

ARHIVIRANJE HRVATSKIH MREŽNIH PUBLIKACIJA :
od projekta do programa arhiviranja u Nacionalnoj i sveučilišnoj knjižnici u
Zagrebu (NSK)

ARCHIVING CROATIAN ONLINE PUBLICATIONS:
From Project to Archiving Program in National and University Library in Zagreb
(NUL)

Dr. Mirna Willer, stručni suradnik za kompjutorizaciju / Professional Assistant for Computerization
Nacionalna i sveučilišna knjižnica, Zagreb / National and University Library, Zagreb
11 000 Zagreb, Hrvatska / Croatia
mwiller@nsk.hr

Sažetak

Danas se više ne postavlja pitanje je li opravdano baviti se u nacionalnoj knjižnici sadržajima na webu. Čak i pojam internet-inkunabula više ne doživljavamo kao novinu u posvemašnjem okruženju ovoga medija. Unatoč tomu, problemi arhiviranja weba izdaleka još nisu riješeni.

Ključne riječi

Web, arhiviranje weba, NUK

Abstract

Should a national library deal with web content is not questioned any more. Even a concept of Internet Incunabulas is not regarded as novelty in comprehensive scope of this medium. Nevertheless, the problems regarding web archiving are not yet solved.

Keywords

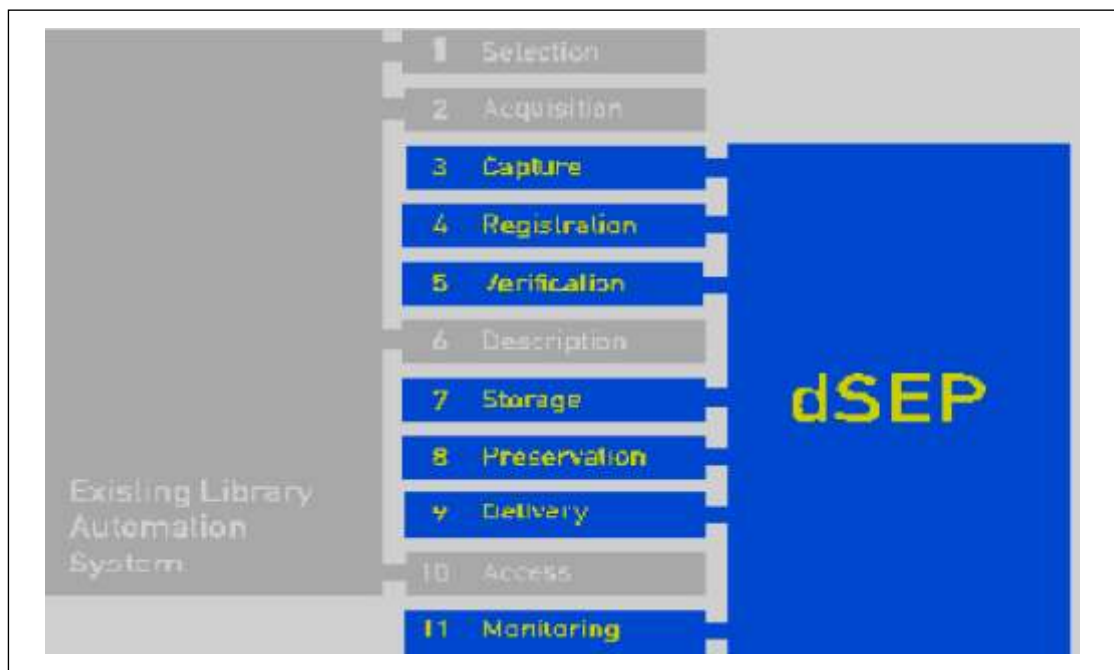
Web, web archiving, NUL

Arhiviranje web sadržaja

Nacionalne su knjižnice različito pristupile arhiviranju weba. Arhiviranje može biti sveobuhvatno što podrazumijeva snimanje web prostora u određenim vremenskim razdobljima s obuhvatom određene, npr. nacionalne domene. Tako arhivirani sadržaj nije podložan bibliografskoj kontroli, već je potrebno nad njim izgraditi pretraživače funkcijom slične pretraživačima na „živom“ internetu. Druga je vrsta arhiviranja weba selektivno arhiviranje koje se može podijeliti na fokusirano i tematsko pobiranje, pri čemu se arhivirana građa može bibliografski obraditi budući da je unaprijed identificirana prema uspostavljenim kriterijima. U tom se kontekstu može govoriti o web ili mrežnim publikacijama po analogiji s definicijom tradicionalne publikacije.

Odluka o metodi arhiviranju weba uvjetuje definiranje izgradnje sustava koji će funkcionalno pratiti poslovne procese upravljanja web sadržajima. U slučaju selektivne metode ti su procesi analogijom s poslovnim procesima upravljanja tradicionalnom građom sljedeći: identifikacija, bibliografska obrada, pobiranje, arhiviranje i davanje na korištenje. Model sustava koji povezuje knjižnično-informacijski sustav (poslovanje s tradicionalnom građom i online katalog) s digitalnim arhivom razvijen je u okviru europskog projekta NEDLIB (Slika 1). Taj je model poslužio Nacionalnoj i sveučilišnoj knjižnici u Zagrebu koja se odlučila za selektivno arhiviranje i za izgradnju vlastita rješenja sustava za arhiviranje, odnosno upravljanje mrežnim publikacijama.





Slika 1. Povezivanje postojećeg automatiziranog knjižničkog sustava s depozitarnim sustavom za digitalne sadržaje (dSEP)

Izgradnjom sustava za arhiviranje i njegova ugradnja u postojeći tradicionalni knjižnično-informacijski sustav ili izgradnjom sustava kao zasebnog arhiva, otvara se jedan od najakutnijih problema, a to je dugoročna zaštita i davanje na korištenje web sadržaja. U tu su svrhu nacionalne knjižnice Australije, Kanade, Danske, Finske, Francuske, Islanda, Italije, Norveške, Švedske i Velike Britanije, te Kongresna knjižnica i Internet Archive 2003. osnovale Međunarodni konzorcij za zaštitu interneta (International Internet Preservation Consortium, IIPC). Konzorciju se do 2007. pridružilo još 26 knjižnica, arhiva, muzeja i baštinskih ustanova, a tijekom 2007. objavljen je poziv za primanje novih članova. Nacionalna i sveučilišna knjižnica u Zagrebu dobila je takav poziv na temelju njena rada na arhiviranju mrežnih publikacija.

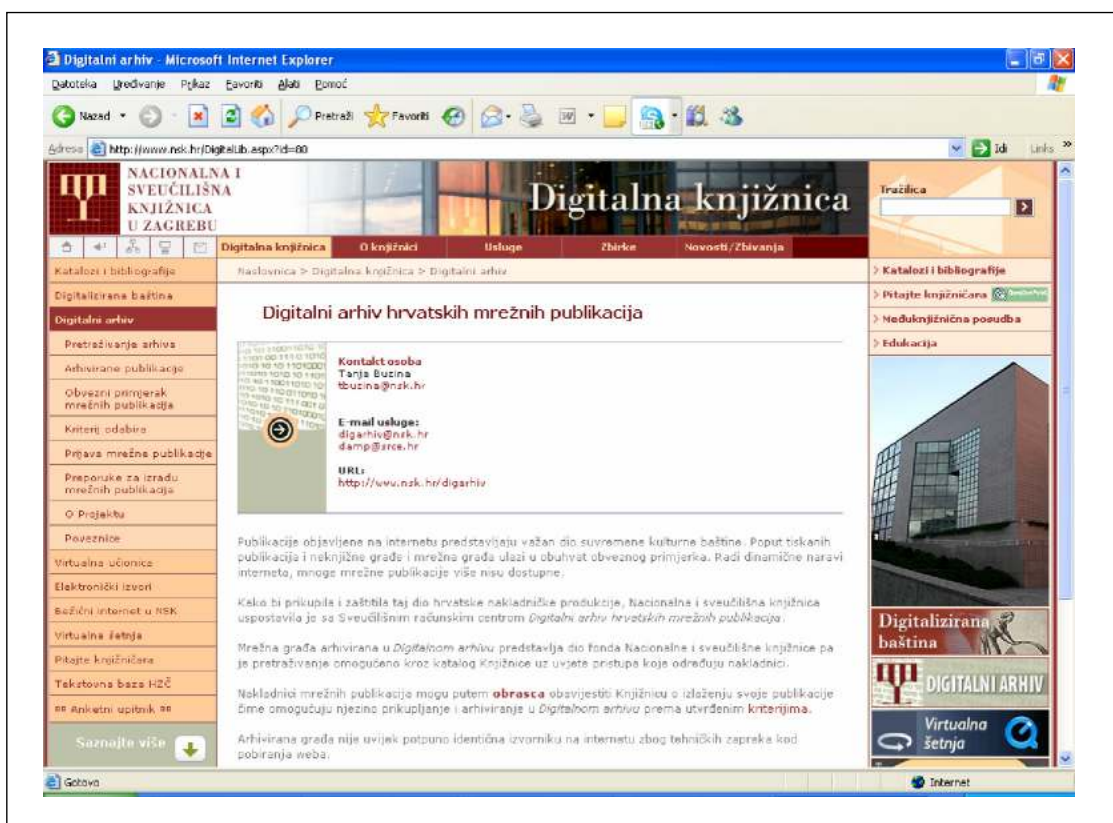
Digitalni arhiv u Nacionalnoj i sveučilišnoj knjižnici u Zagrebu

Nacionalna i sveučilišna knjižnica u Zagrebu (NSK) je, u suradnji sa Sveučilišnim računskim centrom Sveučilišta u Zagrebu (Srce), pokrenula 2003. *Projekt uspostave sustava za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija*. Osnovni cilj projekta bio je uspostava sustava za preuzimanje i arhiviranje obveznog primjerka hrvatskih publikacija na internetu uz očuvanje, u najvećoj mogućoj mjeri, autentičnosti njihova sadržaja, oblika i funkcionalnosti, a u svrhu njihove dugoročne zaštite i korištenja u budućnosti. U razdoblju do 1. listopada 2007. kada završava treća faza projekta, razvijen je sustav Digitalni arhiv mrežnih publikacija – DAMP, koji je povezan s knjižnično-informacijskim sustavom (prvo CROLIST-om, a po promjeni sustava Voyager-om), tj. katalogom NSK u funkcionalni integrirani sustav te definiran poslovni proces upravljanja hrvatskim mrežnim publikacijama. Istodobno s radom na projektu, tijekom 2005. uspostavljena je zasebna služba u okviru Odjela za katalogizaciju koja se isključivo bavi poslovanjem s ovom vrstom građe. Time je projekt prerastao u programsku djelatnost NSK (Slika 2). U tijeku je prihvaćanje razvojnog plana za četvrtu fazu projekta.

Koncept selektivnog arhiviranja statičkih i dinamičkih mrežnih publikacija podrazumijeva postupke identifikacije hrvatskih mrežnih publikacija na webu i njihovog odabira prema postavljenim kriterijima, bibliografske obrade, automatskog pobiranja (jednokratnog ili učestalog), pohrane i davanja pristupa arhiviranim publikacijama i njihovim inačicama putem kataloga NSK. Dva su osnovna razloga odabira ovakvog pristupa arhiviranju mrežnih publikacija.

Prvi je razlog zatečeno stanje u NSK. Slijedom zakonske regulative koja tretira i elektroničku publikaciju kao obvezni primjerak, 1998. se započelo bibliografski obrađivati serijske mrežne publikacije koje su bile identificirane na webu različitim načinima pronalaženja te odabrane prema općim kriterijima za odabir nacionalnog obveznog primjerka i posebnim kriterijima koji se osnivaju na prirodi mrežnih publikacija.





Slika 2. Web stranica programa NSK – Digitalni arhiv hrvatskih mrežnih publikacija:
<http://www.nsk.hr/digarhiv>



BOSNIACA 38

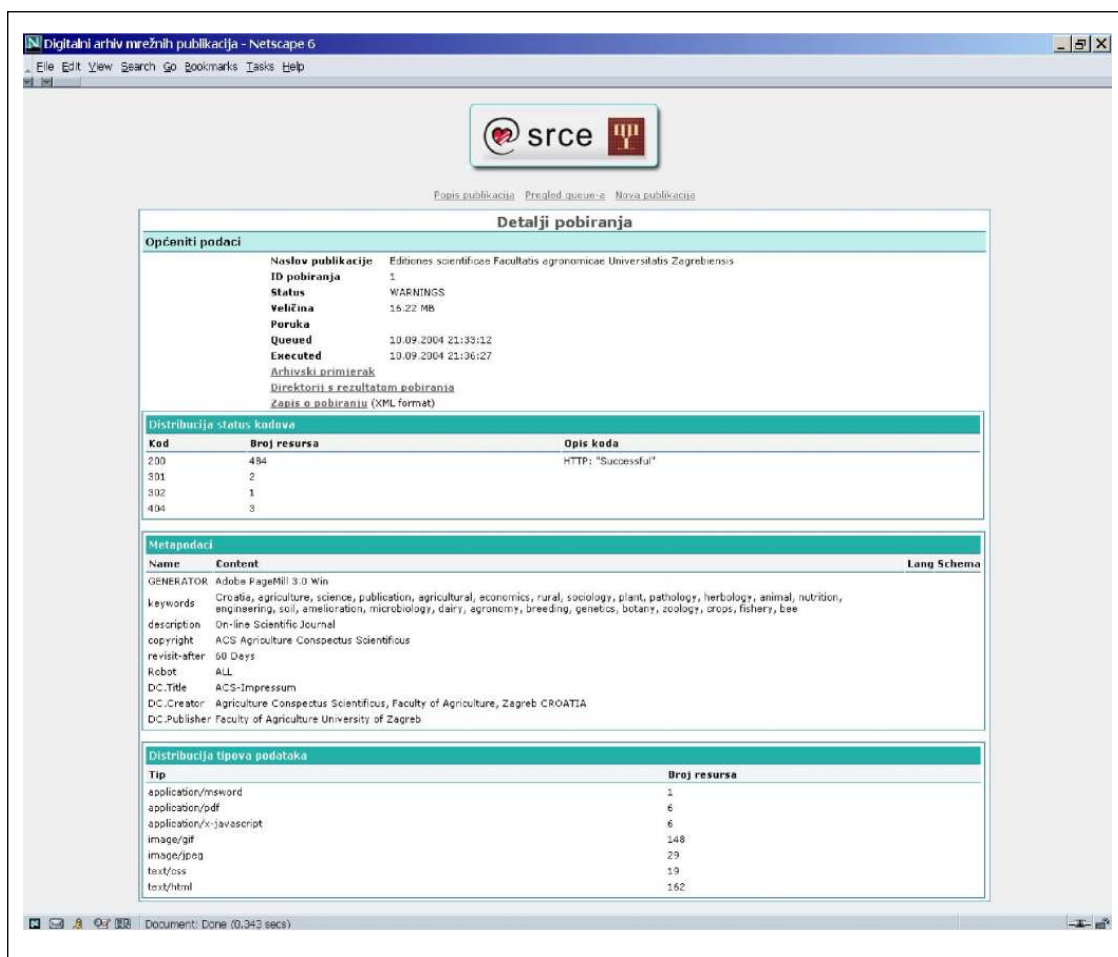
Obrađene publikacije nisu bile arhivirane, a pristup im je bio moguć samo na njihovom izvoru, tj. izvornoj adresi nakladnika. Stoga je bilo neophodno arhivirati već bibliografski identificirane publikacije radi dugoročne pohrane i pristupa, te definirati radni proces poslovanja mrežnom građom u NSK.

Drugi je razlog odabira metode selektivnog arhiviranja analiza međunarodnih iskustava na selektivnom arhiviranju. Ta iskustva upućuju da je takvim arhiviranjem moguće postići sljedeće ciljeve: (1) svaka je jedinica kvalitativno procijenjena i funkcionalna je u najvećoj mogućoj mjeri koju dozvoljava raspoloživa tehnologija; (2) svaka jedinica u arhivu može se u potpunosti katalogizirati i stoga postati dio nacionalne bibliografije; (3) svakoj jedinici u arhivu može se pristupiti, ovisno o tome je li pristup arhivu odobrio nakladnik; i (4) «specifičnosti» jedinica mogu se analizirati i opisati, što pospješuje unapređenje sustava arhiviranja i postavljanja zahtjeva za dugoročnom zaštitom.

Arhiviranje mrežne građe testno je započelo u siječnju 2004. po izgradnji prve radne verzije sustava DAMP, a verzija 1.0 sustava pušten je u rad u studenom iste godine. Početkom listopada 2007. arhiv je sadržavao 1840 jedinica pohranjenih u 17.680 instanci, tj. verzija učestalog pohranjivanja određene jedinice građe. Arhiv zauzima 1 TB, a analizom opsega arhiva ustanovljeno je da raste brzinom od prosječno 300 jedinica građe godišnje, dok je unatrag godinu dana pohranjeno 9.349 instanci jedinica. Građa je dostupna putem kataloga NSK, odnosno integriranim pristupom mrežnim i tradicionalnim publikacijama, te putem web sučelja arhiva (Slika 3). Od navedenog broja publikacija, oko 10 ih je zaštićeno lozinkom pa se mogu pregledavati na za to određenim računalima unutar NSK.

Analiza arhivirane vrste građe, zastupljenosti pojedinih domena i vrsta podataka, tehničkih poteškoća ili nemogućnosti pobiranja određenih mrežnih sadržaja, te načina utvrđivanja granica mrežne publikacije pomoću različitih parametara pobiranja ukazuje da rad s ovom vrstom građe iziskuje ažurno praćenje promjena na već arhiviranim jedinicama građe, prvenstveno na integrirajućoj građi, zatim stalne napore u savladavanju internetskih tehnologija, ali i preispitivanje razumijevanja svojstava i funkcionalnosti web sadržaja – informacijskih objekata!

Sustav DAMP: arhitektura sustava i njegovo funkcioniranje Sustav DAMP je razvijen na temelju javno dostupne programske podrške. Sastoji se od šest cjelina: podsustav za pobiranje web sjedišta (tzv. *gatherer*), podsustav za kontrolu pobiranja i arhiviranje rezultata pobiranja (tzv. *controller*), podsustav za



Slika 5: Web sučelje: Detalji pobiranja

Sustav DAMP ima i određenih ograničenja koja su vezana uz konkretnu izvedbu pojedinog web sjedišta koje se pobire. Uporaba određenih web tehnologija u specifičnim uvjetima otežava, a ponegdje i posve onemogućuje efikasno arhiviranje. U svrhu poboljšanja uspješnosti arhiviranja, ali i upućivanja na standardni način opremanja mrežnih publikacija, izrađene su *Smjernice za nakladnike / autore mrežnih publikacija*.

Svrha je *Smjernica*:

- uputiti u zakonske propise vezane uz obvezni primjerak mrežne građe;
- uputiti u način obavještanja Nacionalne i sveučilišne knjižnice o objavljivanju publikacije na mreži i funkciji digitalnog arhiva za pohranu mrežne građe;
- preporučiti način izrade mrežnih publikacija zbog njihova lakšeg pronalaženja i uspješnog pohranjivanja;
- dati informaciju o programu kojim se NSK služi prilikom pobiranja mrežne građe;
- uputiti na relevantne dokumente vezane uz kriterije odabira, postupke prijave i pohrane;
- dati osnovnu informaciju o projektu uspostavljanja *Digitalnog arhiva hrvatskih mrežnih publikacija*.

Prijedlog unapređenja sustava: DAMP IV

Razvojni poslovi predviđeni za iduće jednogodišnje razdoblje su okvirno sljedeći: uspostava suradnje sa srodnim ustanovama u Hrvatskoj radi koordiniranog arhiviranja hrvatske mrežne građe, analiza postojećih kriterija za selekciju građe i preporuka za nakladnike / autore te njihovo osuvremenjivanje, potpuna funkcionalnost harvestiranja arhiva putem protokola OAI i uključivanje u međunarodne sustave pretraživanja digitalnih sadržaja, kao npr. TEL (The European Library), unapređenje web sučelja na temelju rezultata analize korištenja i indeksiranje arhiva s ciljem omogućavanja pretraživanja sadržaja DAMP-a. Kako hrvatske tekuće nacionalne bibliografije knjiga i serijskih publikacija ne uključuju mrežne publikacije, bit će potrebno razmotriti koncept uključivanja ove vrste građe ili u postojeće nizove ili u zaseban niz hrvatskih mrežnih publikacija.

LITERATURA

Cunnea, Paul. Web arhiviranje u Škotskoj i Velikoj Britaniji: stižu li problemi? // 10. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredile Ivana Marinković Zenić i Mirna Willer. Zagreb : Hrvatsko knjižničarsko društvo, 2007. (u tisku)

Gilliland A. J. Setting the stage. // Gill, T.; Gilliland, A.J.; A.S. Woodley. Introduction to metadata : pathways to digital information / edited by Burtha Baca. Online ed., Ver. 2.1. J. Paul Getty Trust, 2005. http://www.getty.edu/research/conducting_research/standards/intrometadata/ [Printer Friendly PDFs]

Klarin, Sofija. Predmet, motivi i metode arhiviranja sadržaja weba // 8. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2005. Str. 22-35.

Klarin, Sofija; Tomislav Murati. Identifikacija, odabir, obradba i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole. // 4. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u kontekstu globalne informacijske infrastrukture : zbornik radova / uredile Mirna Willer i Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2001. Str. 41-56.

Klarin, Sofija; Sonja Pigac. Hrvatske daljinski dostupne elektroničke serijske publikacije. // Vjesnik bibliotekara Hrvatske 4, 43(2000.[i.e. 2001.]), 156-167.

Milinoić, Miroslav; Nebojša Topolščak. The architecture of DAMP. // Widwisawn, vol. 3, no. 3, 2005. <http://widwisawn.cdrl.strath.ac.uk/Issues/Vol3/issue3_3_1.html>

Nacionalna i sveučilišna knjižnica (Zagreb). Kriteriji odabira obveznog primjerka mrežne građe za obradu i arhiviranje. Dostupno na: <http://www.nsk.hr/DigitalLib.aspx?id=83>.

Nacionalna i sveučilišna knjižnica (Zagreb); Sveučilište u Zagrebu. Sveučilišni računski centar (Srce). Arhiviranje obveznog primjerka mrežne građe : smjernice za nakladnike/autore. Verzija 1.0 (2005.).

D o s t u p n o n a : http://www.nsk.hr/UserFiles/File/dokumenti/Smjernice%20za%20nakladnike_Verzija%201%200.pdf;

Nacionalna i sveučilišna knjižnica (Zagreb). Preporuke za izradu mrežnih publikacija. Dostupno na: <http://www.nsk.hr/DigitalLib.aspx?id=356>

Pigac, Sonja. Digitalni arhiv mrežnih publikacija : rad u sustavu za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija // 8. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2005. Str. 61-67.

Pigac, Sonja; Tanja Buzina. Selektivno arhiviranje hrvatskog weba : rezultati i otvorena pitanja // 9. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredile Mirna Willer i Ivana Zenić. Zagreb : Hrvatsko knjižničarsko društvo, 2006. Str. 28-39.

Steenbakkers, Johan. *Setting up a Deposit System for Electronic Publications. The Hague : Koninklijke Bibliotheek, 2000. (NEDLIB Report Series ; 5). Dostupno i na : <http://nedlib.kb.nl/results/NEDLIBguidelines.pdf>. Projekt NEDLIB dostupan na: <http://nedlib.kb.nl/>*

Willer, Mirna; Miroslav Milinoić. Prema trećoj generaciji knjižnično-informacijskih sustava : hibridna knjižnica za hibridne usluge. // 8. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb: Hrvatsko knjižničarsko društvo, 2005. Str. 36-67.



BOSNIACA 41

Bilješke:

¹ Klarin, Sofija; Sonja Pigac. Hrvatske daljinski dostupne elektroničke serijske publikacije. // Vjesnik bibliotekara Hrvatske 4 43(2000.(i.e. 2001.)), 157.

² Više o metodama arhiviranja sadržaja weba i pristupa nacionalnih knjižnica može pitanju vidjeti u: Klarin, Sofija. Predmet, motivi i metode arhiviranja sadržaja weba // 8. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2005. Str. 22-35.; Cunnea, Paul. Web arhiviranje u Škotskoj i Velikoj Britaniji : što su li problemi? // 10. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredile Ivana Marinković Zenić i Mira Willer. Zagreb : Hrvatsko knjižničarsko društvo, 2007. Str. 27-39.

³ Steenbakkers, Johan. Setting up a Deposit System for Electronic Publications. The Hague : Koninklijke Bibliotheek, 2000. (NEDLIB Report Series ; 5). Dostupno na: <http://nedlib.kb.nl/results/NEDLIBguidelines.pdf>. Projekt NEDLIB dostupan na: <http://nedlib.kb.nl/>

⁴ Willer, Mira; Miroslav Milinović. Prema trećoj generaciji knjižnično-informacijskih sustava : hibridna knjižnica za hibridne usluge // 8. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2005. Str. 36-67.

⁵ International Internet Preservation Consortium – IIPC. Dostupno na: <http://netpreserve.org/about/index.php>. Ciljevi Konzorcija su (1) omogućiti sakupljanje, čuvanje i trajni pristup internetskim sadržajima, (2) razvijati i davati na korištenje zajedničke alate, tehnologije i standarde za uspostavu međunarodnog arhiva, (3) podupirati inicijative i zakone koji potiču sakupljanje, čuvanje i pristup internetskim sadržajima, i (4) poticati knjižnice, arhive, muzeje i ostale baštinske ustanove na prikupljanje sadržaja s interneta.

⁶ Pigac, Sonja. Digitalni arhiv mrežnih publikacija : rad u sustavu za preuzimanje i arhiviranje obveznog primjerka hrvatskih mrežnih publikacija // 8. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredila Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2005. Str. 61-67.

⁷ Nacionalna i sveučilišna knjižnica (Zagreb). Kriteriji odabira obveznog primjerka mrežne građe za obradu i arhiviranje. Dostupno na: <http://www.nsk.hr/DigitalLib.aspx?id=83>.

⁸ Klarin, Sofija; Tomislav Murati. Identifikacija, odabir, obradba i osiguravanje dostupnosti mrežnih publikacija u kontekstu nacionalne bibliografske kontrole. // 4. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u kontekstu globalne informacijske infrastrukture : zbornik radova / uredile Mira Willer i Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2001. Str. 41-56.

⁹ Načini pronalaženja jedinica su: putem portala, mrežnih mjesta, pretraživača, informacije iz medija (novine, TV, radio), informacije iz tiskanog ekvivalenta, informacije od samih izdavača, usmeno obavještanje od korisnika; knjižničara.

¹⁰ Klarin, Sofija; Sonja Pigac. Hrvatske daljinski dostupne elektroničke serijske publikacije. Isto.

¹¹ Pigac, Sonja; Tanja Buzina. Selektivno arhiviranje hrvatskog weba : rezultati i otvorena pitanja // 9. seminar Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredile Mira Willer i Ivana Zenić. Zagreb : Hrvatsko knjižničarsko društvo, 2006. Str. 28-39.

¹² Gilliland A. J. Setting the stage. // Gill, T.; Gilliland, A.J.; A.S. Woodley. Introduction to metadata : pathways to digital information / edited by Burtha Saca. Online ed., Ver. 2.1. J. Paul Getty Trust, 2005. http://www.getty.edu/research/conducting_research/standards/intro/metadata/ [Printer Friendly PDFs]

¹³ Milinović, Miroslav; Nebojša Topolšćak. The architecture of DAMP. // Widisaw, vol. 3, no. 3, 2005. < http://widisaw.cdlib.org/strath.ac.uk/issues/vol3/issue3_3_1.htm. >

¹⁴ Nacionalna i sveučilišna knjižnica (Zagreb). Sveučilište u Zagrebu. Sveučilišni računski centar (Srce). Arhiviranje obveznog primjerka mrežne građe : smjernice za nakladnike/autore. Verzija 1.0 (2005.). Dostupno na: http://www.nsk.hr/UserFiles/File/dokumenti/Smjernice%20za%20nakladnike_Verzija%201%200.pdf.

Nacionalna i sveučilišna knjižnica (Zagreb). Preporuke za izradu mrežnih publikacija. Dostupno na: <http://www.nsk.hr/DigitalLib.aspx?id=356>



BOSNIACA 42

