstnih zajemov je ohranitev značilnosti spleta, torej spletnih strani, ki so med seboj povezane, pogloblitna slabost pa je izguba velike količine informacij, ki so na globljih ravneh spleta. Poleg tega je frekvenca izvedb tovrstnih zajemov precej nižja od tiste, ki jo omogoča selektivni pristop (največ nekajkrat letno). Pomankljivost predstavlja tudi skoraj popolna odsotnost preverjanja kakovosti, saj je količina zajetega gradiva prevelika za ročno pregledovanje.

Zaradi različnih prednosti in slabosti obeh pristopov se mnoge institucije odločajo za kombinacijo obeh. V zadnjih letih smo priča tudi brisanju meje med obema pristopoma. Celostni zajemi zaradi vedno večjih kapacitet strojne opreme lahko posežejo globlje in pridobijo večje količine gradiva, ob enem pa se razvijajo novi postopki avtomatiziranega preverjanja kakovosti zajemov. Zaradi vedno naprednejših aplikacij za delo s programi za samodejni zajem je selektivno zajemanje postalo preprostejše, bolj avtomatizirano in zahteva manj ročnega posredovanja.

Z arhiviranjem vsebin na spletu se pretežno ukvarjajo nacionalne ali večje regionalne knjižnice. Veliko, med njimi tudi NUK, jih je združenih v mednarodnem konzorciju za ohranjanje spleta (International Internet Preservation Consortium – IIPC, 2010), ki združuje 39 institucij z vsega sveta in podpira dejavnost arhiviranja spletnih vsebin ter razvoj različnih orodij za zajemanje, shranjevanje in uporabo arhiviranih spletnih strani. Ena od bolj pomembnih članic je The Internet Archive (2010) – spletni arhiv, ki od leta 1996 deluje v San Franciscu, kot poslanstvo pa si je zadal gradnjo arhiva celotnega svetovnega spleta ali vsaj čim večjega dela. Internet Archive je znan po izdelkih, kot je Wayback Machine (vmesnik za pregledovanje arhiviranih spletnih strani), zelo razširjena pa je tudi uporaba njihovega programa za zajemanje Heritrix.

V NUK smo s sistematičnim zajemanjem slovenskih spletnih strani začeli leta 2008, čeprav začetki raziskovalne dejavnosti s tega področja segajo v leto 2002, ko je NUK v sodelovanju z Institutom Jožef Stefan izpeljal ciljni razvojni projekt Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju (Kodrič - Dačić, 2006). Rezultati projekta so bile predvsem metodološke osnove in strokovna izhodišča za zajemanje spletnih publikacij, pridobili pa smo tudi nekaj praktičnih izkušenj. Šele na podlagi novega *Zakona o obveznem izvodu publikacij* (2006) je postalo zajemanje spletnih vsebin del rednih nalog NUK.

V NUK trenutno izvajamo selektivno zajemanje v skladu s *Pravilnikom o vrstah in izboru elektronskih publikacij za obvezni izvod*. Kriteriji za izbiro so v pravilniku opredeljeni precej ohlapno in tako omogočajo zajemanje precejšnjega dela slovenskega spleta brez posebnega dovoljenja izdelovalcev spletnih strani. Vsaka zajeta spletna stran mora ustrezati vsaj enemu splošnemu in enemu posebnemu

kriteriju iz pravilnika. Za začetek smo se lotili zajemanja spletnih strani s področja javne uprave, visokega šolstva, spletnih strani raziskovalnih institucij in spletnih revij.

Za zajemanje uporabljamo Heritrix, zelo razširjeno in uveljavljeno orodje za zajemanje spletnih strani, kot vmesnik za delo s programom pa orodje Web Curator Tool (WCT), ki sta ga leta 2006 pod okriljem IIPC zasnovali Nacionalna knjižnica Nove Zelandije in Britanska nacionalna knjižnica. Bistvena prednost uporabe WCT je preprosto rokovanje s programom za zajemanje, ki od uporabnika ne zahteva poglobljenega tehničnega znanja, na drugi strani pa naprednemu uporabniku omogoča zelo natančno prilagajanje nastavitvev (Paynter et al., 2008). Uporaba vmesnika se v zadnjih letih povečuje in je v marsikateri instituciji nadomestila tudi lastne rešitve, ki terjajo bistveno več znanja in časa. Ker so zajete vsebine shranjene v posebnem arhivskem formatu, jih ni možno prikazati v običajnem spletnem brskalniku, pač pa za prikaz potrebujemo vmesnik, ki omogoči pregledovanje gradiva v obliki, ki smo je vajeni pri uporabi živih spletnih strani. Za prikaz zajetih vsebin uporabljamo Wayback Machine, ki je zaradi svoje prepoznavnosti postal skoraj sinonim za Internet Archive in omogoča brskanje po spletnem arhivu ter prikaz zajetih spletnih strani v običajnem spletnem brskalniku.

Vsako spletno stran pred zajemom ocenimo glede na njene tehnične, strukturne in oblikovne značilnosti. Sledi vnos spletne strani (tarče) v vmesnik WCT. Vsako tarčo lahko sestavlja eden ali več URL-jev. Določimo tudi frekvenco zajemanja. Tarčo lahko zajemamo v časovno zamejenem ali neomejenem obdobju, zajemi pa se lahko vršijo v enakih intervalih ali le ob točno določenih datumih. Vmesnik omogoča tudi vnašanje nekaterih formalnih in vsebinskih podatkov o tarči v formatu Dublin Core, tehnični metapodatki pa se samodejno beležijo tekom zajema.

Za večino spletnih strani ustreza osnovni profil nastavitvev, medtem ko je v nekaterih primerih potrebno s podrobnejšim ročnim nastavljanjem programa zagotoviti pravilen zajem. Na voljo imamo veliko možnosti, ki omogočajo zelo natančno nastavitvev. Globino zajema lahko določimo z omejitvijo števila povezav, ki vodijo od osnovne strani, ali z omejitvijo števila podmap, ki vodijo od osnovne domene. Zajem lahko omejimo tudi z določitvijo maksimalnega števila zajetih datotek ali skupne zgornje meje velikosti zajema v bajtih. Določimo lahko maksimalno trajanje zajema v sekundah, iz zajema pa lahko izločimo katerekoli tipe datotek.

Ker lahko zajem obremeni internetno povezavo gostitelja spletne strani in tako oteži dostop drugim uporabnikom, moramo upoštevati določene omejitve glede intenzivnosti postopka. Heritrix že v osnovi ne more procesirati več kot enega URL-ja naenkrat. Poleg tega je poskrbljeno tudi za zakasnitev pri prehodu iz za-

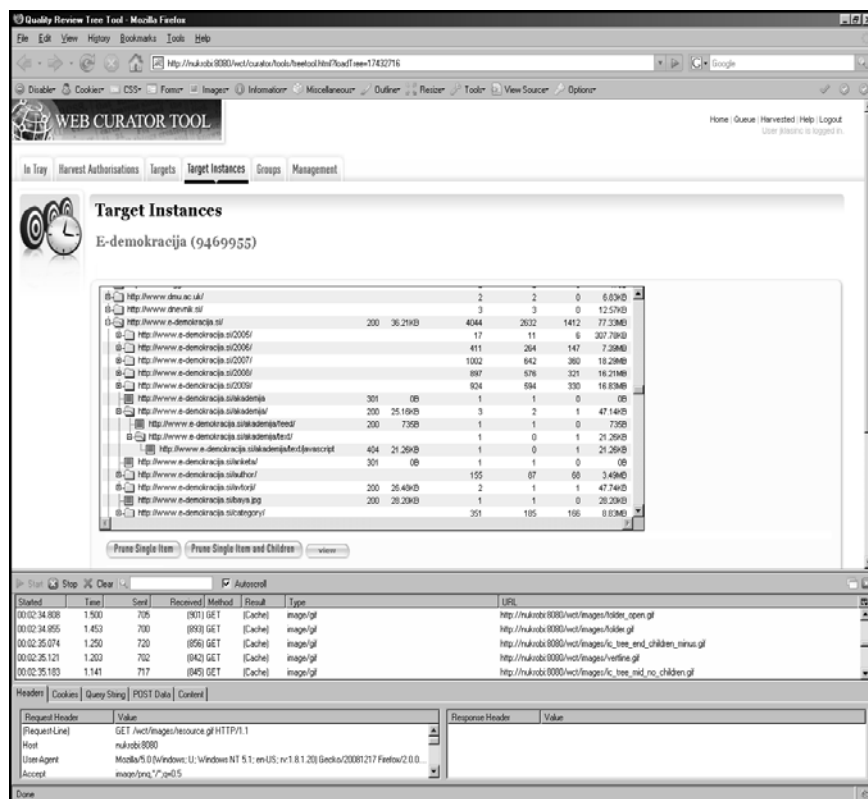
ključnega URL-ja na naslednjega. Določimo lahko tudi omejitev pasovne širine, ki jo lahko uporablja program za zajemanje, in se tako izognemo preveliki obremenitvi spletnih strani, ki imajo na voljo počasnejšo povezavo.

Do neke mere lahko delovanje različnih programov, ki z avtomatiziranimi postopki preiskujejo, beležijo ali zajemajo spletne vsebine, omejijo tudi sami izdelovalci spletnih strani. Za ta namen uporabljajo datoteko robots.txt, ki je običajno v osnovnem direktoriju strani in vsebuje njihova pravila glede možnosti dostopa tovrstnih programov do določenih delov spletne strani. Datoteka običajno vsebuje izključitve – navedbe map, ki jih ti programi ne smejo obiskati – lahko pa tudi popolnoma prepove kakršnokoli delovanje raznih robotskih programov. Pravila, zapisana v robots.txt, sicer veljajo le kot priporočila in ne morejo ustaviti delovanja tistih programov, ki so nastavljeni tako, da ignorirajo ta pravila. Ker želimo našo dejavnost izvajati čim manj invazivno, se držimo pravil, ki jih izdelovalci zapišejo v robots.txt.

Po opravljenem zajemu pregledamo zajeto spletno stran. Pri tem smo pozorni na to, ali je program zajel vse, kar smo želeli, in ali je prikaz zajetih vsebin sprejemljiv s stališča zvestobi originalu. Najbolj pogosta težava je nezmožnost programa za zajem delov spletnih strani, ki predstavljajo t. i. globoki splet (ang. deep web), ki po velikosti zavzema večino celotnega spleta. V teh primerih spletna stran ne vsebuje določene vsebine, pač pa jo šele na zahtevo uporabnika postreže iz ločene baze. Običajno mora uporabnik v neko polje vpisati iskalno zahtevo in tako opraviti poizvedbo v bazi, ki mu vrne ustrezne informacije. Ker program ne more izpolnjevati polj, ne more doseči teh skritih informacij in tako lahko zajame le iskalni vmesnik.

Podobne težave predstavljajo spletne strani z dinamično generirano vsebino. Klasične spletne strani, napisane v jeziku HTML, so sestavljene iz vsebine, ki je vedno prisotna na sami strani, pri dinamičnih spletnih straneh pa se določeni deli strani naložijo šele, ko jih sproži določen dogodek. Za izdelavo tovrstnih spletnih strani se uporabljajo ogrodja, kot so Java, Flash, Ajax ali skriptni jeziki, kot so PHP, Perl, ASP, ASP.NET itd. Heritrix sicer v mnogih primerih uspe zajeti in tudi ohraniti funkcionalnost nekaterih skript v arhiviranih spletnih straneh, vseeno pa količina nedelujočih povezav in napačno prikazanih vsebin predstavlja eno od večjih težav spletnega arhiviranja pri nas in po svetu.

Težave pa ne povzročajo le izpuščene vsebine ali premajhna količina zajete vsebine spletnih strani, saj v nekaterih primerih program zajame preveč oziroma tisto, kar ni relevantno. V teh primerih gre najbolj pogosto za pasti, ki povzročijo, da program za zajem začne izvajati neomejeno število poizvedb za določen URL. Tipičen primer pasti so koledarji, ki generirajo neomejeno število povezav, katerim program sledi dokler ne doseže zgornje meje zajetih dokumentov, ki smo mu jo določili. Na ta način lahko relativno majhna spletna stran povzroči izjemno velik zajem, ker je večina pobranih dokumentov proizvedla past.



Slika 1: Vmesnik v orodju Web Curator Tool za pregledovanje zajetih spletnih strani

Nekatere od omenjenih težav lahko do neke mere odpravimo s podrobnejšim nastavljanjem programa in ponavljanjem zajema dokler ne dosežemo optimalne kakovosti, vseeno pa je na današnji stopnji razvoja arhiviranja spletnih vsebin in zaradi vedno novih tehnologij za gradnjo spletnih strani nemogoče zagotoviti povsem zadovoljive zajeme v čisto vseh primerih. Vsak zajem, ki ustreza kriterijem kakovosti, se pravi, vsebuje vso ali vsaj bistveni del vsebine spletne strani in tudi omogoča uporabo ter prikaz vsebine, arhiviramo v poseben repozitorij. Zajete spletne strani so shranjene v ARC formatu, ki ga je za shranjevanje zajetih spletnih mest ustvaril Internet Archive, ogledovanje arhiviranih vsebin pa je možno z uporabo vmesnika Wayback Machine.

INTERNET ARCHIVE
Wayback Machine

Enter Web Address: All Adv. Search OpenSearch (Beta)

Searched for <http://durs.gov.si/> Set Anchor Window 11 Results

Search Results for Jan 1, 2007 - Jan 1, 2011			
2007	2008	2009	2010
0 pages	6 pages	5 pages	0 pages
	<ul style="list-style-type: none"> Jul 29, 2008 * Jul 29, 2008 Aug 5, 2008 * Oct 5, 2008 * Nov 5, 2008 * Dec 5, 2008 * 	<ul style="list-style-type: none"> Jan 5, 2009 * Feb 5, 2009 * Mar 5, 2009 * Apr 5, 2009 * May 5, 2009 * 	

[Home](#) | [Help](#)

Slika 2: Prikaz seznama zajemov posamezne spletne strani po datumih v vmesniku Wayback Machine

Spletni arhiv, ki ga gradimo v NUK, je za zdaj še zaprt za javnost, v prihodnosti pa načrtujemo tudi javen dostop. Eden od pomembnejših načrtov je tudi poskusna izvedba zajema celotne domene ».si« in spletnih strani iz ostalih domen, ki ustrezajo kriterijem slovenike. Celostne zajeme želimo v prihodnosti vzpostaviti kot stalno prakso NUK in tako dopolniti selektivno metodo, ki jo uporabljamo zdaj. Obenem bomo neprestano širili nabor spletnih strani, ki jih zajemamo selektivno. V sklopu zajemanja vsebin na slovenskem spletu želimo v prihodnosti tudi pokrivati pomembnejše dogodke, kot so državnozborske in lokalne volitve. Veliko izzivov nas čaka tudi na tehničnem področju, kjer si bomo prizadevali za čim višjo raven kakovosti zajemov, izgradnjo arhiva, ki bo ustrezal kriterijem trajnega hranjenja in vmesnikom za javni dostop, ki bo uporabnikom omogočal učinkovito in prijetno uporabo arhiva.

5 Svarog – portal za elektronsko oddajo publikacij

Z avtomatiziranimi postopki zajemanja spletnih vsebin pridobimo veliko količino prosto dostopnega gradiva v obliki spletnih strani. Te so kot objekti dostopne v spletnem arhivu. Pod koncept obveznega izvoda publikacij pa seveda na prvo mesto spadajo publikacije s tradicionalno strukturo, ki so dostopne na spletu ali na različnih omrežjih, torej predvsem elektronske knjige, revije in članki. Na podlagi Pravilnika o vrstah in izboru elektronskih publikacij za obvezni izvod, ki NUK nalaga omogočanje elektronskega načina oddaje publikacij, smo za zbiranje tovrstnega gradiva v samostojnih datotekah vzpostavili portal Svarog – sistem za varno arhiviranje oddanega gradiva (<https://www.nuk.uni-lj.si/svarog/>).

Svarog je preprost vmesnik za oddajanje elektronskih publikacij. Namenjen je predvsem zavezancem za obvezni izvod publikacij in tistim založnikom, ki so sklenili pogodbo z Javno agencijo za raziskovalno dejavnost RS o sofinanciranju domačih znanstvenih in poljudnoznanstvenih periodičnih publikacij ter so po

tej pogodbi zavezani tekoče številke svojih periodičnih publikacij v elektronski obliki pošiljati NUK za namen objave na Digitalni knjižnici Slovenije. Na portal založniki oddajajo predvsem plačljive publikacije, oziroma tiste, ki jih NUK sam s postopki iskanja po spletu ne more poiskati in pridobiti. Seveda pa je namenjen tudi oddaji prosto dostopnih publikacij vseh vrst.

Portal je nastal zaradi potrebe po enotnem načinu posredovanja datotek, ki vsebujejo publikacije, kot so e-knjige ali revije ter članki v PDF in podobnih formatih. Uporaba uveljavljenih poti, kot sta pošiljanje po elektronski ali običajni pošti na fizičnih nosilcih, ni najbolj primerna za zbiranje večje količine elektronskega gradiva, saj precej otežuje vodenje evidenc in arhiviranje publikacij. Z novim Zakonom o obveznem izvodu publikacij, ki je obetal povečano količino oddanih elektronskih publikacij, je bilo potrebno zasnovati sistem, ki bi omogočal preprosto oddajo gradiva, vodenje ustreznih evidenc in arhiviranje oddanega gradiva (Štular Sotošek, 2008).

Svarog je trenutno v fazi, v kateri je namenjen predvsem sprejemanju, pregledovanju in evidentiranju oddanih publikacij. Založniki ob oddaji posredujejo nekaj osnovnih podatkov o svojih publikacijah, ki služijo predvsem za identifikacijo, vsaka oddana publikacija pa je vezana na ustrezen zapis v sistemu COBISS, ki vsebuje ostale formalne metapodatke. Pri oddaji mora založnik označiti tudi možnost dostopa do svoje publikacije. V primeru, da se odloči za prost dostop, s tem dovoli NUK objavo publikacije na Digitalni knjižnici Slovenije, od koder je na daljavo brezplačno dostopna za lastne študijske in raziskovalne namene. Če dostop omeji, NUK publikacijo hrani, na voljo pa jo lahko da le znotraj prostorov knjižnice, kjer mora biti zagotovljena zakonita uporaba.

Z razliko od avtomatiziranega zajemanja vsebin na spletu, kjer kot objekte obravnavamo celotne spletne strani, v tem primeru skupaj s pripadajočimi metapodatki kot objekte arhiviramo posamezne datoteke, ki jih posredujejo založniki. Podobno kot pri oddaji tiskanih publikacij po pregledu oddane publikacije založnik prejme elektronsko potrdilo o prejeti publikaciji, ki je enakovredno potrdilo na papirju.

Svarog je začel delovati v začetku leta 2008. Po uvodni fazi testiranja, ki je trajala dva meseca, smo začeli prejemati prve publikacije. Do časa pisanja članka je bilo na Svarogu registriranih 115 založnikov, ki večinoma oddajajo e-knjige in e-revije v PDF datotekah, pridobili pa smo tudi nekaj večjih plakatov in zvočnih posnetkov.

V letu 2010 načrtujemo posodobitev sistema, ki bo omogočila bolj učinkovito delo moderatorjem in lažji postopek oddaje založnikom, predvsem pa bo uvedla ključno povezavo sistema z repozitorijem digitalnih vsebin, kar nam bo omogočilo še večjo stopnjo standardizacije postopkov pridobivanja in arhiviranja digitalnega gradiva. S časom želimo Svarog vzpostaviti kot centralno točko za zbiranje digitalnega gradiva, ki bi v čim večji meri nadomestila klasične, manj učinkovite poti.

6 Sklep

Zbiranje spletnih publikacij se je po svetu začelo že pred sprejemom zakonskih podlag v okviru različnih projektov, s prostovoljno oddajo ipd. Na ta način so knjižnice ob zbiranju gradiva v praksi postopoma oblikovale sheme in kriterije za zbiranje spletnih publikacij. Zakonski in podzakonski akti so v Sloveniji omogočili začetek zbiranja obveznega izvoda spletnih publikacij, vendar je zbiranje še v začetni fazi. Preko zbiranja dejansko tudi spoznavamo vse različne vrste spletnih publikacij, ki nastajajo in se razširjajo v javnosti. Po določenem času pa bo NUK lahko oblikoval tudi natančnejše kriterije za zbiranje.

S kombinacijo pristopov avtomatiziranega zajemanja vsebin na spletu in pridobivanja posameznih elektronskih publikacij prek portala Svarog skušamo pokriti zelo obsežno področje spletnih publikacij. Seveda je nemogoče pričakovati, da bomo uspeli pridobiti in shraniti čisto vse vsebine, ki bi si to zaslužile, saj se pri sledenju produkcije spletnih publikacij soočamo s težavami, ki jih v tradicionalnem okolju nismo poznali, ali pa smo jih lažje reševali. Tako se vedno znova pojavljajo vprašanja, kako slediti novim relevantnim objavam publikacij, ki niso zavedene v naših evidencah (nimajo zapisa CIP ali številke ISSN), kaj shraniti in česa ne, kako učinkovito katalogizirati vedno večjo količino tega novega gradiva, itd. Prihodnost pa bo prinesla tudi vedno bolj kompleksna vprašanja tehnične narave. V primerjavi s tiskanim gradivom, ki ga je relativno preprosto hraniti, uporaba pa je vedno možna brez posebne tehnične ali druge opreme, se pri spletnih publikacijah srečujemo s težavami, ki jih povzročajo vedno novi datotečni formati, nekompatibilnost novejših aplikacij in operacijskih sistemov s starejšimi formati, neobstoynost in nezanesljivost fizičnih medijev za hranjenje, problematika opisovanja publikacij ter različne metapodatkovne sheme za trajno hranjenje. Da bi tudi v elektronskem okolju uspeli zagotoviti vsaj podobno visoko stopnjo najdljivosti in uporabnosti publikacij čez 50, 100 ali več let, kot smo je vajeni pri papirnatih predhodnikih, bo potrebno veliko truda, predvsem pa vedno novega znanja, s katerim bomo verjetno ves čas vsaj en korak za najnovejšimi iznajdbami digitalne informacijske dobe. Ali bodo prihodnje generacije kdaj odkrile tudi spletne »Brižinske spomenike«?

Navedeni viri

1. Gibby, R., Green, A. (2008). Electronic legal deposit in the United Kingdom. *New Review of Academic Librarianship*, 14, 55–70.
2. Enhancing the culture of reading and books in the digital age. (2009). Pridobljeno 17. 3. 2010 s spletne strani: <http://www.ifla.org/en/events/enhancing-the-culture-of-reading-and-books-in-the-digital-age>

3. *International Internet Preservation Consortium*. (2010). Pridobljeno 25. 1. 2010 s spletne strani: <http://www.netpreserve.org/about/index.php>
4. *Internet Archive: about IA*. (2010). Pridobljeno 25. 1. 2010 s spletne strani: <http://www.archive.org/about/about.php>
5. Kavčič - Čolić, A. (2007). Krmarjenje v srednjem veku informacijske dobe: ali nam bo uspelo ohraniti našo pisno digitalno dediščino? *Knjižnica*, 51 (3–4), 149–161.
6. Kavčič - Čolić, A., Kodrič - Dačić, E., Šolar, R., Jakac - Bizjak, V., Kalčič, D. in Kavčič, I. (2004). *Metodologija zbiranja in arhiviranja slovenskih elektronskih publikacij na medmrežju*. Ljubljana: Narodna in univerzitetna knjižnica.
7. Kodrič - Dačić, E. (2006). Zunaj Gutenbergovega vesolja. *Knjižnica*, 50 (1–2), 173–186.
8. Paynter, G., Joe, S., Lala, V. in Lee, G. (2008). A year of selective web archiving with the web curator at the National Library of New Zealand. *D-Lib Magazine*, 14, (5-6). Pridobljeno 25. 1. 2010 s spletne strani: <http://www.dlib.org/dlib/may08/paynter/05paynter.html>.
9. Pravilnik o vrstah in izboru elektronskih publikacij za obvezni izvod. (2007). *Uradni list RS*, št. 90.
10. Sešek, I. (2006). Strokovne novosti in izvajanje Zakona o obveznem izvodu publikacij. *Knjižnica*, 50 (4), 33–47.
11. Štular Sotošek, K. (2008). Spletni nabiralnik za oddajo elektronskih publikacij: nova spletna storitev za založnike. *Knjižnica*, 52 (1), 123–132.
12. Zakon o obveznem izvodu publikacij. (2006). *Uradni list RS*, št. 69.
13. Zakon o obveznem pošiljanju tiskov. (1972). *Uradni list SRS*, št. 55.
14. Zakon o pogojih za opravljanje reproduktivne video in avdio dejavnosti. (1994). *Uradni list RS*, št. 42.
15. Zakon o ratifikaciji Pogodbe Svetovne organizacije za intelektualno lastnino o avtorski pravici. (1999). *Uradni list RS*, št. 84.
16. Zakonu o spremembah in dopolnitvah Zakona o obveznem izvodu publikacij. (2009). *Uradni list RS*, št. 86.

Janko Klasinc je zaposlen v Narodni in univerzitetni knjižnici.

Naslov: Jakčeva 10, 1000 Ljubljana

Naslov elektronske pošte: janko.klasinc@nuk.uni-lj.si

Irena Sešek je zaposlena v Narodni in univerzitetni knjižnici.

Naslov: Brodska 24, 1000 Ljubljana

Naslov elektronske pošte: irena.sesek@nuk.uni-lj.si