



OD SKENOGRAMA DO DIGITALNEGA ARHIVA: RACIONALIZACIJA PROCESOV DIGITALIZACIJE PUBLIKACIJ NA PODROČJU KULTURE

From Scans to Digital Repository: Digitization Process Rationalization for Culture Public Sector in the Field of Culture

Matjaž Kragelj

Oddano: 4. 6. 2014 – Sprejeto: 18. 6. 2014

1.02 Pregledni znanstveni članek

1.02 Review article

UDK 930.25:004.9

Povzetek

Javne institucije hranijo pomembne zbirke fizičnih gradiv s področja kulture in znanosti. Že več kot desetletje se na podoben način ukvarjajo s problematiko digitalizacije in zahtevami po trajnem ohranjanju nastalih e-vsebin. Hkrati je opazno, da je veliko večja množica tistih javnih institucij, ki se še vedno bolj ali manj uspešno ukvarjajo z začetnimi težavami pri digitalizaciji gradiv. Razlogi za neskladje med enimi in drugimi so lahko v različnih dejavnikih, kot so: pomanjkanje znanja informatike, omejena finančna sredstva in kadri, ki jih javne institucije lahko namenijo prenosu fizičnih gradiv v digitalni medij, predvsem pa omejeni viri za nujno upravljanje procesov arhiviranja, vzdrževanje arhiva in zagotavljanje trajne dostopnosti do e-vsebin uporabnikom. V Narodni in univerzitetni knjižnici se razvija poslovni model, ki bo sorodnim institucijam lahko zagotovil uspešno digitalizacijo in oblikovanje lastnih e-zbirk, ki poleg digitalizacije vključuje tudi model trajnega ohranjanja in zagotavljanja javnega dostopa do e-vsebin. Celotni proces sestavljajo faze, kot so: izbor in priprava fizičnega gradiva, preverjanje avtorskih in drugih upravičenj na gradivu ter oprema digitalnih različic publikacij s pripadajočo licenco Creative Commons, priprava parametrov digitalizacije in bibliografskih podatkov o gradivu, postopek digitalizacije gradiva, vključevanje bibliografskih podatkov v e-publikacijo, postopek avtomatske razpoznavne besedila, arhiviranje gradiva in zagotovitev dostopa do le-tega. Nekatere aktivnosti nastopijo v procesu digitalizacije zgolj v določeni fazi, druge pa skozi ves »življenjski tok« e-publikacije. Zagotavljanje dostopa in trajno ohranjanje morata biti podprti ves čas. V posebnih primerih lahko nastopi potreba po spremembi nekaterih bibliografskih podatkov – npr. licence oz. modalitet dostopnosti. Potreba po spremembi se lahko pojavi zaradi časovne komponente (pretek trajanja licence, publikacija preide v javno uporabo itd.) ali spremenjenega dogovora z



nosilcem avtorske in sorodnih pravic. Z uresničitvijo poslovnega modela želimo prispevati k racionalizaciji porabe javnih sredstev v ta namen, a hkrati spodbuditi večjo ponudbo e-vsebin po spletu uporabnikom ter zagotoviti varno trajno ohranjanje e-vsebin. Pomembno je poudariti, da se z digitalizacijo problematika, povezana z ohranjanjem fizičnega gradiva in njegove e-oblike, ne konča, temveč le razširi iz fizičnega tudi v digitalno okolje.

Ključne besede: *digitalni arhivi, digitalizacija, kultura, digitalne zbirke, upravljanje*

Abstract

Public institutions hold important physical collections of materials from the fields of culture and science. For more than a decade, they have been facing the problem of digitization and archiving requirements as a part of the digitization process when generating e-content. They are using more or less the same procedures when they cope with those difficulties. At the same time, we note a much larger set of institutions, which still deal, more or less successfully, with the initial problems of the digitization process. The main reasons for the gap between the two can be found among different factors, including: lack of knowledge in the field of the science in informatics, limited financial and staff resources provided by public institutions, and in particular, limited resources for the preservation management and the archive maintenance sub processes. The National and University Library is developing a business model that will provide a successful implementation of the digitization process with all main sub processes (submission and/or digitization, archiving, dissemination). Thus, the end product for digital preservation and provision of public access to e-content for similar institutions will be ensured.

The main component parts of the digitization process are as follows: selection and preparation of physical collections, determination of access/use restrictions or copyright, application of Creative Commons license, preparation of project parameters and bibliographic information, digitization, integration of bibliographic data into e-contents, automatic text recognition, archiving and providing access to digitized collections. Some activities, such as access and long-term preservation, need to be ongoing throughout all the phases of the digitization workflow, whereas some of them can be traced only in individual phases of the process. In special cases, it might be necessary to change some bibliographic data, i.e. license terms and conditions or access modality. The digital copy status can change over time (expiration of licence, the digitized materials fall in the Public Domain) or by modification of agreement with the copyright holder.

By implementing a business model, we wish to contribute to the rationalization of the use of public funds for this purpose, but at the same time, to encourage a greater scope of online available e-content and to ensure a safe digital preservation of e-content material. It should be pointed out that the problems, related to long-term preservation of physical collections



and their digital representation are not terminated by the digitization process, they are only expanded from the physical to digital environment.

Keywords: *digital archives, digitization, culture digital collections, management*

1 Uvod

Potreba po digitalizaciji gradiva, z namenom ohraniti kulturno dediščino prihodnjim rodovom, se je v devetdesetih letih prejšnjega stoletja kazala skozi načrtovanje in uresničevanje posameznih projektov, katerih rezultat so zaključene tematske zbirke. Digitalizirani so manjši korpusi posameznih zvrsti gradiva oziroma gradiva določene tematike, skeniranega v visoki ločljivosti, visoki barvni globini ter opisanega z bogatim naborom bibliografskih podatkov. Knjižnice so se običajno odločale za posebej vredno gradivo, torej gradivo, izrednega pomena ne le za institucijo, ampak tudi celotno družbo. Čeprav je bilo na ta način mogoče uporabniku ponuditi digitalne objekte, ki so zelo natančna kopija izvirnikov, so vloženi čas, porabljeni finančna sredstva in drugi viri za izvedbo projekta prav tako omejevali količino gradiva, ki je bila lahko digitalizirana (Sutton, 2012).

Po tradicionalnem umevanju knjižnico sestavljajo tri področja/dimenzije: skladiščni prostor, prostor za uporabnike in upravno območje. Ključna razlika med digitalno in klasično knjižnico ni v njeni pojavnosti. Digitalna knjižnica ni zgolj virtualna podoba klasične knjižnice. H. Hobohm (2012) navaja, da digitalna knjižnica tradicionalno knjižnico razširja, dodaja ji četrto dimenzijo. Ponuja ne le informacije o publikaciji ampak tudi povezave med katalogom in publikacijo ter zagotavlja dostop do same publikacije, običajno prek oddaljenega dostopa. Spletni uporabniki zato postanejo četrto območje knjižnice, značilno le za digitalno knjižnico. R. Maurya (2011) podaja opredelitev digitalne knjižnice kot zbirke dokumentov, v organizirani digitalni obliki, ki je dostopna po (spletnem) omrežju.

K. Tharani (2012) ugotavlja, da se v državah severne Amerike, v visokošolskih knjižnicah, še vedno zaznava povečana potreba po digitalizaciji. Vendar smo še vedno priča dejstvu, da je digitalizacija vezana na posamezne projekte namesto integrirana med redne storitve knjižnice.

Gracy in Kahn (2012) povzemata po Croft (2003), da je digitalna tehnologija v povezavi z ohranjanjem gradiva predvsem vplivala na zamenjavo medija – mikrofilma, ki je bil tradicionalna metoda preformatiranja tiskanih publikacij, predvsem serijskih, zelo malo pozornosti pa je bilo posvečene razvoju teorije in tehnik bodisi za ohranjanje



izvorno digitalnih publikacij, bodisi digitaliziranih, ki so nastale kot konverzija analognih medijev. V zelo kratkem času se je strokovno mnenje spremenilo – digitalizacija ni več zgolj orodje v razvoju, temveč pogosto prva izbira med metodami preformatiranja gradiva. Avtorja ugotavljata, da se knjižnice lotevajo izdelave uporabniških kopij (surogov) tiskanega gradiva iz različnih razlogov, kot so npr:

- zaščita izvornikov pred propadanjem
- zamenjava izvornikov, če poslabšanje fizičnega stanja ogroža uporabnost gradiva
- boljša dostopnost gradiva, ki jo dosežemo z uporabo novih tehnologij (npr. bolj učinkovito iskanje po gradivu zaradi možnosti OCR-postopka, povečava detajlov gradiva itd.).

Conway (2010) je previdno opredelil »digitalizacijo za ohranjanje kot naložbo v ustvarjanje trajnih digitalnih izdelkov« in ne zgolj kot način ohranjanja ogroženih publikacij. Hurst-Wahl (2009) predlaga, da naj stroka razmisli o razvoju splošnih smernic, ki bi bile vodilo pri izbiri gradiva za digitalizacijo, oblikovanju in uporabi metapodatkovnih shem, postopkov skeniranja in preverbe kakovosti ter načrta trajnega ohranjanja digitalnih vsebin. Standardizacija bi imela pozitiven učinek pri poenotenju izvajanja postopkov digitalizacije in bi okrepila interoperabilnost na tem področju. Omeniti velja, da je bila v Narodni in univerzitetni knjižnici leta 2012 sprejeta »Strategija trajnega ohranjanja digitalnih virov v Narodni in univerzitetni knjižnici« (Kavčič-Čolić, 2012), leta 2013 pa je minister za kulturo sprejel »Smernice za zajem, dolgotrajno ohranjanje in dostop do kulturne dediščine v digitalni obliki«, ki so sad skupnih naporov celotnega kulturnega sektorja.¹ Zmotno je mišljenje, da digitalizacija rešuje vsa odprta vprašanja, povezana z ohranjanjem fizičnega gradiva, kakor tudi, da digitalizacija niža ceno hrambe gradiva. Kacy L. Guill (2009) skrbi vidik umikanja fizičnega gradiva. Gradivo, ki zaradi zagotovljenega dostopa do elektronskega nadomestka ni več nujno potrebno na lokaciji, kjer se je nahajalo, se seli v manj primerne prostore (npr. kleti), saj želimo uporabnikom ponuditi večje udobje znotraj ustanov. S tem, ko gradivo selimo v manj primerne prostore, sprožamo dve pasti. Prvič – gradivo je manj varovano, predvsem pred zunanjimi vplivi, kot je npr. poplava, in drugič, čas dostopa do gradiva je daljši. Glede zagotavljanja trajnega dostopa do e-gradiva, pa se pojavlja še vidik potrebnih stroškov za zanesljivo delovanja tega servisa.

¹ <http://nio.gov.si/nio/cms/download/document/bbf2216a1e2283d7c1a5fb57e4fa455e5560634f-1390917025213>, pridobljeno 13. 5. 2014.



2 Primeri metodologij

Zadnja leta je glavni igralec na trgu digitalizacije knjižničnih fondov podjetje Google. Knjižnice so postavljene pred izziv – zagotoviti dostop do zahtevnejšega gradiva s postopki skeniranja, optične razpoznave, indeksiranja in spletne ponudbe uporabnikom (Gracy in Kahn, 2012). Zaradi razvoja informacijske tehnologije so knjižnicam odprte nove poti za digitalizacijo gradiva, ki je bilo dolga leta »skrito« oziroma ga je bilo nemogoče pretvoriti v elektronsko obliko in ponuditi uporabnikom. Med razloge uvrščamo pomanjkanje dokumentacije, zastarele formate, slabo stanje gradiva in druge ovire (Benson, 2009; Evens in Hauttekeete, 2009). Knjižnice in arhivi imajo danes možnost da s sodelovanjem zberejo gradivo istega izvora (provenience), ki se nahaja oz. je v lasti različnih institucij. Takšno prizadevanje lahko opišemo kot virtualna združitev.

Prav tako, kot imamo v ospredju uporabnika pri zagotavljanju storitev klasične, tradicionalne knjižnice, velja analogija v digitalnem okolju. T. K. Huwe (2012) ugotavlja, da je bil še nedolgo nazaj dober knjižničar tisti, ki je bil pred uporabnikom v smislu predvidevanja njegovih potreb in učinkovite uporabe tehničnih možnosti dostopa do informacij. Stanje se je obrnilo. Uporabniki so postali bolj veščji uporabe spletnih iskalnih orodij ter se hitreje prilagajajo novim storitvam. Hkrati ko te storitve s pridom in v polni meri uporabljajo, povsod pričakujejo še možnost interakcije in uporabe oblačnih storitev, tudi v knjižnicah. Na srečo so storitve zasnovane tako, da jih je mogoče integrirati in ponuditi uporabnikom tudi znotraj servisov digitalne knjižnice. Všečkanje, komentiranje, vključevanje dinamičnih komponent, kot je npr. Facebook, Twitter, YouTube, so elementi, za katere uporabnik pričakuje, da so podprti tudi v spletnem okolju knjižnice. Hkrati se zdi, da se vedno bolj briše meja med uporabnikom in knjižničarjem pri oblikovanju storitev spletnega okolja.

K. Tharani (2012) navaja, da je združenje digitalnih knjižnic (Digital Library Federation (DLF)) razvilo storitveno ogrodje (service framework) za digitalne knjižnice z namenom vpogleda v različne postopke in dejavnosti znotraj knjižnic ter njihovo transformacijo v modularne servise za podporo procesov digitalizacije. Akademske knjižnice na področju severne Amerike so lahko zaradi napredka pri razvoju informacijskih tehnologij na novo zasnovale digitalizacijo, kot del knjižnične dejavnosti. Iz digitalizacij, ki so bile načrtovane in izvajane v sklopih različnih projektov, jim vpeljava storitvenega ogrodja omogoča podporo procesov digitalizacije znotraj knjižnične dejavnosti. Ogrodje je sestavljeno iz sedmih glavnih, krovnih procesov, ki jih podpira enainštirideset funkcij. Med glavne procese štejemo: načrtovanje, razporejanje virov, digitalizacijo, metapodatkovno opisovanje, arhiviranje/oddaja, dostop, upravljanje. Entitete, ki nastopajo v procesih in funkcijah, izhajajo iz referenčnega modela OAIS, (Open Archival Information System) (ISO 14721, 2012) kar pomeni, da gre za implementacijo tega modela v



prakso. Referenčni model OAIS podpira tri glavne procese, ki so zajem, ohranjanje in zagotavljanje dostopa do objektov.

Kot primer koraka v smeri platforme, ki pokriva arhiviranje in zagotavljanje dostopa, velja omeniti »sodelovalno« digitalno knjižnico NEUMA (Rigaux et al., 2012). Gre za digitalno knjižnico, posvečeno objavljanju elektronskih notnih zapisov. Odprta in namenjena je ne le specialistom muzikologom, temveč tudi drugim glasbenikom in glasbenim pedagogom. Digitalna knjižnica sestoji iz repozitorija, namenjenega shranjevanju velikih zbirk notnih zapisov, kjer uporabniki (fizični uporabniki ali aplikacijski vmesniki) oddajajo publikacije. Sistem ponuja storitveno orientirano arhitekturo (SOA)² za objavo gradiva, beležk, poizvedb, transformacij in analizo shranjenih objektov – notnih zapisov. Dolgoročni namen projekta je zagotoviti odprt virtualni prostor skupinam uporabnikov, ki želijo shranjevati in deliti zbirke glasbenih vsebin. Cilji, ki bodo s tem doseženi, pa so: podpora znanstveni objavi vsebin z bolj specialnih področjih kulturne dediščine, ki bi bili sicer težko dostopni, razvoj in vzdrževanje zbirke glasbenih del, z možnostmi, ki jih ponuja informacijska tehnologija (brskanje, preiskovanje, možnost analize zvočnih zapisov, bogatenje z novimi metapodatki itd.).

Kot primer omejevanja stroškov razvoja in vzdrževanja digitalnih knjižnic je na voljo primer iz Združenih držav. Wu (2011) ugotavlja, da je v ZDA približno dvesto akreditiranih šol s področja prava (American Bar Association),³ skupni letni strošek za razvoj in vzdrževanje digitalnih knjižnic ocenjujejo na več kot 230 milijonov dolarjev. Pomenljivo je dejstvo, da je med dostopnimi publikacijami precejšnje število duplikatov, tudi med tistimi gradivi, ki so redko uporabljeni. Avtorica predlaga za vzor državo Florida, ki je znižala stroške z uvedbo razvoja zbirke na državni ravni, splošne knjižnice pa so prevzele centralizirane oblike gradnje zbirk. Avtorica prispevek zaključuje z naštevanjem pozitivnih strani centralne digitalne knjižnice. Omenja širši dostop do informacij, zagotavljanje trajnega ohranjanja vsebin, nadzor nad obliko in formatom vsebine. Z zagotavljanjem več in boljših informacij uporabniku, le-tega ščitimo pred morebitnim izkrivljanjem informacij. Če namreč drži, da uporabniki težijo k spletnim virom, moramo poskrbeti, da imajo tam možnost dostopa do enakih informacij kot v tiskani obliki, in ne zgolj do tistih, ki so uporabniku lažje dostopne. Dovolj je razprave o knjižnici prihodnosti, čas je za njen razvoj.

K. Tharani (2012) ugotavlja, da pomenijo največji izziv pri sodelovanju v procesih digitalizacije družbeni vidiki. Le-ti so pomembnejši od tehničnih. Že na institucionalni ravni je potreba po sodelovanju osebja v procesih digitalizacije različnih oddelkov večja

² http://en.wikipedia.org/wiki/Service-oriented_architecture, pridobljeno 12. 5. 2014.

³ <http://www.americanbar.org/aba.html>, pridobljeno 13. 5. 2014.



kot običajno. Odnosi postanejo še kompleksnejši, ko se v digitalizacijo vključi več partnerjev, iz različnih institucij, različnih tako po velikosti, tipu in lokaciji. Z upoštevanjem razlik lahko dosežemo bolj raznovrsten rezultat (zbirko digitaliziranega gradiva), povečamo stopnjo učinkovitosti procesov in pridobimo dodatne vire financiranja.

Spletno okolje 2.0 je omogočilo aktivno vlogo subjektov (uporabnikov) v spletnem okolju. Poleg nekaj primerov, navedenih kot primer dobre prakse, so pomembne še največje in najbolj razširjene implementacije orodij, ki ponazarjajo značilno obliko aktivne vloge posameznika v spletnem okolju. To so npr. Wikipedia (organizacija znanja), Flickr (fotografsko gradivo), iTunes (zvočne vsebine), YouTube (video vsebine).

3 Potreba po procesu digitalizacije

Če povzamem po Maurya, R. (2011), in okolje omejim na dejavnost knjižnic, lahko ocenim, da knjižnice digitalizirajo knjižnično gradivo, ker so še vedno prepričane o trajni vrednosti teh virov za učenje, raziskovanje, poučevanje, dokumentiranje in izvajanje javne službe. Razlogi za prehod na digitalno obliko so med že prej navedenimi tudi:

- dostop do dokumentov brez poškodb izvirnika
- potrebe nove generacije
- večja stroškovna učinkovitost digitalnih knjižnic
- izboljšane možnosti iskanja vsebin tradicionalne knjižnice
- podpora obstoječim funkcijam knjižnice
- ekologija: digitalna knjižnica ohranja okolje, ker porabi manj papirja in omejuje uporabo okolju škodljivih snovi.

Med izzive pri procesu digitalizacije gradiva v knjižnicah R. Maurya (2011) posebej omenja predvsem naslednje:

- pomanjkanje tehničnega osebja
- avtorska in sorodne pravice
- tehnološke spremembe
- varnost podatkov
- nezadostno financiranje in pomanjkljiva infrastruktura
- eksponentna rast količine ustvarjenega (e-)gradiva na internetu.

Čeprav se potrebe in izzivi pri ustvarjanju e-vsebin s postopki digitalizacije, omenjeni v tem prispevku, navezujejo na Indijo, so na drugi strani sveta – v ZDA, prav tako brez jasnih odgovorov. Teper in Shaw (2011) opažata, da so se knjižnice po svetu začele aktivno ukvarjati z množično digitalizacijo. Čeprav dolgoročni učinki množične digitalizacije še niso jasni, knjižnice vpeljujejo in izvajajo digitalizacijo pri rednem delu.



Razlika med »butičnimi« projekti digitalizacije, torej tistimi, ki se večinoma izvajajo v knjižnicah oz. pri zunanjih izvajalcih, in projekti množične digitalizacije (Mass Digitization Projects), kot so npr. Google Books,⁴ the Open Content Alliance⁵ in Million Book Project,⁶ je predvsem v zagotavljanju učinkovite porabe časa in denarja za doseganje cilja. Hitrejši potek dela (workflow) botruje nižji ceni po skenogramu ter posledično večji količini digitaliziranega gradiva.

Drugačno opredelitev digitalizacije ponujajo Katz, Koutroumpis in Callorda (2014). Poleg splošno sprejete in večkrat omenjene razlage, da gre za proces prenosa analogne informacije v digitalno obliko, avtor to razširja in pravi, da je v širšem družbenem kontekstu digitalizacija opredeljena kot ekonomska in socialna transformacija, povzročena z množičnim sprejemom digitalnih tehnologij, ki so osnova ustvarjanja, obdelave, razširjanja in deljenja informacij. V članku avtor uporabi indeks digitalizacije kot merilo napredka, ki ga dosega Evropa pri implementaciji Digitalne agende (Digitalna agenda za Evropo).⁷

Digitalizacija mora biti razpršena, da bi lahko pomembno vplivala na znotraj gospodarske in družbene strukture naroda. Za to mora biti sprejeta na treh ravneh:

- izkoriščena s strani posameznikov, gospodarskih podjetij in družb, vgrajena v procese proizvodnje dobrin in storitev ter vključena v izvajanje javnih služb
- biti mora vseprisotna – dosegljiva čim več prebivalcem na nacionalni ravni
- biti mora zanesljiva, zagotavljati mora zadostno zmogljivost za prenos velike količine informacij ob zahtevanih hitrostih.

Indikatorji, ki jih je avtor uporabil za oblikovanje indeksa digitalizacije, so: *cenovna dostopnost digitalizacije, naložbe v infrastrukturo, dostopnost omrežja, zmogljivost omrežja, uporaba, človeški kapital*. Indikatorji skupaj vsebujejo še 24 podindikatorjev oz. komponent. Kot je razvidno iz Slike 1, je Slovenija uvrščena zelo visoko, med razvitimi državami. Avtor postavlja cenovno dostopnost kot ključni element za doseg dobrega rezultata. Prav tako pomembna pa je ugotovitev, da se lahko države z veliko visoko izobraženimi zaposlenimi hitro prilagodijo spremembam in uporabi novih orodij in tako dohitijo vodilne. Države Evrope zaostajajo zgolj za severno Ameriko, ob tem, da imajo zelo slab indikator – naložbe v infrastrukturo. Po drugi strani pa so države Evrope najboljše pri zagotavljanju *cenovne dostopnosti*.

⁴ <http://books.google.com/>, pridobljeno 20. 5. 2014.

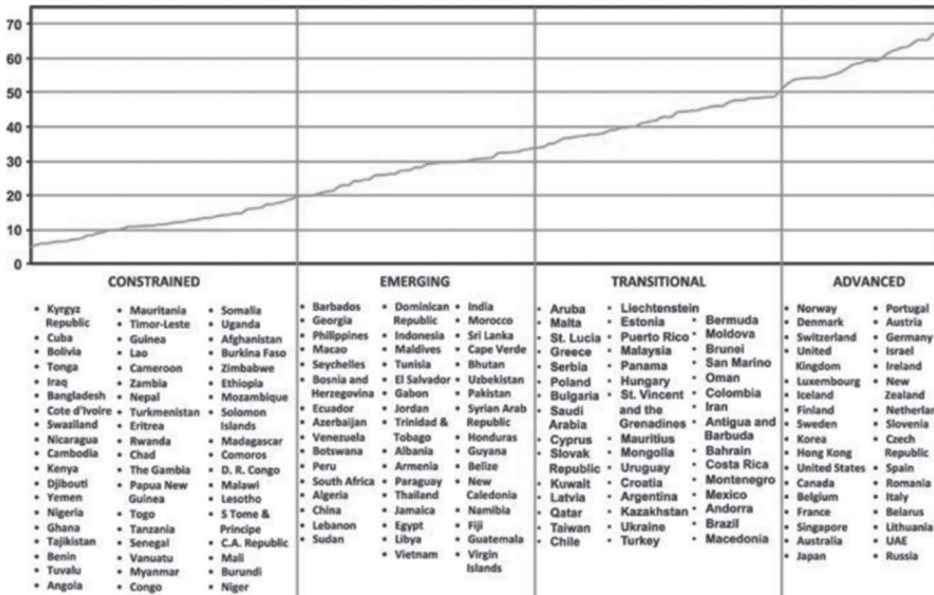
⁵ <http://www.opencontentalliance.org/>, pridobljeno 20. 5. 2014.

⁶ <http://www.ulib.org/index.html>, pridobljeno 20. 5. 2014.

⁷ <http://ec.europa.eu/digital-agenda/digital-agenda-europe>, pridobljeno 21. 5. 2014.



Katz, Koutroumpis in Callorda (2014) države razdelijo v štiri skupine: razvojno omejene, v razvoju, prehodne in napredne – vse v smislu ravni razvoja pri procesih digitalizacije.



Slika 1: Stopnja razvoja in obvladovanja procesov digitalizacije (Vir: Katz, Koutroumpis in Callorda, 2014)

4 Upravljanje digitaliziranih zbirk

Orodij za upravljanje digitalnih zbirk, ki nastanejo v procesu digitalizacije, je mnogo. Na voljo so tako komercialne kot odprtokodne rešitve. Skupina avtorjev (Kamble, Raj in Sangeeta, 2012) ugotavlja, da so v zadnjih letih prav odprtokodni produkti poželi veliko zanimanja v knjižničnih krogih. Po eni izmed opredelitev je odprtokodna programska oprema⁸ programska oprema, za katero je izvorna koda objavljena pod zaščitno licenco (ali pogojem, kot npr. javna domena), ki ustreza kriterijem odprtokodne definicije. Ta dovoljuje vsakršno uporabo in spreminjanje programske opreme, njeno razširjanje v spremenjeni ali nespremenjeni obliki. Velikokrat jo razvijajo v javnosti, s prispevki drugih.

⁸ Odprtokodna programska oprema, http://sl.wikipedia.org/wiki/Odprtokodna_programska_oprema, pridobljeno 20. 5. 2014.



Knjižnice se spoprijemajo s podobnimi izzivi pri zajemu (digitalizaciji), ohranjanju in dostopnosti e-vsebin, zato je pričakovanje, da »nekje že obstaja programska rešitev, prosto dostopna, ki bi bila za nas uporabna« povsem na mestu. Opozoriti je treba le, da je pri komercialnih produktih garancija in tehnična pomoč zagotovljena (strošek je navadno vključen v ceno nakupa, najema), pri odprtokodnih produktih pa se pojavi odvisnost od razvijalcev, ki so zaposleni v knjižnici oz. drugi sorodni instituciji, sicer je organizacija prisiljena iskati razvojno moč na trgu ali v skupnostih, ki so produkt razvile.

Kamble, Raj in Sangeeta (2012) med orodji za upravljanje digitalnih zbirk omenjajo: Koha,⁹ NewGenlib,¹⁰ OpenBiblio,¹¹ Avanti,¹² Greenstone,¹³ Dspace,¹⁴ E-Prints¹⁵ in Fedora.¹⁶

Omenjena programska orodja so večinoma pripravljena za takojšnjo uporabo, a je pri gradnji večjih korpusov različnega tipa podatkov potreben razvoj dodatnih uporabniških in aplikacijskih vmesnikov za doseg avtomatizacije in s tem nižje cene upravljanja enote gradiva.

Z repozitoriji, namenjenimi shranjevanju digitalnih vsebin, publikacij, smo v Narodni in univerzitetni knjižnici v stiku že več kot desetletje. Avtorji članka »Digital repositories ten years on: what do scientific researchers think of them and do they use them?« (Nicholas et al. 2012) na začetku besedila ugotavljajo, da čeprav so tovrstni sistemi že vrsto let prisotni v naših krogih in čeprav je bilo na temo uporabe prelitega veliko črnila ter je bilo mnogo konferenc, zelo malo vemo o tem, kako univerzitetni učitelji uporabljajo te sisteme. V raziskavi so se avtorji dotaknili treh vrst repozitorijev, in sicer:

- institucionalnih (navadno vsebujejo različne vsebine, pridobljene znotraj institucije),
- tematskih (zbiranje je osredotočeno na tematiko, navadno vsebujejo dela več institucij, praviloma so mednarodni),
- določeni s tipom, formatom (npr. disertacije, raziskovalni podatki)

⁹ Koha, <http://www.koha.org/>, pridobljeno 21. 5. 2014.

¹⁰ NewGenlib, <http://www.verussolutions.biz/web/>, pridobljeno 21. 5. 2014.

¹¹ OpenBiblio, <http://obiblio.sourceforge.net/>, pridobljeno 21. 5. 2014.

¹² Avanti, <http://schlaraffamilwaukia.com/>, pridobljeno 21. 5. 2014.

¹³ Greenstone, <http://www.greenstone.org/>, pridobljeno 21. 5. 2014.

¹⁴ Dspace, <http://www.dspace.org/>, pridobljeno 21. 5. 2014.

¹⁵ E-Prints, <http://www.eprints.org/>, pridobljeno 21. 5. 2014.

¹⁶ Fedora Commons, <http://fedorarepository.org/>, pridobljeno 21. 5. 2014.



V mednarodno raziskavo so vključili 85.000 elektronskih naslovnikov. Vsega so pre-
jeli pribl. 2 % odgovorov, večinoma od raziskovalcev tehničnih ved. Med odgovori je
opaziti, da prevladujejo institucionalni in tematski repozitoriji, med tistimi, ki oddajajo
konkretne vsebine, je največ objavljanih znanstvenih člankov. Če imajo avtorji mož-
nost, najraje oddajo gradivo v tematski repozitorij. Med možnimi odgovori na vpraša-
nje o smislu oddaje gradiv v repozitorije prevladuje odgovor, da je najpomembnejše
zagotavljanje javne dostopnosti gradiv.

5 Digitalna knjižnice Slovenije in možnosti razširitve

Digitalna knjižnica Slovenije¹⁷ (v nadaljevanju dLib.si) je spletni portal, knjižnica na
spletu, ki je na voljo vsem – kadarkoli in kjerkoli. Poleg digitalnih vsebin, ki so nastale
s sredstvi in iz gradiva Narodne in univerzitetne knjižnice, vsebuje spletno mesto knji-
žnice tudi gradiva drugih organizacij.¹⁸ Digitalne vsebine teh organizacij niso v lasti
Narodne in univerzitetne knjižnice, ta le skrbi za njihovo dostopnost prek Digitalne
knjižnice Slovenije. Spletno mesto ponuja uporabnikom digitalne vsebine v ogledni
ločljivosti, medtem, ko lastniki gradiva hranijo visoko ločljive skenograme na svoji lo-
kaciji. Na Sliki 2 prikazuje tortni prikaz deležev oddanih gradiv v dLib.si. Ustanove z
največ oddanega gradiva so:

- Narodna in univerzitetna knjižnica (424.565)
- Mestna knjižnica Ljubljana (89.028)
- Knjižnica Ivana Potrča (17.305)
- Osrednja knjižnica Celje (14.869)
- Univerzitetna knjižnica Maribor (9.315).

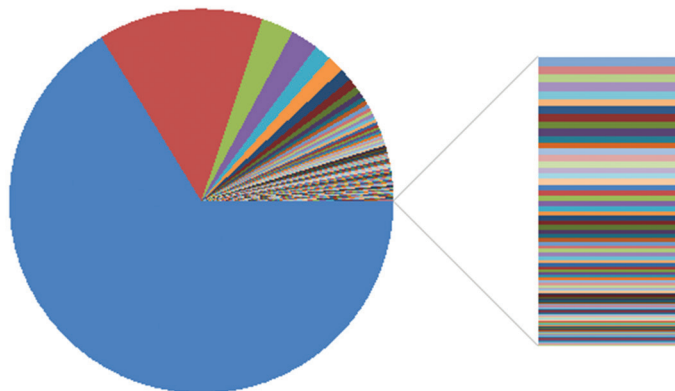
Vseh partnerjev, ki so oddali gradivo na dLib.si, je bilo maja 2014 297. Sto organizacij,
ki so oddale najmanj gradiva na spletni portal, je skupaj oddalo zgolj 324 publikacij.

¹⁷ Digitalna knjižnica Slovenije, <http://www.dlib.si>, pridobljeno, 22. 5. 2014

¹⁸ Partnerji na dLib.si, <http://www.dlib.si/menu/sodelovanje#submenu-442>, pridobljeno 22. 5. 2014

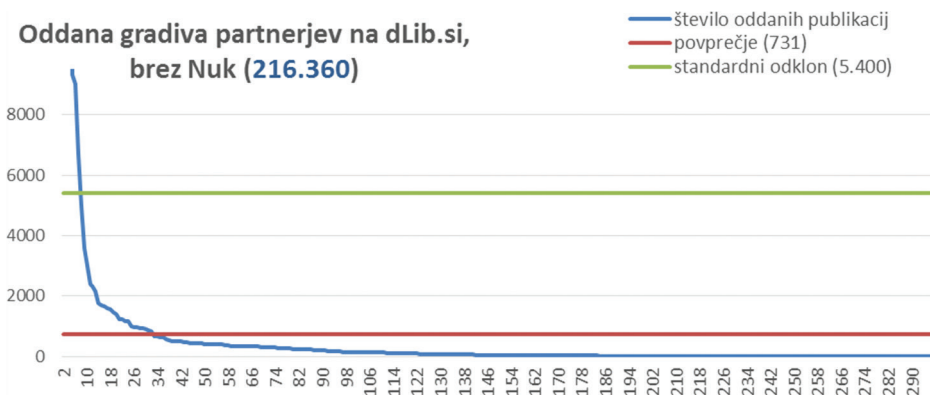


Oddana gradiva na dLib.si (640.925), 297 partnerjev



Slika 2: Tortni prikaz deleža oddanih gradiv na dLib.si

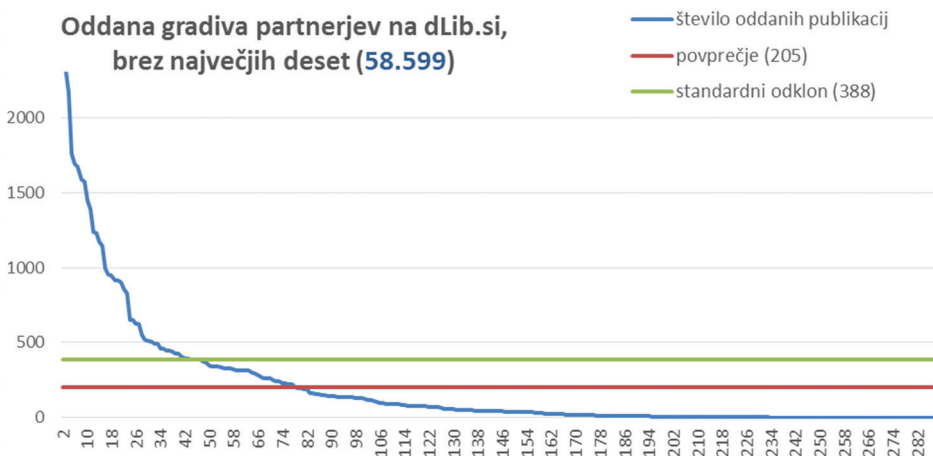
Na Sliki 3 je prikazano razmerje med številom oddanih publikacij na portal, aritmetično sredino števila oddanih publikacij ter standardnim odklonom. Na abscisi je podan indeks knjižnice (manjši indeks prikazuje knjižnico, ki je oddala več gradiva), na ordinati pa število oddanih publikacij. V grafu ni zajete Narodne in univerzitetne knjižnice. Ustanove in partnerji (253-297) so oddali zgolj po eno publikacijo. Na grafu je podana še aritmetična sredina (rdeča barva) in standardni odklon (zeleno barva). Iz grafa in statističnih kazalnikov (aritmetična sredina in standardni odklon) je opaziti, da so velike razlike v količini oddanega gradiva partnerskih organizacij.



Slika 3: Razmerje med oddanim gradivom, povprečjem in standardnim odklonom – brez NUK-a



Na sliki 4 je prikazan graf, na katerem so organizacije, ki oddajajo gradivo na dLib.si, razen največjih deset, ki prispevajo največ gradiva (pribl. 90 %). Čeprav sem umaknil partnerje, ki oddajo največ gradiva, statistična kazalca (aritmetična sredina in standardni odklon) še vedno kažeta, na velike razlike med partnerji.

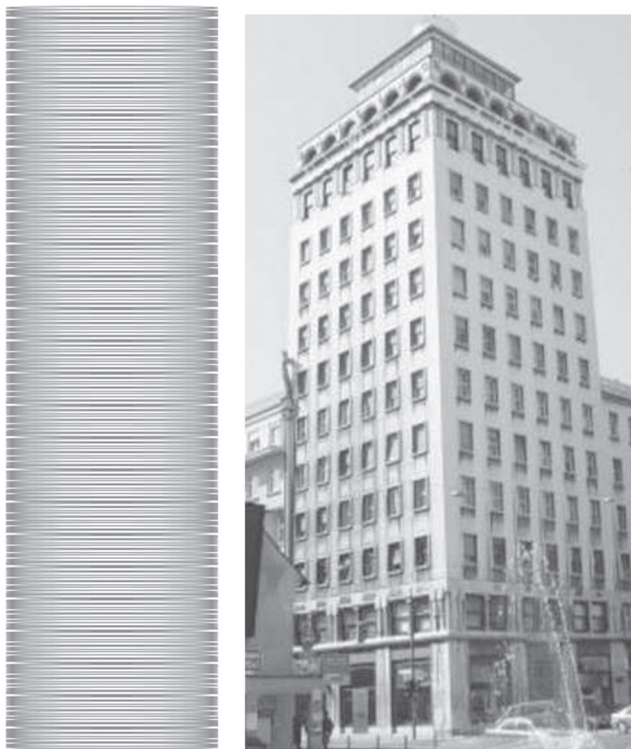


Slika 4: Razmerje med oddanim gradivom, povprečjem in standardnim odklonom za vse partnerje, razen največjih deset

5.1 Trajno ohranjanje publikacij

V raziskavi, opravljeni v podjetju Gartner, ki se ukvarja z raziskavami na področju informacijske tehnologije in svetovanjem, so ugotovili, da strošek razvoja informacijskega sistema obsega zgolj 8 % celotnih stroškov, v primeru, da bomo informacijski sistem uporabljali petnajst let. 92 % obsega vzdrževanje sistema. Med vzdrževanje štejemo tudi nadaljnji razvoj funkcionalnosti, odpravljanje težav, posodabljanje sistema itd.

Celotna digitalna zbirka Narodne in univerzitetne knjižnice presega 50 TB (terabajtov) podatkov. V to niso vključene varnostne kopije, zrcalne slike itd. V tem primeru bi bilo podatkov vsaj še dva- do trikrat toliko. Ponazorjeno s sliko (slika 5), to pomeni 71 metrov v višino zloženih CD-ploščkov. Za primerjavo – Ljubljanski Nebotičnik je visok 70 metrov.



Slika 5: Primerjava stolpa iz CD-ploščkov in Ljubljanskega Nebotičnika (Vir fotografije Nebotičnika: http://baza.givo.si/pics_reference/22a.jpg)

Izhajajoč iz statistike spletnega mesta dLib.si, je podana informacija o količini gradiv na dLib.si, številu subjektov, ki so gradivo oddali, povprečje števila oddanih vsebin na organizacijo in standardni odklon, ki kaže, ogromne razlike v količini oddanega e-gradiva na portal. Pričakovanje, da bo imela vsaka ustanova, ki ustvarja e-vsebine (digitalizira gradivo), možnost obvladovati procese digitalizacije v najširšem smislu v skladu z standardom OAIS, je nerealno.

Bibliographic metadata, DublinCore, MODS, Jhove2, Repository, URN, technical metadata, record, harvesting, hash, RAID, Droid, z39-50, format, METS, converting, pdf/a, jp2, OAI-PMH, curation, OCR, validation, warc, persistent identifier, checksum, OAIS, versioning, preservation ...

Slika 6: Nekaj terminov iz procesov digitalizacije



Na sliki 6 je prikazanih zgolj nekaj naključnih terminov iz procesov digitalizacije. Organizacija, ki želi podpreti vse potrebne procese na področju digitalizacije, mora skrbeti za njihovo trajno upravljanje, za razvoj novih, spreminjanje obstoječih in ukinjanje zastarelih. Ob tem se zastavljata vprašanji:

- *Ali ima organizacija na voljo dovolj virov (finančnih, kadrovskih) za obvladovanje procesov digitalizacije po standardih in priporočilih dobrih praksah?*
- *Ali je smiselno glede na količino e-gradiva, ki ga poseduje, vlagati v razvoj in vzdrževanje potrebnih informacijskih sistemov?*

Kot je zapisano v uvodu – še vedno smo priča precejšnjemu obsegu digitalizacije, ki nastaja znotraj časovno in z viri omejenih projektov. Po končanem projektu se nosilec projekta spoprime z vprašanjem financiranja in vzdrževanja rezultatov projekta (digitaliziranega gradiva).

5.2 Možnost izvedbe modela za rešitve problematike arhiviranja

V javnih ustanovah, ki uporabljajo spletni portal dLib.si za oddajo gradiva, je stanje naslednje:

- **640.925** publikacij je dostopnih na dLib.si
- prispevalo jih je **297** partnerjev (vključno z NUK)
- **2.158** enot je povprečna vrednost oddanih gradiv na partnerja
- **1,3 %** partnerjev je oddalo več kot **10.000** objektov (publikacij)
- **4 %** ustanov presega povprečno vrednost publikacij na dLib.si (oddali več kot 2.158 publikacij).

Po vzoru že omenjenih pristopov (npr. Tharani, 2012; Rigaux et al., 2012; Wu, 2011), upravljavca portala, Narodna in univerzitetna knjižnica, predlaga razvoj modela za upravljanje procesov trajnega ohranjanja kulturnih in raziskovalnih publikacij znotraj enotnega sistema. Za popolno obvladovanje procesov ustvarjanja e-vsebin (digitalizacije, npr. po modelu OAIS) v najširšem obsegu, je treba posebno pozornost nameniti trajnemu ohranjanju publikacij: izbiri arhivskih formatov, izbiri trajnih identifikatorjev, varnemu prenosu podatkov, preverjanju avtentičnosti in verodostojnosti podatkov, lokaciji shranjevanja, varnostnemu kopiranju, duplikaciji (redundančni kopiji), izvajati je treba scenarije povrnitve podatkov iz sekundarnega medija itd.

Narodna in univerzitetna knjižnica je v letu 2009 pridobila nov računalniški center in komunikacijsko vozlišče, ki je eden ključnih pogojev za trajno in varno ohranjanje digitalnega gradiva. ČKot glavne arhitekturne značilnosti velja omeniti naslednje:



1. Visoka razpoložljivost (več kot 99,65 %)

Na letni ravni to pomeni maksimalno 30 ur izpada (predvidenega in nepredvidenega skupaj), pri tem pa posamezni poseg vzdrževanja ne sme trajati več kot 5 ur. Sistemi vsebujejo redundančno napajanje celotnega sistema (UPS, neprekinjeno napajanje). Prostor je hlajen redundančno. Nadzor centra poteka z nadzornim sistemom, kjer se poleg okoljskih parametrov nadzira tudi delovanje strežniške infrastrukture. Za preverjanje pristopa je na voljo več različnih in neodvisnih sistemov vstopa, kar zagotavlja visoko varnost in razpoložljivost podatkov in storitev. Za vse ključne komponente sistema je NUK podpisal ustrezne pogodbe o zagotavljanju servisiranja in vzdrževanja s tipičnimi odzivnimi časi pod 1 uro.

2. Rezervna lokacija

Sistem zagotavlja redundančno – rezervno lokacijo z redundantno optično povezavo.

3. Strežniki in diskovna polja, varnost podatkov

NUK uporablja virtualizirani sistem (VMWARE, HYPER-V) z dinamičnim dodeljevanjem pomnilniškega in procesorskega prostora.

Komponente diskovnega polja so podvojene, z namenom odpraviti SPOF (single point of failure), možen je priklop več tipov diskov hkrati (SSD, FC, SATA), redundanca na diskovnih poljih (RAID), možna menjava diskov med delovanjem (hot-swap), krmilnikov, vmesnikov, predpomnilnika, napajalnika, ventilatorjev, baterijska zaščita predpomnilnika. Na voljo je skladišče (sef) za varovanje zunanjih medijev arhiva (npr. tračnih medijev), zunanjih diskov. Sistem, je varen pred vlomi in požari ter se ponša s standardi, kot so »en14450, S 120 P, EN 1047-1«.

6 Zaključek

V računalniškem centru NUK je mogoče zagotavljati trajno ohranjanje različnih tipov in formatov publikacij. Digitalna knjižnica Slovenije skrbi za *zagotavljanje dostopa oglednih kopij gradiva uporabnikom*. Strojna oprema v NUK-u skrbi za *trajno varovanje in ohranjanje gradiva*. Odprtokodna platforma FedoraCommons in posebej razviti aplikacijski vmesniki skrbijo za *varno oddajo gradiva* (oglednih in arhivskih kopij, bibliografskih in tehničnih metapodatkov).

S priključitvijo in uporabo sistema v NUK-u lahko javni zavodi, druge organizacije, društva itd. gradivo varno oddajo, urejajo in zagotavljajo dostop svojim uporabnikom. Z vidika porabe je mogoče večino javnih sredstev iz razvoja in vzdrževanja informacijskih sistemov znotraj institucij preusmeriti v ustvarjanje digitalnih vsebin. Za enaka sredstva je uporabnikom na voljo več gradiva – dostopnega znanja.



Potrebno je poudariti, da je koncept oz. poslovni model združevanja postopkov oz. procesov digitalizacije in trajnega ohranjanja gradiva šele nujno potrebna platforma, potreben pogoj za vzpostavitev možnosti upravljanja podatkov. Za vzor so lahko npr. rezultati projektov KEEP¹⁹ in SCAPE.²⁰ Naloge, s katerimi se ukvarjajo severne sosede Slovenije, presegajo vzpostavitev ogrodij za trajno ohranjanje vsebin – ti procesi so že več let vtankani v organizacijske sisteme institucij. Izzivi, ki prihajajo na površje, so predvsem analiza in povezovanje podatkov, avtomatizacija procesov, ustvarjanje dodane vrednosti iz podatkov, emulacije starejših računalniških sistemov (C64, Atari, Amstrad, itd)²¹ z namenom dostopati do informacij oz. vsebin, ustvarjenih v prejšnjih desetletjih.

Navedeni viri

- Benson, A. C. (2009). Killed Negatives: The Unseen Photographic Archives. *Archivaria*, št. 68, 1–37.
- Conway, P. (2010). Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas. *Library Quarterly* 80 (1), 65.
- Croft, J. A. (2003). The Preservation Evolution: A Review of Preservation Literature, 1999-2001. *Library Resources & Technical Services*, 47(2), 59–70.
- Evens, T. in Hauttekeete, L. (2009). Unlocking Audio: Towards an Online Repository of Spoken Word Collections in Flanders. *D-Lib Magazine*, 15(5/6). Pridobljeno 15. 6. 2014 s spletne strani: www.dlib.org/dlib/may09/evens/05evens.html
- Gracy, K. F. in Kahn, M. B. (2012). Preservation in the Digital Age. *Library Resources & Technical Services*, 56(1), 25–43.
- Guill, L. L. (2009). Arguing for Space in a User-Focused Environment. *Library & Archival Security*, 22(2), 115–123.
- Hobohm, H. (2012). Can Digital Libraries Generate Knowledge?. *Historical Social Research*, 37(3), 218–229.
- Hurst-Wahl, J. (2009). Digitization: How Many Best Practices, Guidelines, and Standards Do We Need? *Information Standards Quarterly*, 21(4), 22–24.
- Huwe, T. K. (2012). New Technology, New Workflows, New Ways to Collaborate. *Computers in Libraries*, 32(3), 20.
- ISO 14721:2012, Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. Geneva: ISO.
- Kamble, V. T. in Raj, H., in Sangeeta, S. (2012). Open Source Library Management and Digital Library Software. *DESIDOC Journal of Library & Information Technology*, 32(5), 388–392.

¹⁹ <http://www.keep-project.eu/ezpub2/index.php>, pridobljeno 10. 6. 2014.

²⁰ <http://www.scape-project.eu/>, pridobljeno 10. 6. 2014.

²¹ http://liber2014.univie.ac.at/fileadmin/user_upload/DOEVL_events/Kongressfiles/2014-05-20_LIBERworkshop_emulation_en_kb.pdf, pridobljeno 10. 6. 2014.



- Katz, R., Koutroumpis, P. in Callorda, F. (2014). Using a digitization index to measure the economic and social impact of digital agendas. *Info*, 16(1), 32–44. doi:10.1108/info-10-2013-0051
- Kavčič-Čolić, A. (2012). *Strategija trajnega ohranjanja digitalnih virov v Narodni in univerzitetni knjižnici : : 2012–2020 / pripravila Alenka Kavčič-Čolić ; sodelavci Melita Ambrožič ... [et al.]*. Ljubljana : Narodna in univerzitetna knjižnica.
- Maurya, R. (2011). Digital Library and Digitization. *International Journal Of Information Dissemination & Technology*, 1(4), 228–231.
- Nicholas, D., Rowlands, I., Watkinson, A., Brown, D. in Jamali, H. R. (2012). Digital repositories ten years on: what do scientific researchers think of them and how do they use them? *Learned Publishing*, 25(3), 195. doi:10.1087/20120306
- Rigaux, P. P., Abrouk, L. L., Audéon, H. H., Cullot, N. N., Davy-Rigaux, C. C., Faget, Z. in Thion-Goasdoué, V. V. (2012). The design and implementation of Neuma, a collaborative Digital Scores Library. *International Journal on Digital Libraries*, 12(2/3), 73–88. doi:10.1007/s00799-012-0089-9
- Sutton, S. C. (2012). Balancing Boutique-Level Quality and Large-Scale Production: The Impact of „More Product, Less Process“ on Digitization in Archives and Special Collections. *RBM: A Journal of Rare Books, Manuscripts, & Cultural Heritage*, 13(1), 50–63.
- Teper, J. in Shaw, E. F. (2011). Planning for Preservation During Mass Digitization Projects. *Portal: Libraries & The Academy*, 11(2), 717–739.
- Tharani, K. (2012). Collections Digitization Framework: A Service-oriented Approach to Digitization in Academic Libraries. *Partnership: The Canadian Journal Of Library & Information Practice & Research*, 7(2), 1–13.
- Wu, M. M. (2011). Building a Collaborative Digital Collection: A Necessary Evolution in Libraries. *Law Library Journal*, 103(4), 527–551.

Matjaž Kragelj

Narodna in univerzitetna knjižnica, Turjaška 1, 1000 Ljubljana
e-pošta: matjaz.kragelj@nuk.uni-lj.si