

ANALIZA STANJA IN PERSPEKTIVA DIGITALNE HRAMBE V ARHIVU DRUŽBOSLOVNIH PODATKOV (ADP)¹

Janez Štebe
Irena Vipavc Brvar

Oddano: 30. 6. 2010 – Sprejeto: 7. 1. 2011

Izvirni znanstveni članek

UDK 303:004:005.921.1"746"(497.4)

Izvleček

Arhivi družboslovnih podatkov delujejo v različnih državah že od šestdesetih let dalje in so eni od pionirjev digitalne hrambe. Slovenski podatkovni arhiv pri svojem delovanju upošteva, prilagaja in dopolnjuje dobre prakse na svojem posebnem področju: to je pridobivanje raziskovalno zanimivih »surovih podatkov«, njihovo arhiviranje in trajno hranjenje ter zagotavljanje dostopa uporabnikom. Kljub dolgi tradiciji se, tako kot na drugih področjih digitalne hrambe, tudi v svetu podatkovnih arhivov v zadnjem času precej pozornosti posveča izpolnjevanju zahtev dolgotrajne hrambe po standardu OAIS in spremljajočim tehnološkim rešitvam, metapodatkovnim standardom, strategijam in politikam ter dobrim praksam. Namen prispevka je predstaviti postopke v ADP, jih primerjati z izbranimi zgledi dobrih rešitev v organizacijah s podobnim poslanstvom, ter na tej podlagi podati predloge izboljšav. Uporabljeni so zgledi iz partnerskih organizacij v okviru združenja podatkovnih arhivov CESSDA, ki so med najnaprednejšimi pri oblikovanju strategij in organizacijskih rešitev na področju digitalne hrambe, zlasti iz ZDA, Velike Britanije in Nizozemske. Sklicevali smo se tudi na zasnovo skupnih vodil zanesljivega delovanja za tip organizacij arhivov raziskovalnih podatkov, ki je nastala

1 Članek je nastal na osnovi študije v okviru projekta dLib.siPlus – Nove vsebine in storitve Digitalne knjižnice Slovenije (2009–2010), ki ga je sofinanciralo Ministrstvo za visoko šolstvo, znanost in tehnologijo RS.

v okviru istoimenskega (tj. CESSDA PPP) projekta iz 7. okvirnega programa EU. Rezultati analiz so uporabni za vse tiste, ki iščejo praktične rešitve za podobne probleme.

Ključne besede: družboslovne raziskave, podatki, trajna hramba, OAIS, Arhiv družboslovnih podatkov, Slovenija, CESSDA, e-raziskovanje

Original scientific article

UDC 303:004:005.921.1"746"(497.4)

Abstract

Social science data archives have begun to evolve in 1960's, thus being one of the pioneers of digital preservation. The Slovenian Data Archive (ADP) follows and adjusts best practices that have emerged in this specific field. Its tasks comprise acquisition, processing and long term preservation of valuable primary research data, as well as providing accessibility to the content for users. As in other areas of digital preservation much attention is paid to meeting the requirements of long term preservation as specified in the OAIS (Open Archival Information System) standard, and the use of technological solutions, proposed metadata standards, strategies and policies, and best practices. The purpose of this paper is to present procedures in ADP, to compare them with selected solutions of organizations with the similar mission, and to suggest improvements. Examples are taken from partner organizations within the CESSDA data archives association. These organizations (especially those from the USA, Great Britain and the Netherlands) have developed solutions in the field of digital preservation to the highest degree. The paper also reflects on agreed guidelines of reliable services for the archives of research data, which arose in the frame of the FP7 project with the same title (i.e. CESSDA PPP). The results of analysis are useful for all are seeking practical solutions to similar problems.

Key words: social science research, raw data, digital preservation, OAIS, Slovenian Social Science Data Archives, CESSDA, e-research

1 Uvod

Zadnje čase se veliko govori o potencialu *odprte znanosti*, tudi *ljudske oz. državljanske znanosti*, kjer vsakdo lahko črpa iz in prispeva v zakladnico znanja. Vse bolj je razširjeno prepričanje, da bi morali biti rezultati znanstvenih raziskav dostopni v obliki odprto dostopnih člankov v elektronski objavi. Za resnično odprto znanost pa je celo bolj temeljno omogočanje odprtega dostopa do raziskovalnih podatkov, ki so podlaga in dostikrat ključna sestavina zaključkov v raziskavah. Ob odprtem dostopu do podatkov lahko preverimo ugotovitve drug drugega, z večanjem dostopa in dodatnimi analizami se poveča število odkritij, kadar, kot je to s podatki v družboslovju, zaradi njihove kompleksnosti spoznavni potencial ni izčrpan z dotedanjimi analizami, zlasti pa je kolektivno delo na podatkih

odlična spodbuda za dialog, medsebojno kritiko in dopolnjevanje analiz sorodnih problemov. Prednost odprtega dostopa do podatkov je nadalje tudi v tem, da se omogočijo povezave sicer ločenih podatkov, kar kot dodana vrednost znova bogati njihov analitični potencial. Pogoj je, da so podatki dovolj kakovostni in ustrezni glede na namen, kar znova zagotavlja prav princip *odprtega* objavljanja podatkov – po analogiji z objavami člankov. Objava podatkov v odprto dostopnem repozitoriju namreč zahteva ustrezno kakovost in dokumentacijo o nastanku in vsebini, ki je potrebna za njihovo smiselno uporabo. Za zagotavljanje uporabnosti podatkov pa je potrebna celovita storitev skrbništva nad podatki, ki vsebuje tudi vse pomembne elemente dolgotrajne hrambe.

Arhiv družboslovnih podatkov (ADP, <http://www.adp.fdv.uni-lj.si/>) je ena redkih ustanov v Sloveniji, ki se ukvarja s storitvami zagotavljanja dostopa do raziskovalnih podatkov. ADP že več kot desetletje skrbi za pridobivanje, spravilo in razširjanje podatkov, primernih za družboslovne analize. Zavedajoč se pomena shranjenih podatkov smo se, spodbujeni z zgledi in iniciativami v širšem okolju, lotili priprave izhodišč za celovit sistem skrbništva nad podatki. Cilj pričujoče analize je ugotoviti, katere pomanjkljivosti so prisotne v naši organizaciji pri ravnanju s podatkovnimi gradivi in kakšne bi bile boljše rešitve. Enega od zgledov v Sloveniji predstavljajo načrti nam sorodne organizacije z akademskega področja, tj. zamisel digitalnega repozitorija za ohranjanje pisne nacionalne dediščine v okviru projekta dLib.si (Kavčič - Čolić, 2008). Torej, z uporabo terminologije in meril na osnovi modela OAIS narediti okvirna izhodišča načrta dolgotrajne hrambe za ADP. Predstavili bomo tudi širši pregled obstoječih praks in različnih strokovnih vodil, ki so formulirani v dokumentih vodilnih organizacij s področja arhiviranja družboslovnih podatkov (UKDA, ICPSR, DANS ipd.) ter različnih mednarodnih strokovnih iniciativ (DPC – Digital Preservation Coalition, DCC – Digital Curation Centre), ki se jih da uporabiti na našem področju.

Problem priprave ustreznih izhodišč za sisteme dolgotrajne hrambe je tudi uporaba terminologije, ki ni dorečena, saj isti izrazi včasih pomenijo različne stvari. Npr., začne se že pri opredelitvi, kaj so podatki. V splošnem razumevanju so digitalni podatki vsakršen zapis informacij v elektronski obliki, tako bitni zapis slike kot digitalni znaki v besedilu kot zapisi v tabelah ipd (Glossary, 2010). V terminologiji arhivov družboslovnih podatkov pa podatki pomenijo zbirko podatkovnih gradiv, ki so namenjena za znanstvene analize, katerih vsebina so, če govorimo o »kvantitativnih« podatkih, pogosto strukturirane datoteke s podatki iz anketnih raziskav ali pa so to »kvalitativni podatki« v obliki zapisov intervjujev, posnetkov ipd. (Woollard, 2010). Sledeč temu so npr. tudi statistični metapodatki (Gregory, Heus in Ryssevick, 2009) nekaj drugega kot metapodatki, namejnjeni vzdrževanju in upravljanju s sistemom za hrambo.

Pri opredelitvah dokaj abstraktnega izrazoslovja referenčnega modela OAIS (ISO 14721, 2003) in drugega sorodnega izrazoslovja dolgotrajne hrambe smo se veči-

noma ravnali po delu Kavčič - Čolić (2010, 2004) ter pojmovnikih pri ICPSR (Digital, 2007), Digital Curation Centre (DCC, <http://www.dcc.ac.uk>) in Wollard (2010).

2 Pripomočki in modeli za načrtovanje

Namen prikaza stanja in primerjave z zgledi je pripraviti podlage za načrt trajne digitalne hrambe, ki bo upošteval doseženo in zagotovil optimalno alokacijo sredstev (finančnih, strojne opreme) in človeških naporov za doseganje cilja, to je vzpostavitev sistema trajne digitalne hrambe, ki bo zadostil osnovnim profesionalnim kriterijem in bo skladen z dobro prakso drugod.

Preden se naloge lotimo, odgovorimo na dve vprašanji. Kako podroben naj bo načrt: ali naj bo ta izveden na ravni strategije ali tudi na ravni izpeljave v obliki poteka delovnih nalog, spremljajoče dokumentacije in programskih rešitev? Kaj bomo postavili za cilj, ki ga želimo doseči: načrtovanje po modelih dobrih praks v sorodnih organizacijah ali po splošnih modelih, kot je OAIS, DCC-jev *Model življenjskega cikla skrbništva*, ali pa v obliki, primerni za notranjo revizijo (DRAM-BORA) oz. zunanje potrjevanje skladnosti (TRAC, DSA).

Glede prvega vprašanja je dilema povezana z obsegom naloge, ki smo si jo zadali. Če bi želeli razdelati vse podrobnosti načrta do ravni izpeljave, bi nam zmanjkalo prostora. Tako se, denimo, v navodilih za načrtovanje repozitorijev PLATTER omejujejo na zastavitev in sistem doseganja strateških ciljev, s tem da je izpeljava prepuščena nadaljnji razdelavi po specifičnih strokovnih področjih (finančno, kadrovsko, tehnološko načrtovanje). Na drugi strani ADP, kot že delujoč repozitorij, procese obvladuje tudi na ravni izvajanja nalog. Tako bomo v prikazu, na da bi hoteli zajeti vse, izpostavili tudi konkretne postopke, zlasti kjer ti ilustrirajo dobro delujoče rešitve oz. področja, potrebna izboljšave. Načrtovanje in periodična revizija postopkov je tudi eno od načel, ki jih izpostavlja prav omenjeni PLATTER, ne glede na to, ali je organizacija že vzpostavljena ali šele začenja z delom.

Razdelava nalog, kadrov in potrebnih financ za dvig ravni delovanja tako ne bo povsem izčrpana, imamo pa tudi za to področje zelo informativne vodiče. Beagrie s sodelavci (2008a, 2010) ponuja podrobno metodologijo za razdelavo nalog in obsega dejavnosti, ki vključuje tudi fazo pred prevzemom digitalnih gradiv, ter podrobno razdelano fazo razširjanja gradiv, ki med drugim vključuje tudi študije primerov podobnih organizacij, kot je ADP. Pri tem je zanimivo, da vključuje tudi podrobno analizo koristi, ki je pri družboslovnih podatkih povezana z njihovo enkratnostjo v času in prostoru ter uporabnostjo pri načrtovanju socialnih

politik. Hkrati nekatere podrobno opredeljene aktivnosti, npr. prav z ožjega področja hrambe, kot je preverjanje fiksnosti, migracije ipd., potem ko so enkrat avtomatizirane, v delovanju ne zahtevajo dodatnega človeškega časa. Seveda je potrebno imeti prej vzpostavljen sistem, tu pa se podrobnostim ne da izogniti. Podobno je precej bolj, kot si lahko privoščimo v pričajučem pregledu, podrobna študija o kadrovskih potrebah, specializaciji in usposabljanju za delo v različnih vlogah, povezanih z ravnanjem in hrambo podatkov (Swan in Brown, 2008). To besedilo je tudi v izhodišču diskusije in predloga načrta usposabljanja in strokovne izmenjave (Krejci in Cizek, 2009) bodoče Evropske raziskovalne infrastrukture CESSDA ERI (ang. Council of European Social Science Data Archives). Za načrtovanje splošnega profila organizacije pa so na voljo tudi orodja, ki lahko avtomatizirajo izdelavo organizacijskim načrtom podrejenih dokumentov, kot je *OpenDOAR Policies Tool* (<http://www.opendoar.org>).² Tako da bomo, kljub temu da v pričajučem prikazu ne bomo zajeli vseh podrobnosti načrtovanja, vsaj nakazali koncepte in orodja, ki bodo uporabni za kasnejšo izpeljavo rešitev podrobnih problemov.

Ena od nevarnosti za organizacije z majhnim številom zaposlenih, kot je ADP (2 infrastrukturna FTE), je tudi to, da se zaradi tega, ker vsi obvladajo vse naloge in ni potrebno večje koordinacije, zdi formalni zapis podrobnih korakov odveč. Novi ali začasni sodelavci se nalog priučijo ob delu na gradivih. Izkušnja je pokazala, da sistematičen popis vseh faz in formalnega beleženja korakov pri obravnavi pripomore k bolj kakovostnemu izvajanju nalog v arhivu. Pomembna so tudi napisana pravila, ki utemeljujejo posamezne odločitve in zagotavljajo poenoteno obravnavo gradiv, sicer je to prepuščeno trenutni presoji. Tako ima ADP trenutno vzpostavljeno skupino vodičev, ki jih bomo predstavili na ustreznih mestih v nadaljevanju. Zgled tovrstnega pristopa k načrtovanju sprememb v organizaciji in izboljševanju zanesljivosti poslovanja je projekt izdelave detajlnega popisa (ang. data-pipeline) v največjem družboslovnem podatkovnem arhivu na svetu, ameriškem ICPSR (Guttman, Overcashier in Whiteman, 2004). Na tovrsten popis se nato npr. sklicujejo v študijah primerljivosti z OAIS (Vardigan in Whiteman, 2007), pri načrtovanju potreb po razvijanju programske opreme (Guttman et al., 2009) itd.

Za celovito obvladovanje sistema kakovosti digitalnega skrbništva je značilna hierarhija dokumentov, tako kot npr. v UK Data Archive (UK DA, <http://www.data-archive.ac.uk/>; glej tudi ICPSR (Digital, 2009): na vrhu so strateški dokumenti, sledijo dokumenti načrtov po posameznih področjih, njim podrejeni so dokumenti konkretnih navodil in popisov nalog. Vzporedno nastajajo tudi dokumen-

2 Na *OpenDOAR*, portalu repozitorijev z odprto dostopnimi gradivi, je orodje za podporo načrtovanju politike repozitorija, ki avtomatsko ustvarja navzven viden standardni profil organizacije glede dostopa do podatkov in metapodatkov ter digitalne hrambe, ki naj bi bil čim bolj skladen z načeli odprtega dostopa.

ti za splošno dviganje zavedanja glede pomena hrambe in zgodnjega pristopa k digitalni hrambi v širši znanstveni skupnosti. Primer dokumentacije opredeljenih nalog in odgovornosti v UK DA je prikazan v Preglednici 1, vsak dokument ima označeno verzijo in načrt posodabljanja.

Preglednica 1: Prikaz dokumentacije UK Data Archiva (<http://www.data-archive.ac.uk/about/publications>)

Smernice dobrih praks:
- UK Data Archive Data Processing Standards (2010)
- UK Data Archive Data Processing Quick Reference Guide (2010)
- UK Data Archive Documentation Processing Procedures (2010)
- UK Data Archive Qualitative Data Processing Procedures (2010)
- UK Data Archive Quantitative Dataset Processing Procedures (2010)
- UK Data Archive Catalogue Quality Control Procedures (2010)
- Managing and Sharing Data (2009)
- Microdata Handling and Security: Gguide to Good Practice (2010)
Politike:
- UK Data Archive Collections Development Policy (2008)
- UK Data Archive Preservation Policy (2010)

Drug pogled predstavljajo odnosi dokumenta načrta z drugimi dokumenti, kot je to prikazano v Preglednici 2.

Preglednica 2: Primer specifikacije dokumenta UK Data Archiva (Woollard, 2008, str. 2)

Nadomešča: UKDA Preservation Policy (Version 3)
Se nanaša na:
Dokumenti strategij in politik:
- The UKDA Strategic Plan;
- The UKDA Collections Development Policy;
- The UKDA Information Security Policy;
- The UKDA Records Management Policy;
- The UKDA Preservation Strategy.
Pregledi:
- Guidance on Data Management

Podobno informativen zgled podaja Kavčič - Čolić (2008), ki vsebuje podrobne sheme procesov, ter izpostavlja seznam dokumentov spremljajočih navodil in smernic, kot so Splošne smernice za trajno ohranjanje ter njim podrejena navodila za nosilce, formate, varno shranjevanje ipd.

Zelo uporabni so tudi napotki in zgledi v študiji pregleda načrtov digitalne hrambe (Beagrie et al., 2008b), ki vsebuje obsežen dodatek z napotki, kako izkoristiti

krovne nacionalne ali organizacijske strategije za to, da se vanje vnese ustrezne vsebinske poudarke glede pomena digitalne hrambe. Torej se hierarhija dokumentov razteza tudi navzgor, v smeri nacionalnih, organizacijskih strategij, zlasti z univerzitetnega in raziskovalnega področja, pa seveda tehnoloških standardov in zakonodaje, ki jih je potrebno upoštevati.

Drugo vprašanje je, kaj naj bo zgled, kateri model. Prednost vsebinsko specifičnih modelov, ki jim lahko rečemo modeli dobrih praks, je neposredna razumljivost, saj so zgledi s področja vsebine, ki jo poznamo in razumemo sami. Na drugi strani je v tem lahko nevarnost, saj preveč dobesedno posnemanje postopkov ene ustanove oslabi zmožnost za prilagajanje morebitnim organizacijskim posebnostim. Predpostavimo npr. lahko, da bo rešitev problema trajne hrambe zahtevala sodelovanje na nacionalni in mednarodni ravni, torej skrbništvo v distribuiranem modelu disciplinarno specializiranih podatkovnih arhivov (astronomija, biološke vede, kemija, družboslovje, humanistika ...), ki bodo določene storitve centralizirali (npr. poenoteno mesto dostopa do podatkov, enotno mesto fizičnega shranjevanja ...). Na mednarodni ravni bo to povezovanje disciplinarno sorodnih arhivov v mreže, strokovno sodelovanje, poenoteni vmesniki za dostop, upoštevanje priporočil in dobrih praks ... V tem primeru so posamezne faze trajne hrambe elektronskih virov pod skrbništvom različnih institucij, kjer začne dajalec pri ustvarjanju vira v okviru raziskovalnega projekta, vmes je prevzemnik začasnega skrbništva, kot je digitalni repozitorij ustanove (Steinhart, Dietrich in Green, 2009), in na koncu prevzemnik nalog trajne hrambe, kot je za disciplinarno področje pooblaščen arhiv podatkov v sodelovanju z organizacijami hrambe na drugih področjih, ki skupaj zagotavljajo celovito skrbništvo nad za kulturo in znanost pomembnimi gradivi. V slednjem pogledu je potrebno povezovanje z uporabo splošnejših konceptov digitalne hrambe, ki zagotavljajo možnost medsebojnega prevzemanja gradiv v sistemu skrbništva.

Ball (2010), Green, Macdonald in Rice (2009) ter ICPSR (Digital, 2009) ponujajo celovit pregled modelov, navodil, standardov in orodij, vključno s pregledom metapodatkovnih modelov (DDI, PREMIS itd.) za hrambo. Posamezni podatkovni arhiv se bo naslanjal samo na ožji izbor v omenjenih virih prikazanih orodij. S področja načrtovanja in upravljanja skrbništva nad digitalnimi raziskovalnimi podatki sta priporočeni tudi DRAMBORA in TRAC kot konceptualno povezani, vendar vsako specifičnim vidikom preverjanja delovanja namenjeni orodji. Naš pregled se v nadaljevanju mestoma naslanja na klasifikacijo aktivnosti po modelu TRAC, ki so ga za namen načrtovanja kot najbolj uporabnega označili Steinhart, Dietrich in Green (2009), ter DCC-jev *Model življenjskega cikla skrbništva* (ang. curation lifecycle model), ki je podlaga nekaterih zgoraj omenjenih pregledov (zlasti Beagrie et al., 2008b).

3 Izhodišča za digitalno hrambo v ADP

Najvišji dokument, ki opredeljuje načela in naloge digitalne hrambe v organizaciji, je načrt digitalne hrambe. Vlogo takšnega načrta dobro opredeljuje sledeči primer: »Načrt predstavlja zapisano opredelitev dolgo obstoječe dobre arhivske prakse v UKDA« (Woollard, 2010, str. 3). Prakse, ki se je razvijala skozi leta in odraža tako potrebe uporabnikov, notranjo organiziranost kot naravo gradiv, s katerimi ravna. Tej praksi je dodana raven eksplicitne opredelitve na enem mestu. Na ta način je zagotovljena preglednost uporabljenih postopkov, njihove podlage in pa nosilci izvajanja. Obstoj napisanega načrta tudi zagotavlja lažje zunanje preverjanje in ocenjevanje ter periodične notranje revizije in dopolnitve. S tem hkrati sledi načelu preglednosti in javnega dostopa do postopkov, ki jih uporablja, in tudi prenosu teh načel kot dobrih praks širše, v skupnosti podobnih organizacij. Dve izmed vodilnih ustanov na področju arhiviranja družboslovnih podatkov Inter-university Consortium for Political and Social Research (ICPSR - <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>) in UK DA (<http://www.data-archive.ac.uk/>) imata napisani in javno objavljeni načrt (Wollard, 2010; McGovern, 2007).

Našo primerjavo lahko začnemo z ugotovitvijo, da ADP nima napisanega načrta digitalne hrambe. V ADP smo v preteklih letih slediti dobrim praksam v tujini, obenem pa smo tudi sami dodajali številne korake in prilagajali rešitve in orodja. Na splošni ravni načrtov sta oba omenjena zgleđa za nas razumljiva in uporabna kot izhodišče, s tem da je potrebno zagotoviti vključevanje in prilagajanje našim lastnim rešitvam in slediti nadaljnjemu razvoju na področju, tako da se v načrt vključijo najnovejše rešitve. Namen pričujoče študije je prav zagotoviti izhodišča za vzpostavitev zanesljivega sistema dolgotrajne hrambe, prilagojena organizacijskim posebnostim in posebnostim obravnavanih gradiv. Ukrepi, ki jih bo potrebno sprejeti in ki sledijo iz opravljenih analiz, so v besedilu izpostavljeni v obliki ugotovitev. Tako je prva ugotovitev:

- *Ug. 1: Potrebno je pripraviti načrt dejavne dolgotrajne hrambe.*

Podlaga načrta je ugotovitev, da je za doseganje cilja digitalne hrambe, se pravi ohranjanje dostopa do vsebine digitalnih objektov, koristno načrt uskladiti z referenčnim modelom OAIS in drugimi standardi in dobrimi praksami. Model OAIS vsebuje abstraktno terminologijo, ki jo bomo uporabili v nadaljevanju pri opisu podrobnosti delovanja procesov v arhivu, kot so zajeti v pripravi informacijskih paketov (AIP, SIP in DIP, glej http://www.adp.fdv.uni-lj.si/za_uporabnike/o_arhiviranju/slovar_digitalne_hrambe_AD/). Tako UK DA kot ICPSR sta objavila več študij, ki opisujejo vidike usklajenosti načrta z OAIS modelom. Študije so koristne, ker ponujajo zorni kot nam sorodnih področnih organizacij na ta abstraktni model in tako predstavljajo zgled iskanja rešitev za posamezne procese (Beedham et al., 2005, Vardigan in Whiteman, 2007).

V načrtu dolgotrajne hrambe, skladnem z OAIS modelom, je najprej opredeljeno poslanstvo ustanove oziroma organizacije. Poslanstvo ADP (http://www.adp.fdv.uni-lj.si/o_arhivu/) je opredeljeno takole:

Osrednji slovenski Arhiv družboslovnih podatkov je bil ustanovljen leta 1997 kot organizacijska enota v okviru Inštituta za družbene vede na Fakulteti za družbene vede, Univerza v Ljubljani, z nalogami, da pridobiva pomembne podatkovne vire s širokega nabora družboslovnih disciplin, zanimive za slovenske družboslovce, jih shranjuje in varuje pred uničenjem in namenja nadaljnji uporabi za znanstvene, izobraževalne in druge namene.

Tako opredeljeno poslanstvo bi bilo lahko tudi izhodišče načrta dolgotrajne hrambe. Dodatno bi se lahko sklicevali na nacionalne načrte in strategije zagotavljanja dostopa do produktov znanstvenega raziskovanja (npr. Resolucija, 2006) in podatkov iz javno financiranih raziskav (npr. OECD, 2007), kot tudi dokumente in usmeritve matične institucije Univerze v Ljubljani. Za dodatno izkoriščanje usmeritev, vsebovanih v krovnih dokumentih, si lahko pomagamo s predlogi, vsebovanimi v zgoraj omenjenem dodatku britanske študije (Beagrie et al., 2008b).

Ugotovimo lahko, da ADP deluje po načelih samoopredeljenega cilja izpolnjevanja akademskega poslanstva zagotavljanja storitev hrambe in dostopa do gradiv akademski skupnosti. V tem se razlikuje od ustanov, ki delujejo na področju varovanja kulturne dediščine in so zavezane tovrstni zakonodaji (npr. ZOIPub – *Zakon o obveznem izvodu publikacij*), prav tako pa se razlikuje tudi od arhivski zakonodaji zavezanih javnih ustanov (zavezancev ZVDAGA – *Zakona o varstvu dokumentarnega in arhivskega gradiva ter arhivi*). Za dolgoročno delovanje bi bilo to koristno spremeniti. Spodbudili bi pripravljenost dajalcev podatkov za sodelovanje in njihovo skrbnost pri pripravi gradiv, zagotovili pa tudi finančno vzdržnost delovanja in jasno opredeljeno odgovornost. Na eni strani gre pobuda Ministrstvu za visoko šolstvo, znanost in tehnologijo (MVZT), da pripravi strategijo na področju vlog in odgovornosti skrbi za raziskovalne podatke in temu ustrezno predlaga spremembo zakonodaje s področja znanosti³, pri tem pa je lahko v veliko pomoč pregled stanja in perspektiv v razvitem okolju (npr. Lyon, 2007). Na drugi strani je zanimiv model povezav v Britaniji med Statističnim uradom (ang. Office for National Statistics – ONS), Arhivom družboslovnih podatkov (UK DA) ter Nacionalnim arhivom (ang. The National Archive – TNA). UK DA je od TNA pridobil status pooblaščenega ustanove za hrambo arhivskih gradiv (UKDA, 2005), obenem pa od ONS prevzel skrb za ravnanje z in dostop do zaupnih oziroma občutljivih podatkov (Secure Data Service, <http://securedata.ukda.ac.uk/>). ADP sicer tesno strokovno sodeluje s Statističnim uradom RS (SURS), vendar povezave za zdaj še niso določene na ravni dolgoročnih dogovorov o sodelovanju oz. zakonsko opredeljenih obveznosti pri zagotavljanju dostopa do podatkov.

3 Na tem področju zato teče projekt »Odprti podatki – Priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji (2010–2013)«, ki ga po naročilu MVZT izvaja ADP. Glej več na: <http://www.adp.fdv.uni-lj.si/blog/>.

Predvidevamo, da bo korak naprej v to smer prinesel skupni projekt partnerjev CESSDA in statističnih uradov v Evropi z naslovom *Data Without Boundaries* (DwB), ki se bo v okviru posebnega razpisa 7. okvirnega programa (INFRA-2010-1.1.3: European Social Science Data Archives and remote access to Official Statistics) predvidoma začel spomladi 2011 in kjer kot partnerji sodelujemo tako ADP kot SURS. Zgled za tovrstno ureditev je lahko predlog spremembe zakona o statistiki na Finskem, ki zadeva dostop do podatkov starejšega datuma v posesti Statističnega urada Finske.⁴

- *Ug. 2: V zakonodajo s področja znanosti in statistike je potrebno vključiti odgovornosti glede zagotavljanja odprtega dostopa do podatkov.*

Za načrte digitalne hrambe je običajno, da vsebuje tudi podrobnejšo opredelitev ciljev, kjer se izpostavlja zlasti zagotavljanje avtentičnosti in zanesljivosti uporabnikom dostavljenih verzij digitalnih podatkovnih gradiv, učinkovitost, ter spremljanje in izpolnjevanje zahtev standardov ter naročnikov.

Ostali organizacijski pogoji, kot sta finančna in organizacijska vzdržnost, v dosežanih praksi ne dosegajo zadostnega praga za zagotavljanje dolgotrajne hrambe. Financiranje ADP poteka pod naslovom infrastrukturnega raziskovalnega programa v sklopu *Mreže raziskovalnih infrastrukturnih centrov Univerze v Ljubljani*, se pravi, da je vsaj srednjeročno stabilno, saj se program na podlagi razpisa obnavlja vsakih 5 let. Vendar je potrebno glede na funkcionalne zahteve načrta dolgotrajne hrambe zagotoviti ustrezno višino financiranja, da imamo za to primerno opremo in usposobljeno osebje, ki bo pripravljeno gradiva pripraviti tako za hrambo kot distribucijo in obenem pomagati tako ciljnim uporabnikom kot dajalcem. Pri tem mora slediti mednarodnim standardom, spremljati razvoje novih tehnologij in imeti možnosti izobraževanja za te potrebe. Omenjali smo, da se pri ocenjevanju stroškov lahko uporabi predlagana metodologija (Beagrie et al., 2008a, 2010), kjer stroški ustrezajo obsegu dejavnosti, upoštevajo pa tudi morebitno združevanje posameznih storitev in povezovanje z drugimi organizacijami. Prav tako je potrebno predvideti prenos digitalnih objektov v primeru ukinitve organizacije in ugotoviti, katera organizacija je za to vsaj približno primerna. Kandidati so lahko sorodne organizacije doma in v tujini.

ADP sedaj aktivno sodeluje z drugimi raziskovalnimi organizacijami (svojimi dajalci) in se s posameznimi iniciativami povezuje z organizacijami, ki imajo podobno poslanstvo v Sloveniji ter v svetu (CESSDA, IFDO, ICPSR). Na ravni teh omrežij bi bilo potrebno iskati možnosti za združevanje storitev na enem mestu za izboljšanje kakovosti in pocenitev.

- *Ug. 3: Potrebna je presoja finančnih zahtev vzdržnega sistema digitalne hrambe in načrt prenosa digitalnih objektov v primeru ukinitve.*

⁴ Laaksonen, H., osebna korespondenca, 28. 1. 2011.

Naslednji vidik organizacije po kriterijih TRAC (Trustworthy, 2007) je vprašanje ustrezne kadrovske zasedenosti. Če uporabimo zglede drugih organizacij, imajo te zelo razvito kadrovske politiko s podrobnimi opredelitvami zahtev posameznega delovnega mesta. Odgovornost za digitalno hrambo se začne pri predstojniku, nato je opredeljeno mesto vodje digitalne hrambe, pa računalniški center, vodja pridobivanja in prevzema gradiv ter vodja organiziranja dostopa do gradiv in na koncu tudi vsi zaposleni. Glede na to, da so to velike organizacije, je specializacija nalog jasna. V primeru relativno majhne organizacije, kot je ADP, si lahko naloge deli več oseb, vendar je pomembno, da so naloge in odgovornosti jasno opredeljene. Prav tako lahko del nalog posredujemo v izvajanje drugam, drugim oddelkom ali izvajalcem posameznih storitev. Kot poudarja tudi zgoraj omenjena študija, je problem predvsem v izobraževanju kadra tako za specifično obvladovanje procesov pri ravnanju z družboslovnimi podatki kot tudi na področju postopkov in uporabe orodij dolgotrajne hrambe. Z moderno tehnologijo se stvari stalno spreminjajo in prirejajo. Tudi če uporabljamo skupna orodja, moramo nekatere stvari še vedno sami razviti, ker so specifične za naše okolje, splot pa opraviti tudi jezikovne prilagoditve.

- *Ug. 4: Potrebna je opredelitev nalog in odgovornosti za digitalno hrambo ter stalno usposabljanje izvajalcev nalog.*

4 Model procesov v ADP

Model OAIS kot eno od izhodišč postavlja opredelitev ciljne skupnosti uporabnikov. Ta je v ADP (<http://www.adp.fdv.uni-lj.si/>) opredeljena kot:

Ciljni uporabniki so domači in tuji raziskovalci, učitelji in študentje, ki so podatkovno in statistično pismeni za samostojno ravnanje z gradivi. Sem sodi sposobnost samostojnega iskanja podatkov za določen problem, presoje uporabnosti podatkov na podlagi opisov, ki jih nudimo o nastanku raziskave, metodologiji zbiranja in podrobni vsebini posameznih spremenljivk, sposobnost samostojne analize in razlage rezultatov ter oblikovanja zaključkov.

Med uporabniki so tudi novinarji in splošna javnost, vendar ponudba podatkovnih gradiv njim ni posebej prilagojena, pač pa se predpostavlja potrebna seznanitev s posebnostmi narave in uporabe gradiv. Za ta namen, kot tudi za namene zahtevnejše uporabe, je ponujenih več učnih gradiv in nadaljnjih referenc za usposabljanje za delo z gradivi v podatkovnih arhivih.

Sam proces prevzema gradiva se pravzaprav začne že s predhodno fazo presoje in izbora primerne gradiva, čeprav ta faza sodi še v zunanje okolje sistema digitalne hrambe, kot ga opredeljuje model OAIS. To fazo kot samostojno pre-

deljuje tudi DCC-jev *Model življenjskega cikla skrbništva*. Nekateri to opredeljujejo kot predprevzemno funkcijo, kjer je s stališča digitalne hrambe pomembno, da dajalci že ob nastanku zagotovijo dovolj dobro dokumentirana ter ustrezno kakovostno pripravljena podatkovna gradiva, ter preprečijo morebitne zadržke zaradi problemov zaupnosti in drugih pravic do uporabe. Tudi predlog *Data Seal of Approval guidelines* (2009) denimo vsebuje nekaj zahtev, ki se nanašajo neposredno na dajalce, npr. zahteva, da naj »izročijo podatke z ustreznimi spremljajočimi informacijami, ki bodo omogočile drugim oceno znanstvene kakovosti raziskovalnih podatkov, skladno z etičnimi in znanstvenimi normami posamezne discipline«. S področja usposabljanja dajalcev za izročanje je tudi vrsta vodičev in priročnikov, ki podrobno obravnavajo različne vidike priprave, ravnanja s podatki in njihovega varnega shranjevanja (npr. UK DA in ICPSR).

V ADP imamo tako opredeljene kriterije za sprejem v arhiv, zavedajoč se, da dolgotrajno shranjevanje podatkov zahteva dodatne napore in stroške za pripravo v obliki, ki omogoča njihovo nadaljnjo uporabo (Laaksonen, Borg in Štebe, 2006). Ti so upravičeni ob prihranku, ki ga predstavlja večkratna uporaba podatkov. Presoja o prevzemu poteka na podlagi grobega pregleda poslanih gradiv, pri čemer se upošteva sledeče kriterije za izbor:

- vsebinsko bogastvo podatkov v smislu ustreznosti konceptualizacije in tematskega dopolnjevanja zbirke ADP, izpopolnjenost uporabljenih metod, celovitost in ustreznost podatkov in dokumentacije za nadaljnje analize,
- dajalec avtorsko razpolaga s podatki in je pripravljen podatke izročiti arhivu za razširjanje,
- izjemno zanimivost zaradi enkratnosti in dopolnjevanja praznin v obstoječi zbirki.

4.1 Sprejemni informacijski paket (SIP)

Model OAIS loči tri informacijske pakete: *Sprejemni informacijski paket (SIP)*, *Arhivski informacijski paket (AIP)* in *Dostopni informacijski paket (DIP)*, ki predstavljajo stanje in spreminjanje digitalnih vsebin od točke sprejema od dajalca, prek arhivske hrambe in na koncu do posredovanja uporabnikom. Začnimo s SIP.

Za zagotovitev sistemu arhivske hrambe ustrezne priprave digitalnih vsebin pri ustvarjanju in izročanju je priporočljivo, če arhiv s pomočjo navodil usposablja dajalce oz. se prilagaja najbolj razširjenim in v skupnosti ustvarjalcev ustaljenim oblikam. Ob sprejemu je potrebno opredeliti tudi medsebojne pravice in obveznosti glede razpolaganja in posredovanja gradiv.

Za dajalca je ADP zato pripravil sledečo dokumentacijo za izročanje gradiv v arhiv, in sicer :

- *Izjava o izročitvi* – Izročanje podatkov v arhiv poteka na podlagi podpisane izjave o izročitvi, s katero avtor pooblašča arhiv za shranjevanje, pripravo in razširjanje podatkov. Privzeta je licenca Creative Commons Priznanje avtorstva, dodatno je ponujena Creative Commons Priznanje avtorstva + Nekomercialno, prilagodimo pa se lahko željam za začasne dodatne omejitve glede dostopa. Izjava vsebuje zagotovilo avtorstva in urejenih pravic glede zaupnosti podatkov.
- *Obrazec za opis raziskave* – Dajalec kot najboljši poznavalec gradiv pripravi izhodiščni opis bibliografskih in metodskih podrobnosti o raziskavi. Obrazec za opis raziskave je poenostavljena različica deskriptivnih metapodatkov o raziskavi po DDI v prostotekstovni obliki.
- *Pregled izročениh gradiv* – Zabeležka o izročениh podatkovnih in ostalih spremljajočih gradivih, kjer so podane osnovne informacije, kot so imena datotek, velikost in vsebina, po katerih lahko preverimo istovetnost poslanega gradiva.

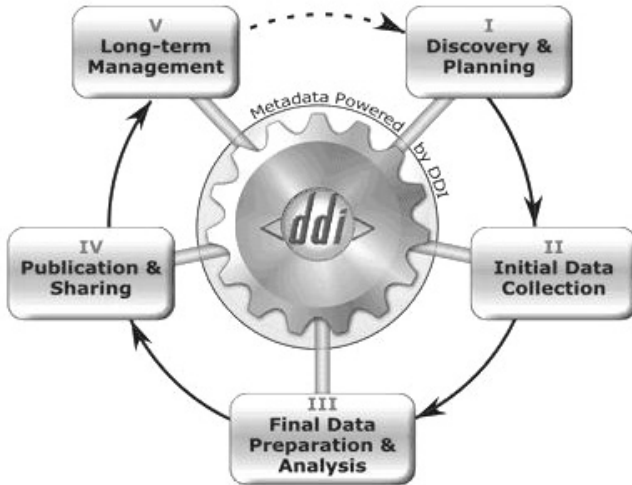
ADP je pripravil tudi seznam priporočenih datotečnih formatov (Preglednica 3).

Preglednica 3: Priporočene in druge oblike zapisov posameznih vrst gradiv za predajo v ADP (primerjaj <http://www.esds.ac.uk/aandp/create/data.asp>)

Vrsta gradiva	Priporočene oblike zapisov	Druge oblike zapisov
Strukturirana besedilna gradiva (Obrazec za opis raziskave, vprašalnik, kodirna knjiga)	Strukturiran metapodatkovni zapis vprašalnika (*.xml) po DDI ali CAI program (*.bmi)*.rtf ali druga tekstovna oblika (*.doc, *.txt, ...)	Tiskana verzija*.pdf ali druga grafična oblika
Strukturirana podatkovna gradiva (Datoteka podatkov)	SPSS (*.por, *.sav)ASCII (*.txt matrična ali z oznakami polj opremljena datoteka + opis za računalniško branje datoteke z imeni spremenljivk in njihovimi kategorijami)	Drugi statistični paketiPreglednice (*.xls ipd.)Baze podatkov
Prosto oblikovana besedilna gradiva za shranjevanje izvirne dokumentacije (Vprašalnik, navodila anketarjem, nagovor raziskovancem, kopije raziskovalnih poročil)	*.pdf ali druga grafična oblika + tiskana verzija	*.rtf ali druga tekstovna oblika (*.doc, *.txt, ...)

Posebnosti pri prevzemu gradiva v ADP so povezane z naravo gradiva, se pravi, večinoma gre za podatke iz anketnih raziskav, ter poslanstvom arhiva, to je ciljnim uporabnikom ponuditi čim bolj uporabne podatke za nadaljnjo analizo. Faza prevzema vključuje tudi že pripravo AIP in SIP, v katero je vključen tudi dajalec v tem smislu, da sproti pomaga razrešiti morebitne dvoumnosti ali napake, od-

krite v poslanih gradivih. Dajalec ob zaključku procesa pregledovanja potrdi vsebinsko ustreznost deskriptivnih metapodatkovnih zapisov v dokumentu DDI datoteke kodirne knjige (Slika 1), arhiv pa mu ob zaključku pregledovanja gradiv potrdi prevzem na seznamu zabeleženih gradiv, ki se nanašajo na dano raziskavo.



Slika 1: Shema življenjskega kroga informacij, povezanih z znanstvenimi podatki, v okviru dokumentacije najnovejše različice DDI standarda (<http://www.ddialliance.org/>)

Ob prevzemu arhiv pripravi zapis v Bazi raziskav z osnovnimi podatki na ravni raziskave, ki vključujejo enkratno šifro raziskave, naslov raziskave, identifikacijo dajalca in datum podpisane izjave o izročitvi. S tem povezan je zapis v Bazi gradiv, ki vključuje osnovne podatke o izvirnih prejetih datotekah, ki se ob prevzemu shranijo v mapi izvirnih prejetih datotek, kjer dobijo svojo identifikacijsko oznako, oznako tipa datoteke in osnovne vsebinske oznake, kot je – ali je to datoteka s podatki, vprašalnik, kodirna knjiga, ali z raziskavo povezana druga gradiva.

Omenjena posebnost obravnave podatkovnih gradiv v znanstvenih arhivih obsega fazo, ki jo imenujemo procesiranje podatkov oz. obogatitev podatkov: ta lahko obsega minimalno prilagajanje gradiv ali pa tudi zelo intenzivno delo, ki podatke nadalje preuredi in opremi z izpopolnjenimi metapodatkovnimi zapisi, s čimer postane nadaljnja uporaba bolj pregledna, lažja in z obogatimi analitičnimi možnostmi. Minimalno poseganje v podatke predstavlja preverjanje ustreznosti anonimizacije podatkov ter celovitosti dokumentacije. Najvišjo raven v ADP predstavljajo datoteke, pri katerih so vsebovane polne standardizirane vsebinske oznake spremenljivk v zapisu DDI skupaj z besedilom vprašanj iz instrumenta. Več ravni obravnave in s tem tudi različic datotek srečamo tudi v arhivih podatkov iz drugih disciplin, kjer denimo ločijo surove podatke iz meril-

ne naprave, ter različne popravljene in izpeljane količine, primerne za bolj ali manj izkušene uporabnike (Ball, 2010).

S tem je povezano tudi vprašanje granularnosti obravnave gradiv in dokumentacije: kako podrobno bomo dokumentirali, npr. raven spremenljivk, njihove povezave s konceptualnimi opredelitvami (ontologije)⁵, povezave s publikacijami, združevanje podatkov iz različnih virov itd.

4.2 Arhivski informacijski paket (AIP)

Z vidika dolgotrajne hrambe je pomembno v fazi prevzema zagotoviti ustvarjanje SIP, ki bo omogočil sledljivost nadaljnjih korakov pri arhivski obravnavi gradiv. Pregled stanja v fazi ustvarjanja verzije za dolgotrajno arhivsko hrambo (AIP) delimo na ravnanje s spremljajočo dokumentacijo ter ravnanje z datoteko s podatki.

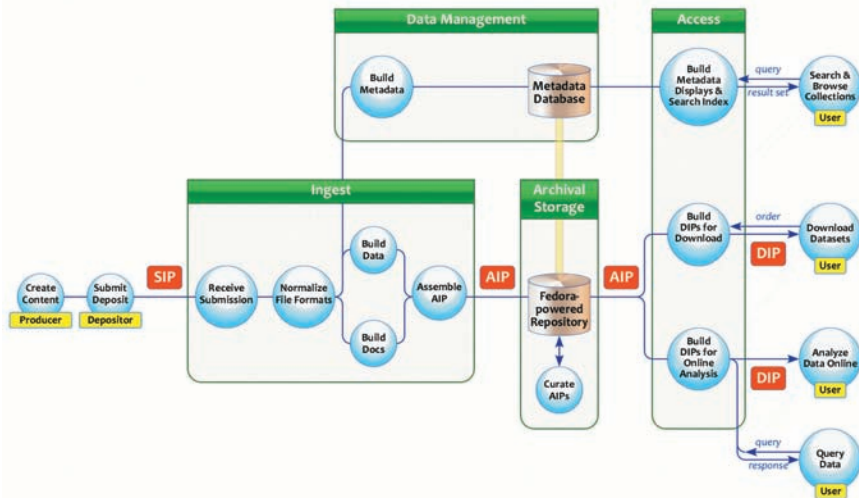
Izvirna dokumentacija na papirju ali v prosti (nestrukturirani) besedilni datoteki se shrani v obliki, kot je prejeta. Iz tega se za sistem trajne hrambe (AIP) pripravi kopija v splošno uporabljanem formatu v elektronski obliki s pomočjo skeniranja oz. konverzije. Primer aplikacije obravnave datoteke po modelu OAIS je prikazan na Sliki 2.

Izboljšanje prakse v ADP se lahko usmeri v presojo in določanje minimalnih informacij o prejetem gradivu z vidika nadzora in upravljanja digitalne hrambe. Pri tem je lahko v oporo uporaba katerega od uveljavljajočih se standardov in orodij za repozitorij. Prav tako je potrebna presoja ustreznosti varnostnih ukrepov, ki so v veljavi na ravni celotne fakultete FDV z vidika zaščite pred vdori v omrežje in virusi.

- *Ug. 5: Fazo prevzema SIP za prostobesedilna gradiva je potrebno standardizirati skladno z uveljavljenimi zgledi dobrih praks. Orodja za avtomatsko identifikacijo formatov in informacijo o fiksni digitalnega gradiva lahko zagotovijo avtomatsko oblikovanje SIP. Za prenos v AIP se izbere ustrezen format za trajno hrambo in izvede konverzija. Zagotoviti je potrebno sledljivost oz. prenos ustreznega vsebinskega dela informacijskega paketa v AIP.*

Originalna gradiva v strukturirani obliki (datoteka s podatki, tipično v formatu SPSS, in vprašalnik za računalniško vodeno anketiranje) se shranijo kot prejeta, iz njih pa se s konverzijo izvede normalizirana verzija datoteke v formatu,

5 BFO (<http://www.ifomis.org/bfo>) je primer projekta, ki se ukvarja z vprašanjem razvoja ontologij.



ICPSR Data Pipeline

Slika 2: Shema procesa obravnave datoteke podatkov po fazah OAIS v ICPSR (Guttman, Overcashier in Whiteman, 2009)

primernem za trajno hrambo in s posegi, ki jih je doživela v poteku procesiranja (SPSS prenosljivi format). Hkrati se s konverzijo iz SPSS datoteke generira informacija o strukturi datoteke z oznakami spremenljivk in vprašanj v zapisu DDI datoteke kodirne knjige. V slednjo dodatno prek urejevalnika za XML vpišemo deskriptivne informacije o vsebini in metodi raziskave, znotraj pa je zabeležena tudi vsebina paketa datotek in zunanjih informacij, ki se nanašajo na določeno raziskavo (povezana gradiva, kot so vprašalniki, podatki o objavah itd.).

Dokumentacija izvirnega formata datoteke s podatki in njene konverzije v format za trajno hrambo in nadaljnega procesiranja je običajno vsebovana v Unicode datoteki, v kateri so zabeleženi ukazi v izrazih sintakse za program SPSS.⁶ Format za trajno hrambo datotek s podatki trenutno v uporabi je *SPSS prenosljivi format* (*.por – SPSS portable format), ki obenem s podatki vsebuje tudi metapodatke o strukturi datoteke z imeni in kratkimi opisi spremenljivk ter pripadajočih kategorij vrednosti. Sprejemljiv bi bil tudi zapis v Unicode datoteki s spremljajočim zapisom o strukturi te datoteke, npr. v dokumentaciji v formatu DDI XML (Tjalsma et al., 2009). SPSS prenosljivi format sicer omogoča od programskega paketa neodvisno branje datoteke in njeno konverzijo v poljubne druge formate. To pomeni, da za vzdrževanje logične predstavitve datoteke podatkov ne potrebujemo licenčnega programskega okolja SPSS. Ob tem pa je usmeritev

6 Primerjaj <http://techaticpsr.blogspot.com/search/label/fedora%20repository> kot zgled za to, da je datoteka sintakse sestavni del paketa za hrambo.

ADP, podobno kot drugod, da sprejema tudi neštevilske podatke, ki nastajajo v okviru kvalitativnih raziskav, tako da se bo v bodoče širila raznovrstnost obravnavanih formatov in z njimi povezanih posebnih navodil za obravnavo.

- *Ug. 6: Paket AIP za družboslovne podatke mora vsebovati izvorno datoteko, njeno normalizirano izvedenko, ter datoteko s podatki o transformaciji.*

Cela raziskava ima pripisan enolični identifikator, prav tako tudi posamezne datoteke (v bazi gradiv in v zapisu DDI XML). Vendar za zdaj ne uporabljamo sistema trajnih digitalnih identifikatorjev. V ICPSR uporabljajo DOI kot najbolj razširjen, med tem ko se arhivi v okviru CESSDA še niso odločili za določen sistem (primerjaj pobudo DataCite, <http://datacite.org/>).

- *Ug. 7: Spremljati je treba in sodelovati pri izbiri optimalnega sistema trajnih identifikatorjev ter ga nato uvesti.*

Če povzamemo, je rezultat po zaključku prevzema in notranjega procesiranja podatkov paket AIP, ki vsebuje izvorno in izvedeno datoteko s podatki, datoteko o njuni transformaciji, datoteko DDI XML z deskriptivnimi in statističnimi strukturnimi metapodatki ter povezana gradiva, kot je vprašalnik in besedilna dokumentacija v elektronski obliki. V paket AIP ni vključen s strani dajalca prejeti Obrazec za opis raziskave, saj je informacija, vključno z redakcijo arhiva in končno potrditvijo dajalca, prenesena v zaključno verzijo DDI datoteke kodirne knjige.

Problem, ki je za zdaj nerazrešen z vidika oblikovanja in upravljanja z AIP, je vprašanje oblikovanja verzij datotek. Izboljššan sistem upravljanja z verzijami datotek bi moral biti zmožen priprave osnovne arhivske kopije datoteke, ki bi se zamenjala samo v primeru, če bi se v izvorni datoteki odkrila napaka, tako da bi osnovno arhivsko kopijo zamenjala popravljena datoteka. Ta verzija datoteke bi vsebovala vse izvirne enote in spremenljivke ter klasifikacije pred nadaljnjim procesiranjem. Glede na različne namene in faze procesiranja bi bilo potrebno uvesti različne verzije datotek in njim prilagojenih pogojev dostopa. V dostopu do datotek z zaupnimi podatki (z nespremenjenimi vrednostmi še pred posegom z anonimizacijo) se uvedejo posebni pogoji (primerjaj <http://securedata.ukda.ac.uk/>). Manj striktno so omejitve dostopa za t. i. SUF (ang. scientific use files), tj. datoteke, namenjene znanstveni rabi, in še manj za PUF (ang. P – Public), tj. za najširšo rabo namenjene datoteke.

Skratka, tako v datotekah s podatki kot v datotekah dokumentacije je potrebno razviti sistem verzij in izdaj, tako da se lahko sledi manjšim ali večjim popravkom, ter da so uporabnikom glede na njihove potrebe na voljo ustrezne zadnje verzije datotek, vključno z informacijami o spremembah. Pri presoji uvedbe ustreznega repozitориjskega programa in metapodatkov za shranjevanje je potrebno zagotoviti tovrstno funkcionalnost. Idealno bi bilo, če bi bilo vzpostavljeno

avtomatsko generiranje iz osnovne arhivske kopije datotek v različne izvedene verzije, vključno tudi v verzijo za DIP (Dostopni informacijski paket). V UK DA posebej poudarjajo specifično odstopanje od modela OAIS zaradi zgodnjega generiranja različice DIP, ki se nato vzporedno hrani tudi kot gradivo v paketu AIP (Wollard, 2008, str. 10). Nova verzija DDI (DDI3.0) ima razvite metapodatkovne modele za tovrstne probleme in je s tega vidika vredna presoje.

- *Ug. 8: Vzpostaviti je potrebno sistem izdaj in verzij elektronskega gradiva ter zagotoviti upravljanje z različicami, tako da bo AIP slonel na osnovni arhivski kopiji datoteke, ki bo po intelektualni vsebini identična datoteki iz SIP, in iz katere bo mogoče v vsakem trenutku izvesti ustrezne izvedene različice za paket DIP.*

Zaključimo lahko, da se priprava gradiv in podatkov v ADP v bistvenih značilnostih ne razlikuje od podobnih procesov v sorodnih organizacijah (glej pod ang. *Data Enhancement* na ICPSR ali ang. *Processing* na http://www.cessda.org/sharing/archival_process).

4.2.1 Fizično shranjevanje

Shranjevanje datotek je eno od kritičnih področij, ki zahtevajo takojšnje ukrepanje, saj trenutno ni uveljavljen noben zanesljiv sistem fizičnega shranjevanja. Začasna rešitev je sistem več kopij vseh obstoječih datotek na diskih na ločenih računalnikih in zunanjem disku. Na ravni FDV se pripravlja sprejem zunanjega ponudnika sistema fizične hrambe, skladno z uveljavljenimi standardi in zahtevami zakonodaje. Ponudnik bo po načrtu poleg dokumentarnega in arhivskega gradiva organizacije prevzel tudi skrb za varno shranjevanje kopij digitalnih gradiv v posesti ADP. Naša naloga je pripraviti izbor datotek, ki se bodo namenile za sistem trajne hrambe, vključno s sistemom izločanja nepotrebne gradiva. Prav tako je potrebno zagotoviti omejen dostop do ključnih datotek in sistem nadzora nad kopijami, zlasti tudi preverjanja istovetnosti. Zagotoviti je potrebno ujemanje lokalnega sistema in sistema ponudnika fizične hrambe glede formatov, paketov povezanih gradiv, verzij in uporabljenih identifikatorjev. Podobno kot ugotavlja jo v poročilu Stainhart et al. (2009), je za distribuiran sistem dolgotrajne hrambe pomembno, da so posamezni koraki med seboj združljivi.

- *Ug. 9: Vzpostaviti je potrebno zanesljiv in združljiv sistem fizičnega shranjevanja.*

4.3 Dostopni informacijski paket (DIP)

Uporabniki lahko do gradiv, hranjenih v ADP, dostopajo na dva načina:

1. Prek kataloga na spletni strani ADP (<http://www.adp.fdv.uni-lj.si/opisi/>) – kjer je prikazan seznam vseh raziskav (te si sledijo po abecednem vrstnem redu),

po želji pa lahko uporabnik izbere tudi izpis po posameznih serijah, vsebinskih področjih in institucijah, v okviru katerih so bile raziskave narejene. Vsaka posamezna raziskava na tem mestu vsebuje tudi osnovne informacije o raziskavi, frekvence vseh spremenljivk ter možnost prenosa vseh pripadajočih gradiv in podatkov. Uporabniku je na voljo tudi povezava na spletni vmesnik Nesstar (www.nesstar.org). Posebej je izpostavljena priporočena oblika bibliografskega navajanja gradiva, saj to uporabnikom dela največ preglavic.

2. Prek spletnega vmesnika Nesstar (<http://nesstar2.adp.fdv.uni-lj.si/webview/>), kjer so na levi strani okna v obliki drevesne strukture prikazane vse raziskave, na desni pa se izpisujejo izbrane informacije o posamezni raziskavi (meta-podatki, opisi spremenljivk). Nesstar poleg natančnega opisa raziskav in prenosa podatkov uporabnikom omogoča tudi izvajanje preprostih spletnih analiz in preurejanja spremenljivk ter enostavno ali napredno iskanje (po posameznih elementih v opisu raziskav).

Datoteka s podatki iz kataloga ADP, ki se po tipologiji COBISS uvrsti v kategorijo 2.20 (Zaključena znanstvena zbirka podatkov), je prek konverzije opisa raziskave iz DDI v COMARC vpisana v bibliografski sistem kataloga COBISS. Leta 2007 sta bila izvedena avtomatska konverzija in prenos večine pomembnejših raziskav v katalog COBISS. Od takrat poskrbijo za vpis posamezne raziskave v osrednji bibliografski katalog prek pooblaščenih bibliotekarjev avtorji sami.⁷

Bistvena razlika med zgoraj navedenimi možnostmi za dostop do gradiv je ta, da Nesstar poleg podrobnejših opisov raziskav uporabnikom ponuja tudi orodje za iskanje ter možnost izvajanja sprotnih analiz prek spleta, kar je prednost za statističnih programov manj večje uporabnike.

Dejstvo, da se vsi opisi raziskave že sedaj pripravljajo v več različicah DDI datoteke kodirne knjige (original in verzija, s konverzijo prirejena za objavo v Seznamu raziskav, ter verzija za katalog Nesstar, vse to še dvojezično v slovenščini in angleščini), predstavlja težavo pri zagotavljanju vsebinske usklajenosti in sledenja v primeru naknadnih sprememb (potreba po vzpostavitvi sistema trajnih identifikatorjev, navezava na Ug. 7). V povezavi s tem je tudi problem sinhronizacije dejanskega stanja objavljenih raziskav z informacijami o stanju raziskav v bazi raziskav. Iskanje rešitve gre v smeri priprave enostavnega programa, s katerim bo možno preveriti najosnovnejše informacije o stanju različic opisov posamezne raziskave.

7 *Za zgled avtomatske konverzije iz DDI v COMARC si lahko ogledate vsak opis raziskave v ADP, ki mu primate podaljšek '/cobiss/', npr. <http://www.adp.fdv.uni-lj.si/opisi/sjm061/cobiss/>, s tem da 'sjm061' zamenjate s pripadajočim ADP identifikatorjem raziskave (ADP IdNo).*

4.3.1 Kategorije uporabnikov

Glede na namen uporabe, ki ga uporabniki navedejo ob registraciji, se jim dodeli pripadajoča kategorija uporabnika. Najpogosteje sta to kategoriji akademski uporabnik (ang. academicuser), ki dovoljuje spletno analizo na Nesstarju ter prenos vseh gradiv, razen datoteke s podatki, in polni akademski uporabnik (ang. fullacademicuser), ki dovoljuje tudi prenos datoteke s podatki. Izjema so gradiva s posebnimi omejitvami glede dostopa, kot so npr. raziskave s posebno občutljivimi vsebinami, kjer je potrebno tudi pisno dovoljenje avtorja raziskave za konkretno uporabo.

Možnost, da bi uporabnik dostopal do izvirnih podatkovnih datotek, trenutno ni predvidena, bi pa bilo morda smiselno uvesti takšno kategorijo uporabnikov ob predpostavki, da se dajalec podatkov s tem strinja, ter da se zelo natančno opredeli pogoje uporabe in dolžnosti uporabnika (primerjaj razpravo zgoraj o različicah bolj ali manj anonimizirane datoteke s podatki).

4.3.2 Navodila za iskanje, dostop in prenos gradiv

Dokumenti, namenjeni uporabnikom, so na spletni strani ADP v meniju Podpora uporabnikom/Usposabljanje. Za pomoč pri uporabi spletnega vmesnika Nesstar je pripravljen Vodič po Nesstarju, ki vsebuje navodila za iskanje, pregledovanje podatkov, spletno analizo ter prenos podatkov. Za iskanje po podatkih je pripravljen tudi video vodič, ki nazorno prikazuje korake iskanja. Podoben vodič bi bilo potrebno narediti tudi za pregledovanje kataloga na spletni strani ADP. Predvidena so tudi navodila za izpolnjevanje registracijskega obrazca, ki se za zdaj prikazujejo le v obliki kratkih opomb ob mirovanju miške na posameznem polju in jih glede na doseganje izkušnje večina uporabnikov spregleda.

- *Ug. 10: Posodobiti in dopolniti je potrebno navodila in vodiče za uporabnike, skupaj z učnimi gradivi za manj veščje uporabnike.*

4.3.3 Spremljanje uporabnikov

Prvi način za spremljanje uporabnikov v ADP je spremljanje obiska spletne strani, ki se izvaja že od leta 2002. V prihodnosti bi bilo potrebno števec nastaviti tako, da bi beležil tudi obiskanost Nesstarja (<http://nesstar2.adp.fdv.uni-lj.si/webview/>), saj zdaj ostanejo nezabeleženi vsi tisti obiskovalci, ki na Nesstar ne vstopajo prek strani ADP. Na podlagi pridobljenih podatkov se ob koncu meseca izdelava krajše poročilo, ki prikaže poenostavljen pregled nad stanjem ter omogoča aktualno prilagajanje potrebam uporabnikov (npr. prikazovanje vsebin v resoluciji, ki jo uporablja večina uporabnikov).

Drugi način za spremljanje uporabnikov pa omogoča Nesstar, ki beleži vse korake posameznega uporabnika. Z analizo Nesstarjevih logov je tako mogoče ugo-

toviti, katere raziskave uporabniki največkrat pregledujejo in analizirajo, katere podatkovne datoteke najpogosteje prenašajo, katere analitične možnosti Ness-tarja uporabljajo ipd. Ta način spremljanja uporabnikov v ADP je potrebno še uvesti.

5 Dokumentacija in vodiči po postopkih

Za notranjo uporabo smo v ADP razvili vrsto navodil za delo. Ta vključujejo navodila, povezana s pripravo opisa raziskave, za delo z dajalci, z uporabniki in številna krajša navodila, ki olajšajo delo v arhivu.

5.1 Navodila za Opis raziskave

Začetek priprave opisa raziskave se začne z vnosom osnovnih informacij o raziskavi v bazo raziskav, pri čemer si pomagamo z *Navodili za vpis podatkov v bazo* in s *Seznamom tematskih blokov*, ki se uporabljajo pri vsebinski kategorizaciji raziskave. Raziskava, spremljajoča dokumentacija, datoteka s podatki ter ostala, s podatki povezana gradiva se ob prihodu v arhiv zapišejo v Bazo raziskav. Ker gre za relacijsko bazo, kjer je potrebno določene informacije zapisati prej kot druge, za to potrebujemo navodila in procedure. V prihodnosti bi se dalo ta postopek delno avtomatizirati zlasti z vidika avtomatskega beleženja nekaterih lastnosti digitalnega objekta na ravni zapisa v Bazi gradiv. Popolna avtomatizacija ne bo mogoča, saj v ADP raziskavi že ob sprejemu (SIP) pripišemo številne karakteristike o izvoru, npr. podatke o naslovu raziskave, dajalcu in v izhodišču raziskavo uvrstimo v posamezni tematski blok po klasifikacijah vsebin CERIF in CESSDA TopClass.

Postopkovno sledijo *Navodila za pregledovanje in spravilo datotek* (preverjanje celovitosti podatkov in dokumentacije). Pripravi se arhivsko kopijo podatkovne datoteke in spremljajočega gradiva. Ustrezno se hranijo vse vmesne podatkovne datoteke. Kadar je datoteke potrebno še dodatno anonimizirati, so na tem koraku zapisana ustrezna navodila. Vse spremembe podatkovne datoteke se izvajajo in zapišejo s programskimi stavki SPSS. Običajni uporabniki SPSS uporabljajo s pomočjo visečih menijev. Arhivsko delo v SPSS z ukazi v programskih stavkih edino omogoča sledenje in ponovljivost dela. S tem namenom smo za sodelavce in študente, ki pri nas opravljajo obvezne študijske prakse, pripravili *Navodila za uporabo programskih stavkov v statističnem paketu SPSS*. Vse spremembe, končne in začetne datoteke se ustrezno hranijo in njihovo mesto in vsebina se ročno zapišeta v Bazo gradiv.

Nato so navodila zaposlenim za oblikovanje izhodiščne DDI datoteke kodirne knjige, ki vsebuje informacije iz baze raziskav: podatki o raziskavi in informacije o ustrezno pripravljenih podatkovnih datotekah in ostalih gradivih. Izpis prve verzije opisa raziskave se naredi neposredno iz baze s konverzijo, rezultat je posebej okolju ADP prilagojen izhodiščni dokument DDI datoteke kodirne knjige.

V bazo dodatno ažurno vpisujemo tudi vse informacije in z raziskavo povezana gradiva, do katerih smo prišli kasneje. To so na primer objave, povezane mednarodne raziskave. Te spremembe se vsakokrat izpisujejo neposredno iz baze in dodajajo v opis raziskave, vsebovan v DDI datoteki kodirne knjige.

Za opis podatkov, ki se nanaša na raziskovalni inštrument, v našem primeru je to v večini anketni vprašalnik, imamo pripravljena *Navodila za izdelavo opisa podatkov v okviru DDI kodirne knjige* na podlagi elektronskega vprašalnika v prosti besedilni obliki. Program je dodatno prilagojen za anketni vprašalnik, kot je v raziskavah Slovenskega javnega mnenja, ki je ena naših pomembnejših raziskav.

Pri pripravi končne verzije opisa raziskave v DDI datoteki kodirne knjige si pomagamo s prilagojenimi navodili za posamezne dele opisa, kot so *Pravila citiranja*. Splošen priročnik za ta del pa je popis vsebine elementov po strukturi DDI.

Za kontrolo pri vpisovanju ključnih besed, ki se pojavljajo v opisu raziskave, smo pripravili manjšo bazo le-teh. S programom izvozimo ključne besede za posamezno raziskavo. V kolikor bi se kasneje odločili za spreminjanje določene ključne besede, bi se zelo enostavno spremenila v vseh povezanih raziskavah. Zlasti to velja za angleške prevode ključnih besed, saj včasih kasneje najdemo ustrežnejši prevod. Pri zapisu in pripravi datoteke ključnih besed za posamezno raziskavo je potrebno slediti postopkom v tej tematiki posvečenem delu navodil.

Vse korake in postopke, povezane z delom na raziskavi, imamo zapisane v obsežnih navodilih, obenem pa smo pripravili tudi *Obrazec za spremljanje dela na raziskavi*. Vpisujemo potek dela, datume, kdaj je bil določen korak opravljen in dodatne informacije, ki bodo pomembne za kasnejše razumevanje dela na pripravi opisa. Na žalost se večkrat zgodi, da je postopek priprave opisa raziskave in gradiv dolgotrajen, saj je mnogo nejasnosti, glede katerih čakamo na odgovore od dajalcev. V vmesnem času običajno obdelujemo že drugo gradivo in lahko se zgodi, da bi kakšno malenkost iz prvega gradiva pozabili, če to ne bi imeli ustrezno zabeleženo.

5.2 Delo z dajalci

Kot smo omenili, ima spletna stran ADP objavljen poziv dajalcem k predaji gradiv ter spremljajoče obrazce za predajo. Ko se dajalec obrne na nas z namenom predaje gradiv, ga seznanimo s postopki in prosimo, da se seznanimo s koraki predaje gradiv in zahtevami glede izpolnjevanja kakovosti ob sprejemu. Dajalcu se nato posreduje *Obrazec za pripravo opisa raziskave*, v katerega mora zabeležiti za nas pomembne informacije o poteku raziskave. Obrazcu je priložen še *Obrazec za popis predanih gradiv*, kamor dajalec natančno zabeleži tako tiskana kot elektronska gradiva, ki nam jih posreduje. Ko je opis raziskave zaključen in so primerno arhivirana vsa gradiva, se pripravi poročilo – Opis raziskave v slovenskem in angleškem jeziku skupaj z gradivi, ki bodo na voljo uporabnikom – in se posreduje dajalcu. Celoten protokol dela z dajalci (pregledovanja izpolnjenih obrazcev, pošiljanja dodatnih in popravljenih verzij gradiv in priprava končnega poročila prevzetih gradiv) je zapisan v temu delu namenjenih navodilih.

5.3 Delo z uporabniki

Naše uporabnike največkrat zanima, kako iskati po gradivih na naših spletnih straneh in kako sneti oz. analizirati podatke. V ta namen imamo pripravljenih nekaj navodil in vodičev, ki smo jih omenili že predhodno v poglavju o DIP. Na spletni strani je dostopen tudi *videovodič* o uporabi iskalnika v vmesniku Nessstar. V pripravi so številni krajši video in tekstovni vodiči. Dejansko se je v preteklosti izkazalo, da uporabniki lažje razumejo potek in možnosti uporabe programov, če njihove zmogljivosti vidijo vizualno in ne samo pisno. Šibka točka priprave vodičev za dajalce je nepoznavanje ustreznih videoprogramov. Tu smo si do sedaj pomagali z zunanji občasni sodelavci.

Poleg prebiranja gradiv raziskav, kot so vprašalniki, in krajših metapodatkovnih informacij uporabniki najbolj povprašujejo po možnostih in postopkih neposrednega prenosa gradiva na svoj osebni računalnik in posledično dela z njim. Pri tem si lahko pomagajo z Navodili za dostop do podatkovne datoteke prek vmesnika Nesstar. Konkretna navodila vsebujejo veliko slikovnega prikaza poti. Z razvojem Nesstarja se spreminja tudi njegova grafična podoba in tako je potrebno vsakih nekaj let spremeniti tudi vsebino navodil.

Za delo s podatkovnimi datotekami, za njihov prenos ali pa neposredno spletno analizo uporabnik potrebuje geslo. Pri tem mora izpolniti registracijski obrazec. V pomoč pri izpolnjevanju so mu navodila. Ko potrdimo registracijo posameznega uporabnika, ta na svoj e-naslov prejme dopis z gesli in natančnimi navodili glede obveznosti, ki jih kot uporabnik ima. Med drugim nas mora obvestiti o vsaki

publikaciji, ki je nastala na podlagi uporabe gradiv. Publikacija se doda opisu raziskave, da jo lahko bodoči uporabniki citirajo.

Tako za uporabnike kot dajalce večkrat letno pripravljamo delavnice, na katerih se lahko seznanijo z našim delom, možnostjo predaje in uporabe gradiv. Dodatno pozornost izobraževanja pa posvečamo dodiplomskim in podiplomskim študentom, saj verjamemo, da so to naši bodoči uporabniki. Prav tako naše uporabnike seznanjamo s ponudbo in usmerjamo na vodiče ter učna gradiva, ki so na voljo v naših partnerskih ustanovah v tujini.

5.4 Interno delo v ADP

Uporaba računalniške tehnologije in omrežja je postala prezahtevna za naše želje in potrebe. Tako sodelujemo s številnimi zunanji sodelavci, ki nam pri tem pomagajo. Za njihovo delo je zelo pomembno, da poznajo naše interne postopke. V pomoč so jim vodiči po posameznih področjih:

- Navodila za dostop do Virtualnega Nesstarja – gre za testno obliko strežnika, ki je nameščena na enem od računalnikov.
- Navodila za izpis baze uporabnikov – kadar nameščamo Nesstar na novem računalniku in ne moremo prenesti baze uporabnikov na drug način. Trenutno imamo uvedeno vmesno bazo in ta je za nas primarna – vsaj z vidika uporabnikov, saj imamo v njej precej več informacij, kot jih lahko sprejema Nesstar.
- Številni manjši programi, kako prenesti datoteke opisov raziskav (npr. xml) v ustrezno obliko v Nesstarju (npr. postopki za konverzijo za na Nesstar v Javi).
- Navodila za objavo gradiv na Nesstarju in na Spletni strani.
- Navodila za delo s spletno stranjo in posameznimi podfunkcijami – npr. objavljanje na Blogu, objavljanje konferenc in novic, e-novice.
- Navodila za dodelitev gesel na Nesstarju – kam vse je potrebno vnesti podatke o uporabniku. Poleg elektronske registracije, ki je bila uvedena nedavno, še vedno hranimo stvari v papirnati različici in v dodatni bazi, zlasti za namen letnega poročanja.
- Navodila za administrativna opravila.
- Shema datotek na Linux strežniku.
- Vzpostavitev slovenske različice Nesstarja.

Sam pristop k digitalni hrambi sistematično skozi korake, kakor so opredeljeni v modelih hrambe, omogoča tudi ugotavljanje pomanjkljivosti v dokumentaciji

in navodilih. Tako denimo v ESDS sproti dopolnjujejo seznam *Smernic za dobro prakso* (<http://www.esds.ac.uk/news/publications.asp>). V ICPSR dopolnjujejo vrsto s sistemom hrambe povezanih strateških dokumentov in vodičev. Celotna arhitektura dokumentov in vodičev ter njihovo periodično obnavljanje in osveževanje je eden od gradnikov zanesljivega organizacijskega delovanja.

Pregled navodil ADP je s tega vidika koristen, saj kaže na razvejanost notranjih postopkov. To je dobra podlaga za naslednji korak sistematičnega in konceptualno usklajenega popisa postopkov in ustreznih navodil in dokumentov. Del navodil je potrebnih revizije, v kolikor vsebujejo že opuščene postopke. Potrebne so uskladitve z novejšimi praktičnimi rešitvami. Pomembno je slediti v uvodu opredeljenim ciljem kakovostnega izvajanja storitev na posameznih področjih, skupaj s preverjanjem doseganja konkretnih razvojnih ciljev. Potrebno je uvesti tudi ustrezen sistem za vodenje dokumentacije in njeno verziranje ter periodično obnavljanje.

- *Ug. 11: Potreben je sistematičen pregled dokumentacije in priprava krajših dokumentov, ki podpirajo določene korake, vzdrževati indeks trenutno veljavnih dokumentov in obnavljati verzije dokumentov ter dopolniti manjkajoča področja.*

6 Zaključki

Značilnost zbirke ADP je, da je po količini gradiv (okrog 500 enot zbirk podatkovnih gradiv, povezanih z raziskavo) relativno majhna. Velik del obravnave gradiv je zelo specializiran in intenziven. Z dolgoletno prakso smo v ADP dosegli ustrezno strokovno raven obravnave, povezane z posebnostmi gradiva in izpolnjevanjem ponudbe gradiv uporabnikom, razvijamo in ohranjamo stike z dajalci in se povezujemo v mednarodna in nacionalna omrežja. Brez poznavanja posebnosti in pomembnih intelektualnih lastnosti gradiv digitalno skrbništvo ni mogoče, tako da je za nadaljnje delo potrebno to raven ohraniti in razvijati naprej. Tisto, kar je potrebno okrepiti, je povezano s splošnim poznavanjem področja digitalne hrambe. Tu se pojavlja potreba po človeku, ki bo strokovno spremljal to področje in povezoval aktivnosti hrambe v arhivu v vseh fazah. Prav tako so potrebne določene investicije v ustrezno opremo, oz. za najem zunanjih izvajalcev, kjer se za fizično hrambo ponuja rešitev najema enega od slovenskih komercialnih izvajalcev na ravni cele fakultete.

Kot kritično pomanjkljivost v tem trenutku vidimo avtomatizirano podporo dokumentacije informacijskih paketov v procesu obravnave. Namreč, tudi če bo za področje fizične hrambe (varnost, kopije, strojna oprema itd.) skrbel zunanji ponudnik ali če se za dolgoročno fizično shranjevanje digitalnih gradiv v Slove-

niji izoblikuje konzorcij, je še vedno potrebno za vsebino in upravljanje informacijskih paketov poskrbeti na mestih delnega skrbništva in zagotoviti integracijo vodenja procesov v takšnem porazdeljenem sistemu. Tako v ADP preučujemo možnost uporabe katerega od repozitorijskih rešitev, pri tem pa iščemo ustrezne prilagoditve za nadzorovanje sistema dolgotrajne hrambe. Ponuja se uporaba okolja Fedora (<http://fedoraproject.org/>) s prilagoditvami, kot so jih uporabili npr. pri UKDA-store (<http://store.data-archive.ac.uk/store/>), DANS Easy (<http://easy.dans.knaw.nl/dms>, primerjaj z ICPSR (2009) ali RODA (<http://roda.di.uminho.pt/>) in SVAROG (<https://www.nuk.uni-lj.si/svarog/>). Glede digitalne hrambe manjkajo nekatere stvari, ki so rešene v okviru tovrstnih repozitorijskih okolij, kot je avtomatski digitalni podpis vsaki stvari, ki pride v ADP. Morali bi imeti avtomatski registrator gradiv.

Prav tako nam med osnovnimi dokumenti manjka načrt v primeru nesreč. Tu bi bilo potrebno poiskati ustrezne splošne načrte na ravni organizacije in dodati ter upoštevati ustrezne vidike digitalne hrambe.

Pomemben vidik nadaljnega razvoja je tudi usposabljanje zaposlenih v ADP. Ti morajo poleg tekočega znanja angleščine (vse stvari se namreč pripravljajo dvojezično – ob tem da bi bilo potrebno zagotoviti lektoriranje angleških različic dokumentacije) obvladati še številne druge računalniške programe, zlasti statistične pakete, kot je SPSS (ki je v Sloveniji za zdaj še popularen) in XML urejevalnike besedil na precej višji ravni, kot jo nudijo slovenske fakultete v okviru obstoječih študijev. Kar dejansko pomeni, da moramo izobraziti vsakega novega zaposlenega. Prav v okviru bodoče CESSDA ERI je zelo izpostavljen ta vidik izmenjave znanj, seminarjev in usposabljanja za delo na skupnih orodjih.

Zaključimo lahko z ugotovitvijo, da je obvladovanje sistemov za dolgotrajno hrambo digitalnih gradiv zelo kompleksno opravilo, ki zahteva poleg poznavanja posebnosti in vsebine vrste gradiv, s katerimi razpolaga repozitorij, tudi strokovno vzpostavitev in upravljanje s sistemom kakovosti, nato načrtovanje in uvajanje različnih, na filozofiji načel dolgotrajne hrambe utemeljenih postopkov, njihovo dokumentacijo in spremljanje ter tehnološko in strojno podporo. Takšen sistem bo v perspektivi lažje razviti ob uporabi izhodišč, ki smo jih predstavili v pričujočem prispevku, zahtevalo pa bo tudi sodelovanje na različnih ravneh med organizacijami in strokovnjaki doma in v tujini.

Navedeni viri

1. Ball, A. (2010). *Review of the state of the art of the digital curation of research data: project report*. Bath: University of Bath. Pridobljeno 3. 8. 2010 s spletne strani:

- <http://opus.bath.ac.uk/19022/>
2. Beagrie, N., Chruszcz, J. in Lavoie, B. (2008a). *Keeping research data safe (Phase 1)*. Bristol: JISC. Pridobljeno 28. 8. 2010 s spletne strani: <http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx>
 3. Beagrie, N., Semple, N., Williams, P. in Wright, R. (2008b). *Digital preservation policy study*. Essex: JISC. Pridobljeno 3. 8. 2010 s spletne strani: <http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008digprespolicystudy.aspx>
 4. Beagrie, N., Lavoie, B. in Woollard, M. (2010). *Keeping research data safe (Phase 2)*. Bristol: JISC. Pridobljeno 28. 8. 2010 s spletne strani: <http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx>
 5. Beedham, H., Missen, J., Palmer, M. in Ruusalepp, R. (2005). *Assessment of UKDA and TNA compliance with OAIS and METS Standards*. Essex: UKDA. Richmond: TNA. Pridobljeno 3. 9. 2010 s spletne strani: <http://www.esds.ac.uk/news/publications/oaismets.pdf>
 6. *Data Seal of Approval guidelines*. (2009). [S. l.]: The Data Seal of Approval (DSA). Pridobljeno 4. 8. 2010 s spletne strani: <http://www.datasealofapproval.org/?q=node/35>
 7. *Digital preservation: glossary*. (2007). Ann Arbor: ICPSR, Institute for Social Research at the University of Michigan. Pridobljeno 10. 9. 2010 s spletne strani: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/preservation/glossary.js>
 8. *Digital preservation policies and planning at ICPSR*. (2009). Ann Arbor: ICPSR, Institute for Social Research at the University of Michigan. Pridobljeno 10. 9. 2010 s spletne strani: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/preservation/policies/>
 9. *Glossary of terms*. (2010) Edinburgh: Digital Curation Centre (DCC). Pridobljeno 19. 8. 2010 s spletne strani: <http://www.dcc.ac.uk/digital-curation/glossary-terms>
 10. Green, A., Macdonald, S. in Rice, R. (2009). *Policy-making for research data in repositories: a guide*. Edinburg: DISC-UK. Pridobljeno 3. 8. 2010 s spletne strani: <http://www.disc-uk.org/docs/guide.pdf>
 11. Gregory, A., Heus, P. in Ryssevik, J. (2009). *Metadata*. Berlin: German Council for Social and Economic Data. Working Paper No. 57. Pridobljeno 12. 6. 2010 s spletne strani: http://www.ratswd.de/download/workingpapers2009/57_09.pdf
 12. Guttman, M., Overcashier, P. in Whiteman, C. (2004). *Reinventing a data archive: process improvement at ICPSR*. Pridobljeno 8. 8. 2010 s spletne strani: <http://www.slideshare.net/colewh/reinventing-a-data-archive-process-improvement-at-icpsr>

13. Gutmann, M., Evans, B., Mitchell, D. in Schürer, K. (2009). *The alliance for data archive technologies: looking towards a common future*. Pridobljeno 1. 8. 2010 s spletne strani: http://www.fsd.uta.fi/iassist2009/presentations/F1_Gutmann.pps
14. ICPSR. (2009). *Principles and good practice for preserving data*. Washington: International Household Survey Network. Pridobljeno 10. 9. 2010 s spletne strani: <http://ihsn.org/home/download.php?file=IHSN-WP003.pdf>
15. Kavčič - Čolić, A. (2004). Metapodatki za trajno ohranjanje elektronskih virov. *Knjižnica*, 48 (3), 97–119. Pridobljeno 18. 2. 2011 s spletne strani: <http://revija-knjiznica.zbds-zveza.si/Izvodi/K0404/kavcic-colic1.pdf>
16. Kavčič - Čolić, A. (2008). *Načrt procesov digitalnega repozitorija za trajno ohranjanje v Narodni in univerzitetni knjižnici*. Ljubljana: NUK. Pridobljeno 23. 8. 2010 s spletne strani: <http://www.dlib.si/todv/hranjenje/file/pdf/NacrtProcesovRepozitorijaNUK.pdf>
17. Kavčič - Čolić, A. (2010). Trajno ohranjanje digitalnih virov: koncepti in metode. *Knjižnica*, 54 (1–2), 99–119. Pridobljeno 18. 2. 2010 s spletne strani: <http://revija-knjiznica.zbds-zveza.si/Izvodi/K1012/Kavcic-Colic.pdf>
18. Krejci, J. in Cizek, T. (2009). *Current training practices and training opportunities for the CESSDA [Report] (v.1.0.)*. Essex: CESSDA PPP. Pridobljeno 23. 8. 2010 s spletne strani: <http://www.cessda.org/project/>
19. Laaksonen, H., Borg, S. in Štebe, J. (2006). Setting up aquisition policies for a new data archive. *IASSIST Quarterly*, 30 (1), 5–12. Pridobljeno 3. 9. 2010 s spletne strani: http://www.iassistdata.org/downloads/iqvol301laaksonen_0.pdf
20. Lyon, L. (2007). *Dealing with data: roles, rights, responsibilities and relationships*. Bath: UKOLN. Pridobljeno 3. 9. 2010 s spletne strani: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf
21. McGovern, N. Y. (2007). *ICPSR Digital preservation policy framework*. Ann Arbor: ICPSR. Pridobljeno 20. 8. 2010 s spletne strani: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/preservation/policies/dpp-framework.jsp>
22. *OECD Principles and guidelines for access to research data from public funding*. (2007). Pridobljeno 4. 8. 2010 s spletne strani: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
23. Resolucija o Nacionalnem raziskovalnem in razvojnem programu za obdobje 2006–2010 (ReNRRP). (2006). *Uradni list RS*, št. 3.
24. Steinhart, G., Dietrich, D. in Green, A. (2009). Establishing trust in a chain of preservation: the TRAC checklist applied to a data staging repository (DataStaR). *D-Lib Magazine*, 15 (9–10). Pridobljeno 3. 8. 2010 s spletne strani: <http://www.dlib.org/dlib/september09/steinhart/09steinhart.html>

25. Swan, A. in Brown, S. (2008). *The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs*. Bristol: JISC. Pridobljeno 20. 8. 2010 s spletne strani: <http://www.jisc.ac.uk/publications/publications/dataskillscareersfinalreport.aspx>
26. Tjalsma, H., Gilissen, V., Koning, H. in Roorda, D. (2009). *Data preservation in the social sciences: recommendations for a CESSDA research infrastructure*. Pridobljeno 3. 9. 2010 s spletne strani: http://www.cessda.org/project/doc/D10.4_Data_Formats.pdf
27. *Trustworthy repositories audit & certification (TRAC)*. (2007). Chicago: CLR; OCLC, 2007. Pridobljeno 23. 8. 2010 s spletne strani: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
28. *UKDA awarded Place of Deposit status from TNA*. (2005, 5. april). Pridobljeno 10. 8. 2010 s spletne strani: <http://www.data-archive.ac.uk/news-events/news.aspx?id=1377>
29. Vardigan, M. in Whiteman, C. (2007). ICPSR meets OAIS: applying the OAIS reference model to the social science archive context. *Archival Science*, 7 (1), 73–87. Pridobljeno 8. 9. 2010 s spletne strani: <http://www.springerlink.com/content/50746212r6g21326/fulltext.pdf>
30. Woollard, M. (2008). *UK Data Archive preservation policy, Ver. 3.08*. Essex: UK Data Archive. Pridobljeno 3. 2. 2010 s spletne strani: <http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf>
31. Woollard, M. (2010). *UK Data Archive preservation policy*. Essex: UK Data Archive. Pridobljeno 12. 7. 2010 s spletne strani: <http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf>

Dr. Janez Štebe je zaposlen v Arhivu družboslovnih podatkov, Fakulteta za družbene vede Univerze v Ljubljani.

Naslov: Kardeljeva ploščad 5, 1000 Ljubljana

Naslov elektronske pošte: janez.stebe@fdv.uni-lj.si

Irena Vipavc Brvar je zaposlena v Arhivu družboslovnih podatkov, Fakulteta za družbene vede Univerze v Ljubljani.

Naslov: Kardeljeva ploščad 5, 1000 Ljubljana

Naslov elektronske pošte: irena.vipavc@fdv.uni-lj.si