

# Moč (meta)podatkov: problematika upravljanja in uporabe (meta)podatkov

Zoran Krstulović, Matjaz Kragelj, Narodna in univerzitetna knjižnica, Turjaška 1, 1000 Ljubljana  
 zoran.krstulovic@nuk.uni-lj.si; matjaz.kragelj@nuk.uni-lj.si

## Izveček

Namen prispevka je opozoriti na nove okoliščine, ki jih v realnem svetu nismo poznali, nastanejo pa zaradi tehnoloških možnosti obdelave podatkov in njihove spletne dostopnosti. Navedene so tudi nekatere uporabniške zahteve glede objavljenih podatkov o njih ali njihovih delih. Spletna dostopnost vsebin s področja kulture in znanosti je danes nepogrešljiva sestavina ponudbe knjižnic, muzejev in galerij. Prav računalniška obdelava analognih vsebin (npr. digitalizacija) in drugo generiranje raziskovalnih podatkov – ankete, meritve itd. – povečuje ne le količino podatkov, ki so na voljo, temveč tudi širi možnosti uporabe le-teh ter omogoča ustvarjanje novega znanja. Metode obdelave besedil omogočajo nove povezave in skoraj trenutno odkrivanje vsebin o predmetu iskanja (npr. o fizičnih osebah), za katere bi v fizičnem svetu potrebovali mesece. Soočamo se s pojavom t. i. »internetne pozabe«, ko uporabniki ne želijo, da so podatki o njih najdljivi z lahkoto. Kljub temu da javni zavodi s področja kulture že več kot desetletje razvijamo spletne storitve, nismo še vzpostavili ustrezne oblike izvajanja javne službe na spletu, ki bi se nanašala na vsa omenjena vprašanja. Ali širše – nujno je opredeliti značilnosti javne službe s področja kulture na spletu, ugotoviti, kako se ta že uresničuje, ter na tej podlagi premisliti izhodišča minimalnih skupnih podlag ustvarjanja, upravljanja in dostopnosti (meta)podatkov.

**Ključne besede:** metapodatki, arhiv, spletni arhiv, pravica do pozabe, trajno ohranjanje.

## Abstract

### The Power of (Meta)Data: Management and Use of (Meta)Data Issues

The aim of this paper is to draw attention to new situations which we are not familiar with in the real world. They arise due to novel technological possibilities of data processing and their online accessibility. We also comment on some user requirements regarding published data about them or their work. Today, web accessibility of cultural and scientific content is an indispensable component of what libraries, museums and galleries offer to their users. Computer processing of analogue content (digitization, etc.) and the generation of research data (surveys, measurements, etc.) increase not only the quantity of available data, but also expand the possibility of using the data, and thus enable the creation of new knowledge. The text processing methods provide new connections and almost immediate retrieval of the content on the search object (e. g. natural persons), for which months would be needed in the physical world. We encounter a new phenomenon of the so-called »online oblivion« when users do not wish that the data about them is easily found. Although web services have been developed by public cultural institutions for more than a decade, no appropriate form of providing public services on the web has been established yet to address all of the above-mentioned issues. Therefore, it is necessary to define the characteristics of public cultural services on the web, to determine their current performance, and on this basis, to reconsider the minimum common grounds of creation, management and accessibility of (meta)data.

**Key words:** metadata, archive, web archive, right to oblivion, long-term preservation.

## 1 UVOD

Zagotavljanje spletne dostopnosti do vsebin s področja kulture in raziskovalne dejavnosti je ena ključnih zahtev sodobnega časa državljanov Slovenije in Evropske skupnosti [5]. Zato so t. i. e-vsebine danes nepogrešljiva sestavina ponudbe knjižnic, muzejev, galerij ter spletnih portalov s področja kulture in raziskovalne dejavnosti. Dostopnost spletnih storitev je izredno pomembna za razvoj družbe, za raziskovalno in izobraževalno dejavnost, pa tudi za osebni razvoj posameznika. Smeri razvoja, ki bodo vplivale na ustvarjanje, dostopnost

in trajno hranjenje e-vsebin, opredeljuje *Evropska digitalna agenda* [1]. Pomembno je, da so spletne storitve, zametek tistega, kar lahko poimenujemo »javna služba na spletu«, že vgrajene v izvajanje javne službe na omenjenih področjih kot komplementarna dejavnost javni službi v fizičnem svetu. Med ključnimi postopki ustvarjanja e-vsebin, *conditio sine qua non* spletnih storitev, je digitalizacija gradiva, ki ga hranijo dediščinske ustanove v fizični obliki. Znotraj nabora procesov digitalizacije je eden pomembnejših postopkov izdelava metapodatkov.

Med metapodatke ne uvrščamo le bibliografskih podatkov – podatkov o fizičnem gradivu, ki služijo identifikaciji vsebine –, temveč tudi podatke, pomembne za trajno hranjenje ustvarjenih digitalnih objektov in za možnost njihove uporabe v prihodnosti. Pomemben del metapodatkov pridobimo iz optično prepoznanih besedil (OCR). Prav ta del metapodatkov na eni strani prispeva dodani vrednosti digitalizacije, po drugi pa ustvarja za posameznike – avtorje oz. osebe, katere se pojavljajo v digitalnih vsebinah – moteče okoliščine.

Za namen tega prispevka se omejimo le na digitalizirane časopise (dnevnike, tednike) in znanstvene ter strokovne revije. Časopis je medij, ki pokriva nabor vsebin od dnevnih novic s področja politike do besedil, ki poročajo o dogodkih, povezanih s posameznimi osebami. Možnosti obdelave (meta)podatkov, ki so bili na novo ustvarjeni iz analognih virov, poraja nove okoliščine, ki jih v realnem svetu nismo poznali. Včasih je bilo treba listati fizične izvode in iskati podatke. Skoraj praviloma nismo vedeli, ali podatki o določenem predmetu zanimanja ali tematiki sploh obstajajo v določeni publikaciji, predvidevanje o njihovem obstoju se je opiralo na znane dogodke (koncerte, gledališke predstave, razstave, politične ali družbene dogodke ipd.). Zapise o posameznikih je bilo skoraj nemogoče odkriti, razen naključno. Nove metode obdelave besedil omogočajo nove povezave med podatki in skoraj trenutno odkrivanje vsebin, ki govorijo o predmetu iskanja (npr. fizičnih osebah), za katere bi v fizičnem svetu potrebovali mesece. V iskalnike vgrajeni jezikovni pregibniki pa omogočajo še dodatne možnosti dostopa do informacij, ki so skrite v digitaliziranih vsebinah.

## 2 OSEBNE ZGODBE SPLETNIH ISKALCEV

Oglejmo si bolj podrobno nekaj značilnih okoliščin, nastalih zaradi nezadovoljstva posameznih uporabnikov z vsebinami, ki so bile (popolnoma legalno) objavljene v tiskanih publikacijah, digitalizirane različice pa so dostopne prek Digitalne knjižnice Slovenije.

### Primer 1

*Na vaši spletni strani sem po naključju odkrila naveden vir iz Planinskega vestnika: Analiza nesreč in reševanja gorske reševalne službe v letih 2001–2011<sup>1</sup> in tam odkrila nave-*

<sup>1</sup> Gre za publikacijo Analiza nesreč in reševalnega dela v letu 2000, *Planinski vestnik* 2001, 6, priloga (URN:NBN:SI:DOC-FYPWC1CB).

*deno tudi svoje ime in priimek, ker se mi je pred desetimi leti zgodila manjša nezgoda v hribih. Ne vem pa, od kod vam dovoljenje, da vi re objavljate s polnimi imeni in priimki, saj so to OSEBNI podatki! Prosim, da nemudoma odstranite moje ime, sicer se bom morala obrniti po pravno pomoč.<sup>2</sup>*

V omenjeni publikaciji so poleg analize reševalnega dela objavljeni tudi pregledi reševalnih, iskalnih in poizvedovalnih akcij, ki prinašajo vrsto podatkov o nesrečah v gorah – datum in kraj nesreče, ime in priimek, starost, kraj bivanja, državljanstvo, vrsto in vzrok poškodbe ponesrečenca. Vsi ti podatki so bili natisnjeni oz. javno objavljeni popolnoma legalno. Ime in priimek uporabnice sta najdljiva na portalu Digitalne knjižnice Slovenije le z uporabo opcije iskanja po celotnem besedilu. Tudi v tem primeru je dobljeno število rezultatov preveliko in je treba uporabiti filtre, ki zožijo rezultate na posamezne naslove revij, med katerimi je tudi *Planinski vestnik*. V primeru uporabe narekovajev pri iskanju (»ime priimek«) in opcije iskanja po celotnem besedilu se omenjena publikacija pojavi kot prvi rezultat iskanja. Enako uporaba narekovajev na iskalniku Google kot rezultat prikaže povezavo na publikacijo na prvi strani.

### Primer 2

*Sem [...] in zahtevam, da se odstrani članek.<sup>3</sup>*

*[...] iz Grosupljega je v rejniško družino prišla še kot dojenček, zato pravih razlogov za odhod od svojih staršev ne pozna. Z njimi tudi nima nobenega stika. »V rejništvu so dobre in slabe stvari, saj se hočeš nočeš zavedaš, da nisi pravi otrok ljudi, pri katerih živiš. Srečan, kot je tokratno, se vedno znova veselim. Spoznala sem, da četudi sem rejenka, nisem zato ne vem kakšna sirota ali potrebna posebnega pomilovanja.«*

*Stran spletne strani je [...].*

*To želim takoj, drugače sledi tožba. Hvala za razumevanje.<sup>4</sup>*

Tudi v tem primeru sta ime in priimek uporabnice najdljiva na Digitalni knjižnici Slovenije le z uporabo opcije iskanja po celotnem besedilu. Prav tako kot v prejšnjem primeru je število rezultatov preveliko in jih je treba omejiti s filtri. V primeru uporabe narekovajev pri iskanju in opcije iskanja po celotnem

<sup>2</sup> Elektronsko sporočilo upravljavcem Digitalne knjižnice Slovenije, 28. 6. 2011.

<sup>3</sup> Prispevek je bil objavljen v publikaciji *Novi tednik* 62/2007 (28. 8.), 68, 6 (URN:NBN:SI:DOC-C8MONORS). Uporabnica problematizira le navedeni del članka, ki govori o njej.

<sup>4</sup> Elektronsko sporočilo upravljavcem Digitalne knjižnice Slovenije, 4. 10. 2011.

besedilu dobimo številko Novega tednika, ki vsebuje omenjeni članek kot edini rezultat. Iskanje prek iskalnika Google povezavo do številke časopisa prikaže kot drugi rezultat, pokaže pa na datoteko v formatu html, ki je namenjena pregledovanju optično prepoznanega besedila.

### Primer 3

*Zanima me, ali lahko moje ime (metapodatke) umaknete iz spodnjega naslova. [...]*

Uporabnikova zahteva se nanaša na omembo avtorskega prispevka v *Bibliografiji sodelavk in sodelavcev SEM za leto 2008*.<sup>5</sup> Prav tako kot v prejšnjih primerih je število rezultatov iskanja uporabnikovega imena preveliko in jih je treba omejiti s filtri. V primeru uporabe narekovajev pri iskanju in možnosti iskanja po celotnem besedilu dobimo številko *Etnologa*, ki vsebuje omenjeni članek kot edini rezultat. Iskanje prek iskalnika Google povezavo do številke časopisa prikaže kot peti rezultat, pokaže pa na datoteko v formatu html.

Stališče Narodne in univerzitetne knjižnice kot upravljavke portala Digitalne knjižnice Slovenije je bilo v vseh treh navedenih primerih enako. Digitalna knjižnica Slovenije zagotavlja dostop do virov v skladu z dogovori z izdajatelji publikacij, in sicer zagotavlja dostop večinoma do digitalnih kopij publikacij, ki so predhodno že bile objavljene v tiskani obliki in so kot take bile in so še dostopne bralcem oz. uporabnikom. Na vsebino teh publikacij digitalna knjižnica nima vpliva, ta je v pristojnosti uredništev, zato tudi za vsebino posamezne publikacije ne more ne odgovarjati, še manj jo spreminjati. Edina možnost, da bi preprečili najdljivost podatkov o konkretnem uporabniku, je, da celotno publikacijo umaknemo s portala digitalne knjižnice, za tak poseg pa bi potrebovali zahtevo uredništva revije oz. časopisa in ne le prizadetega posameznika. Uporabnikom je bilo predlagano, da se z zahtevo obrnejo tudi na uredništva konkretnih posameznih časopisov oz. revij, ki so edino pristojna in odgovorna za vsebino publikacij, ki jih izdajajo. Po odgovoru ni bilo več nobenega odziva, zato ne vemo, ali so se prizadeti uporabniki res obrnili na izdajatelje. Le v tretjem primeru je uporabnik odpisal, naj podatki ostanejo taki, kot so, ali drugače – s pojasnilom se je strinjal.

<sup>5</sup> *Etnolog. Nova vrsta* 20/2010, 311–333 (URN:NBN:SI:DOC-G9CHIQHQ).

Podoben, vendar nekoliko bolj kompleksen primer je bila zahteva, ki smo jo prejeli skupaj z mnenjem informacijskega pooblaščenca.

### Primer 4

[Zahtevek informacijskemu pooblaščenca]: *Pri vnosu [svojega] imena v spletni iskalnik Google se med zadetki na prvi strani pojavi tudi [...] (oz. v prilogi). Menim, da gre pri tem za kršitev osebnih podatkov. Ne želim namreč, da so moji podatki o otroštvu dostopni vsakomur na spletu. Izdajatelj je glede na spletni naslov NUK <http://www.dlib.si/menu/kontakt>. Ali lahko sam ali prek vas zahtevam, da sporno datoteko pdf<sup>6</sup> umaknete s spleta?*<sup>7</sup>

Sledil je zahtevek upravljavcem Digitalna knjižnice Slovenije z enako vsebino in dodatkom:

*Spodaj [je] mnenje informacijskega pooblaščenca o tej zadevi. Prosim vas, da na podlagi 32. člena Zakona o varstvu osebnih podatkov onemogočite oz. umaknete sporno datoteko pdf na način, da le-ta ne bo več dostopna prek iskalnikov! Prosim za odgovor.*<sup>8</sup>

Informacijski pooblaščenec je pojasnil, da se zave da problematike digitalizacije časopisov in spletnih novic, v katerih so bili morda določeni osebni podatki pred časom popolnoma zakonito objavljeni v okviru novinarskega poročanja. Hkrati pa so ti podatki tudi čez leta preprosto dostopni na spletu s pomočjo spletnih iskalnikov tako rekoč komur koli. Trenutno, nadaljuje, evropska zakonodaja opredeljuje le pravico do izbrisa podatkov, kadar so bili ti obdelovani v nasprotju z zakonom, so nepopolni, neažurni ali netočni, kar pa pri novicah na spletu običajno ne pride v poštev (če so bili podatki objavljeni zakonito). Zaradi pomanjkljivega varstva pravic posameznika v takih primerih je tako v predlogu nove *Splošne uredbe o varstvu podatkov* [6], vsebovana nova »pravica do pozabe«, v okviru katere bo posameznik lahko zahteval umik takih podatkov s spleta, čeprav so bili nekoč objavljeni zakonito. Uveljavitev splošne uredbe je v državah članicah EU pričakovana v naslednjih letih. Pooblaščenec je uporabniku v konkretnem primeru lahko svetoval le, da se obrne neposredno na

<sup>6</sup> Sporni članek je objavljen v časopisu *Naša skupnost (Ljubljana)* 22/1981, 2 (URN:NBN:SI:DOC-5787O2S9).

<sup>7</sup> Elektronsko sporočilo informacijskemu pooblaščenca, 23. 11. 2012. Sporočilo je bilo priloženo zahtevku, elektronskemu sporočilu upravljavcem Digitalne knjižnice Slovenije.

<sup>8</sup> Elektronsko sporočilo upravljavcem Digitalne knjižnice Slovenije, 22. 12. 2012.

upravljavca spletne strani in poda zahtevo za umik podatkov, novice, v kateri se pojavlja, na podlagi 32. člena *Zakona o varstvu osebnih podatkov* [8], saj objavljeni podatki niso ažurni, čeprav so bili nekoč objavljeni zakonito. In naprej: »Upravljavca ima na voljo tudi tehnične možnosti preprečevanja dostopa iskalnikom do takih vsebin. Tako je vsebina še vedno dostopna na spletu, a le prek neposrednega dostopa in ne prek iskalnikov.«<sup>9</sup>

Po temeljiti preučitvi prošnje po umiku datoteke na način, da na spletu ne bo več dostopna prek spletnih iskalnikov in mnenja informacijskega pooblaščenca, je NUK kot upravljavca Digitalne knjižnice Slovenije ugotovil, da za tako zahtevo ni pravne podlage. Obravnavana publikacija je bila na portalu Digitalne knjižnice Slovenije (www.dLib.si) objavljena popolnoma zakonito. Člen 32 *Zakona o varstvu osebnih podatkov* (neažurnost sicer zakonito objavljenih podatkov, kot v svojem mnenju navaja informacijski pooblaščenec) v konkretnem primeru ni ustrezna pravna podlaga. Vprašanje ažurnosti oz. neažurnosti objavljenih podatkov v smislu 32. člena tega zakona je v tem primeru popolnoma brezpredmetno, saj je to pri novicah nesmiselno. Da to pri novicah ne pride v poštev, v svojem mnenju navaja tudi informacijski pooblaščenec. Poleg tega upravljavca portala, Narodna in univerzitetna knjižnica, tudi ni na noben način povezan s spletnimi iskalniki in njihovimi lastniki. Kljub temu je NUK z ukrepi, ki so na voljo, v zvezi s prošnjo storil, kar je bilo mogoče. Optično prepoznano in indeksirano besedilo članka je bilo umaknjeno iz baze podatkov. Zdaj ni več mogoče prek Digitalne knjižnice Slovenije z vpisom imena in priimka uporabnika priti do omenjenega članka. Ta postopek prav tako onemogoči indeksiranje besedila v prihodnje s pomočjo spletnih iskalnikov.<sup>10</sup>

Daljši zahtevki nezadovoljnega spletnega iskalca smo dobili v zvezi z dostopnostjo njegovega članka kot posebne bibliografske enote na Digitalni knjižnici Slovenije.

### Primer 5

»Ugotovil sem, da se da priti na internetu s pomočjo iskalnika Google do članka, ki sem ga napisal jaz in je bil objavljen v reviji *Socialno delo*, letnik 1993, št. 3–4<sup>11</sup> [...].

<sup>9</sup> Elektronsko sporočilo upravljavcem Digitalne knjižnice Slovenije, 22. 12. 2012. Iz priloženega mnenja informacijskega pooblaščenca.

<sup>10</sup> Elektronsko sporočilo, odgovor uporabniku, 23. 1. 2013.

<sup>11</sup> *Socialno delo* 32/1993, 3–4 (URN:NBN:SI:DOC-JCA8BDT1).

*Prosti dostop do tega članka pa omogoča izključno vaša spletna stran dlib.si. Ta članek je v obliki pdf in se da priti do njega tako, da se vpiše v iskalnik Google [»ime priimek« avtorja] in iskalnik med prvimi rezultati poda ravno ta članek, ki sem ga napisal jaz. Poda ga takoj v formatu pdf z možnostjo 'downloadanja'. Mogoče so vsi izvodi revije *Socialno delo* prosto dostopni na internetu – tudi brez vaše spletne strani, to ne vem, vem pa, da je ta članek moje avtorsko delo in vas nisem pooblastil ali vam dal dovoljenja, da ga objavljate zunaj konteksta izvoda revije *Socialno delo*. Vi v tem primeru ne objavljate celotnega izvoda revije *Socialno delo*, ampak izključno moj članek [...] kot posebno elektronsko enoto – datoteko pdf in ne kot del izvoda revije *Socialno delo*. [...]*

*Iskalnik google ponudi objavljena dela na internetu na podlagi ključnih besed in te ključne besede so lahko npr. ime avtorja. V mojem primeru se na podlagi ključnih besed avtorja [»ime priimek«] zgodita dve stvari: prvič se najde dve samostojni datoteki pdf, ki ste ju vi ustvarili na podlagi del – člankov, ki sem jih jaz napisal v okviru dveh revij. Drugič: na podlagi ključnih besed avtorja spletni iskalnik Google ponudi vsemu svetu v prosti dostop ti dve samostojni elektronski objavi, ki ste ju ustvarili vi in ne jaz.*

*Mogoče lahko objavljate celotne izvode določenih revij, toda dvomim, da lahko zunaj konteksta teh revij objavljate posamezne članke, ki so vsaj delno moje avtorsko delo. Dvomim torej, da smete ustvarjati te posamezne datoteke pdf kot nekakšne proste objave, in dvomim tudi, da jih lahko ponudite v svetovni splet, da jih lahko spletni iskalniki, kot je Google, najde prek določenih ključnih besed. Na podlagi katerega dovoljenja ste to naredili, ne vem, na podlagi mojega zagotovila ne.*

*Prosil bi vas torej, da na vaši spletni strani odstranite po mojem mnenju nezakonito ustvarjene datoteke pdf, pa tudi onemogočite spletnim iskalnikom, kot je Google, da jih najde npr. s ključnimi besedami imena avtorja.«*

Prošnja spletnega iskalca je zahtevala daljše pojasnilo stališča upravljavca podatkov. Prvič, da je oblika objave znanstvenih in strokovnih revij na portalu Digitalne knjižnice Slovenije v skladu z dogovorom, ki ga upravljavca, NUK, podpiše z vsako posamezno založbo oz. izdajateljem z namenom spletne objave kot najširše mogoče dostopnosti do rezultatov znanstvenega dela. V skladu z dogovorom je založnikova oz. izdajateljeva dolžnost, da z avtorji uredi medsebojna razmerja. Drugič, oba članka, ki ju omenja iskalec, najdemo prek iskalnika Google v formatu html, ki ga Google samodejno ustvari iz različice do-

kumenta, ki ga najde na spletu, in ga samodejno pretvori v format html. Pri tem Google na vidnem mestu navede vir oz. izvorno spletno lokacijo dokumenta (na vrhu Googlove različice dokumenta html). In tretjič, pojasnilo, da samih datotek drugim servisom ne pošiljamo ali jim jih kako drugače posredujemo. Vendar pa so podatki (in tudi dokumenti), ko so enkrat na spletu, dostopni spletnim iskalnikom, upravljalec pa ne more vplivati na Googlovo politiko zajemanja spletnih vsebin, ki jih mimo vednosti lastnikov redno zajema Google, da svojim uporabnikom omogoči hitro najdljivost in neposredno dostopnost vsebin. Predlagali smo iskalcu, da v tem primeru lahko pomagamo le tako, da članek izključimo s portala dLib.si. Avtor se je odločil za umik članka s portala.

### 3 OBVEZNI IZVOD SPLETNIH PUBLIKACIJ IN OBLIKOVANJE SPLETNEGA ARHIVA

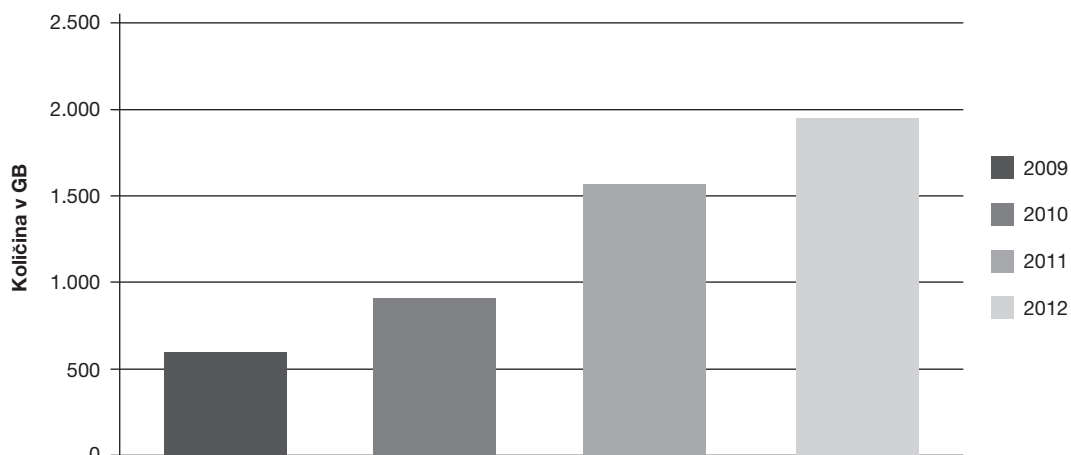
Zajem spleta oz. spletnih publikacij spada v izvajanje nalog, ki jih predpisuje *Zakon o obveznem izvodu publikacij* [7] in *Pravilnik o vrstah in naboru elektronskih publikacij za obvezni izvod* [2]. V skladu z omenjenimi predpisi NUK zajema spletne strani, in sicer selektivno. Vzrok za selektivnost so omejeni človeški viri. Arhiv se nahaja na spletnem naslovu *arhiv.nuk.uni-lj.si* in omogoča preiskovanje arhiviranih slovenskih spletnih strani po besedilu spletnega mesta, domeni spletnega mesta in vsebinskih področjih.

*Zakon o obveznem izvodu publikacij* v 2. členu definira spletne publikacije kot »elektronske knjige, elektronske časopise in časnike, dostopne po spletu ter spletne strani in podobno« in jih v 4. členu opredeli kot predmet obveznega izvoda. *Pravilnik o vrstah in izboru elektronskih publikacij za obvezni izvod* med

spletne publikacije poleg knjig, časnikov, časopisov in člankov, ki so objavljeni na spletu, uvršča tudi »spletna mesta (strani) organizacij, oseb in dogodkov, portale, spletno dostopne storitve, podatkovne zbirke, spletne novice, spletne konference (forumi), spletne biltene (newsletters) ter različne druge elektronske vsebine, kot so video in zvočni zapisi, interaktivni zemljevidi in mestni načrti, računalniški programi, računalniške igre, spletna umetnost (web art), mrežni dnevnik (blogi), hitre strani (wiki), učne na daljavo (e-learning) in podobno«.

Kot je razvidno iz zakona in pravilnika, je poslanstvo zajema spletnih vsebin dodeljeno Narodni in univerzitetni knjižnici, ki se je pred leti lotila naloge zagotavljanja ohranjanja »spletnega obveznega izvoda in kulturne dediščine«, podobno kot to že počne s fizičnimi nosilci vsebin. Narodna in univerzitetna knjižnica je tudi članica konzorcija *International Internet Preservation Consortium* (IIPC), katerega temeljno poslanstvo je ravno to – zbiranje, ohranjanje in zagotavljanje dostopnosti znanja, ustvarjenega na spletu. Zaradi omejenih človeških virov, ki so na voljo za upravljanje tega področja dela, NUK uspe arhivirati le manjši izbor slovenskih spletnih mest oz. spletnih vsebin. Cilj je, kar nam dovoljuje tudi *Zakon o obveznem izvodu publikacij*, arhivirati obsežnejši delež slovenike na spletu. Zato nameravamo letos poskusiti testno zajeti celotno domeno .si.

Leta 2012 je NUK nabor spletnih mest, ki jih zajema, povečal za 20 odstotkov, s 766 na 915. Bil je opravljenih skupno 1855 posameznih zajemov, skupni čas trajanja vseh zajemov je znašal slabih 8921 ur. Skupna količina zajetih podatkov tako znaša nekaj več kot 5 TB.



Slika 1: Količina zajetega gradiva v GB v letih 2009 do 2012 (po letih)

NUK je začel s selektivnim zajemanjem in arhiviranjem »tvitov« v prvi vrsti kot dopolnitev obstoječim spletnim tarčam, ki jih že zajema. Še posebno se zdi smiselno arhiviranje »tvitov« v povezavi s pokrivanjem dogodkov (npr. volitve, pomembnejši družbeni, športni, kulturni dogodki) in drugimi tematskimi zajemi. V takih primerih je treba predvsem zajeti izbor uveljavljenih spletnih mest ustanov ali družbeno bolj prepoznavnih avtorjev. Ker se hkrati zelo velika količina vsebinskih informacij ustvari in objavi tudi na družabnih omrežjih, si današnja generacija ne more privoščiti, da ne bi zbrala in ohranila vsaj del tega gradiva za prihodnje rodove.

S tem ko splet arhiviramo in njegovo vsebino damo na voljo uporabnikom, te soočimo z informacijami, ki na spletu niso več nujno na voljo. Kot primer naj navedemo zajem spletnega mesta neke politične stranke. Njeno delovanje spremljamo ves čas volilne kampanje, robot pa zajema spletno mesto stranke. Politična stranka z namenom ugajanja (ne)opredeljenim volivcem lahko tudi na dnevni bazi spreminja informacije, jih olepšuje, obljublja in opisuje svoje delovanje po volitvah. Podobno smo lahko pričali strankinemu negativnemu stališču oz. mnenju o neki drugi stranki ali politični opredelitvi, stališču. Po volitvah lahko pride do sprememb političnih stališč zaradi na novo definirane političnega polja, novih zavez in koalicij. Upravljalci spletnih mest lahko določene vsebine spremenijo, uskladijo s trenutnimi interesi stranke ali umaknejo, v zajetem arhivu, ki se ohrani, pa ostanejo.

V prihodnosti pričakujemo podobne zaplete pri zajemu twitterja. Tudi pri teh zajemih bo informacija poslana na splet na voljo zanamcem, raziskovalcem, radovednežem. Twitter in druga socialna omrežja bodo tako kot del arhiva v nadaljnji uporabi. V sam namen uporabe, vrednotenja in analizo se Narodna in univerzitetna knjižnica ne more in ne sme vpletati, njena naloga je zagotoviti zajemanje, ohranjanje in dostopnost informacij javnosti.

#### 4 DOSTOPNOST RAZISKOVALNIH PODATKOV

Dostopnost, ponovna uporaba in ohranjanje (digitalnih) podatkov, ki nastajajo v procesu raziskovanja, je problematika, ki postaja vedno bolj pereča. V okviru projekta *Odprti podatki* je bila 5. decembra 2012 organizirana delavnica *Problemi in rešitve na področju podatkovnih storitev v Sloveniji* [4]. Namenjena je bila sedanjim izvajalcem podatkovnih storitev in potencialnim bodočim nosilcem le-teh.

Odprto dostopnost podatkov raziskav priporočata Organizacija za gospodarsko sodelovanje in razvoj (OECD) ter Evropska komisija. Po besedah Neelie Kroes, evropske komisarke za digitalno agendo, je deljenje podatkov raziskav in možnost njihove ponovne uporabe bistvenega pomena za razvoj znanosti. Odprti dostop do podatkov raziskav naj bi bil del nacionalnega okolja odprtega dostopa do rezultatov javno financiranih raziskav, katerega pričakovana zasnova v državah članicah EU je opisana v dokumentu *Priporočila Komisije z dne 17. julija 2012 o dostopu do znanstvenih informacij in njihovem arhiviranju* [3].

Udeleženci delavnice so glede stanja odprtih podatkov v Sloveniji podprli gornja izhodišča in pozvali k izpolnjevanju vseh obveznosti, ki sledijo iz njih, ter ugotovili, da (povzemamo le nekaj ugotovitev):

- raziskovalci večinoma slabo poznajo prednosti in koristi odprtih podatkov, raven njihove seznanjenosti je odvisna od znanstvenega področja, na katerem delujejo;
- zadržki raziskovalcev glede odprtega objavljanja podatkov raziskav se nanašajo v prvi vrsti na vprašanje varovanja intelektualne lastnine in varovanja osebnih podatkov, oviri pa sta odsotnost kulture deljenja podatkov in nestimulativen sistem vrednotenja znanstvenih dosežkov.

Stališča udeležencev so bila (povzemamo le nekaj stališč):

- slovenski financerji raziskav in raziskovalne organizacije morajo v skladu s priporočili OECD ter Evropske komisije opredeliti politike odprtih raziskovalnih podatkov, ki bodo ob upoštevanju pravnih in etičnih obveznosti zagotovile za najširši krog dostopne in uporabne podatke javno financiranih raziskav;
- politike morajo zagotoviti ustrezno motivacijo in nagrado za ustvarjalce podatkov na podlagi njihovega prispevka pri pripravi raziskovalnih podatkov;
- nadaljevati je treba z izgradnjo nacionalne infrastrukture odprtih podatkov, vključno z vzpostavitvijo nacionalnega portala/agregatorja in področnih podatkovnih centrov ter mehanizmov za zajemanje metapodatkov iz mednarodne podatkovne infrastrukture ( uvedba sistema trajnih identifikatorjev, etična obveznost citiranja, sledenje drugi rabi podatkov);
- zlasti je pomembno na nacionalni ravni usklajeno razvijati storitve vrednotenja raziskovalnih podat-

kov, in sicer kot dejavnost pooblaščenih področnih podatkovnih centrov, ter koordinirano razvijati nacionalno povezane storitve digitalne hrambe.

## 5 SKLEP

Dostopnost in uporabnost različnih vsebin na spletu ustvarjata nove okoliščine, ki so lahko s stališča spletnega iskalca problematične in so posledica dodane vrednosti obdelave predvsem metapodatkov. Digitalizacija, arhiviranje in generiranje raziskovalnih podatkov povečujejo ne le količino podatkov, ki so na voljo, temveč tudi širijo možnosti uporabe le-teh, ter omogočajo ustvarjanje novega znanja. Prihaja tudi do novih okoliščin, ki jih v realnem svetu nismo poznali. Iz novih povezav podatkov je mogoče skoraj trenutno odkrivanje vsebin o predmetu iskanja, za katere bi v fizičnem svetu potrebovali mesece. T. i. pojav »internetne pozabe« oz. »pravica do pozabe«, ko uporabniki ne želijo, da so podatki o njih z lahkoto najdljivi, postaja pereč. Pri zajemu spleta se v skladu z Zakonom o obveznem izvodu publikacij soočamo z vprašanji, kako spletni arhiv, obstoj katerega je pomemben za razumevanje današnjega časa v prihodnosti, sploh ponuditi uporabniku, ne da bi prišlo do kršitve intelektualne lastnine, varnosti osebnih podatkov ipd. Poseben problem je prost dostop do znanstvenih objav in – problem s še več vprašanji – ponovna uporaba in prost dostop do (surovih) znanstvenih podatkov, ki so bili zbrani v določeni (pogosto javno (so)financirani) raziskavi. Novost za javni sektor s področja kulture so tudi nove obveznosti, ki mu jih nalaga Direktiva o ponovni uporabi podatkov javnega sektorja. Kljub temu da javni zavodi s področja kulture že več kot desetletje razvijamo spletne storitve, nismo še vzpostavili ustrezne oblike izvajanja javne službe na spletu, ki bi rešila vsa omenjena vprašanja.

Ali širše – nujno je treba opredeliti značilnosti javne službe s področja kulture in znanosti na spletu, ugotoviti, kako se ta že uresničuje, ter na tej podlagi premisliti izhodišča minimalnih skupnih podlag ustvarjanja, upravljanja in dostopnosti do (meta)podatkov.

## 6 VIRI

- [1] *Evropska digitalna agenda*. Bruselj, 26. 8. 2010, COM(2010) 245 koč./2. Pridobljeno 31. 8. 2013 s spletne strani <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:SL:PDF>.
- [2] *Pravilnik o vrstah in naboru elektronskih publikacij za obvezni izvod*. Ur. l. RS, št. 90/2007. Pridobljeno 31. 8. 2013 s spletne strani [http://zakonodaja.gov.si/rpsi/r02/predpis\\_PRAV8482.html](http://zakonodaja.gov.si/rpsi/r02/predpis_PRAV8482.html).
- [3] *Priporočila Komisije z dne 17. julija 2012 o dostopu do znanstvenih informacij in njihovem arhiviranju (2012/417/EU)*. Uradni list Evropske unije, L 194/39, 21. 7. 2012. Pridobljeno 31. 8. 2013 s spletne strani <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:194:0039:0043:SL:PDF>.
- [4] *Priporočila udeležencev delavnice Problemi in rešitve na področju podatkovnih storitev v Sloveniji*. Ljubljana: Fakulteta za družbene vede, 5. 12. 2012. Pridobljeno 31. 8. 2013 s spletne strani: <http://www.adp.fdv.uni-lj.si/odpp10D1/priporocila.html>.
- [5] *Priporočilo komisije z dne 27. 10. 2011 o digitalizaciji in spletni dostopnosti kulturnega gradiva ter njegovi digitalni hrambi*. Bruselj, 27. 10. 2011, (C(2011) 7579 konč. Pridobljeno 31. 8. 2013 s spletne strani [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/recommendation/recom28nov\\_all\\_versions/sl.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/recom28nov_all_versions/sl.pdf).
- [6] *Uredba Evropskega parlamenta in Sveta o varstvu posameznikov pri obdelavi osebnih podatkov in o prostem pretoku takih podatkov (Splošna uredba o varstvu podatkov)*. Predlog. Bruselj, 25. 1. 2012, COM(2012) 11 final. Pridobljeno 31. 8. 2013 s spletne strani <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:SL:PDF>.
- [7] *Zakon o obveznem izvodu publikacij*. Ur. l. RS, št. 69/2006, spremembe Ur. l. RS, št. 86/2009. Pridobljeno 31. 8. 2013 s spletne strani [http://zakonodaja.gov.si/rpsi/r06/predpis\\_ZAKO3606.html](http://zakonodaja.gov.si/rpsi/r06/predpis_ZAKO3606.html).
- [8] *Zakon o varstvu osebnih podatkov*. Ur. l. RS, št. 86/2004, spremembe Ur. l. RS 13/1005, 51/2007, 67/2007, 94, 2007. Pridobljeno 31. 8. 2013 s spletne strani [http://zakonodaja.gov.si/rpsi/r06/predpis\\_ZAKO3906.html](http://zakonodaja.gov.si/rpsi/r06/predpis_ZAKO3906.html).

Zoran Krstulović je v Narodni in univerzitetni knjižnici zaposlen od leta 1988; do leta 1999 kot bibliotekar v glasbeni zbirki NUK, nato v letih 1999 do 2002 vodja bibliografske obdelave NUK, od 2002 do danes pomočnik ravnatelja NUK za vodenje strokovnega dela. Je vodja projekta Digitalna knjižnica Slovenije ter koordinator nalog in projektov NUK v podporo *The European Library* ter *Europeani*, vseevropskemu portalu digitalnih vsebin. Je avtor več člankov in prispevkov s področij problematike razvoja digitalne knjižnice, dostopnosti e-vsebin in trajnega hranjenja elektronskih publikacij.

Matjaž Kragelj je od leta 2006 zaposlen v Narodni in univerzitetni knjižnici, kjer je vodja enote za informacijsko tehnologijo in digitalno knjižnico. Je avtor več člankov in prispevkov s področij problematike razvoja digitalne knjižnice, dostopa in trajnega ohranjanja elektronskih publikacij, dostopnosti e-vsebin, zajema spleta itd. Je podiplomski študent Fakultete za organizacijske vede, smer organizacija in management informacijskih sistemov.