

Univerza v Ljubljani  
Fakulteta *za družbene vede*



# OPIS STANJA NA PODROČJU RAZISKOVALNIH PODATKOV V SLOVENIJI

---

*POROČILO PROJEKTA ODPRTI PODATKI – PRIPRAVA AKCIJSKEGA NAČRTA ZA  
VZPOSTAVITEV SISTEMA ODPRTEGA DOSTOPA DO PODATKOV IZ JAVNO FINANCIRANIH  
RAZISKAV V SLOVENIJI*

Janez Štebe

Sonja Bezjak

Sanja Lužar

Irena Vipavc Brvar

ARHIV DRUŽBOSLOVNIH PODATKOV, UNIVERZA V LJUBLJANI

LJUBLJANA, 2012





Janez Štebe, Sonja Bezjak, Sanja Lužar, Irena Vipavc Brvar

## Opis stanja na področju raziskovalnih podatkov v Sloveniji

Poročilo projekta Odprti podatki – priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji

Izdajatelj in založnik: Arhiv družboslovnih podatkov, Fakulteta za družbene vede, Univerza v Ljubljani



To delo je ponujeno pod [Creative Commons Priznanje avtorstva 2.5 Slovenija](https://creativecommons.org/licenses/by/2.5/si/)

---

Projekt	ODPRTI PODATKI - Priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji  Raziskovalni projekt št. V5-1018 v okviru Ciljnega raziskovalnega programa »KONKURENČNOST SLOVENIJE 2006-2013«
Vodja projekta	Janez Štebe, predstojnik ADP
Vsebinski spremljevalki	Petra Tramte, Direktorat za visoko šolstvo in znanost, MIZKŠ,  Liljana Lučič, ARRS
Financer	MIZKŠ, ARRS

---

CIP - Kataložni zapis o publikaciji  
Narodna in univerzitetna knjižnica, Ljubljana

001.891:004.6

OPIS stanja na področju raziskovalnih podatkov v Sloveniji [Elektronski vir] : poročilo projekta Odprti podatki - priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji / Janez Štebe ... [et al.]. - El. knjiga. - Ljubljana : Arhiv družboslovnih podatkov, Fakulteta za družbene vede, 2012

Način dostopa (URL): [http://adp.fdv.uni-lj.si/o\\_arhivu/publikacije/odpp10\\_opis\\_stanja/](http://adp.fdv.uni-lj.si/o_arhivu/publikacije/odpp10_opis_stanja/)

ISBN 978-961-235-611-8 (pdf)

1. Štebe, Janez

263822848

---

# POVZETEK

---

Življenjski krog podatkov obsega široko področje ravnanja in upravljanja s podatki, od ustvarjanja podatkov, pregledovanja, predelovanja, izbora, prve analize, evidentiranja, hrambe do dostopa in druge uporabe. Različne faze opravljajo akterji, ti imajo vsak svoje cilje, potrebe, navade, ki si med sabo lahko tudi nasprotujejo. Hkrati pa prihajajo iz raznolikih raziskovalnih in znanstvenih področij. Raziskovalci, ki so največkrat omejeni na izkušnje lastnega raziskovalnega področja, pogosto ne vedo, kako širše misliti problematiko. In vendar so razprave o kakovostni hrambi in učinkoviti rabi podatkov v strokovni javnosti že dovolj dolgo in dovolj močno prisotne, da so vprašanja, dileme in odgovori nanje dokaj jasni. V okoljih z razvito raziskovalno infrastrukturo se pojavljajo domišljene in učinkovite rešitve, ki prispevajo k ustvarjanju kakovostnih podatkov ter k učinkoviti izrabi potenciala, ki ga ti nosijo za znanstveno skupnost. Na osnovi živahnih in plodnih razprav takšne izkušnje že dobivajo institucionalni okvir, katerega namen je, da se dobre prakse širijo tudi v okolja, v katerih so problemi manj opredeljeni in dobre prakse precej tuje. In sicer s ciljem, da bo za ustvarjanje podatkov, za njihovo hrambo in dostop poskrbljeno tako, da bo njihov potencial najbolje izkoriščen, s tem pa bodo tudi upravičena vložena sredstva.

S članstvom v mednarodni organizaciji OECD se je Slovenija zavezala, da bo sledila njenim smernicam na področju dostopa do raziskovalnih podatkov, financiranih iz javnih sredstev. Da bi preverili, kakšno je v Sloveniji stanje na področju raziskovalnih podatkov in oblikovali predloge za bodoče načrte, ki bodo utemeljeni na realnih možnostih, je Ministrstvo za izobraževanje, znanost, kulturo in šport v sodelovanju z ARRS na razpisu za predlagani Ciljni raziskovalni projekt izbralo Arhiv družboslovnih podatkov, Univerza v Ljubljani. Pričujoče poročilo je rezultat prvega dela raziskovalnega projekta, se pravi analiza stanja in možnosti v državi. Večji in temeljni del poročila je osnovan na intervjujih s slovenskimi raziskovalci in drugimi, ki imajo pri svojem delu tako ali drugače opraviti z raziskovalnimi podatki. Zanimalo nas je, kdo so že sedaj nosilci dejavnosti in posameznih nalog in kje so potenciali za v bodoče. V nadaljevanju projekta bomo podrobneje preučili tudi mednarodne pobude, ki se ukvarjajo z raziskovalnimi podatki ter dobre prakse in izkušnje drugih držav oziroma posameznih ustanov. Določena izhodišča, ki se naslanjajo na mednarodno sceno, pa so za primerjavo kot možne rešitve nakazana tudi v pričujočem pregledu, ki je sicer bolj usmerjen na zaznavo problemov doma. Poleg tega da smo ugotavljali, kakšno je stanje v Sloveniji, je bil eden od temeljnih ciljev prve faze projekta *Odprti podatki* poskus identifikacije nastavkov bodočega, bolj izpopolnjenega sistema dostopa do raziskovalnih podatkov.

Poročilo je sestavljeno iz osmih poglavij. Prva štiri poglavja osvetljujejo namen in cilje projekta *Odprti podatki* ter podlage, iz katerih ti cilji izhajajo. Na kratko so predstavljeni primeri referenčnih raziskav s Finske, Čila in Velike Britanije, ki odpirajo vpogled v študije potreb, navad in praks raziskovalcev, osvetljuje primere ovir na poti do odprtega dostopa in probleme premalo razvite raziskovalne infrastrukture.

Da bi spoznali in preučili slovenske razmere, smo iskali odgovore s prve roke, torej pri tistih akterjih, ki se s podatki ukvarjajo v svojem vsakodnevem raziskovalnem procesu, bodisi kot ustvarjalci, uporabniki ali skrbniki. Zanimalo nas je, s kakšnimi problemi se soočajo ustvarjalci podatkov in tisti, ki skrbijo za podatke, da bi ti bili širše dostopni. Pri iskanju odgovorov na zastavljena vprašanja, kjer gre za detekcijo problemov v kompleksnem okolju, je najustreznejša metoda zbiranja podatkov 'od spodaj navzgor'. Tako smo opravili 22 polstrukturiranih intervjujev. Pogovore z intervjuvanci z

različnih raziskovalnih področij in iz različnih delovnih okolij (raziskovalni inštituti, knjižnice, fakultete, državni uradi in drugo) smo vodili ob shemi, imenovani življenjski krog podatkov.

V treh poglavjih smo predstavili naslednje faze življenjskega kroga podatkov: ustvarjanje podatkov, dodana vrednost ter iskanje, dostop ter ponovna oz. druga raba podatkov. V intervjujih smo ugotavljali, kako raziskovalci načrtujejo ravnanje s podatki in kako poteka samo ustvarjanje podatkov. Pri tem nas je zanimalo, kakšni tipi podatkov nastajajo, kako je poskrbljeno za kakovost podatkov, ali so oblikovani postopki in merila za doseganje kakovosti. Kako poskrbijo za čiščenje in preoblikovanje podatkov ter kako za ustvarjanje spremne dokumentacije in metapodatkov. Skušali smo izvedeti, kako pomembna je faza digitalne hrambe in skrbništva, kakšne prakse in tehnične rešitve poznajo intervjuvanci. Posebno pozornost smo namenili zagotavljanju možnosti za drugo rabo, pri čemer sodelujejo tako ustvarjalci podatkov kot tudi ustvarjalci storitev, povezanih z življenjskim krogom podatkov, na primer storitev hrambe in organiziranja dostopa do podatkov za širši krog uporabnikov. Zanimalo nas je, kako je poskrbljeno za dostop do podatkov, kakšne ovire in zadržki stojijo na poti, ko raziskovalci skušajo priti do podatkov drugih. Kakšne prakse obstajajo v Sloveniji in kakšne so izkušnje tistih, ki do podatkov dostopajo v mednarodnih zbirkah.

V zadnjem poglavju smo se posvetili odgovornosti v zvezi z reševanjem problemov na področju odprtih podatkov. Skozi pogovore smo identificirali, koliko so raziskovalci sploh osveščeni glede obravnavane problematike, zanimala nas je prevladujoča kultura, prakse, sodelovanje s tujino in pripravljenost na spremembe. Od intervjuvancev smo skušali izvedeti, kje vidijo možnosti za reševanje problemov, kdo bi lahko prispeval k rešitvam, kako bi moral biti urejen sistem, da bi lahko delovali čim bolj kakovostno in vzdržno? Kateri ukrepi (finančni, v obliki politik, strokovne rešitve in podpora) bi spodbudili odpiranje raziskovalnih podatkov?

Da bi ostali čim bolj v stiku s konkretnimi izkušnjami, smo v poročilo vključili dobesedne izjave intervjuvancev. Izjave smo povzeli in jih smiselno vključili v besedilo, v dobesednem navedku pa jih navajamo na koncu vsakega poglavja. Za boljšo preglednost poročila smo zaznane probleme predstavili v okvirjih ob koncu poglavij in na istem mestu nakazali smer reševanja.

Poročilo *Opis stanja na področju raziskovalnih podatkov v Sloveniji* odgovarja na vprašanje, kakšna je problematika raziskovalnih podatkov v Sloveniji leta 2012. Je poskus razumeti situacijo, v kateri se nahajajo slovenski znanstveniki in raziskovalci, v kakšnih strukturnih pogojih delujejo, koliko so seznanjeni s smernicami doma in v svetu, kako široko ali ozko je njihovo poznavanje problematike in kakšna je pripravljenost na sprejemanje rešitev. Opis dejanskega stanja, v danem času in prostoru, je lahko podlaga za oblikovanje bodoče ureditve na področju ravnanja z raziskovalnimi podatki, financiranimi iz javnih sredstev.

Razvoj računalniških tehnologij že danes prinaša nove možnosti, ki kličejo k povezovanju, k deljenju in drugi rabi podatkov in drugih znanstvenih artefaktov znotraj pristopa 'odprte znanosti'. Vendar pa same tehnološke rešitve še niso dovolj, saj moramo najprej doseči strokovni, politični in tudi družbeni dogovor o tem, kaj je vredno hraniti in deliti z drugimi ter kako to urediti, da bodo akterji ustrezno motivirani, nagrajeni, zaščiteni.

# KAZALO

---

1. ŠIRŠI NAMEN IN CILJI PROJEKTA ODPRTI PODATKI .....	1
1.1 VSEBINSKI CILJI PROJEKTA .....	2
1.2 POMEN DEKLARACIJE IN NAČEL OECD .....	4
1.3 PODLAGA ZA IMPLEMENTACIJO NAČEL OECD V SLOVENIJI .....	5
2. KONCEPTUALNI OKVIR ZA PRIKAZ STANJA NA PODROČJU RAZISKOVALNIH PODATKOV V SLOVENIJI	9
2.1 ŠTUDIJE POTREB RAZISKOVALCEV .....	10
2.2 OVIRE NA POTI DO ODPRTEGA DOSTOPA (primer s Finske) .....	13
2.3 PREMALO RAZVITA RAZISKOVALNA INFRASTRUKTURA (primer iz Čila) .....	15
2.4 NAVADE IN PRAKSE RAZISKOVALCEV (primer iz Velike Britanije) .....	16
2.5 POMEN IZOBRAŽEVANJA IN USPOSABLJANJA ZA RAVNANJE S PODATKI .....	17
3. RAZISKOVALNA METODA .....	20
3.3 OSNOVNI PODATKI O VZORCU .....	22
3.4 VPRAŠALNIK .....	24
4. ŽIVLJENJSKI KROG PODATKOV .....	26
4.1 FAZE ŽIVLJENJSKEGA KROGA PODATKOV .....	27
5. PRIKAZ REZULTATOV V FAZI USTVARJANJA PODATKOV .....	29
5.1 TIPI PODATKOV .....	29
5.2 VPRAŠANJA KAKOVOSTI PODATKOV V POGLEDIH INTERVJUJANCEV .....	34
5.3 POMEN KAKOVOSTNIH PODATKOV IN METAPODATKOV .....	35
5.4 POSTOPKI IN MERILA ZA ZAGOTAVLJANJE KAKOVOSTI V FAZI USTVARJANJA .....	36
5.5 IZBOR PODATKOV NA PODLAGI KAKOVOSTI .....	40
5.6 SELEKCIJA PODATKOV V DOSTOPU S SPREMLJAJOČO OBJAVO V ČLANKU ALI MONOGRAFIJI ...	42
5.7 CITATI .....	43
6. FAZA: DODANA VREDNOST .....	50
6.1 ČIŠČENJE, TRANSFORMACIJA PODATKOV IN METAPODATKI .....	52
6.2 SPRAVILO .....	56
6.3 CITATI .....	59
7. FAZA: ISKANJE, DOSTOP (OBJAVA), DRUGA UPORABA .....	66
7.1 DOSTOP DO PODATKOV .....	66
7.2 ZADRŽKI IN OVIRE POVEZANE Z DRUGIMI FAZAMI ŽIVLJENJSKEGA KROGA PODATKOV .....	68

7.3 VREDNOTENJE PODATKOV .....	75
7.5 CITATI.....	80
8. ODGOVORNOST V ZVEZI Z REŠEVANJEM PROBLEMOV .....	91
8.1 POTREBA PO KROVNI POLITIKI .....	91
8.2 USTREZNO VREDNOTENJE PODATKOVNIH VIROV .....	91
8.3 FINANČNA SREDSTVA.....	93
8.4 DOBER SISTEM IZVEDBE: HRAMBE IN DOSTOPA DO PODATKOV .....	93
8.5 CITATI.....	96
ZAKLJUČKI NACIONALNE ANALIZE POLOŽAJA ODPRTIH ZNANSTVENIH PODATKOV .....	103
LITERATURA.....	107
PRILOGE.....	113





# 1. ŠIRŠI NAMEN IN CILJI PROJEKTA ODPRTI PODATKI

---

Hitro naraščanje obsega raziskovalnih podatkov in razvoj tehnologij, ki omogočajo lahek dostop preko interneta, zahtevajo spremembo krovnih politik glede pravic in obveznosti na področju podatkovne izmenjave v znanstveni skupnosti. Ena od 'gurujk' odprte znanosti Christine Borgman (2010) ugotavlja, da je prav v zadnjem času zanimanje za problematiko podatkov izjemno naraslo. Imamo celo vrsto aktualnih poljudnih člankov, ki predstavljajo izzive in možnosti podatkovne izmenjave. V *Nature* so leta 2008 posebno številko namenili problematiki 'velikih podatkov'. Zanimalo jih je, kaj za sodobno znanost pomenijo velike in kompleksne podatkovne strukture ter kako se bodo z novimi izzivi spopadle inštitucije in obstoječa kultura na področju znanosti, saj so še zmeraj ukoreninjene v pred-elektronski dobi.<sup>i</sup> Leto kasneje so v članku *Data's shameful neglect*<sup>ii</sup> obravnavali pomen hrambe in dostopa do podatkov, ki sta predpogoj napredka v znanosti, skrb za to pa naložili financerjem, pa tudi raziskovalcem, ki naj skupaj poskrbijo za strojno in programsko opremo, in za ustanove, ki bodo zagotavljale infrastrukturne storitve za hrambo in dostop do podatkov na način podatkovnih knjižnic in podatkovnih arhivov. Zelo jasno se zavzemajo za to, da mora postati upravljanje s podatki sestavni del vseh korakov v znanosti in eden od temeljev znanja. O podatkovni problematiki govorijo tudi bolj splošno usmerjeni časopisi in revije. Npr. ameriški *WIRED*<sup>iii</sup> preizkuša zamisli o spremembi paradigme pri pridobivanju znanja iz velike količine podatkov. *Economist*<sup>iv</sup> iz leta 2010 govori o vse večjem naraščanju količine podatkov, pa naj gre za podatke v velikih trgovskih verigah, knjižnicah ali na družabnih omrežjih, pri tem pa izpostavlja pomen poklicev bodočnosti, ki se bodo ukvarjali s podatki, kot so podatkovni znanstvenik, podatkovni znanstveni knjižničar ipd. Potrebe po podatkovnih veščinah in znanjih niso več stvar prihodnosti in samo specialistov, pač pa so že del vsakdana znanstvenikov, čeprav bo ponekod sprememba v spremljajoči miselnosti vidna šele z menjavo generacij. Naraščanje podatkov vodi v t. i. 'četrti paradigmo' podatkovno usmerjenih znanosti (Hey in drugi 2009) in botruje nastanku novih področij, kot so 'računalniška biologija' ali pa 'digitalna humanistika' (Borgman 2010).

Cilj projekta *Odprti podatki* je izhajajoč iz predstavljenega jasen: pomagati pri oblikovanju rešitev za čim boljši izkoristek potenciala, ki ga nosijo raziskovalni podatki, nastali z javnimi sredstvi. V bolj odprtem in čim širšem dostopu do podatkov so posredne koristi tako za tiste, ki sodelujejo v znanstvenih projektih, ki ustvarjajo ali uporabljajo podatke, kot tudi za širšo javnost. Veliko se govori o potencialu odprte znanosti, tudi ljudske oz. državljanske znanosti, kjer vsakdo lahko črpa iz zakladnice znanja in prispeva vanjo. Prav nekatere inovacije na področju interneta, npr. sodelovanje skupnosti, so prinesle nove obete za večjo vključenost vseh zainteresiranih pri znanstvenih projektih. Odprta znanost, tudi v povezavi z načeli OECD,<sup>v</sup> je blagovna znamka za različne skupinske pobude in gibanja, ki nasprotujejo vsakovrstnim 'nepravičnim' omejitvam dostopa do znanstvenih objav, podatkov in orodij. V okviru pričujočega projekta želimo ugotoviti, kaj imamo v zametkih v Sloveniji na področju odprtega dostopa v znanosti. Zanima nas, kakšne so sedanje možnosti in potencial za prihodnje glede dostopa do raziskovalnih podatkov za namen nadaljnje znanstvene obravnave. In sicer

---

<sup>i</sup> <http://www.nature.com/nature/journal/v455/n7209/full/455001a.html>, 7. 8. 2012

<sup>ii</sup> <http://www.nature.com/nature/journal/v461/n7261/pdf/461145a.pdf>, 7. 8. 2012

<sup>iii</sup> [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory), 7. 8. 2012

<sup>iv</sup> <http://www.economist.com/node/15557443/print>, 7. 8. 2012

<sup>v</sup> Principles and Guidelines for Access to Research Data from Public Funding (2007). Dokument bomo v poročilu na kratko imenovali Načela OECD ali samo Načela.

je cilj tudi, da začnemo vzpostavljati interesne povezave in da pobudam pomagamo odpirati prostor za uveljavitev. Prepričanje o koristih odprtega dostopa se namreč vse bolj uveljavlja, tako da je pravi čas, da se projektom, povezanim z njimi, omogoči finančne, kadrovske in organizacijske pogoje za resno delovanje.

Osnovna načela so bolj kot ne preprosta. Vsakdo se bo strinjal, da je odprt dostop do podatkov zagotovilo bolj tehtnih zaključkov. Lahko preverimo ugotovitve drug drugega, z večanjem dostopa se poveča množina odkritij, kadar zaradi kompleksnosti podatkov spoznavni potencial ni izčrpan z zaporednimi analizami, zlasti pa je kolektivno delo na podatkih odlična spodbuda za dialog, medsebojno kritiko in dopolnjevanje analiz sorodnih problemov. To pa je - Popperjansko vzeto - zagotovilo napredovanja spoznanj. Na splošno lahko rečemo, da se poveča tako kakovost kot količina spoznanj za isto ceno ( BRTF-SDPA 2010). Prednost odprtega dostopa do podatkov je tudi v tem, da se omogočijo povezave sicer ločenih podatkov, ter povezave podatkov in znanstvenih objav v obliki besedil, tovrstna dodana vrednost pa zopet bogati njihov analitični potencial. Pogoj je, da so podatki za nadaljnjo uporabo ustrezno pripravljene in primerno opisane. Ustrezati morajo principom objavljanja podatkov, oblikovanim po analogiji z objavami člankov. Objava podatkov (v pooblaščenem podatkovnem arhivu ali v podatkovni reviji) namreč zahteva ustrezno dokazilo o kakovosti in dokumentacijo o nastanku in vsebini, ki je potrebna za njihovo smiselno uporabo.

Glede na to samoumevnost o koristih je presenetljivo, zakaj podatki niso javni, še posebej če so nastali s podporo javnih sredstev. Ena pomembnih ovir so vsekakor dodatni napori in čas ter znanje, ki je potrebno za pripravo podatkov in njihovo objavo. Potrebna je podporna infrastruktura. Lahko so to arhivi, specializirani po posameznih disciplinah, nacionalni in povezani mednarodno, lahko so splošni repozitoriji posameznih ustanov, ali začasni podatkovni arhivi posameznih projektov ali posameznikov. Vsaka od teh oblik ima svoje prednosti in slabosti. Potrebne so tudi spodbude, tako finančne kot moralne, ter v politikah zapisane obveznosti za raziskovalce, da bodo pripravljene izročati podatke. Velikokrat predstavlja oviro odsotnost kulture izmenjave podatkov.

Če se zdaj vrnemo k izhodiščnemu vprašanju. Kje v Sloveniji imamo vsaj v zametkih dostop do izvornih podatkov, ki bi se jih dalo uporabljati za nadaljnje analize? Medicina, naravoslovje, biologija, okolje, družboslovje in humanistika? Če da, kako odprt je dostop? Če ne, zakaj ne? In kaj lahko naredimo? V pričujočem projektu CRP oblikovanja akcijskega načrta dostopa do raziskovalnih podatkov iz raziskav, ki se financirajo iz javnih virov, bomo razčistili prav ta vprašanja.

### *1.1 VSEBINSKI CILJI PROJEKTA*

Če povzamemo, so vsebinski cilji projekta Odprti podatki sledeči. Najprej, ugotoviti, kakšno je trenutno stanje na področju ravnanja z raziskovalnimi podatki v Sloveniji. Preveriti, kdo so nosilci dejavnosti in posameznih nalog, ter identificirati nastavke bodočega, bolj izpopolnjenega sistema dostopa do raziskovalnih podatkov. Ugotavljali bomo, kateri so glavni problemi, s katerimi se soočajo ustvarjalci podatkov in tisti, ki dolgoročno skrbijo za podatke, poskušali identificirati vrste podatkov in kaj se s podatki dogaja od nastanka pa do morebitne druge rabe,<sup>vi</sup> kakšne so povezave in medsebojna izmenjava podatkov in storitev doma ter v tujini. Zlasti nas bo zanimalo, kakšna so medsebojna pričakovanja, kako posamezni nosilci vlog razumejo svoje obveznosti do financerjev in medsebojne odvisnosti, kadar nastopajo, včasih tudi v isti osebi kot ustvarjalci, izvajalci storitev povezanih s podatki ter njihovi uporabniki.

---

<sup>vi</sup> V tem poročilu uporabljamo pojem 'druga raba', ki je različica termina 'sekundarna' ali 'ponovna' uporaba, analiza podatkov.

Drugi, vzporedni cilj projekta, je mobilizacija vpletene in zainteresirane javnosti. Odprti dostop je področje, ki predstavlja inovacijo. Da bi sprejeli novost, bo potrebno spremeniti način razmišljanja in delovanja. V okviru projekta si bomo prizadevali za vključevanje tematik, ki zadevajo potencial bodoče ureditve: to je bolj transparenten in široko zajemajoč sistem dostopa do raziskovalnih podatkov, z namenom promocije načel odprtega dostopa. Že danes lahko identificiramo nekatere šampione, ki vlečejo s svojimi zgledi. Te posameznike in skupine bomo še posebej izpostavili in izkoristili njihovo pripravljenost in angažiranost za uveljavljanje novosti na področju organiziranja podatkov in njihovega bolj učinkovitega izkoriščanja. S tem želimo vzbuditi ambicije, pripravljenost za delo in sodelovanje tudi pri tistih, pretežno ustvarjalcih podatkov, ki se do sedaj še niso odločili vstopiti na področje opravljanja storitev in svetovanja glede ravnanja s podatki, za kar je potrebno dodatno usposabljanje. Glede na trenutni ugled poklica 'podatkovnega znanstvenika', 'podatkovnega arhivista' ali 'podatkovnega knjižničarja' ter težavne karijerne odločitve (Swan in Brown 2008b) bodo tudi tukaj potrebne spodbude.

Tretji cilj projekta, to je načrtovanje, bomo uresničili v zadnji fazi projekta, v kateri bomo upoštevali in uporabili obstoječe nastavke dejavnosti ter povezave in zglede iz tujine. Zadnja faza projekta bo obsegala poskus doseganja soglasja med nosilci vlog in včasih tudi nasprotujočih interesov glede strateških ciljev in njihove implementacije. Podali bomo strokovna izhodišča za ocenjevanje stroškov, kadrovskega potreb in dodatnega dela, ter izhodišča za zagotavljanje nagrad in motivacije, predloge za oblikovanje politik in upoštevanje strokovnih standardov. Vse to pa bo prispevalo k zagotavljanju realističnega načrtovanja in optimalnega izkoriščanja kadrovskega, finančnega in podatkovnega virov.

Potek projekta sledi zastavljenim ciljem, pri čemer spodbujamo uporabo družabnih omrežij, katerih se tudi sami poslužujemo. Že v samem začetku smo ustvarili *blog* o projektu, ki je podlaga za medsebojno informiranje in morebitno sodelovanje v razpravah, z ambicijo, da začnemo graditi skupnost in povezovati zainteresirano javnost in angažirane posameznike. Blog je odprt za gostujoče prispevke! Prva faza izvedbe projekta, ki je predstavljena v tem poročilu, sledi ciljem ugotavljanje stanja: kdo in kaj so nosilci dejavnosti oziroma njenih zametkov, kakšne so vloge, medsebojna pričakovanja in kako so opredeljene obveznosti. Gre za relacije med ustvarjalci podatkov in ustvarjalci raziskovalnih politik, financerji ter tistimi, ki že opravljajo ali bi lahko opravljali storitve upravljanja, dolgotrajnega shranjevanja in dostopa do podatkov, ter seveda upošteva potrebe, probleme in interese sedanjih ali bodočih uporabnikov podatkov in z njimi povezanih storitev.

V drugi fazi, katere povzetek bomo predstavili v enem od naslednjih poročil, bomo podali pregled dobrih praks glede dostopa do raziskovalnih podatkov ter številnih pobud in načrtov za reševanje problemov v mednarodnem okolju.

V tretji fazi projekta bomo predstavili sintezo ugotovitev glede stanja doma povezav in študije ter študij dobrih praks in modelov v tujini, iz česar bomo izpeljali izhodiščne predloge za reševanje zaznanih problemov iz prve in druge faze projekta. Temu bo sledila organizirana razprava o predlogih kot preizkus sprejemljivosti za različne javnosti in nosilce vlog, tako na različnih raziskovalnih področjih kot tudi v različnih fazah življenjskega kroga podatkov.<sup>vii</sup> Cilj bo doseganje strokovnega in političnega soglasja glede predlaganih rešitev, čemer bo sledilo končno poročilo o izsledkih projekta in predlog za akcijski načrt.

Pri uresničevanju zastavljenega se bomo posluževali različnih metod za vključevanje ciljnih skupin in za zbiranje informacij, kot so intervjuji in posebej pripravljena poročila o delu posameznih skupin in projektov, kombinirali pa bomo tudi prispevke in komentarje na družabnih omrežjih in organizirali

---

<sup>vii</sup> Več o tem glej 4. Življenjski krog podatkov.

posamezna posvetovanja na način fokusnih skupin po izbranih tematikah. Ob tem si bomo prizadevali za intenzivnejše vključevanje v mednarodna omrežja, situacijo in načrte v mednarodnem okolju pa bomo spremljali s študijem literature, zlasti poročil o tekočih projektih in predstavitev z aktualnih konferenc.

## 1.2 POMEN DEKLARACIJE IN NAČEL OECD

Pri vlaganju v razvoj sistema odprtega dostopa do podatkov je potrebno pazljivo načrtovanje na nivoju strateških odločitev nacionalne porabe sredstev za dodatno delo, ki je potrebno. Načela OECD, ki so podlaga pričujočega projekta, pokrivajo vse najpomembnejše vidike načrtovanja: tehnološke, institucionalne, upravljalne, finančne, proračunske, pravne, vidike nacionalnih politik ter vidik kulture, navad in praks in kot taka predstavljajo odlično izhodišče za razvoj. Ob tem ko poudarjajo celostno zasnovo in obravnavo vseh omenjenih vidikov ter nujnost prilagajanja nacionalnim razmeram, pa sama Načela niso tako podrobno razdelana, da bi nudila konkretne možnosti uresničitve zastavljenih ciljev. Slednje je naloga pričujočega projekta. Zagotavljanje uspeha akcijskega načrta je dolgoročno povezano s tem, kako izhodišča razumejo in sprejemajo različni nosilci vlog v sistemu dostopa do podatkov. Zato si že v okviru izvedbe projekta prizadevamo pridobiti podporo, tako na strani nosilcev moči, ki lahko vplivajo na spremembe, kot na strani nosilcev storitev in njihovih uporabnikov, da spremembe posvojijo. Izhodišče in temeljni dokument za pripravo pregleda stanja v Sloveniji na področju dostopa do odprtih podatkov predstavljajo Načela OECD. Dokument izvira iz Deklaracije,<sup>viii</sup> v kateri so leta 2004 članice OECD pripoznale in potrdile potrebo po dostopu do raziskovalnih podatkov. Temeljni cilj je izboljšati učinkovitost in izkoristek nacionalnih in globalnega znanstvenega sistema. Boljši dostop do podatkov pomaga maksimizirati raziskovalni potencial novih digitalnih tehnologij in omrežij ter zagotavlja večji izkoristek javnih sredstev vloženih v raziskovanje. Pomeni pa tudi potencialni vir znanja, ki je potrebno pri soočanju z izzivi, ki pretijo človeštvu, so zapisali v Načela. Dokument je zasnovan tako, da v temelju določa izhodišča in pogoje, s katerimi bi bilo mogoče zastavljene cilje uresničiti na čim bolj ekonomičen način in ne da bi ovirala razvoj s težavnimi obveznostmi in pravili ali nalagala nove stroške (OECD 2007, 13).

Načela so pomembna kot dokument, ki naj države članice OECD in druge vodi pri oblikovanju strategij, politik in načrtov odprtega dostopa do podatkov. V preambuli dokumenta je jasno opredeljeno, v čem so smisel, pomen in prednosti odprtega dostopa. Dejansko teže tega dokumenta predstavlja soglasje o pomenu cilja najširšega deljenja raziskovalnih podatkov, ki v obliki sprejetih načel države zunaj in znotraj OECD zavezuje, da po svojih zmožnostih težijo k njegovi uresnitvi. Pomen tega dokumenta se tako kaže v novo zasnovanih državnih ali disciplinarnih strategijah in politikah, ki se nanj sklicujejo. Primer implementacije načel OECD, ob eksplicitnem sklicevanju nanj, je denimo dokument Economic & Social Research Council (ESRC) v Veliki Britaniji. Slednji v politikah dostopa do podatkov zahteva, da je v raziskavah, financiranih z javnimi sredstvi, upravljanje s podatki načrtovano že v prijavi projekta, zahtevajo pa tudi dejansko upravljanje s podatki v času izvedbe raziskovalnih projektov, in sicer s ciljem, da bo ob koncu projekta zagotovljen dostop do podatkov (ESRC Research Data Policy, 2010).

Načela OECD predstavljajo pomemben temeljni dokument, ki pomaga državam z različno stopnjo razvitosti, da se začno zavedati problemov in iskati rešitve. Namenjena so financierjem in politikom, raziskovalcem, raziskovalnim ustanovam in državnim raziskovalnim agencijam, da bi razumeli pomen odprtega dostopa do raziskovalnih podatkov. Dokument ne ponuja receptov in univerzalnih rešitev, prinaša pa celosten pristop k reševanju 'problema' odprtih podatkov. Christine Borgman (2011, 2), ena vodilnih avtoric na tem področju, opozarja na kompleksnost vprašanj '*podatkov, raziskovalnih praks,*

---

<sup>viii</sup> Declaration on Access to Research Data from Public Funding (OECD 2004).

*inovacij, spodbud, ekonomike, intelektualne lastnine in javnih politik, ki so povezane z problemom ('conundrum', op.a.). deljenja podatkov – z zahtevnim problemom, ki postavlja izzive".*

S tem želimo reči, da se zavedamo težavnosti tega problema, da želimo pokazati na njegove razsežnosti in kompleksnost, preiskati nekatere ponujene rešitve kot zglede, da pa je na koncu iskanje in doseganje rešitev stvar akterjev, ki so vpleteni v problem, njihovega znanja in motiviranosti ter pogojev, ki omogočajo ravnanje, skladno z omenjenimi načeli.

### 1.3 PODLAGA ZA IMPLEMENTACIJO NAČEL OECD V SLOVENIJI

Zahteve projektne specifikacije iz razpisa se, kot omenjeno, sklicujejo na priporočilo Odbora OECD za znanstveno in tehnološko politiko iz leta 2006,<sup>ix</sup> ki je namenjen nacionalnim vladam. Problemski okvir projekta torej predstavlja dejstvo, da se je Slovenija z vstopom v OECD zavezala k izpolnjevanju določil omenjenih Načel. Pobuda je izjemno pomembna prav zato, ker za razliko od znanih pobud širjenja odprtega dostopa do znanstvenih publikacij<sup>x</sup> v ospredje postavlja raziskovalne podatke. Te dokaj široko opredeli kot primarne zapise iz znanstvenih opazovanj, ki v okviru znanstvene skupnosti predstavljajo ustrezno sredstvo za preverjanje veljavnosti raziskovalnih spoznanj. Poudarek je na primarnih, surovih podatkih, ki jih v različnih disciplinah najdemo v različnih oblikah in jih je mogoče uporabiti za nadaljnje analize. Odprtost pomeni enakopravnost vseh pripravnikov znanstvene skupnosti v dostopu do primarnih podatkov in enakopravnost v možnosti izkoriščanja le-teh za potrebe raziskovanj in usposabljanj.

Dogajanje na področju urejanja problematike dostopa do raziskovalnih podatkov je v zadnjem času zares intenzivno, saj poleg posameznih držav iniciativo prevzemajo tudi razni mednarodni forumi, vključno s *Komisijo EU*,<sup>xi</sup> kar bomo predstavili v ločenem poročilu o mednarodnem dogajanju.

Kot že rečeno, so Načela OECD lahko odlična strateška podlaga za pripravo nacionalnih načrtov, vendar je pri tem potrebno načela izpeljati tako, da bo upoštevana specifična situacija v Sloveniji. Pomembna je tudi prilagoditev posameznim raziskovalnim področjem, o čemer izrecno govori načelo o fleksibilnosti: *"Fleksibilnost zahteva, da upoštevamo hitre in pogosto nepredvidljive spremembe v informacijski tehnologiji, značilnosti posameznih raziskovalnih področij in raznolikost raziskovalnih sistemov: pravnih sistemov in kulture držav članic. Pri oblikovanju dogovorov o dostopu do podatkov, pri razvoju politik za spodbujanje odprtega dostopa in pri implementaciji teh Načel in vodil je treba upoštevati pomen nacionalnih, družbenih, ekonomskih in institucionalnih posebnosti."* (OECD 2007, 15).

Projekt *Odprti podatki - Priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji* smo zasnovali na predpostavki, da je odprt dostop do raziskovalnih podatkov nekaj dobrega. Že v izhodiščih smo se zavedali dejstva, da je vprašanje za veliko vpletenih nekaj novega in da bo treba poiskati področja, storitve in zglede, ki bodo tudi tiste z manj izkušnjami in znanjem spodbudili k uresničevanju odprtega dostopa.

---

<sup>ix</sup> Recommendation of the Council concerning Access to Research Data from Public Funding, 14 December 2006, C (2006)184.

<sup>x</sup> Npr. Berlinska pobuda, Budipeštanska deklaracija ...

<sup>xi</sup> Glej: <http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1301> in [http://cordis.europa.eu/fp7/ict/e-infrastructure/publications\\_en.html](http://cordis.europa.eu/fp7/ict/e-infrastructure/publications_en.html). Zelo znan je zlasti manifest za odprte podatke iz leta 2010 z naslovom: *Riding the wave - How Europe can gain from the rising tide of scientific data - Final report of the High Level Expert Group on Scientific Data*, 7. 8. 2012

Želeli bi zasnovati sistem, ki bo pravičen, učinkovit, ki ne bo povečeval bremen ustvarjalcev in bo izpostavil možnosti integracije storitev, pravil, tehnologij, orodij, hkrati pa ohranil posebnosti pristopov in narave podatkov ter upošteval potrebe znanstvene skupnosti. Kajti kot pravi večkrat citirana avtorica Margaret Henty (v Feijen 2011): *"S strani raziskovalcev se največkrat pojavi zahteva po storitvah, ki so na voljo, ko jih potrebujejo, a so oblikovane tako, da ne posegajo v ustvarjalno delo, ki je bistvo raziskovalnega procesa."*

Pomembno izhodišče za doseganje projektnih ciljev je podpora Ministrstva za izobraževanje, znanost, kulturo in šport, ki razume in podpira načela OECD. Predstavniki Ministrstva sodelujejo v aktivnostih projekta Odprti podatki, vključeni pa bodo zlasti v nadaljnjih fazah zagotavljanja pogojev za rešitve problemov, ki jih bomo identificirali v prvi fazi projekta, konkretno pri oblikovanju ustrezne politike zahtev glede ravnanja s podatki in ustreznega financiranja tako dodatnih naporov ustvarjalcev podatkov, kot storitvene podpore infrastrukturnih dejavnosti.

Ministrstvo je z razpisom projekta presodilo, da za pripravo strategije na tem področju potrebuje strokovna izhodišča. Za pripravo strokovnih izhodišč in osnutka akcijskega načrta so potrebne konkretne izkušnje, ki jih ima izvajalec projekta z izvajanjem storitev dostopa do raziskovalnih podatkov, potrebno je poglobljeno poznavanje situacije v drugih državah in vzpostavljen stik z delujočimi podatkovnimi infrastrukturami na določenem znanstvenem področju. Zlasti pa je potreben akcijsko usmerjen pregled stanja in možnosti, potencialov, ki temelji na vključevanju vseh relevantnih akterjev že v fazi priprave bodoče strategije. Ključni akterji so, prvič, pripadniki znanstvene skupnosti, ki jih je potrebno seznaniti z možnostmi odprtega dostopa do znanstvenih podatkov in prednostmi tega za izboljšanje znanstvenih odkritij, po drugi strani pa jih je potrebno seznaniti z nalogami, ki jih je potrebno opraviti, da bodo prednosti prišle do izraza. Pripadniki znanstvene skupnosti so po tem, ko se seznanijo z načeli odprtega dostopa, v večini primerov pripravljene ta načela podpreti in se dejavno vključiti v njihovo izpeljavo. Drugič, so tu akterji in organizacije, ki opravljajo storitve infrastrukturnih podatkovnih centrov – organiziranih disciplinarno ali v okviru ustanov. Ti zagotavljajo ustrezno podporo ustvarjalcem podatkov, ali s svetovanjem in usmerjanjem ter usposabljanjem za ravnanje s podatki v fazi ustvarjanja in priprave, ali pa z opravljanjem storitev po prevzemu podatkov v svoje skrbništvo. Tretji akter pa so izvajalci in oblikovalci znanstvenih politik, ki morajo zagotoviti ustrezen okvir predpisov in motivacije za izpeljavo nalog zagotavljanja odprtega dostopa, zlasti pa podati ustrezne finančne in organizacijske podlage za delo. Naloga pričujočega projekta je tako pripraviti predlog, ki bo vzdržal s strokovnega vidika, vključeval pa bo tudi zaveze in konsenz glede rešitev s strani omenjenih akterjev.

V okviru pričujočega projekta se bomo dotaknili vsakega od vidikov iz načel OECD, ter izpostavili pregled dobrih praks, izkušenj in zgledov. S tem bomo preizkusili pripravljenost našega sistema na izpolnjevanje teh načel, in sicer z namenom, da se prilagodijo realnosti trenutnega stanja glede finančnih, organizacijskih, strokovnih (znanja) in tehnoloških pogojev. Kot je že zahteva besedila razpisa, bo naloga projekta *"pregled situacije v Sloveniji in strokovna uskladitev med relevantnimi akterji v Sloveniji vključno z nosilci raziskovalne politike glede ključnih korakov v nadaljnjem razvoju (pregled nad zametki storitev odprtega dostopa do podatkov v Sloveniji, organiziranje razprave in povezava med strokovnjaki iz ključnih raziskovalnih okolij zaradi izmenjave informacij, znanj in tehnoloških rešitev za povezovanje storitev na nacionalnem nivoju in mednarodno)"*.

V nadaljevanju povzemamo Načela OECD, ki vsebujejo pomembna izhodišča za zastavitev te naloge:



- Odprtost dostopa do podatkov kot javne dobrine (Kaul, Grunberg in Stern 1999) zahteva posebno pozornost pri ravnanju s podatki že pri izdelavi podatkov, da se zagotovi ustrezna vsebinska obdelava, dokumentacija in priprava za nadaljnjo uporabo.
- Načelo fleksibilnosti (B) bo v fazah projekta priprave nacionalne strategije ključnega pomena. Kakor ugotavljajo tudi avtorji Načel (Arzberger in drugi 2004) je eno nerazdelnih vprašanj, kako uvajati rešitve iz razvitih držav v manj razvitih državah s specifičnimi institucionalnimi pogoji.
- Uvajanje združevalnih katalogov, skupnih metapodatkovnih standardov ali prehodov med njimi, ter intenzivna komunikacija med podatkovnimi arhivi in dajalci podatkov, na splošno opredelijo učinkovito zagotavljanje izpolnjevanja načela preglednosti (C).
- Pri načelih spoštovanja pravnih načel, intelektualne lastnine in formalnih zavez (D., E., F.) je potrebno zagotoviti možnosti za izrabo podatkov v znanstvene namene, s sklicevanjem na etično rabo skladno z etičnim kodeksom, denimo pri varovanju pravice do zasebnosti. Vsa ta načela je potrebno uporabljati na način, da ne predstavljajo nepotrebnih ovir pri dostopu do podatkov, temveč opredeljujejo ustrezne izjeme in poti, kako zagotoviti dostop ob njihovem spoštovanju.<sup>xii</sup>
- Formalne zaveze vključujejo tudi odgovornost za vzdrževanje infrastrukture za dostop do podatkov. Zaradi zagotavljanja pogojev varnega in dolgotrajnega shranjevanja in ohranjanja uporabnosti podatkov morajo te zaveze priti s strani vodstva raziskovalnih ustanov in vladnih organizacij, ki naj opredelijo ustrezne prioritete.
- V prehodu iz sedanjega stanja, to je naključna skrb in nedoločena odgovornost, bo potrebno v pooblaščenih digitalnih arhivih jasno opredeliti, kdo bo skrbel za spravilo in kdo bo prevzel nasledstvo skrbi za zagotavljanje dolgotrajnega dostopa do podatkov (M).
- Načelo profesionalizma (G) predpostavlja kulturo sodelovanja znotraj znanstvene skupnosti, ki se bo ob promociji načel odprtega dostopa lahko dodatno okrepila. Tako bo eden od stranskih, takoj oprijemljivih rezultatov projekta prav povečanje medsebojnega zaupanja in solidarnosti, ki jih predpostavlja odprt dostop do podatkov. Obenem je potrebno omogočiti usposabljanje in profesionalizacijo na področju upravljanja s podatki, kar je nujen korak k zagotavljanju ustrezne dokumentacije in arhiviranja podatkov.
- Tehnološka in semantična interoperabilnost (H) je ključnega pomena za zagotavljanje mednarodnega in interdisciplinarnega dostopa in uporabe raziskovalnih podatkov.
- Na vrednost in kvaliteto raziskovalnega dela vpliva tudi kvaliteta samih podatkov. V okviru izpolnjevanje načela kvalitete (I) je treba določiti primerne metode za zagotavljanje in ocenjevanje kakovosti, ki morajo biti prilagojene posameznim raziskovalnim področjem. Posebej pomemben vidik v kontekstu predlaganega projekta pa je tudi celovita kakovost storitev pridobivanja, shranjevanja in razširjanja podatkov, ki jo običajno presojava na podlagi skladnosti z OAI standardom (OAI 2012; primerjaj tudi Sesink in drugi 2008).
- Varnost (načelo J) podatkov je mogoče doseči s skrbjo za integriteto podatkov in s primernimi načini shranjevanja, zaščite in spreminjanja podatkov, kar vključuje univerzalno stalno identifikacijo podatkov in njihovih različic.
- Načelo učinkovitosti (K) zahteva izkoriščanje prednosti, ki jih prinaša odprti dostop do raziskovalnih podatkov. K izboljšanju lahko pomaga meddisciplinarna in mednarodna izmenjava izkušenj s področja upravljanja s podatki, obsega tudi strategije izbora podatkov z večjim raziskovalnim pomenom in posledično daljšo hrambo, razvoj specializiranih podpornih

---

<sup>xii</sup> <http://www.knowledge-exchange.info/Default.aspx?ID=461>,  
<http://www.surf.nl/en/publicaties/Pages/Juridischestatusvanruwedata.aspx>, 7. 8. 2012

storitev glede na specifične potrebe uporabnikov z različnih področij, uvajanje domiselnih metod, ki raziskovalce motivirajo, da prispevajo podatke v odprti dostop, itn.

- Načelo odgovornosti (L) in preverjanja izpolnjevanja nalog neposredno opredeljuje mehanizme in modele, ki zagotavljajo trajno učinkovitost in optimalno porabo sredstev, kot so delitev porabe v izdelavo novih in hrambo obstoječih podatkov, učinkovitost ustreznih storitev, spremljanje uporabe in registracija novih znanj (npr. v obliki znanstvenih publikacij) nastalih z uporabo obstoječih podatkov ter predvidevanja glede bodočih potreb.

Iz načel OECD posebej povzemamo eno od ugotovitev:

*" Ključna ovira pri doseganju maksimalne vrednosti na podlagi vlaganj v raziskovalne podatke je v posameznih vejah znanosti odsotnost načrtovanja in izvajanja ustreznih nalog, namenjenih dokumentaciji in arhiviranju podatkovnih datotek. Pri načrtovanju projektov in programov je potrebno na vseh ravneh neposredno izpostaviti problematiko podatkov v najzgodnejši fazi zato, da je mogoče že v pripravi projekta predvideti finančne in tehnične zahteve izhodiščne priprave in nadaljnega izvajanja skrbništva nad podatki. Poskrbeti je treba za podporo pobudam in za podporo strokovnemu usposabljanju na vseh področjih ravnanja s podatki" (OECD 2007, 19).*

Ob tem, ko je izpostavljena tudi potreba po usposabljanju za delo s podatki,<sup>xiii</sup> je poudarek na načrtovanju in zagotavljanju pogojev za izvajanje skrbništva nad podatki, in to na ravni konkretnih tekočih raziskovalnih projektov. To je smer, v katero gredo politike glede zahtev dostopa do raziskovalnih podatkov po svetu, očitno v soglasju z načeli OECD. Aktualna je pobuda financerjev znanosti v ZDA<sup>xiv</sup> in Veliki Britaniji, da v prijavih na raziskovalne projekte zahtevajo načrt upravljanja s podatki in dostop do podatkov (Jones 2009). Načrt je kot del prijave podvržen recenzijskemu postopku. Britanski *Digital Curation Centre* (med drugimi podobnimi) pripravlja pripomoček<sup>xv</sup> in navodila za pripravo načrta.<sup>xvi</sup> Britanski arhiv družboslovnih podatkov (UK Data Archive) ponuja gradiva in svetovanje že pri pripravah projektov in v zgodnji fazi njegovega izvajanja.<sup>xvii</sup> Značilnost tega modela je, da financer obenem, ko zahteva načrt upravljanja s podatki, v projektu zagotavi tudi sredstva za to delo. Očitno obstaja vse več soglasja tako na nivoju raziskovalnih politik med državami in med različnimi financerji znotraj držav, da tovrstne rešitve predstavljajo optimalno porazdelitev odgovornosti in zahtev med različnimi vlogami v krogu ravnanja s podatki.

---

<sup>xiii</sup> Problematika je lepo povzeta v enem od poročil Digital Curation Center (DCC) (Swan in Brown 2008).

<sup>xiv</sup> Dissemination and Sharing of Research Results, <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>, 7. 8. 2012

<sup>xv</sup> DMP Online, <http://dmponline.hatii.arts.gla.ac.uk/>, 7. 8. 2012

<sup>xvi</sup> Management Plans, <http://www.dcc.ac.uk/resources/data-management-plans>, 7. 8. 2012

<sup>xvii</sup> CREATE & MANAGE DATA, <http://www.data-archive.ac.uk/create-manage/>, 7. 8. 2012



## 2. KONCEPTUALNI OKVIR ZA PRIKAZ STANJA NA PODROČJU RAZISKOVALNIH PODATKOV V SLOVENIJI

---

Zagotavljanje storitev odprtega dostopa do raziskovalnih podatkov zahteva veliko sodelovanja. Obstaja vrsta modelov, po katerih je mogoče razvijati odprti dostop, ob tem pa je pomembno vprašanje, kako so opredeljene strategije in politike države na področju znanosti, kakšne so strategije raziskovalnih ustanov in ponudnikov podatkovnih storitev. Tudi različna disciplinarna področja imajo različne pristope in zahteve. En sam recept verjetno ne ustreza vsem. Poleg ohranjanja posebnosti raziskovalnih področij pa se je na nivoju oblikovanja nacionalnih in mednarodnih strategij potrebno izogniti podvajanju infastrukturalnih podatkovnih storitev in omogočiti njihovo združljivost, npr. pri agregiranju za namene iskanja in dostopa do gradiv s strani uporabnikov. Analiza trenutnih razmer na področju ustvarjanja in zagotavljanja dostopa do raziskovalnih podatkov bo ponudila uvide glede prednostnih nalog in njihovih potencialnih izvajalcev, upošteva dosežane aktivnosti in kapacitete.

V analizi in prikazu stanja se bomo naslonili na model vlog, pravic, odgovornosti in povezav posameznih nosilcev v fazah življenjskega cikla podatkov (primerjaj Green in Gutman 2006; Lyon 2007; Ruussalepp 2008; Jones 2009). Namen tovrstnega prikaza je poudariti medsebojne odvisnosti, pri katerih izpolnjevanje nalog enega pogojuje možnosti delovanja drugega, izpostaviti dobre rešitve, ki bi se lahko iz enega prenesle na druga področja, opozoriti na pomanjkljivosti in nakazati smernice za odpravo teh. Podporo takšnemu pristopu k presoji delovanja in načrtovanju storitev podatkovnega dostopa najdemo tudi v Načelih OECD in dokumentih britanskega Research Councils (RCUK).<sup>i</sup> Tudi zadnja poročila iz ZDA (Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age 2009; Interagency Working Group on Digital Data 2009) poudarjajo celovito skrb za kakovostno znanstveno produkcijo podatkov ter skrb za shranjevanje in uporabo, s katero so opredeljene vse vloge, tako za strokovnjake s področja ravnanja s podatki, kot za raziskovalce in financerje raziskav.

Pregled je narejen s predpostavko o potrebi po močnih področnih podatkovnih centrih, tam, kjer so obstajajo možnosti za njihovo oblikovanje in delovanje. Zaenkrat razen Arhiva družboslovnih podatkov takega centra v Sloveniji na drugih področjih nimamo. Sicer so mogoče tudi druge rešitve, kot je model samo-arhiviranja, ali pa model repozitorijev digitalnih gradiv ustanov. Naša teza gre v smer, da je obstoj področnih podatkovnih centrov hrbtenica celovitega sistema zagotavljanja dostopa do raziskovalnih podatkov, tako da bomo v prikazu stanja proučevali možnosti in nastavke za oblikovanje tovrstnih osrednjih disciplinarnih podatkovnih storitev v posameznih okoljih. Ostale storitve, kot so inštitucionalni arhivi in mednarodne podatkovne infrastrukture, so lahko dopolnilo ali vmesne rešitve, medtem ko je robusten in vzdržan sistem sodelovanja raziskovalcev in nacionalno dejavnih področnih podatkovnih centrov edini zmožen učinkovito zagotoviti trajno hrambo, dostop in izkoriščanje najbolj kakovostnih podatkovnih virov za znanost.

---

<sup>i</sup> Prvi princip skupnega dokumenta Znanstvenih svetov v Veliki Britaniji (RCUK) se glasi: "Vloga in odgovornost raziskovalcev, raziskovalnih ustanov in financerjev mora biti opredeljena čim bolj jasno (...) za zagotovitev zavedanja in izpolnjevanja odgovornosti ustvarjalcev in uporabnikov raziskovalnih podatkov (...)" (RIN 2008, 8). Podobno razglašajo že omenjena Načela OECD o preglednosti, formalni odgovornosti in profesionalizmu (2007).

Denimo Green in Gutmann (2006) na podlagi situacije v ZDA, kjer je sicer zelo razvita mreža digitalnih repozitorijev po ustanovah, izpostavljata prednosti področnih podatkovnih centrov, kot je Interuniversity Consortium for Political and Social Research (ICPSR), ki delujejo že dolga desetletja in so prepoznavni in uveljavljeni znotraj znanstvene skupnosti. Namreč, repozitoriji ustanov nastopijo svojo vlogo pri koncu raziskovalnega cikla: pretežno prevzemajo že v končni obliki pripravljene publikacije in druga gradiva, nudijo malo ali nič podpore glede priprave podatkov, metapodatkov in njihove izrabe, prevzem je pretežno avtomatiziran in gradiva spremljajo metapodatki brez nujnih podrobnosti za razumevanje konteksta nastanka in vsebinskih lastnosti gradiv, z minimalnim preverjanjem na ravni skladnosti formatov. V Sloveniji sicer močno zaostajamo tudi na področju repozitorijev ustanov. Za razliko od teh pa so področni podatkovni arhivi osredotočeni na izpolnjevanje potreb posameznih disciplinarnih področij tako glede vsebin kot oblik in načina dostopa do podatkovnih gradiv, s svojimi storitvami pa se raztezajo v vse faze življenjskega kroga podatkov, od nastanka, hrambe, objavljanja ter druge rabe. Celota teh storitev pa je poudarjeno tematika pričujočega pregleda.

Obstoj močnega področnega podatkovnega centra kot je ADP zagotavlja kritično maso ohranjanja in razvijanja strokovnega znanja za ravnanje s podatki. To je lahko dobra podlaga za koordiniran nastanek novih področnih podatkovnih centrov za discipline, kjer tega še ni, prilagojen slovenskim razmeram v znanosti. Pomembno je, da na področju dostopa do raziskovalnih podatkov v Sloveniji ne razvijamo nekaj povsem novega, pač pa upoštevamo obstoječe izkušnje nekaterih že delujočih infrastrukturnih projektov. Treba je ohraniti sedanje stanje in ga nadgraditi v smeri kakovosti delovanja v okviru celovitega sistema med seboj prepletenih vlog in odgovornost. Naša naloga je, v okviru izvajanja pričujočega projekta, prenesti izkušnje na druga področja, na katerih so podatkovne storitve še v zametkih, zagotoviti prilagajanje potrebam uporabnikov in posebnostim posameznih disciplin, poskrbeti pa tudi za integracijo strokovnih pristopov, orodij in storitev tam, kjer je to bolj optimalno z vidika porabe sredstev, človeških naporov in nivoja zagotavljanja kakovosti storitev.

## **2.1 ŠTUDIJE POTREB RAZISKOVALCEV**

Preden se lotimo podrobnega prikaza rezultatov naše lastne študije, bomo predstavili nekaj ugotovitev iz podobnih študij potreb in možnosti za oblikovanje storitev odprtih podatkov. V poročilu z naslovom *What researchers want* je Martin Feijen (2011, 4) povzel rezultate različnih raziskav z Nizozemske, Velike Britanije, ZDA, Avstralije in Evrope, zanimale so ga potrebe raziskovalcev na področju ravnanja s podatki in možnosti dostopa do njih. Ena od pomembnejših ugotovitev je, da kljub velikim razlikam med raziskovalnimi področji, raziskovalce tarejo tudi skupni problemi. Npr. zastarela programska oprema in problemi ne-tehnične narave: pomanjkanje zaupanja, strah pred konkurenco, pomanjkanje nadzora ... Zato je razumno in smiselno iskanje skupnih rešitev za tovrstne probleme. Feijen je pustil ob strani vprašanja in dileme, ki zadevajo vodstvo, financerje, podatkovne arhiviste ali knjižničarje in informatike, ter v središče postavil raziskovalčev vidik. Ugotovil je, da imajo raziskovalci velik interes, da uporabijo storitve za podporo pri delu s podatki, vendar morajo te biti pripravljene skladno s posebnimi zahtevami, pri čemer je pomembno:

- da so usklajene z delovnim postopkom, ki je običajno svojevrsten za določeno področje,
- raziskovalci pogosto zavračajo rešitve, ki so 'naložene od zgoraj navzdol', raje uporabijo 'kavarniški' model, pri katerem je mogoče izbirati med različnimi storitvami, po načelu, da vzameš tisto, kar potrebuješ,
- orodja in storitve morajo biti enostavna za rabo,

- tudi če predajo podatke v skrbništvo in posredovanje pooblaščenim podatkovnim arhivom, želijo raziskovalci ohraniti nadzor nad tem, kaj se dogaja s podatki, kdo ima dostop do njih in pod kakšnimi pogoji, ,
- raziskovalci pričakujejo orodja in storitve, ki bodo podpirale vsakodnevno delo, pa tudi dolgoročne zahteve,
- prednosti, ki izvirajo iz podpornih storitev, morajo biti takoj vidne in ne šele čez nekaj let,
- podpora naj bo lokalna, pri roki in na voljo, ko je potrebno.

Druge pomembne ugotovitve, do katerih je prišel avtor ob pregledovanju aktualnih relevantnih poročil iz projektov, kot so Parse (2009), e-IRG (2009) in drugo, so, da je za raziskovalce-ustvarjalce podatkov pogosto v ospredju dostop do podatkov v teku trajanja projekta, in da bi za potrebe dolgotrajne hrambe in dostopnosti po koncu projekta bilo potrebno več vlagati v ravnanje in pripravo podatkov že v teku trajanja (Feijen 2011, 4). Pravne težave in morebitna zloraba ali napačna raba podatkov so med najpogostejše omenjenimi ovirami (Feijen 2011, 11). Zlasti pravne dileme so strokovno področje splošne narave, ki je tipično izven obsega znanj, ki jih posedujejo raziskovalci specialisti, zato bi bilo potrebno posebno svetovanje in poenotena navodila s strani ustanov, ki bi pomagale premostiti te ovire. Kot odgovor na to se v podatkovnih arhivih ali podobnih digitalnih repozitorijih pojavljajo postopki pri prevzemu in ravnanju s podatki, ki so pravno ustrezno utemeljeni. Glede druge, sicer tudi v ostalih študijah večkrat omenjene ovire, to je skrbi ustvarjalcev zaradi možnosti napačne rabe podatkov, če ti pridejo v roke nezadostno seznanjenim uporabnikom, se nakazuje potreba po večji vključenosti raziskovalcev – ustvarjalcev v usposabljanje in svetovanje drugim uporabnikom pri uporabi njihovih podatkov. To so že naloge podatkovnih storitev, ki bi jih glede na svoje kompetence najbolje lahko prevzeli ustvarjalci podatkov, model za izmenjavo znanj pa bi bil lahko oblikovanje (npr. tudi virtualne) skupnosti uporabnikov določene vrste podatkov.

Pri povzemanju ugotovitev iz projekta Parse (2009) je zanimiva perspektiva razlikovanja med pripravljenostjo delovati ter znanjem in informiranostjo glede delovanja. Npr. znanstveniki naravoslovci po njihovih ugotovitvah izražajo večjo pripravljenost za objavljanje podatkov v specifičnih disciplinarnih arhivih kot drugi, vendar hkrati izražajo precejšnjo ne-seznanjenost z obstojem zunanjih storitev za dolgotrajno hrambo, kot so podatkovni arhivi in podobne storitve, prav tako je dokaj nizko zavedanje glede zahtev, ki jih glede dostopa do podatkov postavljajo financerji.

Glede koristi in razlogov za shranjevanje in dostopa do podatkov Feijen (2011, 8) ugotavlja, da za raziskovalce po navadi ni dovolj zgolj postaviti obvezo, pač pa se sami pogosto raje podredijo zahtevam, če jih razumejo in če so izraženi utemeljeni razlogi. Ti naj bodo vidni skozi koristi in prednosti bolj odprtega dostopa do podatkov, npr. vrednost podatkov zaradi težavnega zbiranja in s tem povezani stroški, njihova edinstvenost in možnost preverjanja objavljenih ugotovitev, zaradi česar bi bilo škoda podatke zavreči zgolj po enkratni uporabi. Raziskovalci v večini niso za administrativno določeno obveznost, pač pa podpirajo postavljanje načelnih spodbud. Tudi naša izkušnja v teku intervjujev je bila, da raziskovalci ustvarjalci večinoma pozitivno gledajo na možnost dostopa do podatkov, če ne iz drugih vidikov, zaradi možnosti preverjanja raziskovalne integritete pri objavljenih rezultatih. V nasprotju s tu navedenimi ugotovitvami pa so bili naši intervjuvanci večinoma naklonjeni tudi jasnim obveznostim za izročanje podatkov in dostopu pod pogojem, da so obveznosti enake za vse in da se zagotovi ustrezna finančna in moralna podpora, t.j. priznanje avtorstva.

Ostale relevantne ugotovitve Feijenove pregledne študije zadevajo močno izraženo povpraševanje po možnostih iskanja in ugotavljanja, katere podatke ustvarjajo drugi raziskovalci, kaj vse obstaja, kaj je na voljo in pod kakšnimi pogoji. Tu se kaže temeljna potreba po vzpostavitvi evidence in možnosti osnovnega iskanja po podatkovnih virih. Povsod je mogoče ugotoviti, da je samo manjši del podatkov

deležni obravnave s strani podatkovnega centra, prav tako pa tudi, da je uporabnost podatkov pogosto omejena na ožji krog uporabnikov, ali pa je predvidena dokaj redka bodoča druga uporaba (Feijen 2011, 19 – 20). Na te ugotovitve se bomo vrnil v nadaljevanju, ko bomo obravnavali vprašanja vrednotenja in ocenjevanja potenciala polne uporabe raziskovalnih podatkov (več o tem v 6. Faza: Dodana vrednost). Vsekakor so tu problemi in koristi nekoliko drugačni kot pri klasičnih raziskovalnih poročilih v obliki publikacij. Dostikrat so raziskovalnemu projektu zunanji podatki samo za dodatno evidenco, ali pa primerjavo in umerjanje inštrumentov, vendar tudi kot taki nepogrešljiv del v mozaiku sestavljanja novih kakovostnih spoznanj.<sup>ii</sup>

Na osnovi poročila RIN/BL (2010) Feijen povzema (2011, 21-22), da so potrebe dostikrat povezane s pomanjkanjem zaupanja ustvarjalcev podatkov glede tega, kaj se s podatki dogaja. Kar potrebujejo raziskovalci, je proaktivno vključevanje storitev knjižnic in raziskovalne administracije. Temeljna potreba je čim večja prilagoditev politik, zahtev in storitev obstoječim, pogosto neformalnim omrežjem in potekom raziskovalnih projektov, ki lahko vključujejo orodja in znanja vezana na discipline. Se pravi, da je politika vključevanja in podpore načeloma najbolje sprejeta, če je vezana na uspešna nastajajoča orodja in prakse znotraj raziskovalnih skupnosti samih. Izhajajoč iz teh ugotovitev bomo v nadaljevanju iskali in potrjevali obstoj podobnih iniciativ tudi v okviru naših znanstvenih okolij, s tem da bi bilo potrebno del politik na tem področju usmeriti ravno v spodbujanje inovativnih pristopov podpore podatkovnim storitvam in skupnostnih orodij (kot so *Taverna workflow* orodja),<sup>iii</sup> ki nastajajo v sodelovanju znotraj discipline ali pa v okviru meddisciplinarnega sodelovanja znanstvenikov.<sup>iv</sup> Na splošno je izražena velika potreba po dodatnem usposabljanju delujočih znanstvenikov v veščinah podatkovnega upravljanja in digitalne hrambe (Feijen 2011, 23)<sup>v</sup>. Za bodočo pripravo strategij in akcijskega načrta so koristne tudi ugotovitve, povzete iz poročila UK Research Data Service (2010), ki prav tako sloni na raziskavi o različnih akterjih v življenjskem krogu podatkov. Na tem mestu izpostavimo točke, ki so bile vodilo našega pristopa: potreba po prednostnem vzpostavljanju podatkovnih storitev glede na povpraševanje dajalcev podatkov, potreba po postopnem uvajanju novosti za discipline in področja, kjer je očitna takojšnja korist (npr. sodelovanje in modeli podatkovne izmenjave iz družboslovja na druge discipline, kot je medicina), pragmatične in postopne širitve storitev, ter previdnost pri postavljanju ciljev z vidika izvedljivosti, pomembno je namreč dobro predstaviti ideje ter koristi in zagotoviti potrebno podporo in sodelovanje, zlasti na institucionalni ravni (Feijen 2011, 25).

V zgoraj omenjenem projektu UK Research Data Service so si zadali, da bodo skušali oceniti izvedljivost in stroške razvoja ter vzdrževanja podatkovnih storitev za univerze v Veliki Britaniji, raziskava pa je zanimiva tudi zaradi kombiniranja uporabe več metod zbiranja podatkov. Anketnemu vprašalniku so na vsaki od izbranih univerz sledile fokusne skupine. Namen skupinskih pogovorov je bil ugotoviti, kakšne zahteve imajo univerzitetni raziskovalci glede hrambe podatkov – zato so bolj poglobljeno analizirali odgovore, pridobljene s pomočjo spletnega vprašalnika. Poleg ankete in fokusnih skupin je raziskava obsegala še 37 individualnih intervjujev z raziskovalci z Univerze v Oxfordu in delavnico, na kateri so razpravljali o ugotovitvah, pridobljenih z intervjuji.

---

<sup>ii</sup> Glede distinkcije med podatki v ozadju in v ospredju, ko gre za različne uporabe (kar je podatek v ozadju za nekoga, je za drugega primarna podlaga analiz in objave), glej npr. Wallis in drugi (2012).

<sup>iii</sup> Več o tem glej: <http://www.taverna.org.uk/>, 7. 8. 2012

<sup>iv</sup> Primerjaj projekte digitalne humanistike: <http://www.diggingintodata.org/>, 7. 8. 2012

<sup>v</sup> Glej tudi Williford in Henry (2012, 2.4.2 pogl.).

## 2.2 OVIRE NA POTI DO ODPRTEGA DOSTOPA (primer s Finske)

Arja Kuula in Sami Borg (2008) sta v svojem poročilu priložnostne anketne raziskave z naslovom *Open Access to and Reuse of Research Data – The State of the Art in Finland* prav tako izhajala iz podmen načel OECD. Podobno kot mi v uvodu ugotovljata, da je druga raba podatkov v dokumentih OECD razumljena kot nekaj samo po sebi umevnega. Realno stanje glede podatkovne izmenjave in podporne raziskovalne infrastrukture v manj razvitih okoljih je seveda ravno nasprotno. V tej študiji so ugotovili, da je osnovni problem, ki ga je bilo mogoče pričakovati tudi v okviru naše raziskave, majhna osveščenost in seznanjenost z možnostmi, ki jih ponuja odprti dostop do raziskovalnih podatkov. Morda je v posameznih državah stopnja razvitosti podatkovne infrastrukture in kulture izmenjave podatkov že dosegla stopnjo, na kateri je znanstvena skupnost sprejela odprti dostop in drugo rabo kot nekaj običajnega, vendar avtorja ugotovljata, da na Finskem še zdaleč ni tako. Zaradi teh ugotovitev je situacija v tej državi zanimiva tudi za oblikovanje izhodišč za našo raziskavo.

Finska raziskava je obsegala 150 anketirancev, univerzitetnih profesorjev s širšega področja humanistike in družboslovja. V središče so postavili vprašanje, katere ovire vidijo anketiranci na poti do cilja odprtega dostopa do podatkov. Raziskava je pokazala na tipične ovire in pomisleke, da raziskovalce skrbi nenamerna zloraba podatkov in posledične napake. Pravica glede intelektualnega lastništva podatkov ni jasno opredeljena, ovira je tekmovalnost za akademske položaje in financiranje, dvomi glede uporabnosti in IT težave, odsotnost soglasja in zaupanja, ki zadevata raziskovalno etiko. Vse naštetu lahko pričakujemo tudi v naši situaciji.

Ugotovili so tudi odsotnost politik, navodil ter napotkov na ravni ustanov, kar raziskovalce ustvarjalce podatkov in uporabnike pušča v negotovosti. To kaže na širši problem, da so pri vzpostavitvi in delovanju sistema odprtega dostopa do podatkov pomembne vmesne podporne strukture. To bi lahko bila tudi tema za področje implementacije morebitnih nacionalnih politik, ker je očitno, da bi se takšnih problemov lahko lotili na institucionalni ravni, kjer bi se opravila selekcija pomembnih gradiv, oblikovale spodbude in specialni svetovalni centri ter nekatere vmesne storitve v skrbi za podatke. Npr. v poročilu je omenjen model obveznosti z vidika arhivske zakonodaje, ki že danes določa nekatera pravila glede selekcije in shranjevanja gradiv, ki pa se jih raziskovalci in vodstva ustanov ne zavedajo. Veliko strokovnega znanja glede ravnanja s podatki in na njih zgrajenih storitvah sicer obstaja, a to ostaja razpršeno. Ohranjanje in večja raba tega vira intelektualnega kapitala bi lahko bila predmet oblikovanja nacionalnih usmeritev glede razvoja. Prav človeški kapital je največja vrednost, ki se dostikrat zaradi neustrezne organiziranosti izgublja, poleg tega da odsotnost usposobljenih skrbnikov predstavlja tudi ogrožanje dragocenih podatkovnih virov (primerjaj Ashley 2011). Ta tema, se pravi samoocena strokovne usposobljenosti za različne faze zagotavljanja storitev pri ravnanju s podatki in izražena potreba po tovrstni podpori, je ena od izpostavljenih tem pri določanju potencialov tudi za stanje v naši državi. V smeri načrtovanja bodoče ureditve gre s tem v zvezi tudi za vprašanje, kako zagotoviti prenos te lokalne strokovnosti na širši krog uporabnikov, npr. na način svetovanja, ali pa na osnovi modela skupnosti, ki jo gradijo uporabniki določenih podatkov. V naši raziskavi je načelno naklonjenost in pripravljenost za prevzem svetovalne vloge pokazal denimo intervjuvanec s področja medicinskih ved in statistike (več o tem glej 6.2 Spravilo).

Kuula in Borg (2008, 27) posebej izpostavita vrsto koristi in problemov, ki bi sledili implementaciji odprtega dostopa in so bile prepoznane s strani anketirancev:

- finančno: bolj učinkovita izraba raziskovalnih podatkov vodi v maksimiranje finančnega izkoristka glede na velike vložke v ustvarjanje podatkov;
- znanstveno: odpiranje znanosti, kontinuirano ustvarjanje podatkov za zagotovitev njihove ažurnosti, ponovljivost in kontrola raziskav, večje sodelovanje v raziskovalni skupnosti;

- raziskovalno-etično: npr. upoštevati pravice sodelujočih posameznikov med ustvarjanjem in rabo podatkov, kar je zlasti izpostavljeno v medicinskih in družboslovnih raziskavah;
- pravno: upoštevati pravo med ustvarjanjem, rabo in hrambo podatkov;
- demokratično: natančno opredeliti položaj in odgovornosti raziskovalnih subjektov, raziskovalcev, raziskovalnih organizacij in financerjev;
- enakost: promocija enakega dostopa do raziskovalnih podatkov;
- družbeno: izboljšati kakovost in stroškovno učinkovitost informacijskih virov;
- politično: OECD obvezuje države članice.

Med predlogi za implementacijo odprtega dostopa je tudi ustrezno vrednotenje raziskovalnih podatkov, se pravi, da bi raziskovalni podatki dosegli podoben status, kot ga imajo raziskovalne objave. Kuula in Borg opozorita naj bodo razlike med objavami in podatki vodilo za oblikovanje za vsako področje prilagojenih politik. Raziskovalci so večinoma naklonjeni temu, da se podatke obravnava kot samostojni izdelek, zanje je to dodatna motivacija za zgledno pripravo podatkov in nagrado za vloženo znanje, ki ga sami po sebi vsebujejo raziskovalni podatki, neodvisno od publikacije ali v povezavi z njo, saj predstavljajo dodatni vir informacij, ki jih publikacija ne vsebuje. Raziskovalna publikacija je samo-razlagalna, medtem ko raziskovalni podatki zahtevajo dodatno razlago in dokumentacijo, zlasti če niso ustrezno arhivirani. Za razliko od objav lahko podatki vključujejo občutljive in zaupne informacije. Raba podatkov običajno zahteva dovoljenje, publikacija ne. Pri publikacijah so lastništvo in avtorske pravice običajno razjasnjene, pri podatkih so pogosto nejasne. Objave so v znanstveni skupnosti dostopne: prosto ali proti plačilu. Pri podatkih gre za več stopenj odprtosti (od popolne odprtosti do tega, da so podatki zaprti, zaradi posebnosti povezanih s specifičnim podatkovnim virom). Publikacija velja za znanstveni rezultat, podatki, tudi če so objavljeni, običajno nimajo statusa znanstvenega dosežka. Publikacija je že pripravljena za drugo rabo, podatki terjajo dodatne postopke (Kuula in Borg 2008, 28). Tema je zelo aktualna in zasluži podrobnejšo obravnavo, saj je povezana s specifičnimi vprašanji izbora in znanstvene kakovosti podatkov, ustrezne infrastrukture za registracijo identifikatorjev podatkov in podporo citiranju podatkov. Prav zato predstavlja enega od izzivov za reševanje tudi v okviru preiskovanja pogojev in možnosti v Sloveniji.<sup>vi</sup>

---

<sup>vi</sup> Primerjaj opis ciljev projekta Open AIRE Plus v mednarodnem pregledu in druge iniciative v zvezi z infrastrukturo za citiranje podatkov.

### 2.3 PREMALO RAZVITA RAZISKOVALNA INFRASTRUKTURA (primer iz Čila)

Izhajajoč iz načel OECD so se vprašanja odprtih podatkov lotili tudi v Čilu. Njihov osnovni cilj je bil urediti, okrepiti in zagotoviti odprti dostop do podatkov in nasploh do raziskovalnih informacij, zbranih in ustvarjenih v Čilu, financiranih z javnimi sredstvi. Z namenom, da spodbudijo izmenjavo podatkov, ustvarjenih v državi in da ohranijo raziskovalno podatkovno zapuščino. Pod drobnogled so postavili institucionalni okvir, človeški kapital, zapuščino in politike in prišli do sledečih ugotovitev:<sup>vii</sup>

Institucionalni okvir:

- pomanjkanje institucionalne zavesti o pomenu teme, začne se pri vodstvu,
- odgovornost za upravljanje s podatki in znanstvenimi informacijami je razpršena po različnih organizacijah, odsotnost konkretne organizacije, ki bi imela to zadolžitev,
- sredstva za upravljanje s podatki so razpršena po projektih in nimajo statusa stalne dejavnosti,
- raziskovalni podatkovni zapuščini ni priznana institucionalna in nacionalna vrednost, pač pa kot da gre za raziskovalčev osebni vir,
- pomanjkanje standardizacije podatkov in formatov.

Človeški kapital:

- pomanjkanje kritične mase človeških virov specializiranih za upravljanje s podatki,
- pičila bera uradnih in specializiranih akademskih programov, usmerjenih na upravljanje s podatki,
- pomanjkanje spodbud za oblikovanje kariernih poti na tem področju,
- brez vodstva ali nacionalnih referenc,
- delovna mesta so negotova in kratkoročna.

Zapuščina:

- pomanjkanje ali šibko vrednotenje raziskovalne podatkovne zapuščine,
- zapuščina je razpršena med različnimi organizacijami,
- zapuščina velja za osebno in ne za institucionalno ali nacionalno dobro,
- hramba in formati so stvar individualne presoje,
- pomanjkanje spodbud za izmenjavo zapuščine,
- ni kulture prostega dostopa.

Politike:

- odsotnost politik, ki določajo dostop do raziskovalnih podatkov in znanstvenih informacij.

Poročilo je zanimivo, ker kaže tipične ugotovitve za okolja, ki nimajo razvite raziskovalne podatkovne infrastrukture, kot je po vseh predvidevanjih tudi naše. Mi smo pri pregledu situacije izhajali iz tovrstnih pričakovanj, ki so nekatera tudi eksplicitno potrjena.

---

<sup>vii</sup> Izvlečke te raziskave je Patricia Muñoz Palma predstavila na Board on International Scientific Organizations, aprila 2011, v Washingtonu, [http://sites.nationalacademies.org/PGA/biso/PGA\\_061353](http://sites.nationalacademies.org/PGA/biso/PGA_061353), 7. 8. 2012



## 2.4 NAVADE IN PRAKSE RAZISKOVALCEV (primer iz Velike Britanije)

Britanski RIN, Joint Information Systems Committee (JISC) in Natural Environment Research Council (NERC) so leta 2007 naročili raziskavo, ki se je osredotočila na navade in prakse raziskovalcev v odnosu do raziskovalnih podatkov. Poročilo temelji na več kot 100 podrobnih intervjujih z raziskovalci z različnih področij: astronomija, kemijska kristalografija, književnost, klimatologija, genomika, družbene vede in javno zdravstvo ter z dveh interdisciplinarnih področij, t.j. sistemska biologija in ruralna ekonomija in raba zemlje v Veliki Britaniji. Posebej jih je zanimalo, v kakšnem stanju je kultura deljenja podatkov, kakšne so infrastrukturne omejitve pri objavah podatkov, kak je učinek pobud, katere politike spodbujajo objave podatkov in kakšna je nasploh pripravljenost za objavljanje podatkovnih zbirk.

Ugotovili so, da se področja med seboj precej razlikujejo. Kultura deljenja podatkov je visoka na področju astronomije, genomike in književnosti, srednja na področju kemijske kristalografije, sistemske biologije ter ruralne ekonomije in rabe zemlje v Veliki Britaniji, ter nizka na področju družbenih ved in javnega zdravja ter klimatologije. Splošna pripravljenost za objave podatkovnih zbirk (z ustreznimi metapodatki in spremljajočo dokumentacijo) je nizka na področju družbenih ved, srednja na področju ruralne ekonomije in rabe zemlje, sistemske biologije, klimatologije in književnosti, ter visoka na področju astronomije, kemijske kristalografije in genomike (Swan in Brown 2007, 54). Skratka, ena najpomembnejših ugotovitev tega poročila je pomanjkanje usklajenosti med disciplinami, v smislu ravnanja, politik in potreb. Swan in Brown (2007, 13) sta zapisala: *"Kakršno koli rešitev problema bomo opredelili, bo zahteve in prakse potrebno prilagoditi vsaki posamezni raziskovalni disciplini."*

Posebno pozornost so med drugim posvetili tudi vprašanju, zakaj nekateri raziskovalci objavljajo svoje podatke, drugi pa ne. Raziskovalci, ki objavljajo podatke, so kot spodbude navedli sledeče:

- altruizem in prizadevanje za javno dobro,
- prisotnost kulture deljenja podatkov v okolju, v katerem delujejo,
- večja prepoznavnost raziskovalne skupine in ustanove,
- možnost za soavtorstvo pri objavah,
- možnost za sodelovanje z drugimi tudi izven njihovega lastnega področja,
- dejavnik upoštevanja in pričakovanj s strani financerja,
- tisti, ki delijo svoje podatke, težijo k priznanju svojih dosežkov (Swan in Brown 2007, 25-26).

Med ovirami, zaradi katerih ne objavljajo podatkov, so navedli sledeče:

- pomanjkanje časa in sredstev za upravljanje s podatki,
- pomanjkanje časa, da bi se ukvarjal z vsakim posebej, ki bi želel njegove podatke,
- pomanjkanje izkušenj ali ekspertize v upravljanju s podatki,
- razpoložljivost, dostopnost, uporabnost,
- pravne ali etične ovire,
- nepoznavanje možnosti arhiviranja,
- tekmovalnost,
- strah pred zlorabo ali napačno rabo,
- prepričanje, da nihče ne potrebuje njegovih podatkov,
- omejene nagrade ali nagrad sploh ni (Swan in Brown 2007, 26-28).



Prisotnost podobnih razlogov, ki ovirajo deljenje podatkov, bomo ugotovili tudi med slovenskimi raziskovalci. Ugotovili pa bomo tudi, da v naših razmerah ne navajajo veliko razlogov, ki bi bili v prid deljenju, kar je vsekakor posledica odsotnosti delujočih primerov in pozitivnih zgledov, ki bi bili širše uveljavljeni.

## 2.5 POMEN IZOBRAŽEVANJA IN USPOSABLJANJA ZA RAVNANJE S PODATKI

Za premagovanje ovir na poti k odprtim podatkom je še kako pomembno poznavanje navad, praks in kulture raziskovalcev.<sup>viii</sup> Poleg kakovostnih podatkov in kakovostnih postopkov za upravljanje s podatki je pomemben del pozornosti treba usmeriti tudi na usposabljanje strokovnjakov, ki ravnajo z njimi. Posamezne ustanove že govorijo o rešitvah. Na Univerzi v Oxfordu so v svojih politikah<sup>ix</sup> zapisali, da sta ustrezna odgovornost in ozaveščenost zaposlenih pomembna cilja za doseganje kakovostnih podatkov. V središče so postavili idejo, da morajo vsi zaposleni prepoznati potrebo po skrbi za kakovost podatkov in razumeti, da lahko k temu sami prispevajo. Da se morajo zavedati individualne odgovornosti, ki mora biti vključena v ustvarjanje, hrambo, analizo in poročanje o podatkih. Prav vsak zaposlen se mora zavedati posledic slabe kakovosti podatkov, ki vplivajo ne le na ozko področje, pač pa tudi na delo drugih oddelkov in ustanov, ki so z njimi povezane. Vsak je svojemu vodji dolžan poročati o težavah v zvezi s kakovostjo podatkov in vodja je dolžan ukrepati. Vsak zaposlen se mora zavedati politik, povezanih s kakovostjo in zaščito podatkov. Z institucionalizacijo praks postanejo zahteve in odgovornosti transparentne in njihovo uresničevanje del vsakodnevnih praks raziskovalnega procesa.

Čeprav se morajo vsi zavedati pomena in odgovornosti za doseganje kakovostnih podatkov, gre posebna vloga tistim, ki se poleg znanstvenega ustvarjanja ukvarjajo tudi s storitvami, povezanimi s podatki. Če bi dejavnosti ločili po področjih in temu primerno ovrednotili naloge, se pokaže možnost bodočega poklica – *podatkovni znanstvenik*. V obstoječih pogojih je zato smotno poiskati centre in nosilce znanja, ki bi lahko bili ustrezna podlaga za oblikovanje in sistematizacijo novega poklica ter tovrstna znanja vključiti v univerzitetno izobraževanje znanstvenikov, prilagojeni njihovi disciplini.

Eno od področij, o katerem se dosti razpravlja v zvezi s spremenjenimi poklicnimi usmeritvami, je tudi vloga knjižničarjev v novem okolju politik in zahtev odprtega dostopa do podatkov. Ameriško združenje knjižničarjev (Association of Research Libraries, v nadaljevanju ARL) se v poročilu *Agenda for Developing e-science in research libraries- ARL* (Agenda 2007) odkrito zavzema za to, da bi knjižničarji postali pomemben člen oblikovanja strategij na področju e-znanosti in nasploh na področju dostopa do podatkov. Knjižnice bi s svojim znanjem in izkušnjami že danes lahko služile kot infrastrukturna podpora, pravijo pri ARL, kar argumentirajo z naslednjim:

- posedujejo ekspertno razumevanje politik in načel povezanih z odprto izmenjavo znanstvenih in izobraževalnih informacij, ter poznajo vloge, ki bi jih lahko opravljali institucionalni repozitoriji za zagotavljanje izmenjave;
- imajo izkušnje z razvojem in podporo orodij za distribucijo in iskanje informacij, tako znotraj discipline kot tudi preko;
- imajo izkušnje z razvojem in podporo poslovnih in tehničnih strategij za dolgoročno arhiviranje;
- razumejo tiste vidike znanstvene informacije, ki zadevajo arhiviranje in življenjski krog podatkov ter pomen zagotavljanja dostopa in uporabnosti na dolgi rok (Agenda 2007, 13).

<sup>viii</sup> Za več informacij je zanimiva študija uporabnikov, ki so jo na vzorcu 1329 znanstvenikov, pripravili Tenopir in drugi (2011).

<sup>ix</sup> Data Quality and Data Quality Assurance Policy, [http://www.admin.ox.ac.uk/pras/aboutus/data\\_quality/](http://www.admin.ox.ac.uk/pras/aboutus/data_quality/), 7. 8. 2012

Čeprav se knjižničarji zavedajo svojih omejitev, se zavzemajo za to, da bi v prihodnosti razvili orodja in veščine, specifične za področje ustvarjanja podatkov, dostopa in druge rabe. Zavzemajo se, da bi bili vse bolj udeleženi v postopkih nastajanja podatkov, saj bi tako lažje in bolje prispevali k izvajanju programa storitev infrastrukturne podatkovne podpore. Konkreten cilj je razviti znanja, sposobnosti in veščine, da bodo knjižničarji mogli prispevati k e-znanosti in k oblikovanju novih vlog in modelov storitev ter postali aktivni udeleženci v procesu zasnove in razvoja raziskovalne infrastrukture, vključujoč sistem in storitve, ki bodo podpirale celoten življenjski krog raziskovanja (Agenda 2007, 16-17; glej tudi Lossau 2012). Ta dokument je zanimiv tudi zato, ker obsega strukturo ciljev, strategij in akcij o tem, kaj konkretno bi bilo potrebno storiti. Npr. ustanoviti delovno skupino, organizirati sestanke, pripraviti razpise za pilotske projekte, oblikovati učne načrte in usposabljanja.

Izoblikovali so delovno skupino in leta 2009 izvedli raziskavo med članicami ARL. Namen je bil *"spoznati, kako lahko knjižnice prispevajo k e-znanstvenim dejavnostim"* ter *"opredeliti organizacije in ustanove, ki imajo podoben interes"* (Soehner in drugi 2010, 7). V raziskavi je sodelovalo 57 od 123 članic ARL iz ZDA in Kanade. V 21 knjižnicah že zagotavljajo infrastrukturo ali podpirne storitve s področja e-znanosti, 23 jih je v fazi načrtovanja, v 13 knjižnicah pa ne zagotavljajo podpore e-znanosti. Rezultati raziskave so pokazali, da bo potrebno namesto enega samega sistema storitev razviti koordiniran in sintetiziran sklop, ki bo šele tvoril podporni sistem e-znanosti (Soehner in drugi 2010, 7-8). Med najbolj perečimi problemi, v tej raziskavi, izpostavljajo: pomanjkanje sredstev, ustrezno usposobljenost kadra in strokovnjakov s področja upravljanja s podatki ali hrambe in pomanjkanje poenotenih smernic. Čeprav so izrazili zaskrbljenost zaradi težav in omejitev, pa je med ameriškimi in kanadskimi knjižničarji mogoče zaznati navdušenje nad novimi vlogami v akademskem raziskovalnem procesu (Soehner in drugi 2010, 8). V primerjavi z ugotovitvami te študije je vtis, da gre pri nas za izrazitejšo pomanjkanje institucionalnih politik na področju vprašanj e-znanosti ter o marginalni vlogi knjižnic pri podpori projektov e-znanosti. Knjižnice ostajajo omejene na tradicionalni servis dostopa do kupljenih revij in drugih tiskanih in e-publikacij ter vodenje raziskovalnih bibliografij.

Po drugi strani se tudi pri nas soočamo s podobnimi problemi, kot so ti, ki so jih prepoznali med ameriškimi in kanadskimi knjižničarji: neurejeno financiranje razvoja in projektov na teh področjih, nezadostno strokovno znanje in usposabljanje kadrov pri upravljanju in skrbi za podatke. Veliko se pričakuje od rezultatov projektov, kot je DataNet,<sup>x</sup> se pravi razvojnih projektov in projektov usmerjenih v oblikovanje strategij, politik ter rešitev dobrih praks. Razvijajo se programi in študijske smeri na fakultetah.

S tem problematičnim področjem, kjer gre za iskanje obstoječih znanj splošnih informacijskih strokovnjakov, kot so knjižničarji, in njihovo transformacijo za podporo podatkovnim storitvam, se bomo v poročilu še ukvarjali, saj se kaže kot premalo izkoriščen potencial, ki obenem predstavlja novost in kreativnost pri iskanju in določanju vlog v raziskovalni podpori: raziskovalci in ustvarjalci prevzemajo del vlog podatkovnih storitev, knjižničarji se vključujejo v aktivnosti raziskovalnih projektov, postajajo del raziskovalnega tima. Dolgoročno to predstavlja tudi izziv za oblikovanje izobraževalnih programov in študijskih smeri na fakultetah. Vzorčen primer predstavljajo učni moduli, financirani s strani britanske organizacije JISC.<sup>xi</sup> Gre za pilotske primere tečajev in učnih gradiv za različna področja znanstvenega raziskovanja. Namenjeni so podiplomskim študentom, ki v okviru tečajev pridobijo osnovna znanja s področja raziskovalnih podatkov, največji del pozornosti je

<sup>x</sup> Sustainable Digital Data Preservation and Access Network Partner, NSF.

<sup>xi</sup> Več o tem glej JISC: <http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain.aspx>, 7. 8. 2012

usmerjen na upravljanje s podatki. Zaenkrat so na voljo moduli za področje: zdravstvenih ved/medicine,<sup>xii</sup> umetnosti,<sup>xiii</sup> arheologije,<sup>xiv</sup> psihologije,<sup>xv</sup> geoznanosti, družbene in politične vede.<sup>xvi</sup> Vzorčni primeri predstavljajo zgled, kako je mogoče problematiko vključiti v redno izobraževanje novih raziskovalnih kadrov. Za splošno področje informacijskih ved podoben program razvijajo v okviru projekta RDMRose,<sup>xvii</sup> programi usposabljanja pa se pojavljajo tudi med knjižničarji v ZDA.<sup>xviii</sup>

---

<sup>xii</sup> DATUM for Health

(<http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/health/materials/?view=Standard>, 7. 8. 2012)

<sup>xiii</sup> Cairo je podiplomski učni modul, namenjen za ustvarjalne umetniške discipline. (<http://www.projectcairo.org/node/9>, 7. 8. 2012)

<sup>xiv</sup> DataTrain projekt za arheologijo. (<http://archaeologydataservice.ac.uk/learning/DataTrain>, 7. 8. 2012)

<sup>xv</sup> Data Management Training za psihologe. (<http://www.dmtpsych.york.ac.uk/>, 7. 8. 2012)

<sup>xvi</sup> MANTRA, Research Data Management Training je tečaj namenjen podiplomskim študentom s področja geoznanosti, družbenih in političnih ved ter klinične psihologije. Več o tem glej: <http://datalib.edina.ac.uk/mantra/introduction.html>, 7. 8. 2012.

<sup>xvii</sup> <http://www.sheffield.ac.uk/is/research/projects>, 7. 8. 2012.

<sup>xviii</sup> <http://americanlibrariesmagazine.org/features/08212012/what-s-new-lis-schools>, 7. 8. 2012, govori o .

## 3. RAZISKOVALNA METODA

---

Mark Parsons (2010) poudarja, da ni vprašanje, ali odpreti dostop do podatkov ali ne, pač pa da je potrebno bolj podrobno razdelati kdaj, v katerih primerih in kako odpreti podatke, se pravi pripraviti politike, ki upoštevajo stvarno situacijo na posameznih področjih. Kakor je izraženo v načelu OECD o fleksibilnosti, je pri pripravi potrebno upoštevati raznolikost različnih ravni in področij. V tem duhu Parsons izpostavi potrebo po pristopu "od spodaj navzgor", po politikah, ki bodo razumele stanje, saj bodo izoblikovane skozi razumevanje problematike, kot jo vidijo raziskovalci in drugi, ki se ukvarjajo s podatki.

Tako bo mogoče spoznati raznovrstnost ravnanja z znanstvenimi podatki, ki je pogojeno in omejeno s posameznimi področji. Učinkovit akcijski načrt je mogoče pripraviti le tako, da poznamo razmišljanje in delovanje akterjev na različnih raziskovalnih področjih ter v različnih fazah življenjskega kroga podatkov. Zato je za oblikovanje politik in strategij na tem področju potrebno koordinirano naslavljanje problemov in njihovo reševanje iz različnih perspektiv. V enem od zaključkov poročila *Shared Responsibilities in Sharing Research Data* iz Padove (ESF 2008) je eksplicitno poudarjeno, da morajo politike izmenjave raziskovalnih podatkov 'ostati blizu znanosti in znanstvenikom', sicer se lahko primeri, da bodo podatkovne storitve same sebi namen.

V izogib temu si je treba prizadevati za čim večjo vključitev raziskovalcev v razvoj politik in zmogljivosti na področju podatkov. Odgovoriti si je potrebno na pomembna vprašanja. Ali bodo raziskovalci izročili svoje podatke, če ni na voljo ustrezne infrastrukture, ki bi znala ravnati z njimi? Ali je smiselno veliko in trajno vlagati v raziskovalno infrastrukturo, če repozitoriji ostajajo prazni zaradi pomanjkanja pripravljenosti, da bi raziskovalci deponirali svoje podatke? Takšna vprašanja nakazujejo, da morajo politike in financiranje hoditi z roko v roki z raziskovalci, ustvarjalci in uporabniki podatkov (ESF 2008, 24-25). Ne glede na raznolikost politik in pristopov pa Parsons (2010) poudarja, da morajo vse nujno vključevati odprt dostop in dolgotrajno digitalno hrambo.

Da bi dosegli odprto in etično ter zakonsko podprto podatkovno ureditev, bo treba uporabiti oba pristopa: 'od zgoraj navzdol' in 'od spodaj navzgor'. Od 'spodaj navzgor' bodo raziskovalne skupnosti morale opredeliti nove etične zahteve čim bolj odprtega deljenja podatkov, ki bo temeljilo na sodelovanju in pravičnosti glede na prispevke. 'Od zgoraj navzdol' bo politike treba uskladiti z nacionalnimi zakonodajami, izhajajoč iz primata načel odprtosti in etične rabe, ki v znanosti temelji predvsem na zaupanju, se pravi, da bo potrebno oblikovati ustrezno interpretacijo in prilagoditev zakonodaje (Parsons 2010, 4).

V nadaljevanju predstavljamo še nekaj relevantnih raziskav, ki so nam služile kot zgled za preučevanje 'od spodaj navzgor'.

Američanka Francine Berman (2010) predlaga, kako bi bilo mogoče podatke popisati na način popisa prebivalstva. Tako zbrani oziroma popisani podatki bi po njenem mnenju omogočili učinkovitejše načrtovanje javne infrastrukture za hranjenje digitalnih podatkov. Pri popisu podatkov iz raziskav financiranih iz javnih sredstev, naj bi si pomagali z naslednjimi vprašanji:

- Koliko je posameznih podatkovnih datotek in kakšna je njihova velikost (v bitih)?
- Kakšne vrste podatkovnih datotek (besedilo/video/avdio ...) se pojavljajo na posameznih (znanstvenih) področjih?
- Kakšne so potrebe po hrambi (potrebe uporabnikov/obveznosti)?

- Kakšne storitve in orodja se uporabljajo v povezavi s temi podatki (npr. sekvenčna analiza genov, vizualizacija podatkov in drugo)?

Zastavljena vprašanja predstavljajo izhodišče naše raziskave, ob strani smo pustili le prvo vprašanje, saj smo se z izborom metode in eksplorativnim pristopom bolj osredotočili na primere kot pa na kvantitativen popis stanja.

Raziskava, ki so jo Harley in drugi (2010) izvedli na Univerzi v Kaliforniji, opiše pristop k izvajanju intervjujev in fokusnih skupin. Odločili so se za priložnostno vzorčenje. Intervjuje so posneli, prepisali, povzeli in shranili. Udeležencem so pred izvedbo intervjuja poslali kratek povzetek vprašanj, na katera so kasneje odgovarjali; med intervjujem so dovolili, da so intervjuvanci govorili tudi o drugih temah, če so s tem dodatno osvetlili obravnavano tematiko. Izvedli so 160 intervjujev z osebami iz 45 različnih ustanov, 28 udeležencev se je v nadaljevanju srečalo v fokusnih skupinah. Z obsežno raziskavo, katere začetek sega v leto 2005, so pokrili različna raziskovalna področja: arheologija, astrofizika, biologija, ekonomija, zgodovina, glasba in politične vede. Vsako so predstavili kot študijo primera, ki odgovarja na naslednja vprašanja:

- Kaj so prednosti dobrega raziskovalca?
- Po katerih merilih izbirate medij?
- Na kak način delite svoje delo z drugimi in se seznanjate z novostmi na svojem področju?
- Kako sodelujete z drugimi? Na katerih stopnjah raziskave?
- Kaj potrebujete za izvajanje svojih raziskav?
- Do kakšne stopnje ste vi ali vaši kolegi povezani z javnostjo? Zakaj in zakaj ne?

Tudi mi smo za temeljno metodo ugotavljanja izhodiščnega stanja na področju ustvarjanja, vrednotenja, shranjevanja, objavljanja in uporabe raziskovalnih podatkov izbrali metodo intervjuja s strokovnjaki z različnih znanstvenih področij. Da bi čim bolj spoznali razmere, pogoje in okoliščine, v katerih nastajajo podatki, se hranijo in ponovno uporabijo, smo izbrali metodo pol-strukturiranega intervjuja. Prednost izbrane metode je možnost prilagajanja vprašanj in medsebojnega pojasnjevanja za boljše razumevanje. Izpostavljena je perspektiva pristopa 'od spodaj navzgor', s tem so imeli intervjuvanci možnost izraziti pomisleke in svoje predloge glede obravnavane problematike.

Osnovna zahteva je bila pokriti več različnih znanstvenih disciplin in spoznati različne faze življenjskega kroga podatkov. Zato je bilo treba ustvariti nabor sodelujočih, ki prihajajo iz različnih raziskovalnih okolij in imajo s podatki različne izkušnje, kot ustvarjalci, skrbniki ali uporabniki. Prvi korak k izboru intervjuvancev je bil sestanek s predstavniki raziskovalnih inštitutov, ki ga je 19. januarja 2011 organiziralo Ministrstvo za visoko šolstvo, znanost in tehnologijo. Sestanka se je udeležilo okoli 30 udeležencev, ki so na seznam vpisali svoje kontaktne podatke. Ti so bili en vir za pripravo seznama potencialnih intervjuvancev, druge smo poiskali z namiznim raziskovanjem, tretje preko osebnih poznanstev in s priporočili drugih. Končno smo izoblikovali seznam 85 oseb, ki so bile potencialno zanimive za sodelovanje.

Med temi smo izbrali 47 kandidatov, in sicer tako, da smo pokrili čim bolj različna raziskovalna področja. Kandidate za intervju smo pisno obvestili o vsebini projekta Odprti podatki, o predvidenem poteku raziskave ter o ciljih. Zaposili smo jih, da po svojih močeh čim dejavnije sodelujejo in tako pomagajo pri odstranjevanju ovir na poti k prostemu dostopu do podatkov in k spodbujanju čim širše rabe podatkov.<sup>1</sup> Pisno prošnjo je spremljalo priporočilo Ministrstva za visoko šolstvo, znanost in

---

<sup>1</sup> Priloga 1 – Prošnja za sodelovanje v projektu Odprti podatki.

tehnologijo z dne 17.12.2010.<sup>ii</sup> Po pisnem obvestilu smo stopili v stik s kandidati še osebno in se z njimi skušali dogovoriti za intervju.

Nekateri so sodelovanje zavrnilo, druge smo izpustili, ko se je izpolnila kvota predstavnikov z njihovega področja. Spet tretje smo dodatno vključili po priporočilih intervjuvancev, če so jih ti označili za kompetentne za določeno področje, ki ga do tedaj še nismo pokrili. Intervjuje smo izvedli z 22 kandidati.

### *Protokol intervjuja*

Intervjuji so potekali v prisotnosti dveh predstavnikov ADP, eden je vodil pogovor, drugi po potrebi dodatno povprašal, beležil, skrbel za delovanje snemalne naprave in drugo. Večina intervjujev je potekala individualno in brez motenj. Le v enem primeru je bil del intervjuja opravljen v prisotnosti nadrejene osebe, v nekaj primerih je prišlo do manjših nenačrtovanih motenj (obiski), ki so za krajši čas prekinili potek pogovora, a ga niso zmotili. V začetku intervjuja smo intervjuvancem dali v podpis *Izjavo o strinjanju s pogoji sodelovanja*,<sup>iii</sup> s katero so potrdili, da soglašajo s sodelovanjem v projektu.

Sledilo je izpolnjevanje osnovnih podatkov o intervjuvancu. Vpisali smo ime in priimek, organizacijo, na kateri intervjuvanec deluje, njegovo znanstveno področje, položaj, izobrazbo, leta dela in leto rojstva. Potem smo z nekaj uvodnimi vprašanji intervjuvanca vpeljali v temo pogovora. V začetnem delu intervjuja smo ga oziroma jo prosili, da na razpredelnici označi, katere vloge opravlja na področju podatkov in v katerih fazah življenjskega kroga podatkov sodeluje.<sup>iv</sup>

Uvodna vprašanja so bila iztočnica za nadaljevanje pogovora o fazah življenjskega kroga podatkov. Intervju je v povprečju trajal eno uro. Intervjuvanci so se pozitivno odzvali na pogovor in večinoma izkazali interes za morebitno nadaljnje sodelovanje.

Vse intervjuje smo zvočno posneli, vse posnetke prepisali, očistili in opremili z internetnimi povezavami na obravnavane projekte, formate, primere, zglede... 11 intervjuvancev se je odločilo, da želijo ohraniti anonimnost,<sup>v</sup> trije so ostali neopredeljeni. Intervjuji se pripravljajo za spravilo kot raziskovalni podatki projekta v podatkovnih arhivih ADP, kjer bodo po koncu projekta v ustrezno anonimizirani obliki dostopni uporabnikom za druge raziskovalne ali izobraževalne namene.

### **3.3 OSNOVNI PODATKI O VZORCU**

Od 31.3.2011 do 10.6.2011 smo opravili 22 intervjujev. Intervjuvanci prihajajo iz 17 različnih disciplin:

- biokemija in molekularna biologija,
- biologija,
- fizika,
- medicinske veze in statistika,
- geodezija,
- gradbeništvo,
- ekonomija,
- jezikovne tehnologije,

<sup>ii</sup> Priloga 2 – Priporočilno pismo za sodelovanje v projektu Odprti podatki.

<sup>iii</sup> Priloga 3 - Izjavo o strinjanju s pogoji sodelovanja.

<sup>iv</sup> Priloga 4 – Vprašalnik.

<sup>v</sup> "Potrdite, ali izrecno želite, da je vaše ime navedeno v poročilih in drugih gradivih projekta, ki se navezujejo na izrečeno v okviru intervjuja. "

- slovenistična literarna veda,
- glasbena akustika,
- zgodovina,
- arheologija,
- geografija, kartografija,
- urbanizem (specializirana knjižnica),
- socialna antropologija,
- socialno delo,
- statistika.

Večina intervjuvancev je zaposlenih na Univerzi v Ljubljani in na posameznih specializiranih raziskovalnih inštitutih. Vsi intervjuji so bili opravljeni v Ljubljani.

#### **Univerza v Ljubljani:**

- Filozofska fakulteta,
- Biotehniška fakulteta,
- Fakulteta za socialno delo,
- Ekonomska fakulteta,
- Fakulteta za matematiko in fiziko in
- Fakulteta za gradbeništvo in geodezijo.

#### **Inštituti:**

- Inštitut za novejšo zgodovino,
- Urbanistični inštitut RS,
- ZRC SAZU, Inštitut za slovenski jezik,
- ZRC SAZU, Inštitut za arheologijo,
- ZRC SAZU, Glasbenonarodopisni inštitut,
- Inštitut za biostatistiko in medicinsko informatiko, Medicinska fakulteta,
- Institut Jožef Stefan,
- Geodetski inštitut Slovenije in
- Inštitut za ekonomska raziskovanja.

Glede na spol so bili intervjuvanci približno enako razdeljeni. Več kot polovica vprašanih je pri izobrazbi dosegla stopnjo doktorata, ostali imajo končano najmanj univerzitetno izobrazbo. Najmlajši intervjuvanec je leta 2011 štel 35 let, najstarejši pa 75. V povprečju so bili intervjuvanci stari nekaj manj kot 47 let.

Povprašali smo jih tudi po položaju, ki ga zasedajo v organizaciji, kjer so zaposleni. Med intervjuvanci je sledeča struktura:

- 2 asistenta,
- 1 izredni profesor,
- 2 redna profesorja,
- 2 prodekana,
- 1 samostojni strokovni sodelavec,
- 1 raziskovalec,
- 2 znanstvena sodelavca,

- 1 višji znanstveni sodelavec,
- 1 znanstveni svetnik,
- 1 direktor inštituta,
- 1 pomočnik direktorja inštituta,
- 1 svetovalec v direktorjevi pisarni,
- 2 vodji skupine,
- 1 vodja znanstveno-informacijskega centra,
- 1 vodja projektov in
- 2 vodji knjižnice.

Izkazalo se je, da so intervjuvanci pogosto hkrati nosilci različnih vlog. Med 18, ki so odgovorili na vprašanje, katere vloge opravljajo na področju podatkov, se jih je 12 opredelilo za ustvarjalce, 14 jih je reklo, da opravljajo storitve s področja podatkov, 11 se jih je opredelilo za uporabnike podatkov in 9 jih dela na področju vodstva in oblikovanja politik.

Med 18, ki so opredelili svoje vloge, je 5 takih, ki so nosilci izključno ene same vloge – in sicer opravljajo storitve na področju podatkov. 3 so odgovorili, da opravljajo dve od naštetih vlog, 5 jih je reklo, da opravljajo tri vloge in 5, da so nosilci vseh naštetih vlog.

### 3.4 VPRAŠALNIK

V zasnovi vprašalnika smo izhajali iz naslednjih vprašanj. Kje so glavne pomanjkljivosti ustvarjanja, dostopa in hrambe raziskovalnih podatkov, ki bi jih bilo potrebno izboljšati? Ali gre za pomanjkanje znanja, vodil, pripravljenosti, odgovornosti? Kako akterji vidijo drug drugega: identificiranje medsebojnih odvisnosti in vlog, odgovornosti. Zanimalo nas je poznavanje in uporaba dobrih praks in standardov, ki bi lahko služili za določanje pomanjkljivosti pri ustvarjanju podatkov in medsebojno izmenjavo izkušenj za izboljšave.

Pri intervjujih smo izhajali iz pričakovanj, da so na nekaterih področjih odgovori na vprašanja že zelo jasno izoblikovani, medtem ko se bo drugje pokazala zelo nizka stopnja ozaveščenosti in razgledanosti. Na videz preprosto vprašanje, kje bi lahko objavili svoje podatke, lahko ima jasen odgovor - institucionalni repozitorij, domači ali svetovni disciplinarni repozitorij, revije, na domači strani raziskovalca, lahko pa ugotovimo, da intervjuvanci o tem nič ne vedo. Denimo eno od področij, ki je v razvitem svetu vse bolj izpostavljeno, to je zahteva, da objavo v reviji spremljajo tudi podatki oziroma da mora biti omogočen dostop do podatkov, je poznana le redkim ustvarjalcem podatkov, in sicer tistim, ki so ob objavi ali branju publikacije naleteli na tovrstna spremljajoča gradiva. Tudi če za določeno področje obstaja možnost spravila in storitev dostopa do podatkov, sta zahtevnost priprave podatkov pripravljenih po določenih standardih in pomanjkanje časa oviri za pripravljenost izkoristiti to možnost. Pogosto je pripravljenost za omogočanje dostopa do podatkov povezana tudi s transparentnostjo vseh postopkov in analiz. Kadar so problematični statistični postopki in drugi vidiki priprave ali dokumentacije podatkov, je nizka tudi pripravljenost izpostaviti in narediti vidno celoto vsega, kar je v ozadju ugotovitev v člankih.<sup>vi</sup>

Zaradi tega smo v intervjuje vključevali tudi hipotetična vprašanja, ali bi bili sposobni in ali bi jim bilo v interesu prevzeti vlogo servisiranja drugih področij, ali v smislu centrov za svetovanje, npr. pri pripravi DMP,<sup>vii</sup> ali izmenjave dobrih praks, ali pa tudi kot storitve hrambe in dostopa do podatkov. Kdo bi se lahko s tem ukvarjal? Knjižnica, raziskovalci, specializirani centri, splošen repozitorij za

<sup>vi</sup> Za več o tem, ko so statistične slabosti razlog za neobjavo podatkov, glej Heather Piwowar (2011).

<sup>vii</sup> Data Management Planning oziroma Načrtovano upravljanje s podatki.



podatke? Kakšni so nastavki te dejavnosti: kdo so favoriti, kje so že vzpostavljene povezave doma in v tujini, uporaba dobrih praks in standardov?

Pri oblikovanju vprašalnika smo izhajali iz podobnih študij drugje, večinoma omenjenih zgoraj, in imeli pred seboj seznam vprašanj in problemov, ki smo se jih hoteli dotakniti. Nismo vnaprej pričakovali, da bomo dobili zelo drugačne odgovore kot drugod. Bolj je bilo pomembno, da iz prve roke dobimo predstavo o stanju in možnih zadržkih ter ovirah pri vzpostavitvi bodočega sistema in politik odprtih podatkov. S čimer smo sledili tudi Parsonsovemu (2010, 4) stališču, da je ena od najpomembnejših kratkoročnih strategij to, da morajo podatkovne politike vključevati jasno opredelitev vlog, odgovornosti in virov. Zato je potrebno opredeliti mehanizme za deponiranje podatkov v arhive ter identificirati strokovnjake za hrambo in deljenje podatkov. Prvi korak do oblikovanja kakovostnih politik je temeljito poznavanje dejanskega stanja na področju: potrebe, ovire, zadržki, predstave, pričakovanja, izkušnje, razmerja med akterji...

## 4. ŽIVLJENJSKI KROG PODATKOV

---

Področje medsebojne izmenjave oziroma deljenja podatkov predpostavlja tudi odgovornost različnih nosilcev vlog. Ne zadostuje želja razvijati in spodbujati dejavnost storitev prevzema, skrb za podatke in dostop do njih, saj lahko tovrstna infrastruktura ostane neizkoriščena, če ni spodbud za raziskovalce - ustvarjalce podatkov. Zato je pomembno, da oblikovanje politik in strategij na tem področju zajema koordinirano naslavljanje problemov in njihovo reševanje z različnih perspektiv. Poleg tega mora vključevati vse akterje na tem področju, tudi financerje.

Idealnotipsko bomo na področju sistema dostopa do podatkov ločili naslednje nosilce aktivnosti: na eni strani so to nosilci in izvajalci raziskovalnih politik, kot sta na nacionalni ravni Ministrstvo za izobraževanje, znanost, kulturo in šport (MIZKŠ) in Javna agencija za raziskovalno dejavnost RS (ARRS) ter vodstva inštitucij in drugi organi odločanja ter na drugi strani člani znanstvene skupnosti, ki nastopajo v različnih vlogah kot uporabniki ali kot ustvarjalci in dajalci podatkov, kot povezovalni člen pa nastopajo nosilci podatkovnih infrastrukturnih storitev.

Preden preidemo k predstavitvi vprašalnika, moramo opredeliti pojme, s katerimi bomo poskušali pojasniti stanje na področju raziskovalnih podatkov in izpostaviti kategorije, za katere po zgledih tujih študij sklepamo, da pomagajo soustvarjati življenjski potek raziskovalnih podatkov: kaj se dogaja s podatki od načrtovanja in konceptualizacije naprej. Za čim bolj jasno predstavo o življenjski poti podatkov bomo opredelili vloge in njihove nosilce, objekte oziroma predmete in pojme, ki se nanašajo na podatke ter življenjski krog podatkov.

**Vloga** se uresničuje skozi navade in običaje, omejujejo jo pravila, norme, standardi, protokoli, licence, ki določajo in upravljajo raziskovalno dejavnost na posameznem področju. Od tega, kako so opredeljene in strukturirane vloge, je odvisna raven kulture in organiziranosti deljenja podatkov.

Zgled je lahko stališče Research Information Network (RIN), predstavljeno v *Stewardship of digital research data - principles and guidelines* (2008), da morajo biti vloge in odgovornosti raziskovalcev, raziskovalnih ustanov in financerjev, kolikor je mogoče, podrobno in natančno opredeljene. Akterji morajo skupaj oblikovati kodeks delovanja, ki bo zagotavljal, da bodo ustvarjalci in uporabniki podatkov izpolnjevali svoje odgovornosti skladno z opredeljenimi načeli.

Swan in Brown (2008, 2) ugotavljata, da se je večina znanstvenikov priučila vloge vezane na podatke šele ob delu. Pridobiti so si morali veliko različnih veščin. Ob področni strokovni usposobljenosti in računalniškem znanju, ki sta za znanstvenika bolj ali manj samoumevna, za najpomembnejšo štejejo sposobnost 'prevajanja' oziroma prenosa potreb in praks znanstvenikov v jezik tehnoloških orodij in potem spet nazaj v prakso.

Vloge se izražajo in uresničujejo z akterji, ki delujejo na področju raziskovalnih podatkov. Nosilci vlog uresničujejo pravice, dolžnosti in obveznosti ali pa tudi ne. Iz položaja nosilca vloge v znanstveno-raziskovalni skupnosti izvira stopnja moči ter njegove prednosti in slabosti pri izvajanju nalog. Glavni nosilci vlog so, po zgledu Liz Lyon (2007) in Interagency Working Group on Digital Data (2009):

- ustvarjalec podatkov: podatkovni znanstvenik, raziskovalna enota, inštitut;

- izvajalec storitev: skrb za podatke, katalogiziranje, opremljanje z metapodatki za uporabo, spravilo, povezovanje, specializirani področni podatkovni arhivi, mednarodni arhivi, digitalni institucionalni ali nacionalni repozitoriji, knjižnice;
- uporabnik (drugi raziskovalci, študentje, širša javnost);
- vodstvo ustanov in oblikovalci politik, financerji;

Eden od projektnih ciljev je ugotoviti dobre strani, jih podpreti in okrepiti, ter v njihovi smeri načrtovati oblikovanje bodočih storitev. Pri pregledu stanja smo pozornost namenili razmerju moč/nemoč v povezavi s pristojnostmi in medsebojnimi pričakovanji o tem, kdo lahko izpolni določene naloge. Zgledi dobrih praks naj služijo v procesu opolnomočenja manj razvitih področij.

**Objekti.** Za razumevanje in razlago procesov, ki spremljajo življenjski krog podatkov, je pomembnih več objektov, pojmov, vsebin. Najprej seveda *podatki*, ki jih delimo na: surove, predelane in izvedene. Podatke morajo spremljati *metapodatki*, ki so neke vrste dokumentacija o podatkih in so ključnega pomena za razumevanje podatkov in za drugo rabo. (Vipavc Brvar in Klep 2004) *Publikacije* so pomembna institucija za predstavitev ugotovitev, temelječih na podatkih. Vse bolj postajajo pomembna institucija tudi za predstavitev podatkov kot takih, npr. kot podatkovna objava. *Orodja* za ustvarjanje, prikaz in analizo podatkov zadevajo tehnični vidik ustvarjanja podatkov in ravnanja z njimi. *Standardi, dobre prakse, navodila* o kakovosti, metapodatkih in hrambi obsegajo normativni vidik, pisana in nepisana pravila znanstvene skupnosti, znanstvenih skupin ali posameznih znanstvenikov. Ter so neposredno pomembne za ustvarjanje in drugo rabo podatkov. Nad vsem tem pa visijo *politike, strategije, načrti, projekti*, ki kratkoročno in/ali dolgoročno vplivajo na celo področje problemov povezanih s podatki. Iz njih izvirajo *pravice, dolžnosti in odgovornosti* akterjev. Iz njih pa izhajajo tudi *nagrade, spodbude in sankcije*, ki lahko okrepijo določene aktivnosti in zmanjšajo druge. Na kratko rečeno, tehnični vidik in kulturo ustvarjanja in rabe podatkov, ki se zrcali v življenjskem krogu podatkov, je mogoče razložiti le s poznavanjem vsebin in razmerij med nosilci vlog in objekti.

#### 4.1 FAZE ŽIVLJENJSKEGA KROGA PODATKOV

Življenjska doba podatkov je običajno daljša, kot je trajanje projekta, v okviru katerega so nastali. Saj lahko raziskovalci podatke uporabijo ali nadgradijo v sorodnih drugih projektih, lahko jih pa uporabijo tudi povsem drugi raziskovalci, so zapisali v UK Data Archive. Življenjsko pot podatkov vidijo kot preprost krog, v katerem si sledijo naslednje faze: ustvarjanje podatkov, obravnava podatkov, analiza, hramba, zagotavljanje dostopa do podatkov in druga raba podatkov.<sup>1</sup>

V Digital Curation Centre (DCC) so oblikovali bolj podroben življenjski krog podatkov, ki vključuje več stopenj življenjske poti podatkov in upošteva različen značaj aktivnosti. V ospredje so postavili stopnje, ki zagotavljajo uspešno hrambo in zaščito podatkov, od zasnove do postopkov hrambe. Njihov model je namenjen načrtovanju aktivnosti v organizacijah.

Opredelili so aktivnosti, ki so posebej pomembne zato, ker potekajo ves čas življenjskega kroga podatkov: zagotavljanje metapodatkov, načrtovanje digitalne hrambe, spremljanje skupnosti in udeležba pri razvoju standardov, orodij ter shranjevanje in digitalno skrbništvo v teku celotnega življenjskega kroga podatkov.

Poleg dejavnosti, ki potekajo ves čas, so pomembne te, ki potekajo v določenem zaporedju. Prva med njimi je konceptualizacija (načrtovanje ustvarjanja podatkov, vključno s hrambo). Sledijo ji:

---

<sup>1</sup> Za shematski prikaz glej: Research Data Lifecycle. <http://www.data-archive.ac.uk/create-manage/life-cycle>, 7.8.2012

ustvarjanje podatkov ali prejem podatkov od drugih; ovrednotenje in izbor; sprejem podatkov v hrambo; aktivnosti hrambe, spravilo; dostop, raba in druga raba; predelava podatkov.

Spregledati pa ne smemo niti občasnih aktivnosti, ki so prav tako pomembne za ustrežno hrambo in vzdrževanje podatkov: prenos podatkov, ki niso bili izbrani za hrambo; ponovna ocena vrednosti podatkov in dodatne aktivnosti za izboljšanje kakovosti; pretvorba podatkov v druge formate ipd.<sup>ii</sup>

Na osnovi tovrstnih modelov smo pripravili poenostavljen slikovni prikaz, ki nam je služil kot opora za oblikovanje vprašalnika in izvedbo polstrukturiranega intervjuja.



Življenjski krog podatkov, kot smo ga obravnavali v intervjujih, je strukturiran po fazah, kot sledi: načrtovanje, financiranje, ustvarjanje, dodana vrednost, izbor/vrednotenje kakovosti in uporabnosti, spravilo, analiza/objava rezultatov v članku, dostop/iskanje/druge raba.

Posamezne faze so služile kot osnova za oblikovanje področnih sklopov vprašanj, ki jih bomo predstavili v nadaljevanju.

Med oblikovanjem vprašalnika se je izkazalo, da je določene sklope smotno združiti, med pogovori pa je postalo jasno, da intervjuvanci večinoma govorijo o več fazah hkrati, zato se je nemalokrat struktura pogovora sprostita in je sledila temam, ki so bile intervjuvancu bliže, ali jih je zmožgel bolje opredeliti, pojasniti ali preprosto izraziti svoje stališče.

<sup>ii</sup> Za shematski prikaz glej: The DCC Curation Lifecycle Model.  
<http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>, 7.8.2012

## 5. PRIKAZ REZULTATOV V FAZI USTVARJANJA PODATKOV

---

Prva faza življenjskega kroga podatkov obsega *opis nastajanja podatkov*, z njimi povezana orodja, postopke in navodila. Dvanajst od osemnajstih intervjuvancev je odgovorilo, da je ustvarjanje podatkov ena od njihovih vlog. Izkazalo se je, da se intervjuvanci med seboj zelo razlikujejo po tem, kakšne vrste in tipe podatkov ustvarjajo in kakšne metode pri tem uporabljajo ter s kakšno infrastrukturo razpolagajo.

Zanimalo nas je, kako nastajajo podatki in metapodatki: za kakšne vrste podatkov gre in kaj se z njimi dogaja? Kako jih ustvarjajo in/ali kako z njimi ravnajo? Od intervjuvancev smo skušali izvedeti, *kako vidijo odgovornost* (lastno in drugih) v zvezi z reševanjem problemov? Čemu so zavezani s pravili: zahteve projekta glede kakovosti, eksplicitna merila kakovosti, ki jih upoštevajo pri ustvarjanju podatkov; beleženje karakteristik glede kakovosti pri nastajanju podatkov? Zanimalo nas je, kakšna je odgovornost do naročnika, financerja, vodstva, uporabnika: ali že v načrtu opredelijo, kako in kakšne vrste podatkov bodo ustvarili? Za koga proizvajajo podatke: zase, za ožji krog, za druge? Ali odgovornost do drugih uporabnikov razumejo v smislu skrbi za kakovost? Intervjuvance smo prosili, naj navedejo primere problemov: zanimalo nas je, ali identificira probleme, pozna rešitve, vidi vlogo standardov, ukrepov različnih akterjev? Na tej osnovi smo izluščili nosilce problemov in nosilce zgledov reševanja, ki bodo v nadaljevanju osnova za preučevanje možnosti in pogojev za uveljavljanje dobrih praks.

Zanimalo nas je, kateri *problemi* iz faze ustvarjanja podatkov so povezani z drugimi fazami življenjskega kroga, ali jih akterji upoštevajo že v tej fazi ali ne, ali načrtujejo, kaj se bo zbiralo in za kak namen, ali upoštevajo vrednotenje in izbor, da bi z načrtovanjem dosegli kakovostne podatke? Ali upoštevajo zahteve hrambe in ali razrešujejo probleme na poti do zagotavljanja odprtega dostopa?

### 5.1 TIPI PODATKOV

Že na uvodnem srečanju predstavnikov različnih slovenskih inštitutov, ki ga je januarja 2011 v sodelovanju z Ministrstvom za visoko šolstvo, znanost in tehnologijo organiziral izvajalec projekta, in tudi kasneje v začetnem delu intervjujev, smo sogovornikom vsakič znova zastavili osnovno vprašanje, kaj so podatki za različne akterje in na različnih področjih. Ali denimo numerične simulacije predstavljajo podatke? Kako je s podatki v humanistiki, kjer običajno uporabljajo drugačno terminologijo, denimo zgodovinarji govorijo o virih? Po začetni negotovosti smo se običajno hitro sporazumeli glede namenov projekta, tako na omenjenem sestanku kot pri intervjujih so izražali naklonjenost in podporo ciljem. Udeleženci prvega sestanka so na nas naslovili razmeroma veliko vprašanj, pripomb in predlogov, ki smo jih v nadaljevanju naslovili tudi v okviru intervjujev, kot so določitev meril kakovosti podatkov, vprašanja plačljivosti in dostopa do individualnih podatkov, dostopnost naravoslovnih podatkov v primerjavi z družboslovnimi, izraženi so bili pomisleki glede avtorskih pravic.

Nekaj osnovnih dilem se je nanašalo, kot omenjeno, na samo izhodiščno definicijo 'raziskovalnih podatkov', ki bi predstavnikom raziskovalnih ustanov pomagala osredotočiti se na probleme. Z naše strani je bil podan predlog, da naj raziskovalci sami opredelijo, katere podatke na svojih področjih štejejo kot pomembne in pod kakšnimi pogoji so jih pripravljene deliti. Prav tako pa smo nosilce aktivnosti na začetku intervjujev zaprosili, da sami podrobneje opišejo, s kakšnimi podatki ravnajo v svoji dejavnosti, izhajajoč iz zelo splošnih opredelitev.

Za boljšo predstavitev problematike opredeljevanja raznolikih vrst podatkov v nadaljevanju povzemamo nekaj tovrstnih splošnih definicij podatkov, ki sta jih v *IASSIST Quarterly* povzela Yoon in Tibbo (2011):

V The National Research Council (1999, 15) podatke opredelijo kot "dejstva, številke, črke in simbole, ki opisujejo predmete, zamisli, pogoje, situacije ali druge dejavnike".

V The National Science Board (2005, 13) pojem 'podatki' rabijo za "kakršno koli informacijo, vključujoč besedilo, številke, podobe, video ali filme, avdio, programsko opremo, algoritme, enačbe, animacije, modele, simulacije in drugo."

V The National Science Foundation razlikujejo štiri tipe podatkov:

- podatki opazovanj (npr. vreme, človeško ravnanje),
- računalniško ustvarjeni podatki (rezultati računalniških modelov in simulacij),
- eksperimentalni (iz laboratorijskih študij),
- zapisi (s strani vlad, podjetij, javnega in zasebnega življenja) (povzeto po Borgman 2010, 19).

Tu je še nekaj različnih definicij raziskovalnih podatkov, s katerimi si avstralske univerze pomagajo pri upravljanju z raziskovalnimi podatki (ANDS 2011):

"Pojem raziskovalni podatki obsega dejstva, opažanja, podobe, rezultate računalniških programov, zapisi, meritve ali izkustva, na katerih temelji argument, teorija, preizkus, hipoteza ali kak drug raziskovalni izsledok." so zapisali v svojih politikah o upravljanju s podatki na *The Queensland University of Technology*.

Na *The University of Melbourne* so raziskovalne podatke opredelili kot "dejstva, opažanja ali izkustva, na katerih temelji argument, teorija ali preizkus. Podatki so lahko numerični, opisni ali vizualni. Podatki so lahko surovi ali analizirani, eksperimentalni ali podatki opazovanj. Podatki vključujejo: laboratorijske in terenske zapiske, primarne raziskovalne podatke (vključujoč raziskovalne podatke v tiskani ali računalniški obliki), vprašalnike, avdiokasete, videokasete, modele, fotografije, filme, testne odgovore. Raziskovalne zbirke lahko vključujejo: diapozitive, artefakte, primerke in vzorce. Izvirne informacije o podatkih lahko vključujejo: kako, kdaj in kje so bili zbrani podatki in s čim (npr. naprava). Obsega lahko tudi programsko kodo, s katero so ustvarili, komentirali ali analizirali podatke."

Na *The Monash University* raziskovalne podatke razumejo kot: "Podatki, posnetki, dokumenti ali drugi dokazi, ne glede na vsebino ali obliko (to je tiskana, digitalna, fizična ali druga oblika), ki vključuje raziskovalna opažanja, ugotovitve ali izsledke, vključujoč primarno gradivo in analizirane podatke."

Na *Griffith University* razlikujejo med raziskovalnimi podatki in primarnim gradivom. Raziskovalni podatki so: zapiski v obliki števil, simbolov, besedila, podob ali zvokov, uporabljeni kot primarni vir za raziskavo in so sprejeti v raziskovalni skupnosti kot nujni za vrednotenje raziskovalnih ugotovitev.

Pri oblikovanju definicije, kaj so raziskovalni podatki, si je mogoče pomagati tudi z opredelitvijo, kateri podatki vanje ne sodijo:

- administrativni podatki: četudi se nekateri nanašajo na raziskovalne projekte in jih je mogoče videti kot raziskovalne podatke, zanje večinoma poskrbijo v okviru upravnih postopkov znotraj inštitucije (menedžment, strategije);
- učni podatki: so del učnega procesa na univerzi, običajno zanje skrbijo neodvisno;
- raziskovalne objave: mogoče jih je videti kot podatke, vendar je zanje večinoma dobro poskrbljeno s strani založnikov (ANDS 2011).

Na kratko bomo strnili, da se definicije nanašajo na obliko podatkov in na nosilce podatkov. V nekaterih primerih opišejo tudi, na kak način so podatki vključeni v raziskovalni proces. V nadaljevanju pogledimo, kako raziskovalne podatke vidijo udeleženci naše raziskave.

*Jaz si ne predstavljam, kaj je za vas podatek. Običajno si pod to predstavljam neko surovo številko, ali pa neke vrednosti, ker pač s tem največ delam.* (znanstvenik s področja geografije, NO8)

*Mi nimamo podatkov, mi se ukvarjamo z viri.* (znanstvenik s področja zgodovine, NO10)

*Podatek ni samo grob podatek. Lahko je knjiga, ki jo je Charles Dickinson napisal, lahko je študija, ki jo je Charles Dickinson napisal, lahko so njegove lingvistične analize ... Nekaj, kar jaz uporabim kot podatek in naredim analizo, bo nekdo drug uporabil kot podatek.* (Nina Vodopivec, raziskovalka s področja socialne antropologije)

Z intervjuji smo uspešno pokrili več različnih področij tako v družboslovju in humanistiki kot v naravoslovju, kar se odraža v širokem naboru zelo raznolikih tipov podatkov, ki jih zbirajo v posameznih inštitucijah.<sup>iii</sup>

Izkazalo se je, da že sama opredelitev, kaj je podatek, v različnih znanstvenih disciplinah ni povsem samoumevna. Nabor zabeleženih tipov podatkov je temu primerno obsežen. Intervjuvanci naše raziskave so navedli različne tipe podatkov, s katerimi se srečujejo pri raziskovalnem delu:

- numerični podatki, meritve,
- rezultati numeričnih modelov,
- ekonomski modeli,
- anketni podatki,
- intervjuji,
- dnevniški zapisi,
- delovna poročila,
- slike,
- grafi,
- risbe,
- besede (besedilni korpus),
- pisna gradiva,
- statistični letopisi,
- popisi prebivalstva,
- bibliografske podatkovne zbirke,
- zvočni zapisi,

---

<sup>iii</sup> Iz tega sledi sklep, da bo bodoče "izvajanje storitev skrbništva nad podatki potrebno prilagoditi močno raznolikim pod-disciplinarnim karakteristikam podatkov in navadam deljenja podatkov." (Cragin in drugi 2010)

- video,
- programska oprema,
- strojna oprema,
- aplikacije,
- spektri,
- senzogrami,
- lidarski podatki,<sup>iv</sup>
- herbariji,
- zbirke živali,
- vzorci tkiv.

Na osnovi odgovorov intervjuvancev na vprašanje, s kakšnimi podatki se ukvarjajo, smo ugotovili, da podatke pogosto predstavljajo običajni numerični podatki v obliki tabel, podatkovnih baz ipd., veliko pa je tudi pisnega, slikovnega, avdio in video gradiva. Ugotovili smo, da obstaja tudi kar nekaj bolj posebnih zbirk podatkov, kot so npr. herbariji, zbirke živali, vzorci tkiv, lidarski podatki, spektri, numerični modeli ipd. Kar nekaj intervjuvancev je v povezavi z opredelitvijo podatkov navedlo tudi svoje produkte v obliki programske ali strojne opreme ter aplikacije, s pomočjo katerih se podatki generirajo.

Tako kot v National Science Board report bi bilo mogoče tudi v našem naboru podatke razdeliti v naslednje štiri kategorije: opazovanja, simulacije, eksperimenti in zapisi. Pogosto se podatke deli tudi na t. i. referenčne zbirke in ostale podatke (Long-Lived Digital Data Collections 2005). Tudi nekateri naši intervjuvanci so "imeli težave" pri razumevanju, da so podatki lahko tudi še kaj drugega kot so referenčne zbirke,<sup>v</sup> ki so na določenih področjih najbolj opazne. Velikokrat se je razumevanje pojma podatek izčistilo in razjasnilo šele tekom intervjuja, med pogovorom so se ostrile razlike med tem, kaj je podatek, kaj podatkovna zbirka.<sup>1,2,3</sup> Na različnih področjih so se pokazale razlike v pojmovanju primarnega gradiva, razlike med surovimi podatki in predelanimi podatki, dokumentacijo, zapiski... Na primer razlikovanje med arheološkim najdiščem in podatki povezanimi z lokacijo, najdbo, izkopavanjem in drugo.<sup>4</sup> Včasih je bilo potrebno zastaviti dodatno vprašanje, da smo v pogovoru prišli do osnovne ravni raziskovanja, do primarnih podatkov.<sup>5</sup> Tovrstne razlike v pojmovanju bo potrebno upoštevati v pripravi akcijskega načrta.

Glede na metodo zbiranja je tipe podatkov mogoče razdeliti v naslednje skupine:

- podatki, ki jih zberejo raziskovalci sami,
- podatki narejeni z aparaturami in
- podatki, katerih osnova za generiranje so drugi viri npr. vzorci, kartoteke.<sup>6</sup>

<sup>iv</sup> Podatki pridobljeni z metodo zračnega laserskega skeniranja.

<sup>v</sup> Npr. SARNAS (Seznam Arheoloških najdišč Slovenije) in druge zbirke, dostopne preko spletnega modela ARZENAL, več o tem glej: <http://www.arzenal.si/sobe/zbirke>, 7. 8. 2012.



### Okvirček 5.1

#### **Medicina: Inštitut za biomedicinsko informatiko (IBMI)**

IBMI je ustanovila Medicinska fakulteta<sup>vi</sup>, in sicer leta 1973 s ciljem, da se bo ustanova posvečala informacijam, ki so pomembne za raziskovalno delo. Že od samih začetkov v inštitutu ločijo strokovni in izobraževalni oziroma raziskovalni program. Jedro strokovnega programa tvori slovenska nacionalna bibliografska zbirka Biomedicina Slovenica, ki vsebuje podatke o slovenskih znanstvenih in strokovnih dokumentih s področja biomedicine.

V okviru inštituta deluje biostatistični center, ki vodi raziskovalni program in sodeluje pri raziskovalnih projektih s poudarjeno metodološko in statistično podporo pri delu na raziskovalnih podatkih s področja medicine, svetuje tako pri ustvarjanju kot pri analizi.

Vir: <http://ibmi.mf.uni-lj.si/ibmi/index.html>

V nekaterih inštitucijah je poleg ustvarjanja lastnih podatkov pogosta praksa tudi druga raba podatkov, zlasti je razširjena raba podatkov, ki so jih ustvarile za to uradno zadolžene javne ustanove (SURS, ARSO, AJPES, Geodetska uprava RS, različni javni arhivi) ter nadaljnje delo z njimi, tudi preurejanje in združevanje. Intervjuvanci so omenili več problemov povezanih s tem področjem. Denimo, kakšna je dostopnost do podatkov po tem, ko jih je raziskovalec dopolnil in preuredil za svoje namene. V tem pogledu je pogosto sklicevanje na stroge pogoje, pod katerimi so raziskovalci prišli do podatkov. Te pogoje je mogoče razumeti tudi kot izgovor, da podatkov v principu ni mogoče posredovati naprej. Npr. na področju kartografije in geografije so se zavezali, da bodo podatke uporabili samo za lastne potrebe. Kar lahko delijo z drugimi, so novo nastali produkti (npr. karte), ne smejo pa deliti oziroma posredovati surovih podatkov.<sup>7</sup> Podobno velja tudi na področju ekonomije, kjer imajo dostop do določenih podatkov, vendar jih ne smejo prenašati naokoli, kot se je izrazil intervjuvanec.<sup>8</sup> Poleg tovrstnih omejitev pa raziskovalce omejujejo tudi drugi pogoji. Na področju geodezije raziskovalci npr. lahko dobijo koordinate. V kolikor bi želeli podatke, ki segajo na področje varovanja osebnih podatkov, se pogoji zaostrijo.<sup>9</sup> Podobno velja tudi na področju ekonomije, kjer se raziskovalci ukvarjajo s preučevanjem podatkov o posameznikih ali podjetjih, za kar so posebej določeni pogoji.<sup>10</sup>

Rešitve problemov oziroma vprašanj, kot npr. pogoji dostopa do podatkov v javnih in državnih ustanovah ter do samih registrov obstoječih podatkov, po katerih je mogoče iskati, za to je denimo mogoče pooblastiti področne podatkovne arhive, ki bodo posredovali pri dostopu do podatkov. Primer takega sodelovanja že poteka med ADP in SURS.

Intervjuvanci deloma poznajo tudi rešitve, ki so uveljavljene v tujini. Npr. intervjuvanec s področja ekonomije pravi, da imajo v ZDA raziskovalci takojšen dostop do statističnih podatkov, do katerih morajo pri nas čakati tudi do pol leta. Želi si, da bi tudi v Sloveniji bile mogoče takšne rešitve.<sup>11</sup>

Poseben sklop predstavljajo podatki različnih (mednarodnih) centrov za hrambo in distribucijo specifičnih vrst podatkov, kot so GenBank, UniProt, Protein Data Bank, Amadeus, ipd. Nekateri podatki iz takšnih baz so prosto dostopni, do drugih lahko raziskovalci dostopajo le proti plačilu.

V Sloveniji je pomembno upoštevati tudi razlike med področji oziroma delitev na 'velike' ('big science') in 'majhne' znanosti ('small science').<sup>vii</sup> Veliko vrst podatkov izhaja iz domen 'majhnih'

<sup>vi</sup> <http://www.mf.uni-lj.si/>, 7. 8. 2012

<sup>vii</sup> "Trenutno obstoječa znanja s področij ravnanja s podatki in skrbništva v akademskem okolju zadevata dva problema. Prvič, stalna raba fizike delcev visokih energij in/ali astronomije kot referenčne in primerjalne točke

znanosti in so zaradi tega pogosto zapostavljeni v okviru specializiranih podatkovnih arhivov. Raziskovalec, ki ima izkušnjo z dvema različnima raziskovalnima področjema (jezikoslovje in fizika) odgovarja, da je na področju fizike delcev zavest glede hrambe in ravnanja s podatki visoka in problemi drugačni kot na področju družboslovja.<sup>12</sup> Za 'velike' znanosti je ponavadi dobro poskrbljeno, uveljavljeni so standardizirani zapisi in skrb za podatke je dobro vzpostavljena, podatki so zmeraj na voljo v elektronski obliki. Področje je zelo napredno, hkrati pa zelo ozko opredeljeno, samo za določen niz podatkov.

#### **Okvir 5.1: P-R1**

##### **PROBLEM:**

Podatki so zelo raznoliki, raziskovalci različno dojemajo, kaj so podatki in kateri so pomembni v njihovem raziskovalnem postopku. Opaziti je odsotnost jasne definicije, kaj podatki sploh so.

Področje 'velikih' znanosti je precej urejeno in ne terjaja iskanja novih rešitev, še zlasti, ker so uveljavljene in sprejete mednarodne rešitve. Opazno manj je urejeno področje 'majhnih' znanosti.

##### **REŠITEV:**

Pri oblikovanju bodočih načrtov bo potrebno upoštevati kompleksnost in različne zahteve različnih področij.

V iskanju rešitev se bo potrebno posebej osredotočiti na 'majhne' znanosti, saj v nacionalnem okolju, ki ga pokrivamo, prevladujejo prav te.

Pri iskanju rešitev so smiselne razprave med raziskovalci in nosilci storitev iz različnih področij. Področno povezovanje je treba dopolniti z naprednimi zgledi in izkušnjami 'velikih' znanosti, pri katerih so definicije, postopki in rešitve že preverjene in uveljavljene, v kolikor so prenosljivi na druga področja.

Odvisno od tega, ali storitve upravljanja in dostopa do podatkov že nastajajo ali ne, je smiselno oblikovanje razvojnih strategij različnih hitrosti. Na področjih, na katerih so akterji izrazili interes, da so pripravljene prevzeti naloge povezane s podatkovnimi storitvami, je potrebno omogočiti razvoj in ne čakati na področja, ki zaostajajo.

## **5.2 VPRAŠANJA KAKOVOSTI PODATKOV V POGLEDIH INTERVJUANCEV**

Vprašanje kakovosti je še eden od poudarkov, ki smo si jih izbrali za pregled problemov odprtega dostopa do raziskovalnih podatkov. Kakovost se nanaša na tako na elementarno kakovost samih podatkovnih gradiv, kot tudi kakovost metapodatkov in storitev povezanih z gradivi (npr. časovni rok za dostop, stopnja čiščenja in dodane vrednosti podatkov).

Ocena kakovosti podatkov je lahko odvisna tudi od tega, ali so podatki ustvarjeni za ozek namen, npr.

---

prikrije kompleksnost ustvarjanja in komuniciranja, ki je značilno za majhne znanosti in pod-discipline. (npr. Chompalov in drugi 2002) Drugič, osredotočanje na univerzalne rešitve glede nalog skrbništva nad podatki običajno prinaša zgolj modele skupnih storitev in modele na najbolj splošnem nivoju." (Cragin in drugi 2010, 4025) Podoben je tudi zaključek drugih avtorjev: "Veliki preostanek ('The long tail', op.a.) malih podatkovnih datotek predstavlja poseben izziv – to, kako heterogene podatkovne datoteke privesti skupaj." (Theesen in Patterson 2011, 43)

za izdelavo raziskovalnega poročila v okviru lastnega projekta ali pa so namenjeni za širšo uporabo, npr. za odprt dostop. Kakovost podatkov kot taka obsega več vidikov, lahko pomeni temeljno zmanjševanje napake, s ciljem, da bi se z meritvami približali pravi vrednosti. Zajema pa tudi širše organizacijske vidike, ki vključujejo poznavanje pravil in postopkov, načrtovanje in usposabljanje za kakovostne storitve v okolju ravnanja s podatki: od nastanka preko dodane vrednosti in do zagotavljanja dostopa in rabe pri generiranju raziskovalnih poročil (Primerjaj Biemer in Lyberg, 2003).

### **5.3 POMEN KAKOVOSTNIH PODATKOV IN METAPODATKOV**

Tukaj navedena izhodišča za ocenjevanje kakovosti podatkov in metapodatkov bodo služila za primerjavo z zaključki, ki smo jih dobili na podlagi intervjujev, kjer je bil vidik kakovosti v ospredju vseh fazah izdelave in skrbi za podatke.

Najširša opredelitev kakovosti podatkov v okviru sistema odprtega dostopa je po kriteriju uporabnosti podatkov za določen namen. Neuporabnost je lahko posledica napak pri ustvarjanju podatkov, neustrezne dokumentacije ali neočiščenih podatkov. Kakovostna storitev zagotavljanja dostopa pomeni, da je uporabnost gradiv maksimizirana, pri tem pa so upoštevani različni pogoji: kulturni, pravni in tehnični (van der Graaf in Waaijers 2011).

Prioritetna skrb tistih, ki si prizadevajo za čim večji in tem boljši izkoristek podatkov, mora biti kakovost podatkov. Zagotovilo, da postane velika količina podatkov dostopnih in da so pripravljene za analizo, je dobro upravljanje z njimi. Ravnanje s podatki mora biti načrtovano, vodeno, o njem teče komunikacija in je tudi ovrednoteno. Prav tako je pomembno vprašanje ocenjevanja vrednosti podatkov za nadaljnjo izrabo in selekcije, ki upošteva povezljivost različnih tipov podatkov, potencial, ki ga podatki nosijo s seboj - nekateri izčrpajo svojo vrednost že v objavi in nato služijo bolj kot dokument, ki spremlja objavo v smislu preverljivosti.

Več mednarodnih institucionalnih pobud se neposredno ukvarja z vsebino in formati, in sicer s ciljem, da bi zagotovili čim bolj kakovostne podatke in kakovostne podatkovne storitve. *European Strategy Forum on Research Infrastructure* v Poročilu o upravljanju s podatki iz leta 2009 kot enega pomembnejših ciljev na poti do E-Znanosti izpostavi kakovostne podatke in tudi kakovostne metapodatke. Poudarjajo, da mora biti zapisovanje metapodatkov dovolj prilagodljivo, da je mogoče dodajati nove elemente, različne tipe, ki bodo omogočali raznoliko rabo. Poročilo se podrobno posveča temam, kot so: raba, obseg, izvor, trajanje, agregacija, standardizacija, povezljivost, kakovost, čas in razpoložljivost, s ciljem, da bi dosegli boljši izkoristek podatkovnega potenciala.

Podobne zaključke glede vprašanj kakovosti povzema tudi ekspertiza Kevina Ashleya, direktorja Digital Curation Center, Velika Britanija. Med drugim izpostavlja pogosto neusklajenost ali parcialnost meril kakovosti, uporabljenih na različnih vrstah podatkov. "Vsi akterji v vseh fazah življenjskega kroga podatkov skrbijo za kakovost podatkov, študije pa kažejo, da imajo le redko enaka merila kakovosti: jaz želim obsežne podatke, ti jih želiš pravočasno, nekdo tretji želi točne podatke in četrti hoče prosti dostop." Velikokrat je sicer te zahteve mogoče uskladiti, kdaj pa so si tudi v nasprotju, saj na primer obsežnih podatkov ni mogoče ustvariti hitro in poceni (Ashley 2011).

Pomemben vidik tako razumljene kakovosti storitev odprtega dostopa predstavlja upoštevanje in vključevanje strokovnih kriterijev na več področjih: določanje vsebine zbirk podatkov (načrtovanje in selekcija, kateri vidiki bodo zajeti), ustvarjanje podatkov (kako bodo generirani podatki) in dodana vrednost (čiščenje, dokumentacija, priprava na shranjevanje).

Merila strokovnosti je potrebno upoštevati tudi pri določanju pogojev dostopa. V določenem segmentu storitev je potrebno vsebinsko, ožje poznavanje načel stroke, ki ga je mogoče pridobiti s strokovnim usposabljanjem in v sodelovanju z izkušenimi raziskovalci. V tem se kaže velik pomen ožjih strokovnih področij pri opredeljevanju in izvajanju ukrepov za doseganje kakovosti.

#### *5.4 POSTOPKI IN MERILA ZA ZAGOTAVLJANJE KAKOVOSTI V FAZI USTVARJANJA*

Na splošno lahko rečemo, da so udeleženci te raziskave izrazili visoko skrb za kakovost v fazi ustvarjanja podatkov, tako v zavedanju glede pomena kakovosti kot glede poznavanja postopkov in meril. Na standardiziranih področjih ustvarjalci natančno poznajo postopke in se pri ustvarjanju držijo predpisanih protokolov.<sup>13, 14</sup> Konsenz, ki so ga sprejeli v stroki, ni le ustno pravilo, tradicija ali običaj, pač pa je opredeljen v pisni obliki kot neke vrste pravilnik ali protokol, ki ga je strokovnjak dolžan izpolniti. To pomeni, da pravila niso odvisna zgolj od osebnih navad, prepričanj, izkušenj posameznih nosilcev vlog, pač pa so formalizirana in institucionalizirana. V takšnih primerih imajo ustvarjalci preverjeno podlago, iz katere črpajo. Imajo odgovore na vprašanja: kaj je kakovosten postopek (npr. eksperiment), kako ga kakovostno izvesti ter kaj so kakovostni podatki in kako jih ustvariti. Zanimiv je primer s polja biologije, kjer je skrb za kakovost izjemno dodelana, do ravni certificiranja laboratorijev za določene meritve. V takem primeru je sistemsko poskrbljeno, da podatki dosežejo zahtevano kakovostno raven.<sup>15</sup>

Posamezniki, skupine, laboratoriji uporabljajo tudi znanja in postopke, ki niso vselej formalizirane, pač pa se jih mora raziskovalec priučiti ali jih pridobiti z lastnimi izkušnjami. Nepisana pravila so skrb vsakokratnega ustvarjalca in/ali njegove skupine.<sup>16, 17</sup> Primer s področja ekonomije kaže, da so ponekod vzpostavljeni načini ocenjevanja kakovosti podatkov v smislu napake. Razumevanje je vzpostavljeno na ravni bližnjega poznavanja samega ustvarjalca podatkov, ki pa temelji na merilih stroke in v tem okviru opredeljenih in uveljavljenih načinov preverjanja in zagotavljanja kakovosti.<sup>18</sup>

Med intervjuvanci so tudi takšni, ki imajo izkušnje z mednarodnimi zbirkami, npr. zbirka genskih podatkov. Na področju, s katerega prihaja intervjuvanec, so razvili dodatne aplikacije za zagotavljanje kakovosti podatkov, namenjenih hrambi. Pred oddajo podatkov je raziskovalec primoran iti skozi postopek, v katerem podatke opremi z ustreznimi metapodatki.<sup>19</sup> Ta postopek je v pomoč pri preverjanju kakovosti razpoložljivih podatkov tako ustvarjalcem kot tudi uporabnikom.<sup>20</sup>

Okvir 5.2.: P-R 2

**PROBLEM:**

Opazen je velik razkorak med ustvarjalci, ki delajo v okolju formaliziranih, področno in mednarodno standardiziranih pogojev in drugimi, ki so pri delu prepuščeni lastni iznajdljivosti in izkušnjam.

Po drugi strani obstaja relativna visoka ozaveščenost o pomenu kakovosti na vseh ravneh ustvarjanja podatkov.

**REŠITEV:**

Neenake stopnje standardiziranosti so priložnost za deljenje izkušenj, dobrih praks in zgledov, ki lahko ustvarjalce z manj standardiziranih področij motivirajo k večji standardizaciji. Izmenjava izkušenj bo spodbudila proces povečanja kakovosti, ki je eden od pogojev odprtega dostopa, in povečala možnosti bolj širokega izkoriščanja ustvarjenih podatkov. Ena od možnosti za zblíževanje in izmenjavo dobrih praks je oblikovati področne mreže okoli problematike 'odprtih podatkov', ki naj presegajo posamezne laboratorije, oddelke, katedre idr. (primer spletišče SIDIH).

Medpodročna srečanja lahko spodbudijo izmenjavo izkušenj in povečajo ozaveščenost o pomenu hrambe in dostopa do podatkov, ki presegata interes in potrebe vsakega posameznega raziskovalca in skupine.

Sicer pa je odnos do kakovosti pogosto izrazito usmerjen v ciljne produkte, denimo na področju jezikoslovja na slovarske produkte. Gradivo izbirajo premišljeno in ga sistematično čistijo.<sup>21, 22</sup> Na področju hrambe glasbenih posnetkov ustvarjajo metapodatke vzporedno z nastajanjem in urejanjem podatkov<sup>23</sup> in upoštevajo priporočila sprejeta na mednarodni ravni. Poleg pomena kakovosti se zavedajo tudi dejstva, da je ta zmeraj kompromis med tehničnimi in vsebinskimi zmožnostmi.<sup>24</sup> Kakovost podatkov je ponavadi ocenjena realistično, je nek 'kompromis' v okviru možnega.<sup>25</sup>

Na področju socialnega dela so podatki praviloma namenjeni konkretnemu projektu, zato v fazi ustvarjanja praviloma na razmišljajo o drugi rabi niti nimajo razvitih orodij, ki bi jim bila pri tem v pomoč. Metapodatke in drugo dokumentacijo razumejo le tisti, ki so neposredno vključeni v postopek ustvarjanja podatkov. V takih primerih je kakovost podatkov nizka oziroma omejena, kar zadeva drugo rabo. Za kontrolo poskrbijo tako, da zbrane in urejene podatke pred rabo dajo v preverbo intervjuvancem oziroma respondentom in tako odpravijo morebitne napake, ki bi bile posledica neustreznega razumevanja med spraševanim in spraševalcem.<sup>26</sup>

Merila, kaj so dobri podatki, temeljijo na določeni samoumevnosti, prepričanju o dobrih praksah, ki so vzpostavljene v okviru znanstvene skupnosti. Nekdo se priuči ravnanj, ki zagotavljajo dobre podatke, z izkušnjami in z usposabljanjem za člana znanstvene skupnosti.<sup>27</sup> Zagotovilo kakovosti podatkov je sama kakovost raziskovalca – njegova usposobljenost in znanje sta osnova zaupanja, da dela kakovostno.<sup>28</sup>

*PROBLEM:*

Čeprav so na določenih področjih raziskovalci oskrbljeni s standardiziranimi orodji in načini, so na drugih področjih prepuščeni lastni iznajdljivosti. Prisotne so velike razlike v ravni poznavanja postopkov in meril za zagotavljanje kakovosti. Ovire v dostopu, napake, neustrezna ali pomanjkljiva dokumentacija, neočiščeni podatki idr. zmanjšujejo kakovost podatkov in otežujejo rabo ustvarjenih podatkov.

Na nizko kakovost vplivajo tudi tehnični, kulturni in pravni pogoji, ki so nemalokrat prepuščeni v presojo posamezni raziskovalni skupnosti ali posameznemu raziskovalcu.

*REŠITEV:*

Tako kot se morajo ustvarjalci podatkov spoznati z vsebino preučevanega področja, da bi lahko ustvarjali podatke, bo potrebno osvojiti tudi znanja in veščine, ki zadevajo samo ravnanje s podatki. Čeprav so med področji, kot denimo družboslovje, humanistika, naravoslovje, medicina, velike in na videz nepremostljive razlike, kar zadeva vsebine in metode preučevanja, jih na področju ustvarjanja in upravljanja s podatki zadevajo vsaj deloma podobni problemi (kakovost, selekcija, hramba, dostop...).

Rešitev problemov je treba iskati v smeri strokovnega usposabljanja, v obliki delavnic, tečajev, poletnih šol idr.

Dolgoročno je treba razmišljati o tem, da naj upravljanje s podatki postane obvezen del izobraževanja ustvarjalcev in uporabnikov podatkov ter strokovnega usposabljanja.

Zanimiv je tudi pristop, ki ga uveljavlja pobuda za pridobivanje potrdila o ustreznosti podatkov za dolgotrajno uporabo - Data Seal of Approval (DSA) (tabela 5.1). Za razliko od nekaterih drugih tovrstnih postopkov preverjanja, ki so usmerjeni na procese digitalne hrambe v ožjem smislu (glej pregled v Štebe in Vipavc Brvar 2011), ta pristop izraziteje poudarja vlogo dajalca pri ustrezni pripravi podatkov, tako da ti izpolnjujejo pogoje za vstop v sistem dolgotrajne hrambe.

Tabela 5.1: Prva tri vodila DSA

- |  |
|--|
| <ol style="list-style-type: none"> <li>1. Ustvarjalec podatkov v podatkovni repozitorij izroči raziskovalne podatke, opremljene s spremljajočimi informacijami, na osnovi katerih je mogoče ovrednotiti znanstveno kakovost oddanega gradiva, skladno z merili posameznih disciplin in etičnimi zahtevami.</li> <li>2. Ustvarjalec podatkov poskrbi, da je gradivo oddano v obliki, ki jo priporoča podatkovni repozitorij.</li> <li>3. Ustvarjalec podatkov poskrbi, da so raziskovalni podatki opremljeni z metapodatki in da izpolnjujejo zahteve podatkovnega repozitorija.</li> </ol> |
|--|

Vir: Data Seal of Approval<sup>viii</sup>

V Sloveniji predstavlja primer podpore pri izročanju podatkov delovanje ADP. ADP je za dajalce podatkov namenil podrobna navodila in obrazce za pripravo in oddajanje podatkov v arhiv. Usposabljanje za pripravo podatkov je tematika občasnih seminarjev, ki so namenjeni dajalcem (Štebe in Vipavc Brvar 2008). V sodelovanju s partnerskimi organizacijami ADP pomaga pri pripravi vodičev za dajalce (CESSDA<sup>ix</sup>, ICPSR<sup>x</sup>, DATA PASS<sup>xi</sup>, UK DA<sup>xii</sup>). Eden od značilnih problemov, ki jih rešujemo s pomočjo tovrstnih navodil, so težave pri izročanju podatkov ob nejasnih določilih obsega prostovoljnega pristanka in zaupnosti podatkov v okviru kvalitativnih družboslovnih raziskav (primerjaj Štebe in drugi 2011). Vprašanje zaupnosti je v širšem smislu povezano tudi z dostopnostjo podatkov statističnih uradov. Na eni strani je v zvezi s tem naloga dajalca, da (skupaj s podatkovnih arhivom) opredeli postopke za anonimizacijo in kriterije glede zahtev zaupnosti. Kadar dajalec podatkov določi omejitve glede namena uporabe ali kategorije uporabnika, je naloga arhiva podatkov, da omogoča ustrezno registracijo uporabnikov in kontrolira dostop do gradiv, upoštevajoč opredeljene omejitve gradiv glede dostopa in kako se te skladajo s pravicami različnih kategorij uporabnikov. Tovrsten sistem zahteva nadaljnjo uskladitev in sodelovanje na ravni mednarodne izmenjave in dostopa do podatkov, kar je ena od nalog projekta Data Without Boundaries (DwB), pri katerem sodelujejo partnerji CESSDA in Statističnih uradov v Evropi. Možna rešitev pri integraciji sistemov dostopnosti in različnih režimov dostopa do podatkov je, da se poenoti in med sabo poveže samo metapodatkovne opise, namenjene iskanju in presoji uporabnosti podatkovnih gradiv za sekundarno analizo s strani uporabnika (glej LeClere 2006), kar se nanaša na nekatere podatke statističnega urada. Pri tem pa fizično mesto dostopa še zmeraj ostane pri dajalcu.

<sup>viii</sup> <http://www.datasealofapproval.org/>, 7.8.2012.

<sup>ix</sup> <http://www.cessda.org/sharing/>, 7. 8. 2012

<sup>x</sup> <http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/deposit/>, 7. 8. 2012

<sup>xi</sup> <http://www.icpsr.umich.edu/icpsrweb/content/DATAPASS/citations.html>, 7. 8. 2012

<sup>xii</sup> <http://www.data-archive.ac.uk/deposit>, 7. 8. 2012



### **Družboslovje: Arhiv družboslovnih podatkov (ADP)**

ADP je področni podatkovni center za družboslovje, ki hrani zbirko podatkov, zanimivih za družboslovne analize, s poudarkom na problemih, povezanih s slovensko družbo. Prednost imajo vsebinsko pomembnejše in metodološko dobro izdelane raziskave, ob tem še posebej podatki, ki jih je mogoče primerjati v različnih časovnih obdobjih, in mednarodni primerjalni podatki, pri katerih je upoštevana Slovenija. Ciljni uporabniki so domači in tuji raziskovalci, učitelji in študentje, ki so podatkovno in statistično pismeni<sup>xiii</sup> za samostojno ravnanje z gradivi. Za dostop do gradiv, ki jih hrani Arhiv družboslovnih podatkov se je potrebno registrirati.<sup>xiv</sup> Arhiv od dajalcev sprejema podatke, ki dosegajo osnovna merila kakovosti in vsebinske zanimivosti za nadaljnje analize. Izročeni podatki štejejo kot znanstvena objava po merilih ARRS.

ADP hrani približno 600 raziskav iz različnih družboslovnih področij. V glavnem gre za kvantitativne, nekaj pa je tudi kvalitativnih raziskav. ADP poleg podatkovne datoteke za posamezno raziskavo hrani tudi druga spremljajoča gradiva, potrebna za razumevanje podatkov. Običajno so to kopije originalnih vprašalnikov, lahko pa tudi šifranti (npr. standardna klasifikacija poklicev), razne pokazne kartice, obvestilna pisma vzorčnim enotam in obvestila o obisku anketarja, navodila anketarjem, sumarniki itd.

Od leta 1998 je ADP član evropskega združenja družboslovnih arhivov – CESSDA<sup>xv</sup>. Osnutek dejavnosti mreže CESSDA – ERIC predvideva novo obliko delovanja, to je organizacijsko in strokovno nadgradnjo podatkovne infrastrukture na področju družboslovja, s ciljem zagotoviti pogoje za vzpostavitev dobro opredeljene, celovite in med seboj povezane podatkovne infrastrukture za družboslovje tako na nacionalni kot na mednarodni ravni.

Vir: <http://adp.fdv.uni-lj.si/>

### **5.5 IZBOR PODATKOV NA PODLAGI KAKOVOSTI**

Vrednotenje in izbor podatkov na podlagi kakovosti nista samoumevna na način, kot je uveljavljen pri publikacijah, kjer obstajajo uveljavljene uredniške politike za revije ali monografije in kjer recenzenti s svojimi ocenami zagotovijo kakovosten izbor. Problema smo se dotaknili tudi v teku intervjujev, in sicer smo spraševali, ali je mogoče izbor na podlagi kakovosti podatkov formalizirati po analogiji z recenzijskim postopkom, ki velja za znanstvene članke. Vtis je, da je o tem, kaj so dobri oziroma uporabni podatki, doseženo neko splošno soglasje v okviru znanstvene skupnosti.

Intervjuvanci so izpostavili tudi pogled, da so lahko zanimivi tudi slabi podatki, saj povedo več, kot če podatka ni in je potrebno znova investirati čas in energijo v iskanje, odkrivanje, ustvarjanje ...<sup>29</sup> Zato odgovor, da je upravičeno hraniti zgolj kakovostne podatke, ni za vse sprejemljiv. Izreženo je bilo mnenje, da je tudi npr. neformalen nadzor nad prispevki na način delovanja Wikipedije lahko zagotovilo kakovosti. V takem primeru je kakovost stvar konsenza, ki ga je mogoče doseči v diskusiji med zainteresiranimi in z uporabo moči strokovnih argumentov, to pomeni, da zanjo ni osnova stroga uredniška politika, recenziranje ipd., kot je uveljavljeno na področju znanstvenega publiciranja.<sup>30, 31</sup>

Med intervjuvanci je prisotna skepsa do zunanjega ocenjevanja in vrednotenja, kaj so dobri podatki. Nekateri so zadržani do tega, da bi podatke obravnavala inštitutska knjižnica, na način urednikovanja

<sup>xiii</sup> [http://sl.wikipedia.org/wiki/Podatkovna\\_pismenost](http://sl.wikipedia.org/wiki/Podatkovna_pismenost), 7. 8. 2012

<sup>xiv</sup> <http://adp.fdv.uni-lj.si/registracija/>, 7. 8. 2012

<sup>xv</sup> <http://www.cessda.org/>, 7. 8. 2012



publikacij. Menijo, da mora to biti prepuščeno samemu raziskovalcu, razen v primerih kadar spremlja objavo, ko sprejem članka v objavo avtomatsko pomeni tudi potrditev veljavnosti podatkov. Tako kot v fazi ustvarjanja podatkov so merila za ocenjevanje kakovosti podatkov stvar ožje strokovne kompetence. Knjižničarji sami menijo, da niso usposobljeni za vrednotenje, ki zadeva vsebinsko kakovost podatkov.<sup>32</sup> Po drugi strani pa vidijo realno možnost v tem, da oblikujejo in vzdržujejo formalne kriterije in omejitve.<sup>33</sup>

Izkazalo se je, da bi knjižničarji kvečjemu lahko bili serviserji sistema, sami pa tega projekta ne morejo začeti, saj se ne čutijo kompetentne, niti ne poznajo raznolikosti področij in se ne morejo odločati o vsebini in kakovosti.<sup>34, 35, 36, 37</sup> Prispevek knjižnice je lahko v tem, da so strokovnjaki za splošne metapodatke. V smislu, da zahtevajo minimalen, vendar popoln formalni opis podatkov, zgodovino nastanka in življenja podatkov v okviru življenjskega cikla, se pravi dokumentacijo, po kateri je podatkovno gradivo določeno po svoji enkratnosti, ne presojujejo pa drugih vidikov kakovosti.

Najpomembnejši zaključek glede kakovosti podatkov v fazi ustvarjanja je, da so podatki lahko tudi slabši, toda še vedno koristni. Podatki so lahko samo toliko dobri, kot je to možno doseči ob omejitvah glede usposobljenosti raziskovalca in pomočnikov, aparature in finančnih možnosti. Poleg tega je pomembno zagotoviti razumevanje doseženega nivoja kakovosti podatkov kot konteksta, ki pogojuje njegovo uporabo. Prav neočiščeni in pomanjkljivo dokumentirani podatki pogosto predstavljajo omejitev za dostop, saj zahtevajo poglobljeno razumevanje nastanka in specifične narave podatkov.

#### Okvir 5.4: P-R 4

##### *PROBLEM:*

Vrednotenje in izbor podatkov med intervjuvanci nista nekaj samoumevnega. Nekateri zagovarjajo selekcioniranje na osnovi konsenza v znanstveni skupnosti, drugi bi konsenz razširili in uredniško politiko nadomestili z izmenjavo argumentov v zainteresirani javnosti.

Intervjuvanci niso enotni glede tega, ali bi lahko nekdo zunanji ocenjeval in vrednotil kakovost podatkov. Knjižničarji, ki bi lahko bili potencialni nosilci vloge, se zaenkrat čutijo nemočne in nepripravljene za vrednotenje podatkov.

##### *REŠITEV:*

Rešitve je smotrno iskati v zgledih iz tujine, saj je razprava o recenziranju in selekcioniranju podatkov trenutno zelo aktualna. Narašča število podatkovnih revij, v katerih sta izbor in sprejem objav omejena na formalne zahteve, ki jim morajo podatki zadostiti. To pomeni, da obstaja konsenz glede zahtev, ki morajo biti izpolnjene, da so podatki ustrezni za objavo, s tem pa tudi za drugo rabo.

Ena od možnih rešitev je razmislek v smeri prilagajanja uredniških politik, kot jih poznajo tradicionalne revije. Novi modeli recenziranja, ki jih preizkušajo v nekaterih revijah z odprtim dostopom in v podatkovnih revijah, npr. PLOS, kažejo na potrebe po splošnem spreminjanju tradicionalnih uredniških politik in recenzentskih postopkov, ne samo na področju vrednotenja podatkov.

V tem trenutku je mogoče detektirati problem, vendar rešitve zaenkrat še ni mogoče ponuditi, saj je debata še odprta. Povezana je s širšim vprašanjem vrednotenja v znanosti, ki se prevečkrat omejuje na golo preštevanje objav, citiranja idr. Potrebno bo spodbuditi tovrstne razprave, se vanje vključevati in uporabiti modele, ko bodo ti vzpostavljeni.

## 5.6 SELEKCIJA PODATKOV V DOSTOPU S SPREMLJAJOČO OBJAVO V ČLANKU ALI MONOGRAFIJI

Ta vidik je povezan z diferenciacijo tipov podatkov v gradnji zbirk. Vseh podatkov ni mogoče pripraviti do ravni, da so prečiščeni in razumljivi brez dodatne razlage (npr. muzikološka zbirka). Zato se teži k reprezentativnemu izboru podatkov in k temu, da se izbrano posreduje v prečiščeni obliki, z dodano vrednostjo, se pravi z interpretacijami glede okoliščin in dodatnimi razlagami, ki gradivo postavijo v širši kontekst, potreben za razumevanje. Ta oblika dostopa do gradiv je že monografska publikacija, ki jo spremlja izbrano in prečiščeno podatkovno gradivo. To pomeni, da poleg zbiranja podatkov ustvarjalci pokrijejo tudi druge faze, vse do objave. Ne gre samo za ustvarjanje podatkov, pač pa tudi za izdelavo končnih proizvodov.<sup>38</sup>

Kadar podatke spremljajo pomanjkljivi metapodatki, ki so nastali sproti ob raziskovalnem delu, naj bi bil dostop omejen. Ker gre za gradiva, ki so nastala v preteklosti, ne da bi med raziskovanjem razmišljali, da jih bodo nekoč dali v rabo širši javnosti, so razumljivi pomisleki glede tega.<sup>39, 40</sup> Ena od možnosti je, da se selekcionira, katera gradiva ostanejo za interno rabo in katera bodo dostopna javnosti. Intervjuvanec s področja hrambe glasbenih gradiv je izpostavil pomen sodelovanja z uporabniki. Gradivo, ki ni ali ne more biti ustrezno opremljeno z metapodatki, da bi omogočalo samostojno drugo rabo, terja, da se ustvarjalci podatkov in zbirk vključijo tudi v naslednje faze življenjskega kroga podatkov.<sup>41</sup> Gre za podoben problem, kot smo ga omenili v uvodu pri razpravi o pomislekih, ovirah zaradi strahu pred zlorabami, napačnimi interpretacijami in so ga zaznali tudi v drugih sorodnih študijah.<sup>xvi</sup> Več o tem bomo spregovorili v poglavju 7.1 Dostop do podatkov.

Na podlagi pogovorov z intervjuvanci ugotavljamo, da se ustvarjalci podatkov in zbirk zavedajo odgovornosti v zvezi s kakovostjo podatkov. Kakor se zavedajo tudi tega, da je zaradi odsotnosti standardnih postopkov druga raba precej omejena.

---

<sup>xvi</sup> "V razpravi o deljenju podatkov predstavljajo tveganja o zlorabi podatkov (in druge ovire) pomembno temo." (Cragin in drugi 2010)

## Okvir 5.5: P-R 5

**PROBLEM:**

Ustvarjalci podatkov in zbirk opravljajo svoje vloge skladno z razpoložljivimi možnostmi in se zavedajo, da so rezultati kompromis.

V obstoječih okoliščinah so zahteve hrambe v veliki meri stvar posameznikov in skupin. Razen v primerih, kadar svoje podatke oddajo v obstoječe zbirke, ki imajo svoje standarde in protokole.

**REŠITEV:**

Rešitev je oblikovanje kriterijev za vrednotenje pomena podatkov, kadar so podatki deležni posebne obravnave, npr. da se uvrstijo v zbirko podatkovnega centra. V okviru podpore Načrtom ravnanja s podatki (DMP) je potrebno stremeti h temu, da se zagotovi sredstva in pogoje za delovanje področnih podatkovnih centrov, ki bodo skrbeli za koordinirano pripravo 'referenčnih' podatkov in opravljali storitev njihovega skrbništva ter posredovanja zainteresiranim drugim uporabnikom.

Zagotoviti je treba tudi rešitve in modele za podatke, ki so manjšega pomena, kar zadeva večkratno rabo. Npr. da se za revije oblikuje skupen repozitorij za podatke povezane z objavami, kot je na primer DRYAD.<sup>xvii</sup>

## 5.7 CITATI

<sup>1</sup> Podatki so eno, ko pa to vključiš recimo v COBISS, so pa vključeni v zbirko podatkov. (vodja znanstvenoinformacijskega centra, NO5)

<sup>2</sup> Ja, ampak spekter podatkov je stvar, ki ima strukturo, načrt, definirano z vpisnimi polji, z orodji za iskanje, z vsem tem. To je zbirka podatkov. Medtem ko sami podatki, so pa lahko številke 'file-ov'. In to je zelo daleč od zbirke podatkov. (vodja znanstvenoinformacijskega centra, NO5)

<sup>3</sup> Aha, eno je 'data', drugo je 'data base'. To, kar ti govoriš [*"anketa, ki jo naredimo, kjer je na primer tisoč enot in sto vprašanj"*], je zame 'data'. 'Data base' je pa aplikacija, ki vzame v poštev strukturo podatkov in doda neko dodano vrednost s pomočjo iskalnikov in tega. Tipični primeri so poslovne 'data base-e'. (vodja znanstvenoinformacijskega centra, NO5)

<sup>4</sup> Eno je najdišče, ki predstavlja izvorno lokacijo najdb in raznih drugih ostalin človekove dejavnosti v preteklosti. Opredeljeno je kot nepremična dediščina, ki je običajno pokopana pod današnjim površjem in pride na dan šele z zemeljskimi deli in terenskimi raziskavami, s katerimi pa so arheološke ostaline največkrat tudi odstranjene oziroma pretekle sledi zabrisane. Zato so za stroko izjemnega pomena podatki o lokaciji in obsegu najdišča, o odkritih plasteh, strukturah in najdbah, o metodi terenskih raziskav. Na podlagi teh podatkov je najdišče mogoče prostorsko umestiti ter kronološko in tipološko ovrednotiti. Drugo so geografske in kartografske reference in podatki o tem, kje se hranijo najdbe in dokumentacija terenskih raziskav (elaborati in poročila, ki vključujejo tudi razne analize, fotografsko in risarsko dokumentacijo, meritve oz. načrte) ter sklici na dosedanje objave. Vsi ti podatki so opremljeni z enotno kodo in strukturirani v podatkovni zbirki, ki omogoča poizvedovanje prek različnih iskalnikov. Podatkovna zbirka arheoloških najdišč je podprta tudi z GIS karto, v katero so vključeni razni sloji prostorskih in geografskih podatkov. Ta karta omogoča

<sup>xvii</sup> <http://datadryad.org/>, 7. 8. 2012

---

pregledovanje arheoloških najdišč v prostorskem kontekstu in je povezana s tekstovno bazo najdišč. (raziskovalec s področja arheologije, NO14)

<sup>5</sup> Arheološko dediščino delimo na nepremične ostaline, to so antropogene plasti, vkopi in ohranjene strukture, in na premične najdbe, to so razni predmeti iz preteklosti, ki so z najdišča preneseni v muzejske ali kake druge ustanove v trajno hrambo. V muzejih so ti predmeti zabeleženi v inventarnih knjigah in zbirkah podatkov o predmetih, skupaj z dokumentacijo terenskih raziskav v klasični in/ali elektronski obliki. Zbirke podatkov o predmetih niso povsem enotno koncipirane, kar otežuje povezovanje raznih podatkovnih zbirk. Prav tako (še) niso javno dostopne prek spleta. (raziskovalec s področja arheologije, NO14)

<sup>6</sup> Za eksperimentalne študije je zelo standardizirano. Torej če nekdo izvede nek klinični poskus, potem je ta obrazec zelo standardiziran, ima vse dele raziskave zelo natančno opisane. Če pa gledate neko opazovalno študijo, na primer retrospektivno študijo, zelo pogosto ne bodo napisali čisto nič od začetka, ampak to je slabo. /.../ Na primer, da gredo k zdravniku v arhiv in poiščejo bolnike, ki so imeli v zadnjih petih letih neko bolezen in potem se sprašujejo, ali je bila ta bolezen bolj pogosta pri ženskah ali pri moških, ali je bila bolj pogosta med kadilci, kot nekadilci ... To so raziskave, ki jih lahko hitro izvedemo. *Je to arhiv?* To so lahko kar kartoteke pacientov, na primer. Odvisno od ustanove, ki imajo lahko različne vrste informatizacije vseh teh podatkov. Obstajajo nekatere klinike, kjer so vsi podatki že elektronski, kot na primer pediatrična klinika, medtem ko imajo druge klinike podatke na papirju, kar pomeni, da mora raziskovalec, ko želi narediti neko raziskavo, listati po kartonih in pridobiti podatke. (znanstvenik s področja medicinskih vez in statistike, NO15)

<sup>7</sup> Podatki so na kartah. Ampak teh podatkov mi ne smemo posredovati naprej, tako smo se obvezali takrat, ko smo jih dobili. /.../ Za podatke, ki nam jih posredujejo s točno določenim namenom, da naredimo potem nek produkt iz tega, imamo v vseh pogodbah opredeljeno uporabo podatkov izključno samo za tisti namen. (znanstvenik s področja geografije, NO8)

<sup>8</sup> To so statistični podatki na individualni ravni, ki niso prosto dostopni komur koli, dostopni so raziskovalcem. Do sedaj smo jih imeli tudi fizično pri nas, zdaj pa se prenašajo na oddaljen dostop in ravno prejšnji teden smo se uspeli povezati z njihovim strežnikom. To sicer ni najbolj učinkovita rešitev, ampak ni druge možnosti. V bistvu nam dovoljujejo uporabo deindividualiziranih in individualnih podatkov, ne dovoljujejo pa jih prosto prenašati naokoli. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>9</sup> Če potrebuješ samo koordinate, bi jih načeloma lahko dobil proti nekemu plačilu. Ko pa imaš zraven kataster, zemljiško knjigo, ali podatke o nepremičninah, iz česar so razvidni lastniki, je pa malo bolj zaprto zaradi varstva osebnih podatkov. Drugače pa, kolikor vem, lahko na GURS-u tudi tujci dobijo podatke za raziskovalne namene brezplačno. (znanstvenik s področja geodezije, NO19)

<sup>10</sup> Kar je morda specifično je, da so nekateri podatki deindividualizirani, in sicer podatki na ravni posameznika. V tej fazi imajo tisti, ki so določeni, da jih lahko uporabljajo, oddaljen dostop do njih. Tukaj smo prišli do situacije, ko želi naročnik imeti ne le podatke, ampak tudi orodja za uporabo teh podatkov. To je recimo mikrosimulacijski model za dohodnino, oziroma davke in subvencije, ki smo ga pripravili. Seveda Statistični urad ne dovoli, da bi ta model skupaj s podatki prosto predajali uporabnikom. Ideja, ki je sedaj dogovorjena in potrjena je, da se ti podatki in orodja postavijo na njihovem strežniku in da lahko zainteresirani uporabniki, ki dobijo dostop, to uporabljajo. Ne vidijo podatkov, ampak uporabljajo grafični vmesnik, ki jim omogoča na primer simuliranje določenih ukrepov. (Boris Majcen, znanstvenik s področja ekonomije)

---

<sup>11</sup> Ameriške univerze imajo takojšen dostop do vseh mikro podatkov, recimo do takšnih, za katere moramo mi prositi na primer SURS, čakati pol leta, ter plačevati iz svojih lastnih projektov. /.../ Čisto parcialno imajo urejeno. Na Češkem, na 'Seržu', imajo recimo varno sobo, na njihovi instituciji imajo konstanten on-line dostop do mikro podatkov. Tako kot bi vi kot Arhiv družboslovnih podatkov imeli pogodbo s SURS-om in bi nekdo, ki bi želel delati na teh podatkih, prišel do vas, sedel za računalnik in tam delal. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>12</sup> V obeh primerih [op. CERN in BELL] gre za podatke, ki so v surovi obliki, razen za specialiste, povsem nezanimivi. Imajo pa naslednjo značilnost: merijo se v petabajtih. Da so zanimivi za analize, jih je treba večkrat in zelo zahtevno računsko obdelati. V večini eksperimentalnih, pa tudi teoretičnih fizikalnih projektih, se danes dela tako, da se primerja podatke, ki so generirani z naključnimi metodami na osnovi teoretičnih modelov s podatki, ki so simulacije - temu se reče Monte Carlo simulacije, z dejanskimi podatki iz detektorja in se potem to uporablja na več načinov: lahko se metode analize izboljšuje, lahko se primerja podatke in ugotavlja odstopanja med tem, kar je bilo teoretično predvideno in med tem, kar je bilo praktično ugotovljeno, potem se pa ugotavlja ali odstopanja kažejo, da je nekaj narobe z detektorjem, ali je nekaj narobe s teorijo. Tukaj so problemi malo drugačni kot pri družboslovnih podatkih. Po eni strani je bolj prisotna zavest znanstvenikov, da je treba podatke skrbno hraniti, da so postopki ponovljivi, ker ko pride do objave nekega resnega odkritja, je treba dat podatke na voljo, da lahko druga skupina ponovi cel postopek. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>13</sup> Obstajajo postopki za preverjanje kakovosti, recimo, če so podatki nehomogeni ali pa preveč razpršeni. To je neka majhna ploščica, na njej so lahko vseh vrst napake in se meri kakovost odziva. En bi bolj reagiral, drug pa zaradi tehničnih kriterijev ne. (znanstvenik s področja biologije, NO6)

<sup>14</sup> Z vizualnimi statističnimi metodami oceni, ali je tisti podatek relevanten ali ne. (znanstvenik s področja biologije, NO6)

<sup>15</sup> Takšen laboratorij ima akreditacijo nekajkrat na leto in če ustreza standardom za pridobivanje podatkov iz tovrstnih eksperimentov, potem so tisti podatki na določenem kakovostnem nivoju. (znanstvenik s področja biologije, NO6)

<sup>16</sup> V znanstveni srenji z našega področja so sprejeta neka splošna merila. Ve se, kaj pomeni, da je čist protein, ali ve se, kdaj je dobro izveden eksperiment. V pisni obliki je sprejeto, kaj pomeni, da je eksperiment dobro izveden, kako dobro izvesti eksperiment in potem kakšna mora biti kvaliteta podatkov, da lahko rečeš, temu pa res verjamem. Tega se tudi držimo ali pa se čim bolj poskušamo držati. Medtem ko za takšne rutinske stvari v laboratoriju, kot so ti geli, to je pa bolj modrost laboratorija. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>17</sup> Spet je malo odvisno od podatka. Nekje je eksperimentalno zajemanje podatkov tako trivialno, da ne vpliva dosti na kvaliteto podatka, nekje pa ni, in takrat so zraven napisani še dodatni parametri, tako da lahko nato sam presodiš. Ampak v tem primeru moraš imeti malo več znanja. Nekje je dokumentirano zraven podatka, kdo ga je ustvaril, kdaj, s kakšno metodo ipd. Malo je odvisno od zbirke do zbirke. Te omenjene, so stare zbirke, recimo dvajset let. In tekom let so se standardi že toliko postavili, da se za tisto, kar gre v zbirko, ve, kakšne kvalitete je. Je pa res, da je v biologiji veliko novih pristopov, čisto tako, eksperimentalnih. Tam pa imajo včasih ljudje probleme s tem, da niso vse informacije zraven pripisane. Dostikrat tudi kvaliteta tistih podatkov ni zadovoljiva, kakšni rezultati so recimo lažno pozitivni. Neizkušen uporabnik ima lahko velike probleme z uporabo takšnih podatkov. Ampak starejše podatkovne zbirke, te pa so ... Na tem področju se je v znanstvenem okolju sčasoma pojavila

---

potreba po vzpostavitvi standardov, da se to nekam zapiše in se reče, da bomo zdaj vsi tako delali. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>18</sup> Saj vedno prej preverimo podatke. Te serije podatkov zbiramo že približno 10 let in vedno, ko se vračamo v podjetja, prosimo za nove podatke ter pogledamo, kaj se je dogajalo v časovni vrsti. Tako vidimo, ali so kakšna odstopanja ali niso, in če so, kontaktiramo nazaj podjetje. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>19</sup> Ponavadi je tako, da imajo podatkovne zbirke, ki hranijo te podatke, narejene aplikacije. Aplikacije pregledajo tisto, kar hočeš oddati, še dodatno malo poanalizirajo, če je vse v redu, in ko greš skozi njihovo kontrolo kvalitete, potem je lahko deponirano v zbirki. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>20</sup> Raziskovalci, ki gredo potem pogledat podatkovno zbirko, takoj vidijo, ali je podatek v redu ali ne. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>21</sup> Če iz besedilnega korpusa, ki ga niste sistematično očistili smeti, naredite slovar besednih oblik, se lahko dobre frekventne besede držijo celi grozdi napačno zapisanih besed. In tisto potem vsakogar takoj zelo odbije. (znanstvenik s področja jezikoslovja, NO13)

<sup>22</sup> Eno je, da ni smeti, drugo pa, da skušamo zajeti čim bolj kvalitetno gradivo. (znanstvenik s področja jezikoslovja, NO13)

<sup>23</sup> *Kje so ovire, da bi katalog, tak kot je zdaj, se pravi nek ožji nabor metapodatkov, naredili dostopen? Zdaj gredo pobude verjetno tudi preko Europeane v to smer, da se naredijo neki združevalni katalogi?* Po eni strani to, po eni strani pa ima Europeana željo po avdio gradivu, da je dostop do konkretnega avdio gradiva. In mi smo že nekaj dali, tako da nekaj je že dostopnega, nekaj pa še bo, tudi avdio gradiva. Težava je pri nas ravno pripraviti metapodatke, ki bi bili nekemu naključnemu obiskovalcu v pomoč in tudi ponuditi avdio posnetek takšen, da bo nekomu koristen. Na primer če pojejo pesem, da ni to samo eno kašljanje, nerazumljivo ... Mi smo tako kot vi, spustimo trak in je lahko marsikaj na tem traku in je treba izbrat ... (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>24</sup> Tretja tema je tehnična kakovost, ki je mogoče malo bolj v moji domeni. Kakšno opremo imaš, to je prvo, po drugi strani pa potem postavitev opreme na licu mesta, konkretno mikrofonom, situacija in tako naprej. Pri čemer seveda pazimo, da ne postavljamo informatorje v nenormalne pogoje, ker seveda ni zmeraj najlepši posnetek tudi najboljši. Da jih ne odbijemo, da potem vsebinsko ne okrnimo posnetka. Zmeraj je nek kompromis med tehnično kakovostjo in vsebinsko. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>25</sup> So pa druge vrste podatkov, kjer je kvaliteta odvisna predvsem od tega, kako dobro je bil eksperiment narejen. Tisto, kar daš v zbirko, je lahko brez kontaminacij, brez nekih artefaktov, ampak je kvaliteta podatka slabša, ne zaradi slabše izvedenega eksperimenta, temveč zato, ker se ni dalo boljše narediti. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>26</sup> Tudi jaz imam takšne izkušnje, da je takrat, ko sama zbiram podatke, tudi analiza lažja. Pri nas včasih prosimo študente, da naredijo kakšne intervjuje, večinoma za raziskave, ki jih potem niti ne objavljamo širše, bolj za vpogled v situacije. In tisto je včasih zelo skopo, ker tisti, ki je delal intervju, ni postavljajl podvprašanj, ni poskušal pravzaprav zares razumeti, kaj mu je oseba povedala in tam imamo težave. (znanstvenik s področja socialnega dela, NO7)

---

<sup>27</sup> Osnovne podatke o nekem najdišču (geografska lokacija, vrsta ostalin, obseg in metoda terenske raziskave, dosedanje ugotovitve) nadgrajujemo z ovrednotenji teh podatkov s strani specialistov za določena arheološka obdobja in za določeno vrsto virov, s čimer dopolnjujemo védenje o najdišču. Odkrite ostaline, strukture in najdeni predmeti postanejo arheološka dediščina šele v luči interpretacij tega, kaj predstavljajo in iz katerega časa so. To pa je vsekakor znanstveno-strokovno delo. (raziskovalec s področja arheologije, NO14)

<sup>28</sup> Preprosto, to je zelo preprosto. Dober raziskovalec ima dobre podatke, slab raziskovalec ima slabe podatke. Ve se, po čem se meri dobrega raziskovalca. Ljudje na področju vedo, kdo je dober, kdo ni dober in glede na to, če ima opravka z meritvami, če je dober, je gotovo naredil dobre meritve. (vodja znanstvenoinformacijskega centra, NO5)

<sup>29</sup> Vsi podatki, kvalitetni in nekvalitetni, so vključeni, zato ker tudi slab podatek nekaj pove. Če podatek manjka, potem vedno znova iščeš in odkrivaš, zakaj ga nekdo ni vključil. Bolje slab podatek kot pa, da ga ni. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>30</sup> Sam zelo vzpodbujam svojo 1400-glavo javnost, ki jo tvori forum Slovlit, naj sama poseže in opozori na napake, če jih vidi. Skratka delam reklamo tudi za to, da dobimo opozorilo, kje je prišlo do neumnosti in da jih potem popravimo. /.../ Moj namen je, čisto pedagoški, pa tudi civilizacijski, da podatkovne zbirke uporabljamo kot prostor za artikulacijo lastnega znanja. Da, v tem vidim glavno težo oziroma prednost zbirk, ki jih gradi skupnost sama. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>31</sup> Nič ni organizirano, gre za eno vrsto konsenza. Tisti, ki se pogosto pojavlja, počasi pridobi renome, mogoče tudi status administratorja, in to ga zaveže, da potem redno spremlja dejavnost. Če pride do težav, se med seboj pogovorimo ... (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>32</sup> Če nekdo poroča o rezultatih novih meritev, ni načina, da preveriš, ali je to res meril ali ne. Moraš verjeti, da je to meril in da ni popravljajl števil, ko je pisal članek. In konec koncev lahko kdo drug še enkrat izmeri. V tem smislu 'peer review' ne ugotavlja kvalitete novih prispevkov. /.../ In če že 'peer review' ne gre v ocenjevanje kvalitete, nek knjižničar sploh nima 'šans' pri tem. (vodja znanstvenoinformacijskega centra, NO5)

<sup>33</sup> Formalni kriteriji pa so, seveda. Ampak formalni kriteriji, ki jih jaz vidim, so kriteriji primerljivosti z metapodatki. Toliko in toliko podatkov mi moraš dati, da jih bom lahko vključil v svojo bazo. Če mi jih ne daš, jih ne morem vključiti, žal. (vodja znanstvenoinformacijskega centra, NO5)

<sup>34</sup> Mi ne gremo tako daleč, to je na nivoju tega: "Glej, ti si ta članek sicer objavil, pa si prepričan, da je to dobro?" Kot prvo bo užaljen, kot drugo, bo vedno prepričan, da je dober. Če je to nek človek, ki je tukaj na inštitutu, bo rekel: "Kaj me sedaj to sprašuješ? Ali misliš, da jaz neumnosti objavljam?!" Tako da v to že ne morem iti in jaz si običajno rečem, njegovo ime je gor, ne moje. Njemu se bodo smejali, ne meni. Bo že vedel, ali se spleča to objaviti, ali je to nekaj, s čimer si bo slabo reklamo delal po svetu. (vodja znanstvenoinformacijskega centra, NO5)

<sup>35</sup> Ampak mislim, da nek knjižničar nima možnosti, da bi šel tako daleč, da bi recimo neko tabelo številčk toliko opremil z metapodatki, da bi lahko nek uporabnik 'ven potegnill' že točno fizikalne podatke. Mislim, da tako daleč knjižničar ne more iti. (vodja znanstvenoinformacijskega centra, NO5)

---

<sup>36</sup> Na nivoju podatkov do zdaj pravzaprav nimamo izkušenj, medtem ko z objavami jih imamo. Tako da se pri objavah v bistvu na nek način že marsikaj dogaja, če nič drugega, gradimo COBISS in SICRIS. Ne vem, če lahko nastane kak repozitorij iz tega, ampak načeloma bi mogoče celo lahko. Pri podatkih pa ne vem, mogoče vidim večjo vlogo knjižnice pri prosto dostopnih objavah, kakor pa pri prosto dostopnih podatkih. Ker podatki so le nekako material, ki ni tako blizu knjižnicam, mi imamo zmeraj opravke z nekimi že obdelanimi stvarmi. *Se vam zdi, da bi v tem primeru morali bolj delati sami laboratoriji ali pa raziskovalci?* Ja, jaz mislim, da je v bistvu odgovornost v prvi vrsti na njih, kaj dajo iz rok in kako pridobijo te podatke. (vodja knjižnice, NO18)

<sup>37</sup> Hočem reči, da se bo po mojem mnenju premik na tem področju zgodil, če bo financer spoznal, da je to potrebno. Podobno kot se sedaj v tej politiki že omenja prosti dostop do rezultatov. Če se bo to začelo omenjati v politiki, potem bo to možno. Če bo recimo v neki sferi, na nekem področju, to postal standard, če bo od tebe pričakovano, da to objaviš, potem bo to tudi prodrlo. Kot je A. govoril, ti to imajo na svojem področju in zato se s tem ukvarjajo in vedo, kaj je to. Mi ostali, po eni strani nekateri tega sploh nimajo, po drugi strani pa ni politika, da bi bilo to odprto. Ta dva pristopa sta, ali bo baza na neki sferi to dosegla, ali pa bo financer na splošno. Da bi pa knjižničarji to pokrenili, to pa ne, po mojem mnenju. (vodja znanstvenoinformacijskega centra, NO5)

<sup>38</sup> Oni so bili takrat plačani direktno od države, tega sedaj ni več, mi se projektno financiramo. Res pa je, da je večina projektov še zmeraj teh, ki jih naročajo Ministrstva. Ti oddelki so bili ustanovljeni z namenom, da zbirajo gradivo, ki je nacionalnega pomena, kulturna ali pa naravna dediščina. To so bile etnografske, arheološke stvari itd. Potem, ko je prišla nova politika, se je ustanovil Javni zavod, ki ga je ustanovil SAZU, da je te oddelke nekako ločil od strukture SAZU-ja in je postal ZRC, potem pa je nastajalo vedno več novih inštitutov. Tako da je bilo naše prvo poslanstvo morda res zbiranje, ampak zmeraj so vsi ti ljudje, večinoma akademiki, iz tega delali študije in pisali. Njihov namen je zagotovo bil, da bodo to gradivo tudi obdelali in objavili rezultate. In to je v bistvu naše poslanstvo še sedaj. Mi konkretno smo delali recimo tudi popisni atlas, naročil ga je Statistični urad Republike Slovenije. Podatke so nam res dali oni, ampak mi smo potem te podatke obdelali in naredili knjigo do tiska. (znanstvenik s področja geografije, NO8)

<sup>39</sup> Mogoče še samo opozorim glede vrste podatkov, če imamo neko publicirano gradivo, na primer neko ploščo, potem je jasno: ker je to gradivo že namenjeno publiki, ponavadi zadaj stoji avtor in izvajalec, to je vse navedeno in je gradivo tudi pripravljeno za objavo. To je tudi relativno preprosto prenesti, samo preneseš v digitalno obliko. V tovrstnem gradivu, kakršna imamo v naših arhivih, ko gradivo ni bilo v osnovi namenjeno javnosti, je pa rezultat nekega raziskovalčevega raziskovanja, dela, dokumentiranja za kasnejšo znanstveno analizo in zato malo težje. Nekateri ne razumejo, rečejo, saj imate toliko in toliko trakov, enostavno digitalizirate in daste na mrežo. Pa rečemo, to pa ni čisto tako preprosto, ker večkrat je vmes to, da se ljudje motijo, se popravljajo. Ne vem, koliko je to zanimivo za nekega naključnega poslušalca, po drugi strani pa so včasih tudi kakšni takšni podatki ali pa pogovori ... Namreč zmeraj, kadar se snema, se pove, mi smo od te inštitucije, delamo za to in to ... In v tej luči se je potem zbiralo gradivo. Do pred parih let nikoli niso imeli v mislih, niti raziskovalci niti tisti, ki so snemali, da bo to nekje javno dostopno nekemu naključnemu obiskovalcu. Po eni strani bi rekel, da znajo biti tudi neki moralni zadržki, kako to ponuditi. /.../ Potem je treba neko selekcijo narediti. Saj mi tudi objavljamo gradivo, ampak to gradivo je potem izbrano in je dano v nek kontekst, mogoče tudi z nekim opisom, na primer na CD-ju ali pa na ploščah in se nekako pove, zakaj, kako... (znanstvenik s področja zvočnega arhiviranja, NO21)



---

<sup>40</sup> To je zmerom bilo neko dodatno delo, ki je bilo potrebno, da si sploh prišel do gradiva. Razlagam malo zgodovino, da boste vedeli, kakšna je danes realnost. Zato je bila nekako druga pot objavljane gradiva. Seveda se danes to sliši nesmiselno, ampak po drugi strani pa je glede tovrstnega gradiva to še zmeraj zelo dober način vračanja gradiva nazaj. Se pravi obdelati neko temo, področje, recimo Ribniško dolino, v obliki neke zgoščenke, knjige in vrniti potem gradivo nazaj. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>41</sup> Seveda subjektivno, objektivno, to je stvar avtorja, ampak tu se poskuša na nek znanstveni ali pa poljudnoznanstveni način, če je namenjen širši publikii, predstaviti tovrstno gradivo javnosti. In tega je veliko. Tako da se mi zdi, da je bila v tem smislu dobra pot. Še danes je precej uspešna, kot primer naj navedem, pride nekdo iz nekega kraja ali pa iz neke pokrajine in reče: "Joj, vi imate toliko gradiva, ste snemali 50 let, ste tukaj raziskovali, ali bi nam lahko naredili kopijo tega gradiva? Na primer avdio, mi bi tudi poslušali te posnetke." Pa je odgovor po navadi tak: "Veste, saj samo poslušanje vam veliko ne bo pomenilo, to ni publicirano gradivo, to ni zbirka plošč. Potrebujete tudi vse metapodatke, naše terenske zvezke in tako naprej, da boste lahko iz tega iztržili to, kar bi lahko. Poleg tega se pa priporoča še poznavanje širšega konteksta." Pa ne, da bi hoteli zapirati to gradivo. Tako se ponavadi potem dogovorimo, da skupaj, v sodelovanju z njimi naredimo publikacijo, v kateri se to gradivo predstavi, na primer konkretno nek pesemski repertoar neke pokrajine, s komentarji in tako naprej. (znanstvenik s področja zvočnega arhiviranja, NO21)

## 6. FAZA: DODANA VREDNOST

---

Fazi zbiranja podatkov običajno sledijo določeni postopki, s katerimi zbrane surove podatke ustrezno pripravimo za nadaljnjo uporabo. Nanaša na predelovanje podatkov ter na čiščenje oziroma odstranjevanje nepotrebne, odvečnega, motečega, vse to pa po možnosti spremlja dokumentacija vsega, kar se s podatki dogaja v metapodatkovnih zapisih. To fazo imenujemo dodana vrednost, kjer torej en del predstavljata predvsem čiščenje in transformacija podatkov, za kar raziskovalci poskrbijo že zaradi lastnega interesa, da podatke lahko analizirajo, drugi del pa je opremljanje podatkov z metapodatki, v čemer večina raziskovalcev ne vidi neposredne koristi. To se kaže v slabši skupni kakovosti podatkov z vidika možnosti druge rabe. Zato je še posebej pomembno, da se ustvarjalci podatkov zavedajo, katere spremljajoče informacije se potrebne za drugo rabo.

Intervjuvance smo prosili, da nam opišejo *stanje* v tej fazi življenjskega kroga podatkov ter katera orodja, postopke in navodila uporabljajo v tej fazi. Bolj natančno smo skušali izvedeti, kako poteka proces spreminjanja podatkov in dodajanja metapodatkov ter kakšne standarde pri tem upoštevajo, če sploh katere. Ali uporabljajo surove podatke, ali jim dodajo novo vrednost? Kaj se dogaja s podatki od nastanka do točke spravila z vidika dodane vrednosti?

Zanimalo nas je, kakšni problemi pestijo fazo dodane vrednosti in ali jih upoštevajo in rešujejo ali ne? Ali upoštevajo načrtovanje kakovosti, zahteve hrambe, ali je zagotovljena sledljivost postopkov in transformacij? Ali razrešujejo probleme za zagotavljanje dostopa: lažja dostopnost, odstranjene ovire zaupnosti in podobno.

Da bi opredelili *odgovornosti* v zvezi z reševanjem problemov, smo jih prosili, da opišejo svoje dolžnosti, pravice, medsebojna pričakovanja in odvisnost nosilcev vlog. Kdo opravlja? Sam ali pooblaščen podatkovni center, repozitorij ali specialist v okviru organizacije, centralni nacionalni digitalni arhiv, mednarodni centri ...

Zanimalo nas je, kakšen je pomen politik, vodil, standardov pri opredeljevanju odgovornosti ter kakšne so ovire, sankcije, spodbude in nagrade. Ali dodana vrednost prinese nagrado? Ali dobi izvajalec povrnjene stroške? Ali so opredeljene obveznosti glede dodane vrednosti, metapodatkov (na nivoju organizacije, discipline, nacionalno)? Kdo bi moral opredeliti obveznosti? Ali je za doseganje ciljev potrebno opredeliti obveznosti – enotna raziskovalna politika? Skušali smo identificirati *nosilce dobrih praks*, zato smo jih vprašali, kje dobijo navodila oziroma napotke za usposabljanje za delo na tem področju? So to morda področni centri za strokovno svetovanje, kdo je vodilni na tem področju, kje so nastavki dejavnosti?

Pokazalo se je, da na nekaterih področjih obstajajo že uveljavljeni standardi, na drugih morda bolj neformalni dogovori, na tretjih so stvari popolnoma prepuščene raziskovalcu. Intervjuvanec s področja zgodovine pravi, da so pogovori o metapodatkih na tem področju precej novi, o tem se govori šele kratek čas.<sup>1</sup> Raziskovalec s področja naravoslovnih ved pravi, da v ustanovi nimajo enotnega predpisa o dokumentiranju, da se raziskovalec tega znanja priuči od mentorja tekom uvajanja v raziskovalno delo<sup>2</sup> in dodaja, da so na tem področju podatki brez ustrezne dokumentacije za drugega neuporabni.<sup>3</sup> Podobno ugotavlja tudi znanstvenik s področja ekonomije.<sup>4</sup> Na področju zvočnega arhiviranja se zavedajo prednosti standardizacije, ampak še niso našli prave rešitve.<sup>5</sup> Znanstvenik s področja gradbene informatike pravi, da je XML tehnologija dovolj standardizirana, da je uporabna za večino namenov,<sup>6</sup> hkrati pa vidi potrebo, da bi država na tem področju spodbujala rabo standardov.<sup>7</sup> Na področju biologije so v splošnem že sprejeta pravila in standardi glede tega, kaj so kakovostni

podatki,<sup>8</sup> k upoštevanju le-teh pa dodatno prispevajo akreditirani laboratoriji, ki od ustvarjalcev podatkov zahtevajo določeno kakovostno raven.<sup>9</sup> Tudi na področju meteorologije je mogoče govoriti o ustrezni standardizaciji podatkov,<sup>10</sup> na področju fizike delcev pa poteka vzpostavitev infrastrukture, ki bo omogočala učinkovito hrambo in izmenjavo podatkov velikega obsega.<sup>11</sup>

Znanstvenik s področja zvočnega arhiviranja ugotavlja, da raziskovalci deloma že sami po svoji vesti in skladno z utečenimi navadami pripravljajo spremno dokumentacijo, vendar pa jim za kaj bolj doslednega zmanjkuje časa.<sup>12</sup> Pomanjkanje časa, pa tudi ustreznih znanj, je ena od ključnih ovir, ki znižuje kakovost dokumentacije v fazi ustvarjanja. Za kakovostno in kontrolirano pripravo podatkov bi potrebovali veliko ur in ustrezno tehnologijo, česar v obstoječem sistemu preprosto nimajo, dodaja znanstvenik s področja zvočnega arhiviranja.<sup>13</sup>

Intervjuvanci so izrazili potrebo po strokovni pomoči oziroma po kadrih, ki bi jih razbremenili potrebnega in nujnega dela. Znanstvenik s področja biologije ugotavlja, da bi za to potrebovali dodatno službo,<sup>14</sup> podobno zaključuje tudi intervjuvanec s področja ekonomije.<sup>15</sup> Več o tem bomo spregovorili nekoliko kasneje.

Splošno prevladujoče stanje je, da znotraj inštitucije običajno ni službe, ki bi skrbela za arhiviranje niti ni skupne politike glede postopkov dolgotrajne digitalne hrambe. Skrb za hrambo podatkov in ostalih gradiv je tako večinoma stvar metodologije posameznih raziskovalcev ali raziskovalne skupine, ki tekom dela običajno sami oblikujejo svoja neformalna pravila. Je pa izražen interes, da se podatki hranijo čim bližje nastanku in da raziskovalec ohrani nadzor nad tem, kaj se s podatki dogaja.<sup>16</sup> Takšen model skrbništva nad podatki s strani specialistov se v določenem pogledu uveljavlja tudi v okviru disciplinarnih arhivov, tak primer je britanski Natural Environment Research Council (NERC), ki deluje kot mreža pooblaščenih inštitucij za posamezna področja.

## Okvirček 6.1: Primer interdisciplinarne ustanove

### ***Natural Environment Research Council***

Vzorčen primer za zasnovo interdisciplinarne ustanove – v kateri so povezani različni raziskovalni centri - je britanski NERC. Ustanova skrbi za financiranje, za upravljanje z raziskavami, za usposabljanje in menjavo znanja na področju okoljskih vprašanj. Ena od njenih nalog je tudi arhiviranje podatkov in hramba publikacij. Pod okriljem NERC delajo *British Geological Survey* (BGS), *British Antarctic Survey* (BAS), *Centre for Ecology and Hydrology* (CEH) in *National Oceanographic Centre* (NOC). Pri izvajanju raziskav pa sodelujejo še številni drugi britanski centri.

Za NERC delajo številni znanstveniki in inženirji: biologi, kemiki, ekologi, geologi, matematiki, meteorologi, oceanografi, fiziki in zoologi. Raziskovanje vključuje atmosferske, zemeljske, biološke, kopenske in vodne znanosti, od globin oceanov do vrhnjih plasti atmosfere, od tečajev do ekvatorja. Gre za visoko tehnološke raziskave, pri katerih uporabljajo raziskovalne ladje, letala, satelite, super računalnike in drugo specializirano opremo.

V predstavitvi na svoji spletni strani so zapisali, da so njihovo največje premoženje podatki, ki jih sami ustvarijo. Oblikovali so vrsto podatkovnih središč, ki hranijo in upravljajo podatke NERC-a. Ta središča zagotavljajo podporo in vodstvo pri upravljanju s podatki, odgovorni so za dolgoročno upravljanje in zagotavljajo dostop do podatkov. Politika je oblikovana tako, da so vsi podatki, ki so financirani s strani NERC, javno dostopni in brez omejitev, kar zadeva rabo. Ustanova od raziskovalcev zahteva ustrezno načrtovanje in upravljanje s podatki. Jasno je tudi določena pravica 'prve rabe', po kateri bo imel ustvarjalec podatkov največ dve leti časa prednost pri rabi podatkov, ki jih je sam ustvaril. Oblikovali so seznam vrednot, po katerem je mogoče določiti, kateri podatki so smiselni za dolgoročno hrambo.

Vir: <http://www.nerc.ac.uk/>

## 6.1 ČIŠČENJE, TRANSFORMACIJA PODATKOV IN METAPODATKI

Z intervjuvanci smo se pogovarjali o tem, ali morajo surove podatke, predno jih lahko uporabijo, kakor koli obdelati ter s kakšnimi postopki to storijo. Iz odgovorov sledi, da je večino podatkov potrebno vsaj minimalno obdelati, predno so uporabni za analizo, na nekaterih področjih pa so brez dodatne obdelave celo povsem neuporabni.

Surovi podatki so osnova, v nadaljnjih raziskovalnih postopkih, to je z rabo različnih metod in tehnologij, se ti spreminjajo, preoblikujejo, čistijo. Končni proizvod so lahko podatki v različnih verzijah, ki so uporabni za različne raziskovalce. Za hrambo podatkov, torej za arhiviranje, niso pomembni le surovi podatki in transformirani podatki, pač pa tudi zapisi o transformaciji, programska oprema in druge spremljajoče informacije, ki zagotavljajo sledenje spremembam. Problem je tudi uporaba večinoma licenčnih programov, saj bi se transparentnost v smislu dolgotrajne hrambe povečala z rabo odprtokodnih programov, pri katerih lahko arhiv lažje preveri in ohrani izvorno kodo. Sledenje smernicam ISO OAI standarda in na njih temelječih dobrih praksah pri oblikovanju in vzdrževanju informacijskih paketov prav tako velja kot garancija za kakovostno hrambo (Griffin in drugi 2012).

Kakšni so ti postopki, je zelo odvisno od tipa podatkov, so pa večinoma pravila za čiščenje in transformacijo podatkov vsaj neformalno opredeljena kot uveljavljene prakse na posameznih inštitutih. Podatke, ki jih generira neka naprava, ki mora že sama po sebi zadostiti določenim standardom, je običajno potrebno le vizualno očistiti, pri čemer se izločijo artefakti, mora pa biti

oseba, ki to počne, zadostno usposobljena. Na področjih, kjer pogosto kombinirajo več različnih podatkovnih baz, se pred analizo redno preverja konsistentnost podatkov in v primeru nenavadnih odstopanj se za konkretne primere poskuša ugotoviti razlog odstopanja.<sup>17</sup> Pri zgodovinarjih je zelo uveljavljena praksa preverjanja pravilnosti posameznega podatka iz različnih zgodovinskih virov. Raziskovalci, ki izvajajo poglobljene kvalitativne intervjuje, potem ko naredijo prepis intervjuja, intervjuvance pogosto prosijo za avtorizacijo besedila (medtem ko protokola za natančnost pri transformaciji zvočnega zapisa v besedilo nimajo) in s tem zagotovijo kar najmanj izgub informacij, ki nastanejo v procesu transformacije.<sup>18</sup> Postopek čiščenja je tako posebej zahteven pri avdio-vizualnih gradivih in besedilih (besedilni korpusi).

Na nekaterih področjih pa so standardi za kakovost podatkov še posebno ustaljeni in jasni, npr. v naravoslovju obstajajo politike obravnavanja podatkov in je kakovost podatkov mogoče preverjati tudi za nazaj, ko imamo na podlagi analize podatkov že rezultat.<sup>19</sup> Za zbiranje prostorskih podatkov obstaja ISO standard za kakovost,<sup>20</sup> za zbiranje zvočnih in video gradiv obstajajo priporočila za delo na terenu, ISO, AES, IASA,<sup>21</sup> ponekod pa standarde za kakovost določi kar skupina strokovnjakov v sami metodologiji, ki jo pripravijo na začetku projekta.<sup>22</sup> Vsi intervjuvanci dobro razumejo, da je za zagotavljanje uporabnosti podatkov v dostopu potrebna dodatna dokumentacija.

## Okvirček 6.2

### **Zgodovina: spletna stran Zgodovina Slovenije - Sistory**

Zgodovina Slovenije - Sistory je spletna stran Inštituta za novejšo zgodovino (INZ) v Ljubljani. Izobraževalni in raziskovalni portal slovenskega zgodovinopisja je javno dostopen od konca leta 2008. V štirih vsebinskih/področnih sklopih je uporabnikom na voljo obsežen nabor zgodovinskih in zgodovinopisnih vsebin z adekvatnimi analitičnimi sistemi in ustreznim znanstvenim aparatom.

Gradnja portala še poteka in vključuje prenos gradiva v novo medijsko in tehnološko obliko oziroma digitalizacijo in prezentacijo gradiva na spletni strani, s tem pa tudi skrbi za podaljšanje njegove relevance. V letu 2011 so v INZ skenirali 13.000 strani, v fazi digitalizacije pa je 43.000 strani arhivskega gradiva.

Leta 2011 je INZ portal nadgradil v spletni sistem, ki je primerljiv z evropskimi in svetovnimi trendi in sistemi, ki omogočajo ustvarjanje raziskovalnih podatkov in njihovo uporabo v raziskovalnem okolju. Spletna stran je kompatibilna z ostalimi sorodnimi evropskimi spletnimi portali oziroma z arhitekturo evropske mreže za humanistiko (DARIAH), v prvi fazi pa predstavlja predvsem izhodišče sodelovanja z ZRC SAZU pri vzpostavitvi nacionalne digitalne infrastrukture za humanistiko in umetnost z akronimom SI-DIH, ki je opredeljen z osnutkom Načrta razvoja RI 2011-2020 Ministrstva za visoko šolstvo, znanost in tehnologijo.

Sami pravijo, da je spletni portal Zgodovina Slovenije - Sistory po nadgradnji dokončno presegel vlogo repozitorija, knjižnice ali arhiva gradiv v klasičnem pomenu besede in je povsem primerljiv z evropskimi in svetovnimi trendi in sistemi, ki omogočajo ustvarjanje raziskovalnih podatkov in njihovo uporabo v raziskovalnem okolju. Verificirane vsebine, objavljene na spletnem portalu Zgodovina Slovenije - Sistory, so pritegnile raziskovalne in izobraževalne kadre različnih disciplin doma in v tujini, hkrati pa so vzbudile zanimanje tudi med osnovnošolci ter dijaki in nenazadnje med laično publiko.

Vir: <http://www.sistory.si/>

Za določen segment podatkov, tistih, ki so namenjeni za mednarodne zbirke (npr. GenBank, UniProt, Protein Data Bank, Amadeus), je zagotovljena popolna dokumentacija nastajanja podatkov. Postopki dodane vrednosti (normalizacije, čiščenja) podatkov so dokumentirani; oddajo se tako surovi podatki kot normalizirani podatki. Spremljajoča kakovostna, izčrpna in standardizirana informacija o poteku eksperimentov je v tem segmentu del vzpostavljene prakse. Tovrstne mednarodne povezave so pomembne tudi zaradi izmenjav znanja pri ustvarjanju in dokumentaciji ter vrednotenju raziskovalnih podatkov.

V določenih krogih, v katerih je kultura izmenjave podatkov že dalj časa prisotna, npr. genomske raziskave, je ravno potreba po izmenjavi podatkov spodbudila vzpostavitev in uporabo metapodatkovnih standardov.<sup>1</sup> Raziskovalci takšnih znanosti se zato zelo dobro zavedajo problemov protokolov in točnega zapisa, poleg tega pa so tudi zavezani k njihovi uporabi pri objavi podatkov v mednarodnih zbirkah. Takšne podatkovne zbirke imajo običajno za vnos podatkov že narejene aplikacije, ki zahtevajo določene metapodatke.<sup>23</sup>

Podoben primer dobre prakse je tudi področje meteorologije, kjer so protokoli za zbiranje podatkov in standardi za metapodatke že zelo dolgo splošno sprejeti in uporabljeni.<sup>24</sup> Njihovo uveljavitev in vsesplošno uporabo je prav tako spodbudila potreba po medsebojni izmenjavi tovrstnih podatkov, ki pa se kaže tudi na drugih področjih, npr. v gradbeništvu, kjer informacijsko bolj razvite države zahtevajo dokumentacijo v standardnih oblikah (ISO-Step, IFC).<sup>25</sup>

Pomembnosti metapodatkov se dobro zavedajo tudi intervjuvanci, ki se ukvarjajo z avdio in video gradivom. Tovrstne podatke običajno spremljajo vsebinsko zelo bogati metapodatki v obliki zapisnikov, ki vsebujejo bolj ali manj strukturirana polja za tehnične in vsebinske informacije. Brez tega bi bila uporabnost gradiva znatno manjša.<sup>26</sup>

Na nekaterih področjih, na katerih še ni uveljavljenih standardov raziskovanja in dokumentacije podatkov, so se posamezne inštitucije znašle na različne načine: za zapis metapodatkov uporabljajo kar Dublin Core, ki ga za potrebe svoje zbirke podatkov dopolnijo z dodatnimi polji, metapodatke zapisujejo v nestandarden xml zapis, ki ga je razmeroma enostavno brati, v eni od inštitucij pa so se lotili tudi priprave internega pravilnika o organizaciji raziskovalnega dela, ki izhaja iz načel projektnega menedžmenta.<sup>27</sup>

V povezavi z zagotavljanjem metapodatkov smo naleteli tudi na več zanimivih manjših projektov, kot sta npr. projekt Fototeka, ki je vključeval digitalizacijo njihovega lastnega arhiva slik in opremljanje z metapodatki,<sup>28</sup> ter jezikovni korpus, ki uporablja metapodatkovni standard TEI.<sup>29</sup>

Kljub temu, da že obstaja kar nekaj dobrih praks za specifične tipe podatkov, pa je dokumentacija podatkov v fazi nastajanja najpogosteje še vedno prepuščena navadam in potrebam samega raziskovalca. Znanje in strogost glede priprave in sprotne natančne dokumentacije podatkov je priučeno znanje, stvar notranje kulture raziskovalne skupine ali celo pazljivosti, discipline posameznega raziskovalca; pogosto metapodatke predstavlja raziskovalčev prosti opis eksperimenta, ki podatke spremlja v obliki delovnega poročila ali programske dokumentacije in nima vnaprej definiranih polj, po katerih bi podatke lahko klasificiral in iskal v katalogu.

Podatki, ki jih generirajo naprave, so že ob nastanku opremljeni z osnovnimi metapodatki, pri novejših aparataturah je možno vpisati tudi veliko dodatnih informacij, česar pa se raziskovalci le redko poslužujejo.

---

<sup>1</sup> Standard MIAME, <http://www.mged.org/Workgroups/MIAME/miame.html>, 7. 8. 2012

Eden od razlogov, da večina organizacij nima predpisanih standardov za metapodatke je, da so področja pogosto zelo kompleksna in si odgovorni ne predstavljajo, kako bi zagotovili poenotene predpise za različne discipline. Za mnenje smo povprašali tudi knjižničarje, kot strokovnjake na področju bibliografije, ki pa rešitve pri opisovanju podatkov prepuščajo raziskovalcem, sami se namreč ne vidijo v kakšni bolj aktivni vlogi, ne pri opisovanju in pripravi podatkov, ne pri njihovem vrednotenju.<sup>30</sup> Prispevek knjižničarjev bi lahko bil, da se usposobijo za sprejemanje različnih vrst gradiv. COMARC standard je po njihovem mnenju dovolj širok, da bi lahko pokrival tudi raziskovalne podatke.

Očitno je, da znanja in izkušnje s pripravo podatkov za podatkovni arhiv v znanstveni skupnosti obstajajo, saj smo našli na kar nekaj dobrih praks iz različnih področij. Področja, kjer so te prakse vzpostavljene, lahko služijo kot nastavek za reševanje problemov na drugih področjih. Te dobre prakse bi bilo potrebno izpostaviti in na njih graditi, vključno z usposabljanjem za uporabo standardov pri opisovanju, kot je MIAME. Za določene, t. i. bolj 'tradicionalne' eksperimente, pa je pristop bolj 'akademski' v tem smislu, da postopki niso tako strogi, marsikaj se v hitenju k analizi zapiše bolj na hitro, se opusti.

#### Okvir 6.1: P-R 6

##### *PROBLEM:*

Inštitucije večinoma (še) niso oblikovale in sprejele pravil za uporabo metapodatkovnih standardov, tako je dokumentacija podatkov največkrat prepuščena samim raziskovalcem.

Večina raziskovalcev dokumentiranje podatkov še vedno vidi le kot dodatno delo, za katerega nimajo ne časa ne sredstev. Nekateri ne vidijo prednosti zaradi uporabe standarda oziroma menijo da zanje niso dovolj velike, da bi v to vlagali svoj čas in znanje.

Ugotovili smo tudi, da znanje sicer obstaja, vendar ni spodbude s strani institucij in financerja, da bi se to znanje dejansko uporabilo.

##### *REŠITEV:*

Rešitev je treba iskati v smeri povezovanja skupin ekspertov na disciplinarnih področjih, ki bodo svetovali (dobri zgledi in prakse) pri pripravi metapodatkov že v fazi ustvarjanja podatkov. Hkrati pa bodo svetovali tudi pri iskanju specializiranih storitev hrambe in dostopa do podatkov. Potrebno bo spodbuditi področja, na katerih obstajajo znanja in razvojne dejavnosti ter vzpostaviti evidenco ekspertnih skupin in njihovih kompetenc in poskrbeti za institucionalizacijo zagotavljanja njihovih svetovalnih storitev.

S tem se odpira potreba po oblikovanju novega poklica in novega kadra. Hkrati bo treba preučiti, kakšen del nove storitve bodo lahko prevzeli knjižničarji. Knjižničarje, kot prejemnike gradiva v repozitorije, bi bilo potrebno dodatno usposobiti glede klasifikacij, tipov podatkovnih gradiv, in za podpora pri pripravi podatkov za objavo.



## 6.2 SPRAVILO

Produkti ustvarjanje in transformacija podatkov skupaj z njihovo zgodovino so predmet spravila podatkov, skrbi za njihovo dolgotrajno razpoložljivost tako za ustvarjalce kot tudi za drugo rabo. Intervjuvance smo prosili, da nam opišejo korake v fazi spravila. Zanimalo nas je, kaj se dogaja s podatki, kakšne vrste in formati podatkov nastopajo; kje podatke hranijo; kakšna znanja obstajajo na področju hrambe; katera orodja, metapodatke, standarde in postopke uporabljajo? Ali imajo načrte za hrambo? Ali poznajo Open Archival Information System (OAIS)?

Nadalje nas je zanimalo, kako opredeljujejo probleme, povezane z drugimi fazami življenjskega kroga podatkov. Ali je shranjevanje namenjeno bolj za lastne potrebe ali pa upoštevajo možnost, da ob koncu projekta podatki služijo tudi drugim? Kakšne so prednosti hrambe na obstoječi lokaciji ali centralizirano po disciplinah? Katere ciljne vsebine (*intellectual content*) so namenjene hrambi? Ali je pomen podatkov strokovno ovrednoten, v smislu dolgotrajne vrednosti za projekt in za druge? Kaj se zgodi s podatki majhne vrednosti? Ali gre za "enkratne" podatke, ki jih je mogoče z lahkoto reproducirati? Ali je zagotovljeno ohranjanje vrednosti podatkov? Ali je razpoložljivost podatkov ogrožena zaradi odsotnosti hrambe? Ali je v okviru sistema hrambe zagotovljena sledljivost izvora podatkov in postopkov v predhodnih fazah?

Intervjuvance smo prosili, naj pojasnijo, kako vidijo odgovornost v zvezi z reševanjem problemov v fazi spravila podatkov. Kje vidijo primerno mesto za odlaganje podatkov? Kakšen je pomen politik, vodil in standardov pri opredeljevanju odgovornosti? Ali imajo merila in vrednotenje kakovostne hrambe? Kaj so ovire, sankcije, spodbude in nagrade? Kdo plača hrambo, kdo skrbi za usposabljanje glede hrambe? Kakšni so pravni problemi hrambe: prenos pravic, zaščita podatkov? Ter primeri dobrih praks, zgledov, ki bi lahko bili izhodišča za razvoj te faze podatkov.

V ustanovah, iz katerih prihajajo naši intervjuvanci, ne samo da ustvarijo ogromno raznovrstnih podatkov, pač pa je tudi množstvo različnih sistemov hrambe tako z vidika formatov kot same organizacije spravila gradiva in skrbi za dolgotrajno hrambo. Če govorimo o elektronskih gradivih, se ta največkrat hranijo na osebnih računalnikih, včasih pa se dodatno prenesejo še na kakšen drug medij ali na strežnik, za katerega se periodično dela varnostna kopija.<sup>31</sup> Ponekod imajo tudi vzpostavljen sistem sistematičnega shranjevanja gradiv na skupen strežnik, do katerega lahko dostopajo vsi zaposleni, običajno je urejen po sistemu kronološkega shranjevanja.<sup>32,33</sup> Pri gradivih v fizični obliki, teh je bilo manj, pa prevladuje praksa, da se podatki hranijo v lastnih prostorih ali pa v za to namenjenih ustanovah (npr. muzejske zbirke, vzorci tkiv, zbirka jamskih živali ipd.).<sup>34,35</sup>

Če se vrnemo nazaj k podatkom v elektronski obliki, se v povezavi z različnimi mediji, na katerih te podatke hranijo, pojavlja velik problem zastarelosti medijev. Pri starejšem gradivu, ki obstaja ali v fizični obliki ali pa na zastarelem mediju, so se marsikje lotili tudi digitalizacije, ki pa je zelo dolgotrajna in zahtevna.<sup>36, 37, 38</sup>

Tudi na področju formatov največji problem predstavlja zastarelost, zato marsikje poskušajo uporabljati čim bolj standardne formate (ASCII) ter programe in orodja, ki so uporabni na dolgi rok.<sup>39</sup> Pogosto se podatki hranijo v formatu, ki je vezan na program, v katerem podatke analizirajo. Podatkovne baze ponekod hranijo v tekstovnem formatu ali Excelu, zraven pa hranijo tudi sintakso in merski instrument.<sup>40</sup> Včasih so formati specifični za projekt, ali pa so lastni posamezni napravi za generiranje podatkov. Ko se z aparaturo zamenja tudi format, običajno proizvajalci poskrbijo za možnost izvoza v bolj univerzalen format, npr. ASCII.<sup>41</sup>



## Okvirček 6.3

**Umetnost: Arhiv in zbirke Glasbenonarodopisnega inštituta ZRC SAZU**

Glasbenonarodopisni inštitut je bil ustanovljen leta 1934 in je kot tak najstarejši inštitut ZRC SAZU. Ena od temeljnih nalog tega inštituta je *"sestavljanje čim popolnejše zbirke slovenske glasbene folklore"*. Poleg slovenskih ljudskih pesmi v arhivu hranijo raznovrstne zapise ljudske glasbe drugih narodov in razne dokumentarne posnetke. V njihovih zbirkah je moč najti: **pisni arhiv** (rokopisi, pesmarice, zapuščine, transkripcije zvočnih posnetkov, zapisniki, dokumentacijo delovanja inštituta in starejših akcij zbiranja ljudskih pesmi), **zvočni arhiv** (posnetki na voščeni valjih, na gramofonskih ploščah, na magnetofonskih trakovih, na kasetah, na digitalnih zvočnih kasetah (DAT) in posnetki na zgoščenkah), **vizualno dokumentacijo** (filmi s folklorno tematiko, videokasete različnih formatov, fototeka in diateka) ter **zbirko ljudskih inštrumentov**.

Bogatenje zbirke je bilo in je še zmeraj močno povezano z razvojem snemalnih naprav in možnosti, ki jih je imel na voljo inštitut, hkrati pa je rezultat povezovanja inštituta navzven.

*"Danes je temeljna dejavnost Glasbenonarodopisnega inštituta poleg rednega raziskovalnega dela in objavljanja raziskav predvsem izdajanje temeljnih in drugih zbirk ljudskih pesmi, sistematično izdajanje ljudskih plesov, izpopolnjevanje, hranjenje in objavljanje izjemnega arhiva na kasetah in zgoščenkah, spremljanje sodobne dejavnosti, povezane z glasbenim in plesnim izročilom, in strokovno posredovanje na tistih področjih, kjer se folklorizem oplaja ob pesemski, glasbeni in plesni dediščini."*

Vir: <http://gni.zrc-sazu.si/#v>

Na nekaterih področjih so za potrebe lastnega shranjevanja sami razvili (odprtokodne) aplikacije, na primer program Neva (aplikacija za shranjevanje, iskanje in prikazovanje gesel), ki deluje vzporedno s spletno aplikacijo Eva, ki podatke v obliki html-ja prikazuje na spletu.<sup>42</sup>

V redkih primerih, kakršno je področje meteorologije, za hrambo podatkov poskrbijo operativne nacionalne službe (ARSO), shranjevanje podatkov pa je urejeno na svetovni ravni (WMO).<sup>43</sup>

Pri spravilu gradiva težavo pogosto predstavljajo izredno velike datoteke, ki zavzamejo ogromno prostora, npr. podatki v obliki slik v visoki ločljivosti, digitalizirano besedilo v TIF formatu, avdio datoteke, ipd. Zaradi prostorskih omejitev podatke tudi brišejo.<sup>44</sup>

Največji problem, ki so ga intervjuvanci izpostavljali, je ta, da v okviru projekta trenutno ni mogoče poskrbeti za dolgoročno hrambo, saj se z zaključkom projekta zaključi tudi financiranje.<sup>45, 46</sup> Več intervjuvancev je izrazilo željo, da bi bila na višji ravni (nacionalni, mednarodni) vzpostavljena storitve dolgotrajne hrambe, kot nek centraliziran sistem, sami namreč nimajo ustreznega znanja za to. V povezavi s tem se bojijo prevelikih stroškov. Nekateri izražajo pričakovanje, da bi lahko razmere glede hrambe digitalnega gradiva, tudi kar zadeva raziskovalne podatke, urejal obstoječi krovni Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih (ZVDAGA).<sup>47</sup>

Intervjuvanci se zavedajo, da je vprašanje digitalne hrambe in skrbništva, če je zastavljeno na trdnih strokovnih temeljih, precej kompleksno področje, ki presega kompetence delujočih znanstvenikov ter tudi časovne in finančne zmožnosti tekočih projektov.<sup>48</sup> Iz tega izhaja potreba po centraliziranih in strokovno vodenih storitvah za zagotavljanje ustreznih pristopov in uporabe orodij ter medijev. Tovrstno vlogo bi lahko zagotavljali strokovnjaki, ki so sodelovali na Slovenski konferenci o trajnem

ohranjanju digitalnih virov,<sup>ii</sup> tehnološke rešitve in strojna oprema pa so vsakič stvar optimalnih izbir danega trenutka. Tudi v primerjavi z drugimi raziskavami, ki smo jih omenili v uvodu (npr. Parse.Insight), je splošni vtis, da je vprašanje digitalnega skrbništva in digitalne hrambe šele začelo postajati ena od pomembnih sestavin upravljanja s podatki znotraj življenjskega kroga podatkov. Razvoj gre v smeri, da bodo raziskovalci potrebovali pomoč tistih strokovnjakov, ki imajo posebna znanja in izkušnje s področja upravljanja s podatki in ki razumejo obseg in vsebine življenjskega kroga podatkov. Se pravi, da je potrebno razlikovanje med strokovnimi kompetencami ustvarjalca podatkov in strokovnim znanjem za opravljanje storitev skrbništva in ravnanja s podatki. Slednja so posebna znanja, čeprav pogosto v tesni povezavi z ožjimi strokovnimi področji posameznih znanstvenih disciplin – kako dokumentirati, urejati, hraniti – da bodo podatki dostopni in uporabni tudi za druge, ne le za ustvarjalce. Pri tem se pojavlja potreba po dveh vlogah: prva se nanaša na samo izvajanje storitve, druga pa kot pomoč v smislu svetovanja o dobrih praksah in zgledih.

Sicer pa smo na posameznih področjih identificirali kar nekaj zbirk, ki so nastale v okviru različnih projektov, primarno z namenom, da lahko do njih dostopajo različni uporabniki, in jih je mogoče razumeti kot zametke dobrih praks:

- v Centru za kartografijo favne in flore<sup>iii</sup> generirajo biološke podatke (razširjenost živalskih in rastlinskih vrst), lahko bi jih vzeli kot referenco za to, kako poskrbeti za posredovanje podatkov drugim.
- Eva in Neva<sup>iv</sup> - jezikoslovci so razvili lastno aplikacijo za shranjevanje, iskanje in prikazovanje gesel, sestavljeno iz dveh delov: Neva teče na strežniku, drugi del, Eva pa skrbi za prikazovanje v oblik html-ja preko spletnih brskalnikov (koda je odprta, lahko se dodajajo nova polja, če je potrebno). Prikaz za splošno rabo (on-line dostop) je omejen, za lokalno rabo pa lahko raziskovalci dobijo tudi daljša besedila, Program Eva uporabljajo tudi biologi, arheologi na Inštitutu za raziskovanje Krasa,<sup>v</sup>
- jezikoslovci uporabljajo tudi Wikiverzo (kot vadnico, ni dostopno javnosti), Wikivir<sup>vi</sup> (za tekste) in Wikipedijo<sup>vii</sup> (za besedilno zbirko),
- slovarske in besedilne zbirke, objavljene na spletu,<sup>viii</sup>
- FidaPLUS – korpus slovenskega jezika,<sup>ix</sup>
- Slovensko leposlovje na spletu,<sup>x</sup>
- Slovenski zgodovinski roman,<sup>xi</sup>
- Slovenska kmečka povest<sup>xii</sup>
- Zgodovina Slovenije - Sistory<sup>xiii</sup> je portal, na katerem so poleg digitalnih objavljene tudi digitalizirane zbirke podatkov (slike, tiskano gradivo, ustno) s področja zgodovine. Pri objavljanju določenih gradiv je pomembno vprašanje avtorskih pravic.

<sup>ii</sup> Glej: <http://www.nuk.uni-lj.si/todv/konferenca/>, 7. 8. 2012

<sup>iii</sup> <http://www.ckff.si/>, 7. 8. 2012

<sup>iv</sup> <http://www.laze.org/>, 7. 8. 2012

<sup>v</sup> <http://izrk.zrc-sazu.si/#v>, 7. 8. 2012

<sup>vi</sup> [http://sl.wikisource.org/wiki/Glavna\\_stran](http://sl.wikisource.org/wiki/Glavna_stran), 7. 8. 2012

<sup>vii</sup> [http://sl.wikipedia.org/wiki/Glavna\\_stran](http://sl.wikipedia.org/wiki/Glavna_stran), 7. 8. 2012

<sup>viii</sup> <http://bos.zrc-sazu.si/>, 7. 8. 2012

<sup>ix</sup> <http://www.fidaplus.net/>, 7. 8. 2012

<sup>x</sup> <http://slovenskaliteratura.ff.uni-lj.si/sl.html>, 7. 8. 2012

<sup>xi</sup> [http://lit.ijs.si/zgr\\_dons.html](http://lit.ijs.si/zgr_dons.html), 7. 8. 2012

<sup>xii</sup> <http://lit.ijs.si/skp1.html>, 7. 8. 2012

<sup>xiii</sup> <http://www.sistory.si/>, 7. 8. 2012

- ARKAS<sup>xiv</sup> je osnovna baza arheoloških najdišč Slovenije, v njej so združene podatkovne, tekstovne baze in kartografske podlage. Enota je arheološko najdišče, ki ga opišejo na različne načine, da ga lahko povežejo z drugimi bazami, ki niso njihove. Ponuja možnosti iskanja. So eni prvih v Sloveniji, ki so se lotili zametka informacijskega sistema v digitalni obliki in ki bi naj predstavljal eno infrastrukturo za celotno arheološko dejavnost.
- Na področju humanistike in umetnosti je povezovanje digitaliziranih podatkov v precejšnjem zagonu. Mednarodna pobuda DARIAH<sup>xv</sup> je projekt, s katerim skušajo podpreti in okrepi digitalno raziskovanje.
- V povezavi z DARIAH se razvija portal slovenske znanstvene digitalne humanistike, imenovan SIDIH.<sup>xvi</sup> In sicer s ciljem, povezati različne baze podatkov in poenotiti metapodatke za različna gradiva.
- Primer specializiranega spletišča s področja humanistike je Etnoinfolab,<sup>xvii</sup> ki predstavlja slikovno, zvočno in filmsko gradivo s področja etnologije in kulturne antropologije.
- DEDI<sup>xviii</sup> - Enciklopedija naravne in kulturne dediščine na Slovenskem.
- Arzenal<sup>xix</sup> je portal, na katerem inštituti ZRC SAZU-ja objavljajo vsebine, ki jih želijo dati v uporabo javnosti; metapodatki za gradiva so v besedilni obliki.

Takšne zbirke običajno obsegajo tudi različne vidike podatkovnih storitev (npr. prevzemajo gradiva več avtorjev, omogočajo dostop, itd.). Sredstva za financiranje vzdrževanja in nadgrajevanja teh zbirk so v glavnem pridobljena s projekti, zaradi česar ni dolgoročno zagotovljenega financiranja in je ogrožena kontinuiteta, kar skrbniki izpostavljajo kot hud problem.

#### Okvir 6.2: P-R 7

##### **PROBLEM:**

Največji problem na področju hrambe je, da stvari niso urejene na ravni institucionalnih politik in je skrb za hrambo v največji meri stvar metodologije vsakega posameznika.

Večina raziskovalcev je povedala, da v okviru projekta nimajo zagotovljenih finančnih sredstev, s katerimi bi lahko zagotovili dolgoročno hrambo.

Poleg tega se srečujejo še z drugimi ovirami, kot so hitra zastarelost medijev in formatov, pomanjkanje prostora ipd.

##### **REŠITEV:**

Potrebno bo oblikovati politike, ki bodo upoštevale načrtovanje upravljanja s podatki (DMP). Tako bo financiranje upravljanja s podatki že vnaprej načrtovano in kot tako tudi zagotovljeno.

Pri premagovanju tehnoloških rešitev je smiselna mednarodna integracija: katalogi, orodja itd. Na to temo bi lahko vodili kontinuirano evidentiranje orodij in pristopov ter uvajanje novosti na področjih: tehnološka podpora za upravljanje in razširjanje podatkov.

### 6.3 CITATI

<sup>xiv</sup> <http://arkas.zrc-sazu.si/>, 7. 8. 2012

<sup>xv</sup> Digitalna infrastruktura za humanistiko in umetnost: <http://www.dariah.eu/>, 7. 8. 2012

<sup>xvi</sup> [http://www.sidih.si/kdo\\_smo](http://www.sidih.si/kdo_smo), 7. 8. 2012

<sup>xvii</sup> <http://www.etnoinfolab.org/>, 7. 8. 2012

<sup>xviii</sup> <http://www.dedi.si/>, 7. 8. 2012

<sup>xix</sup> <http://www.arzenal.si/>, 7. 8. 2012

---

<sup>1</sup> Metapodatki so bili do nedavnega pri nas pozabljena stvar. Najprej se je poskušalo postaviti bazo in o metapodatkih se je začelo razmišljati naknadno, dokler so stvari še obvladljive. (znanstvenik s področja zgodovine, NO10)

<sup>2</sup> Kot hiša nimamo nekega enotnega predpisa o tem, na kakšen način bi moral vsak raziskovalec hraniti in dokumentirati svoje izvirne podatke. To je bolj stvar neke metodike, metodologije dela, kot se je je raziskovalec navadil v času sodelovanja s svojim mentorjem in priprave doktorata. (raziskovalec s področja naravoslovnih ved, NO3)

<sup>3</sup> Rezultati v obliki nekih tabel, grafov, ali pa fotokopij zapiskov, ki jih raziskovalci beležijo v svojih beležnicah, imajo zelo majhno vrednost, ali praktično nobene, če niso ustrezno komentirani. (raziskovalec s področja naravoslovnih ved, NO3)

<sup>4</sup> Tisti, ki delamo že več let na tem, imamo neko formo, ki jo vsi poznamo in jo vsi uporabljamo. Problem je, če pride kdo nov, ker tega nismo napisali, da bi mu lahko točno rekli, to je ta podatek in točno tako smo to zmerili. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>5</sup> Ideje in želje po standardih na vsak način so, ampak zmerom se je izkazalo, da smo med sabo preveč različni. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>6</sup> Danes z xml tehnologijo standardni niti niso več tako pomembni. Če je zapisano v xml-ju, ki je nekak meta standard, je zapis transparenten. Strukturiran, ampak transparenten. Sicer ne standarden, ampak vsak študent zna razmeroma hitro to razbit in uporabiti. (znanstvenik s področja gradbene informatike, NO20)

<sup>7</sup> Na industrijskem področju bi bilo koristno, če bi država bolj spodbujala uveljavljanje standardov za predstavitev gradbenih objektov. (znanstvenik s področja gradbene informatike, NO20)

<sup>8</sup> Sprejeta so neka splošna merila v znanstveni srenji z našega področja. Ve se, kaj pomeni, da je čist protein, ali pa, ve se, kdaj je eksperiment dobro izveden. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>9</sup> Tak laboratorij ima akreditacijo nekajkrat na leto in če ustreza standardom za pridobivanje podatkov iz tovrstnih eksperimentov, potem so tisti podatki pač na določenem kakovostnem nivoju. (znanstvenik s področja biologije, NO6)

<sup>10</sup> Ja, vse je standardizirano. V operativni meteorologiji so stvari zelo dobro urejene, ker je bilo za potrebe spremljanja in napovedovanja vremena že pred časi potrebno urediti metode shranjevanja podatkov in načine njihove izmenjave. Vreme je pomembno. (Nedjeljka Žagar, znanstvenica s področja meteorologije)

<sup>11</sup> Treba je vzpostaviti resno infrastrukturo, da se lahko shrani podatke, in se odločiti, v kakšni obliki se jih bo shranilo, da bo to mogoče, saj so v surovi obliki preveliki. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>12</sup> Glejte, pri nas ni težav. Samoumevno je, da da gradivo, ki so ga raziskovali inštitutski raziskovalci, pride v arhiv in da se tu tudi obdela. Seveda, eni so malo bolj vestni, drugi malo manj. /.../ Gre morda res za nekaj težav pri pripravi opisov in metapodatkov, ker je konkreten raziskovalec največji vir metapodatkov in zaradi pomanjkanja časa, drugih obveznosti, večkrat ta stvar zamuja. Ampak nekako smo že vpeljali pravilo, da morajo biti tiste osnovne informacije oddane ob prihodu. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>13</sup> V vsakem avdio-vizualnem arhivu vam bodo povedali, da je tukaj vloženo delo lahko zelo dolgotrajno, ocenjuje se vsaj ena proti tri, če je relativno dobro obdelano gradivo. Če pa želiš to še

---

malo urediti, pripraviti z metapodatki, je ena proti deset. Če imaš nekaj ur gradiva, je to nekaj sto tisoč ur dela, ki jih je treba plačati, treba je imeti specialno opremo, ki se kviri in tako naprej. Finančno in časovno je to lahko zalogaj. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>14</sup> Problem je, da stvari niso tako enostavne, da raziskovalci ne bi imeli občutka, da je tukaj neko dodatno delo, ker niso prepričani, če se splača, če bo zato kaj boljše. V hiši pa trenutno nimam nikogar, trenutno zagotovo ne. Če bi hoteli to urejati, bi morali vpeljati neko službo. (znanstvenik s področja biologije, NO6)

<sup>15</sup> Ne, metapodatkovnih opisov ne pripravljamo, ker nimamo časa. Imamo toliko projektov in smo tako obremenjeni, da ponavadi drvimmo z enega v drugo. To je malo problem in to moramo izboljšati. Ampak na žalost ni kakšnega tehničnega osebja, tukaj bi rabili zelo veliko pomoči. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>16</sup> Podatke bi moral hraniti in vzdrževati tisti, ki jih producira. Ne morejo biti kar dosegljivi vsepovsod. Potem je samo vprašanje, kako definirati protokol zagotavljanja dostopa do teh podatkov. (znanstvenik s področja geografije, NO8)

<sup>17</sup> V obeh primerih gre za podatke, ki so v surovi obliki, razen za specialiste, povsem nezanimivi. Imajo pa naslednje značilnosti: merijo se v petabajtih. Da so zanimivi za analize, jih je treba večkrat in zelo zahtevno računsko obdelati. V večini eksperimentalnih, pa tudi teoretičnih fizikalnih projektih, se danes dela tako, da se primerja podatke, ki so zgenerirani z naključnimi metodami na osnovi teoretičnih modelov, to so simulacije - temu se reče Monte Carlo simulacije, z dejanskimi podatki iz detektorja. Potem se to uporablja na več načinov: lahko se metode analize izboljšuje, lahko se primerja podatke in ugotavlja odstopanja med tem, kar je bilo teoretično predvideno in med tem, kar je bilo praktično ugotovljeno ter se nato ugotavlja ali odstopanja kažejo, da je nekaj narobe z detektorjem ali je nekaj narobe s teorijo. Tukaj so problemi malo drugačni kot pri družboslovnih podatkih. Po eni strani je bolj prisotna zavest znanstvenikov, da je treba podatke skrbno hraniti zato, da so postopki ponovljivi, ker ko pride do objave nekega resnega odkritja, je treba dati podatke na voljo, da lahko neka druga skupina ponovi cel postopek. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>18</sup> Ampak škoda bi bilo čisto zavreči vrednost katerega koli gradiva, ki je zbrano. Sploh kadar imamo tudi avtorizacijo, potrditev, da smo pravilno razumeli neke stvari, ki so jih ljudje povedali. (znanstvenik s področja socialnega dela, NO7)

<sup>19</sup> Kaj konkretno je naredil, je spet odvisno od posameznika do posameznika in od tega, kako posameznik to dokumentira. Generirani, izvorni podatki so zmeraj shranjeni nekje posebej. Se ne izgubijo, vedno lahko gremo za nazaj pogledat, kaj je bilo narejeno. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>20</sup> *Prej smo se pogovarjali o vrednotenju oziroma o oceni kakovosti. Kakšne kriterije ali pa merila imate, za to, da rečete, to so pa kakovostni podatki, npr. v smislu da so točni ali kaj podobnega? Za to pa imamo neke ISO standarde.* (raziskovalec s področja geodezije, NO19)

<sup>21</sup> Po tehnični plati si največ pomagamo z AES, Audio Engeniring Society. Je tudi za radio in na splošno o snemalni tehniki in tako naprej. Za zvočne arhive in avdio-vizualne pa mislim, da je zelo močna organizacija IASA, International Association of Sound and Audiovisual Archives, ki ni samo evropska, ampak tudi ameriška in tudi druge organizacije so članice IASA-e. IASA je članica UNESCO. Ampak res smo specializirani, govorim v prvi osebi, ker smo tudi mi člani in sam sem bil v tehnični enoti zelo aktiven. Kadar se piše neka priporočila, se potem preveri recimo razne standarde,

---

med članicami so npr. British Library, BBC in tako naprej. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>22</sup> Metodološki aparat je bil izdelan na začetku. /.../ Odvisno je od tega, kam hodiš na seminar. Jaz sem hodila na seminar, kjer smo zelo lepo obdelali recimo tudi arhivsko gradivo, pogovarjali smo se o matičnih knjigah. (znanstvenik s področja zgodovine, NO10)

<sup>23</sup> Podatki morajo biti pripravljene v posebni obliki, ampak ponavadi je tako, da imajo podatkovne zbirke, ki hranijo te podatke, narejene aplikacije, s katerimi pregledajo, kar hočeš oddati, še dodatno malo po-analizirajo, če je vse v redu in ko greš skozi njihovo kontrolo kvalitete, je potem lahko deponirano v zbirki. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>24</sup> Zelo natančno je definiran format za izmenjavo aktualnih in klimatoloških meritev in za izmenjavo ostalih podatkov, kot so napovedi vremena, posebej v letalski meteorologiji zaradi varnosti zračnega prometa. V operativni meteorologiji so stvari zelo dobro urejene, ker je bilo za potrebe napovedovanja vremena potrebno urediti podatke in načine njihove izmenjave. /.../ V operativni meteorologiji prognostiki vedno spremljajo, kaj se dogaja v ozračju in svojo zadnjo napoved, če je potrebno, spremenijo ali popravijo. Posebej je to pomembno pri opisu stanja in napovedih za letalstvo, ki se popravljajo vsaj enkrat na uro. (Nedjeljka Žagar, znanstvenica s področja meteorologije)

<sup>25</sup> Obstaja nek standard, kako se ti objekti zapisujejo, ki se mu reče ISO-Step. In obstaja standard, ki se mu reče IFC, Industry Foundation Classes. (znanstvenik s področja gradbene informatike, NO20)

<sup>26</sup> Recimo pri takšnem avdiovizualnem gradivu, ko pridemo domov, naredimo, mi temu rečemo zapisnik, v tujini imajo izraz protokol, kjer so v glavi tega zapisnika osnovni podatki: kdo, kdaj, kje ter ti osnovni podatki, recimo informatorjev, čim več takšnih podatkov. Samo za ilustracijo, nas zanima tudi, če je nek pevec starejši, kje je služil vojsko, od kod se je preselil, kako se je žena priženila ... Ker so to zanimive relacije, potovanja, pesmi, repertoarji. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>27</sup> Gremo v to smer. Sedaj se veliko dogovarjamo o tem, oblikujemo tudi pravilnik o organizaciji raziskovalnega dela. /.../ Ampak to so zelo težki koraki. Je precej upiranja temu, da bi se stvari dale iz rok in na vpogled vsem. Začeli smo iz načina projektnega vodenja oziroma vzeli smo znanja projektnega menedžmenta, sedaj pa poskušamo ugotovljati, koliko bi to ustrezalo naši situaciji. Vsi raziskovalci imajo priložnost to videti, povedati, pokomentirati. (znanstvenik s področja socialnega dela, NO7)

<sup>28</sup> Ja, mi imamo en projekt, ki sicer malo stoji, ampak je pa nekoliko podoben temu, namreč Fototeka. Na inštitutu imamo fotografa, že 40 let istega, in smo se nekako lotili digitalizacije njegovega arhiva. Seveda, poskenira se hitro, ampak potem je pa treba 'keyword-e' dodati. Načeloma je pomembno, kateri človek je na fotografiji, datum, ki ga ni problem dobiti, recimo, za kakšen dogodek je šlo ... Potem pa je treba s temi podatki opremiti tudi fotografije. (vodja znanstvenoinformacijskega centra, NO5)

<sup>29</sup> Od verzije 4 se pri nas uporablja TEI. V bistvu se težave pojavljajo zato, ker je stvar kompleksna, ampak s samim formatom smo pa doslej popolnoma zadovoljni. Največ težav se je pojavilo zaradi tega, ker včasih človek ne ve čisto točno, za katero rešitev bi se odločil, ker je možnih več rešitev. Pa kdaj bi razbil kakšen projekt na več dokumentov, kdaj bi ga razbil samo na več datotek in bi ohranil

---

eno TEI strukturo. Te stvari so včasih kar komplicirane, ampak pri korpusih je stvar precej jasna. Korpus je zbirka dokumentov, posamezni dokument je zapisan v TEI formatu, metapodatki so lepo v 'headerju', sledijo pa potem stavki, standard za označevanje stavkov je jasen, morfosintaktične oznake so pa specifične za posamezen korpus in ločeno dokumentirane. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>30</sup> Mislim, da knjižničar nima možnosti, da bi šel tako daleč, da bi recimo neko tabelo številke toliko opremil z metapodatki, da bi nek uporabnik lahko iz njih črpal že fizikalne podatke. Mislim, da tako daleč knjižničar ne more iti. (vodja znanstvenoinformacijskega centra, NO5)

<sup>31</sup> Da, to imamo potem na svojih računalnikih. Vedno imamo tudi eno varnostno kopijo, ampak načeloma nas ima kar nekaj celotno bazo, tako da minimiziramo tveganje, če se komu kaj zgodi. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>32</sup> Ja, sedaj v hiši je večinoma tako, da so razen na področju dinamike, o katerem sva prej govorila, podatki relativno slabo arhivirani. Tam, kjer nimajo velikih količin podatkov, je to tako ad hoc, za čas raziskovanja, potem pa nekje stoji in bomo mogoče še kdaj našli. Ni nobenega posebnega sistema. Tam, kjer imajo več podatkov, to so mikroskopski ali pa zvočni, imajo sisteme kronološkega shranjevanja ali pa za mikroskopijo imajo sisteme za pregledovanje slik. Ampak to je neke sorte Picassa, malo bolj profesionalni programi, ampak v tem smislu. (znanstvenik s področja biologije, NO6)

<sup>33</sup> Spletna podatkovna zbirka arheoloških najdišč je nameščena na lokalnem strežniku, dopolnjujemo in nadgrajujemo jo s pomočjo spletnih aplikacij, za delovanje in shranjevanje podatkov ter varnostnih kopij pa skrbi računalniški center pri naši ustanovi. (raziskovalec s področja arheologije, NO14)

<sup>34</sup> Sproti, ko gremo na teren, nabiramo te vzorce. Ne nabiramo posameznih živali, ampak vse, kar je na dnu nekega potoka, v mrežo ujamemo vse živali, ki jih voda odnese. Temu rečemo vzorec. Potem se vzorec tukaj v laboratoriju prebere in to popišemo, tako da vemo, kaj imamo. Enkrat ko je čas in se tega dovolj nabere, pa taksonomsko, se pravi sistematsko, identificiramo. Od nabiranja pa do vpisa v tisti register lahko traja deset let. Tega niti približno ni mogoče narediti sproti. (znanstvenik s področja favnistike jamskih nevretenčarjev, NO22)

<sup>35</sup> Dobre inštitucije imajo zelo dobre arhive za hranjenje, tako da lahko na primer analiziramo podatke iz tkiv, ki so bili zbrani recimo 20 let nazaj. (znanstvenik s področja medicinskih vez in statistike, NO15)

<sup>36</sup> Saj zdaj prehaja vse v digitalno obliko, tudi starejša gradiva. Ampak kar ste prej omenili, ena težava je dostopnost, ker enostavno veliko gradiva še ni dostopnega v digitalni obliki. Mi smo sicer naredili en projekt hitre digitalizacije, tako imenovan 'quick and dirty'. Ne bom šel v podrobnosti, kako se presnemava, restavrira te posnetke, ampak vse tisto, kar je bilo mogoče relativno preprosto predvajati, smo preprosto predvajali in načeloma digitalizirali. Ocenjujem, da je 90 % uporabnikov, tudi raziskovalcev, znanstvenikov, zadovoljnih, da le pridejo do gradiva. Ne zanima pa jih, ali je posnetek tudi arhivsko in tehnično gledano kakovosten, kar je pomembno za dolgoročno hrambo. Preprosto zato, ker če bi naredili vse, kot bi zahtevala stroka, traja predolgo in ni dostopa. Tako da nekaj tega gradiva imamo in je danes lažje dostopno, s tem da se zavedamo, da ta kopija pač ni arhivska. Arhivske kopije še nimamo. Seveda pa v današnjem času prihajajo podatki v digitalni obliki, tako

---

zvočni kot avdio-vizualni, metapodatke pa ustvarjamo hkrati, ko pride gradivo s terena. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>37</sup> Z ARRS-jem ne pridemo skupaj, ker v bistvu ARRS spodbuja k digitalizaciji in seveda k spletni objavi, ne vidi pa problemov v ozadju, ne nazadnje tistega fizičnega dela, zbiranja soglasij. Imamo nek manjši nabor stvari, ki so avtorskoppravno še sporne in to urejamo, ampak ARRS bi moral tudi tukaj narediti korak naproti k temu. Ali pa JAK, da z nekim uredništvom sklene pogodbe na ta način, da se potem na primer avtor ali odreče ali kako drugače. Ne pa da nek urednik nekega portala ureja, tako z uredništvu kot s posameznimi avtorji in tako naprej. Dejansko bi morali urediti na višji stopnji. (znanstvenik s področja zgodovine, NO10)

<sup>38</sup> Inštitut je v 60. letih minulega stoletja vodil vseslovenski projekt zbiranja podatkov o arheoloških najdiščih Slovenije. Zbrani podatki so bili objavljeni v leksikonski obliki leta 1975. To publikacijo smo nato prepisali v elektronsko obliko in jo vključili v spletno podatkovno zbirko ARKAS, kar je potekalo v okviru inštitutske infrastrukturne dejavnosti. V 70. in 80. letih so sledile akcije arheološke topografije, ko so se podatki o najdiščih preverjali in dopolnjevali na terenu, beležili pa so se v standardizirani obliki topografskih zapisnikov. Topografski pregledi z nekaterih območij so bili tudi objavljeni. Celoten fond teh zapisnikov je arhiviran na inštitutu, iz njega pa so črpani tudi podatki za spletno zbirko ARKAS, ki je pričela v elektronski obliki nastajati v 90. letih. V inštitutskem arhivu hranimo tudi najrazličnejšo dokumentacijo terenskih raziskav, nekaj od tega smo uspeli digitalizirati (npr. velik del fonda načrtov najdišč), v zadnjem desetletju pa ta dokumentacija nastaja v digitalni obliki. Razen tega so med zaposlenimi na inštitutu specialisti za posamezna arheološka obdobja (prazgodovino, rimsko dobo, srednji vek) in za druge vire (antična zgodovina in epigrafika, arheobotanika, arheozoologija), ki so in še ustvarjajo svoje specialne podatkovne zbirke. Prav tako so prek spleta v celoti dostopne inštitutske publikacije iz obdobja zadnjih desetih let. (raziskovalec s področja arheologije, NO14)

<sup>39</sup> Kar se tiče eksotičnih formatov ali pa plačljive programske opreme, ki neha delovati takoj, ko se izteče licenca, to seveda niso perspektivna orodja, zato se odločam za tista, ki so uporabna na dolgi rok. Se pravi, da so zapisi v čim bolj standardnem formatu, po možnosti ASCII, in da so na takih mestih, kjer si obetamo, da bodo trajali. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>40</sup> *Vse skupaj v enem direktoriju?* Tako. V podmapah, da je mogoče slediti, da je, razvidno in ponavadi v raziskovalnih poročilih ne predložimo vedno celih zapisov, prav gotovo pa merski instrument. Pri nas je še ena težava, sploh takrat, ko gre za študij primera, je zelo težko zagotoviti anonimnost. Ker če si v nekem konkretnem okolju, se da že iz teh zgodb, ki so napletene, zelo hitro razbrati, za koga gre. Tako da ne moremo vsega dati v javnost. In tukaj se najbolj sprašujem, kako to zagotoviti. Če daš podatke, ki so širše dostopni, ne govorim zgolj znotraj hiše, ker tukaj smo tako ali tako vsi temu zavezani, ampak ko bi bili pa malo širše dostopni, je pa že vprašanje. (znanstvenik s področja socialnega dela, NO7)

<sup>41</sup> Formati so specifični za proizvajalca posamezne aparature, največkrat je tako. Nato pa lahko to izvoziš v kar koli drugega in uporabljaš še naprej. Ampak večinoma uporabljamo formate proizvajalcev in tudi njihovo programsko opremo za analize. /.../ Ne, ponavadi ne spremenijo samih meritev podatkov, ki jih dobiš med meritvijo. Spremenijo samo format tako, da lahko drugi programi to berejo. Tipično je recimo, da poskušaš pripraviti stvari za publikacijo, ali pa hočeš narediti neko dodatno analizo, ki je ne moreš narediti v programski opremi, ki pride zraven z aparaturo, in potem to



---

pretvoriš v take formate, da lahko delaš recimo v Excelu ali v Origin-u. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>42</sup> Obstajata dva dela software-a sta, en teče na strojih. /.../ To sem jaz naredil, da tukaj dela. Če ne, bi pač kaj drugega. Namreč stvar je iz dveh delov, ena je taka, da lahko tole vi tukaj gledate na vsakem stroju posebej, en del je pa na strežniku, ki servisira ali Internet Explorer ali Firefox ali Chrome. Odvisno s čim gledate, dobi od Chrome ali pa od nekega pregledovalnika neko zahtevo, naredi poizvedbo in vrne html stran, ki jo potem on prikaže. Ta del, ki teče na teh strojih, se imenuje Eva, naslov je [www.laze.org](http://www.laze.org), tam je o Evi vse napisano. Tisti del na strežniku pa je Neva, kot Eva, tisto je pa tudi na isti strani dokumentirano. In to sta dva software-a, ki to poganjata. (znanstvenik s področja jezikoslovja, NO13)

<sup>43</sup> Na univerzi se ne ukvarjamo s shranjevanjem operativnih podatkov. Meritve stanja ozračja, tal in vod so izjemnega pomena za napovedovanje in za klimatologijo in nam omogočajo spremljati, kako se vreme in klima spreminjata skozi zgodovino. Za shranjevanje in osnovno obdelavo meritev skrbijo operativne nacionalne službe. Pri nas je to ARSO, Agencija za okolje Republike Slovenije. Meteorologija je bila prva veda na svetu, ki je uredila formate in načine izmenjave podatkov med različnimi državami. Vse skupaj regulira svetovna meteorološka organizacija-WMO, ki ima sedež v Ženevi in v katero so vključene praktično vse države sveta. Ustanovljena je bila veliko pred ZN, po ustanovitvi ZN pa je svetovna meteorološka organizacija postala ena od njihovih specializiranih agencij. Pomembno je, da se vse v zvezi z opazovanji ureja na enak način po celem svetu, da bi vse meritve lahko uporabljali na enak način in da bi bile primerljive. (Nedjeljka Žagar, znanstvenica s področja meteorologije)

<sup>44</sup> Znotraj posameznih projektov so definirane metode, ki se pa tudi po večkrat spreminjajo. Ker imamo ves čas probleme s prostorom, z razvojem detektorja in programske opreme, se je podatkovni format napihnil, ljudje pa pridno kupujejo računske enote, torej računalnike ne pa toliko diskov, ker se potem na mesečnem poročilu vidi, koliko je računalnikov in tam ni očitno ali je star pet let ali dve, pri disku pa se vidi kapaciteta in to je zelo očitno, zato 'diskovja' kupujejo čim kasneje. Tako da imamo tukaj s prostorom resne probleme in podatke tudi brišemo. Podatke, iz katerih ni nastal članek in so bili nadomeščeni z novejšimi, boljšimi, tudi brišemo. Tehnično ni druge možnosti. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>45</sup> Seveda ne dobimo zmeraj projekta, s katerim bi lahko zagotavljali ažuriranje nekih podatkov, ki smo jih delali pred desetimi leti. (znanstvenik s področja geografije, NO8)

<sup>46</sup> Pri nas je bilo vložena ogromno denarja v razne podatke, digitalizacijo. Ampak če se to ne vzdržuje sproti, če ni denarja za vzdrževanje, potem nima sploh nobenega smisla. (znanstvenik s področja geografije, NO8)

<sup>47</sup> Glede shranjevanja oziroma arhiviranja digitalnega gradiva ravnamo po lastni presoji, vendar bi si želeli uradnih priporočil in navodil, ki bi urejala to področje. (raziskovalec s področja arheologije, NO14)

<sup>48</sup> Dejansko bo enkrat potrebno vzpostaviti močan sistem strežnikov, ki bodo po vseh standardih, na 100 kilometrski razdalji in ki bo državni, javni. Ker gre v to smer, da bo tega vedno več, enostavno pa si noben ne bo mogel privoščiti, da bi zakupil prostor pri komercialnem ponudniku. (znanstvenik s področja zgodovine, NO10)

## 7. FAZA: ISKANJE, DOSTOP (OBJAVA), DRUGA UPORABA

---

### 7.1 DOSTOP DO PODATKOV

Intervjuvance smo prosili, da nam opišejo, kateri podatki so dostopni. Zanimalo nas je komu, pod kakšnimi pogoji so dostopni ter kakšnih oblik, formatov in v kakšnih verzijah? Razumljivo je, da odvisno od tega, kako več je uporabnik, lahko prilagodimo ponudbo podatkovnih gradiv, pripravimo orodja in vmesnike, ki olajšajo iskanje in presojo relevantnosti podatkov ter njihovo uporabo. Katera orodja, metapodatki, standardi, postopki so uporabljeni za iskanje in dostop? Ali so podatki evidentirani in kako podrobno so opisani v katalogu? Ter po drugi strani kakšna je praksa glede omejevanja dostopa?

Zanimalo nas je, ali so v predhodnih fazah priprave podatkov izpolnjeni pogoji za zagotavljanje dostopa do podatkov? Ali imajo podatki vrednost za druge uporabnike: preverjanje objav, raziskovalni potencial za nova odkritja in objave, kombiniranje več podatkovnih virov med sabo? Ali se spremlja in beleži druga raba? Kakšen je postopek in merila? Ali ima uporabnik možnost oceniti vrednost in uporabnost podatkov? Ali obstaja interes za podatke drugih? Če da, za kakšen namen? Ali je zagotovljena sledljivost izvora podatkov in postopkov v predhodnih fazah? S čimer bi bila zagotovljena uporabnost podatkov za druge, dostop do metapodatkov, dokumentacije.

Vprašali smo jih tudi, kako vidijo odgovornosti v zvezi z reševanjem težav na tem področju. Kje bi bilo primerno mesto za dostop do podatkov? Kdo bi lahko opravljal storitve zagotavljanja dostopa (knjižnice)? Katere funkcije bi moral izpolnjevati? Zanimalo nas je, kakšen je pomen politik, vodil, standardov pri opredeljevanju odgovornosti. Ali imamo merila kakovostne podatkovne objave? Ali poznajo in podpirajo strategije in politike za zagotavljanje dostopa.

Kaj so ovire in sankcije za premagovanje: nezmožnost lociranja podatkov s strani uporabnikov, zaustavljeno financiranje projekta s strani naročnika? In kaj so spodbude ter nagrade: kaj pričakuje kot avtor? Citiranje, zabeleženo objavo. Kaj so pravni problemi dostopa: avtorske pravice, etična občutljivost podatkov. Kakšen je pogled na obveznost dostopa do podatkov s strani naročnika: javno financirani projekti, zahteve EU, revije ipd. Posebej nas je zanimala kultura izmenjave ter pogled glede omejevanja in pogojev dostopa. Ali ima interesent za podatke opredeljene pravice glede dostopa? Kakšen je odnos do načel OECD: kdo in kako zagotavlja izpolnjevanje na nacionalni ravni? Ali vidi sebe in druge v tej vlogi? Je to nekaj, kar velja za samoumevno ali neizvršljivo?

Dostop do podatkov je na različnih področjih raziskovanja neenako razvit, kar se kaže v različni ozaveščenosti glede pomena te faze življenjskega kroga, pa tudi v različnih rešitvah. Tako imamo na eni strani visoko profesionalizirane in standardizirane rešitve, povezane z mednarodnim sistemom. Na drugi strani so iskanje, dostop in ponovna raba podatkov prepuščeni iznajdljivosti posameznega raziskovalca ali skupine, rešitve pa so pogojene tudi z močjo in prednostmi, ki izvirajo iz njegovega položaja. Opaziti pa je mogoče tudi zametke dostopa, ki so jih razvili znotraj posamičnih ustanov in temeljijo na delu in znanju posameznikov, ki so se zgledovali po tujini ali iznašli svoje rešitve. Na primer v nadaljevanju govorimo o izkušnji Glasbenonarodopisnega inštituta ZRC SAZU, ki znanje črpa iz mednarodnih izkušenj, vendar je pri svojem delu samostojen in inventiven pri razvoju postopkov in storitev.

Na področju naravoslovja, kjer je mednarodno sodelovanje pogosta praksa, so uveljavljena rešitev osrednji mednarodni podatkovni arhivi, specializirani za določeno področje ali mednarodne podatkovne banke, kot je npr. Worldwide Protein Databank.<sup>i,1</sup> Gre za repozitorije s podrobno določenimi protokoli za delo in rabo, iz tega pa izhajajo utečeni postopki in uveljavljeni standardi hrambe in dostopa.<sup>2,3,4</sup> Mednarodne banke so najboljša rešitev za določene vrste podatkov, vendar je taka rešitev znana le ozkemu krogu poznavalcev oziroma specialistov, tako kadar gre za ustvarjanje ter s tem povezano možnost deponiranja podatkov, kot kadar gre za njihovo ponovno uporabo in možnost pridobivanja.<sup>ii</sup> Tudi na področju naravoslovja so znane in bolj pogoste prakse posameznikov in skupin, da s podatki, ki so jih sami ustvarili, ravnajo po lastni presoji in se vedejo kot neke vrste lastniki podatkov. Torej se sami odločajo, s kom in kdaj jih bodo delili ter v kakšni obliki in kakšnem obsegu.

Na področju slovenistike že obstajajo prakse odprtega dostopa. Nastale so na pobudo posameznih ustvarjalcev zbirk, ki so v delo vložili velik osebni angažma.<sup>5,6</sup> Gre za zanesenjake, ki delajo tudi brez finančne podpore. Iščejo lastne rešitve, tudi po sistemu t. i. robinhudovstva, kadar ni jasne politike ali obstoječih rešitev, se odločajo v smeri večjega dostopa.<sup>7</sup> Težijo k večji promociji svojih podatkov, da ti niso rezervirani za ekspertno skupino, ampak so v širši rabi, tako med študenti kot tudi v laični javnosti.<sup>8</sup> Določena besedila, gre večinoma za žanrska literarna besedila starejšega datuma, ki se množično ustvarjajo v digitalni obliki (skeniranje, prepisovanje), so tudi namenjena širšemu javnemu dostopu – pri tem pa je kompromis med količino podatkov in kakovostjo (odsotnost napak) bolj na strani prvega, lahko so tudi slabše narejena. Druga besedila so namenjena pretežno za znanstveno rabo in predstavljajo reprezentativne 'kritične' izdaje, so opremljena in pregledana po t. i. visokih znanstvenih merilih s težnjo po kar se da doslednem odpravljanju napak, tam je stopnja natančnosti neprimerno večja, a je veliko manj besedil.<sup>9,10</sup>

Na področju humanistike je mogoče opaziti zametke posamičnih podatkovnih središč, na primer: Arkas - arheologija, Zgodovina Slovenije - SIStory - zgodovina, DEDI – naravna in kulturna dediščina, Arzenal – virtualna zakladnica nacionalne dediščine... Ta področja se soočajo s težavami vzpostavljanja ali vzdrževanja sistema storitev generiranja in izkoriščanja gradiv v digitalni obliki. Identificirali smo velik interes tako vodstev kot posameznih centrov, ki opravljajo storitve v zamestkih, da se delo nadaljuje in da se nadgradi v smeri nacionalnih in mednarodnih povezav v okviru projektov DARIAH, SI-DIH ipd., z ambicijami intenzivneje vstopiti tudi na področje nove paradigme 'digitalne humanistike'.<sup>iii</sup> Da gre za živahno in ambiciozno zastavljene načrte in ideje, kaže tudi predlagani projekt za Center odličnosti – Humanistično središče za novo družbo iz leta 2009, ki je bil interdisciplinarno zasnovan, toda žal ni bil odobren. Na področju socialnega dela je faza iskanja, dostopa in ponovne uporabe podatkov manj razvita.<sup>11</sup> Omejena je s praktičnimi razsežnostmi kvalitativnih raziskav in podatkov, ki jih je težko standardizirati. Poleg tega veliko metapodatkov sploh ne beležijo, so izgubljeni ali založeni po predalih. Omejujoče je tudi prepričanje raziskovalcev, da podatkov ni mogoče rabiti, če nisi sodeloval pri ustvarjanju.<sup>12</sup>

<sup>i</sup> Hierarhija naraščanja vrednosti in stalnosti podatkovnih arhivov je videti tako, da so čisto na vrhu glavni mednarodni viri (npr. World Wide Protein Banka), drugo raven, ki pomeni tudi nižjo stopnjo, predstavljajo nacionalna podatkovna središča (npr. UK Data Archive), še nižje, na tretji ravni se nahajajo inštitutski repozitoriji (npr. ePrints Soton), najnižje pa so umeščene individualne zbirke. Vrednost zbirk narašča od dna proti vrhu piramide, od individualnega pomena k skupnostnemu. Od spodaj navzgor narašča tudi odgovornost in zahteve, povezane z dostopom do podatkov. V isti smeri narašča tudi obseg infrastrukture, s tem pa tudi pomen standardov, stalnosti in vzdržljivosti sistema. (The Royal Society Science Policy Centre, 2012)

<sup>ii</sup> Na tem področju je zgled dobre prakse spletna stran Databib, <http://databib.org/editors.php>, ki je namenjena iskanju podatkovnih središč. Gre za storitev identificiranja in lociranja repozitorijev raziskovalnih podatkov.

<sup>iii</sup> Glej: [http://www.sidih.si/digitalna\\_humanistika](http://www.sidih.si/digitalna_humanistika), 7. 8. 2012

### Okvirček 7.1

#### **Humanistika: spletna stran SI-DIH**

Na področju humanistike predstavlja pomembno novost portal SI-DIH, ki je skupni projekt Inštituta za novejšo zgodovino in Znanstveno raziskovalnega centra Slovenske akademije znanosti in umetnosti. Spletna stran je še v izgradnji, osnovo pa v tej fazi predstavljajo vsebine portalov Zgodovina Slovenije – Sistory (INZ) in Arzenal (ZRC SAZU). V prihodnje bodo mrežo partnerjev širili, saj so si zadali postati vozlišče slovenske znanstvene digitalne humanistike.

SI-DIH primarno vključuje digitalna gradiva, hkrati pa ponuja tudi gradiva starejšega nastanka, ki so bila digitalizirana z namenom, da bodo javno objavljena in tako dostopna širšemu krogu uporabnikov. Zaenkrat je raziskovalcem na voljo spletni iskalnik po digitalnih vsebinah, v prihodnosti pa ga bodo skušali čim bolj prilagoditi potrebam humanistike. Poleg tega je njihov namen spodbuditi raziskovalce k uporabi novih tehnologij ter izkoriščanju možnosti, ki jih te ponujajo. In sicer s ciljem, da bo *"mogoče zbrano znanje ponovno interpretirati ali inovativno pristopiti k novim raziskavam."* Kar zadeva tehnologijo zajema so v predstavitvi strani zapisali tako: *"s pomočjo protokola Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH), v iskanju najmanjšega možnega skupnega imenovalca pa zajem poteka preko metapodatkov, ustvarjenih po standardu Dublin Core Metadata Initiative. Na podlagi teh osnovnih standardov portal omogoča zajem skoraj vseh digitalnih vsebin, saj je kreiranje metapodatkov in njihovo zajemanje avtomatizirano in nezahtevno. Na podlagi nenehnega izboljševanja standardov opremljenosti digitalnih vsebin in povečevanja njihovega števila bo portal omogočal nadgradnjo iskalnikov in uporabo novih orodij."*

Portal SI-DIH je partner nastajajoče evropske infrastrukturne mreže za digitalno humanistiko DARIAH in aktivno sodeluje pri njeni izgradnji.

Vir: <http://www.sidih.si/>

Ker ni enotnega sistema, po katerem bi bili določeni postopki, pravice in odgovornosti v fazi iskanja, dostopa in ponovne uporabe, so na določenih področjih podatki slabše dostopni in ostajajo neizkoriščen potencial, ki je prepuščen znanju, zmogljivostim in sposobnostim posameznih ustvarjalcev podatkov. Ker dosedanji sistem menjave temelji na medsebojnem zaupanju, poznanstvih in osebni presoji raziskovalcev, so pogoji menjave stvar individualnih dogovorov, posledično pa so podatki pridržani za ožji segment uporabnikov. To je značilna ugotovitev, ki se pojavlja tudi v drugih podobnih študijah: odsotnost institucionalizacije storitev izmenjave podatkov z opredeljenimi protokoli, politikami, opredeljenimi pravicami in dolžnostmi pusti prostor za podatkovne izmenjave na ozki interesni podlagi neposrednih koristi vpletenih posameznikov.

### **7.2 ZADRŽKI IN OVIRE POVEZANE Z DRUGIMI FAZAMI ŽIVLJENJSKEGA KROGA PODATKOV**

Ustvarjalci in dajalci podatkov se v fazi življenjskega kroga, ki obsega iskanje, dostop in ponovno uporabo raziskovalnih podatkov, srečujejo z različnimi zadržki in ovirami, ki jim preprečujejo, da bi podatke dali v ponovno rabo ali jim onemogočajo, da bi prišli do podatkov, ki so jih ustvarili drugi. V tem podpoglavju smo podrobneje opredelili zaznane težave in predstavili, kako jih vidijo uporabniki in ustvarjalci. Ustvarjalci in uporabniki z različnih področij so kot zadržke in ovire navajali sledeče: konkurenčnost, tekmovalnost, slab zgled državnih ustanov, pomen velikega osebnega angažmaja ustvarjalcev, vprašanje avtorstva, vprašanje kakovosti.

### ***Tekmovalnost in konkurenčnost***

Poleg tehničnih, finančnih, pravnih in strokovnih zadržkov prosto rabo podatkov oziroma odprt dostop omejuje tudi kompetitivna kultura, ki je posledica poudarjeno točkovnega sistema pridobivanja sredstev za raziskovanje. Številni ustvarjalci in uporabniki druge vidijo kot tekmece, zato sodelovanje med njimi ni običajna praksa. Med razlogi, zakaj podatkov, ki so jih sami ustvarili, ne želijo dati, izpostavljajo vrednost podatkov oziroma neizkoriščen potencial. Za nekatere niti denarna nagrada ne bi bila dovolj, da bi podatke delili, saj menijo, da so pomembnejše objave, ki jih lahko ustvarjalec generira na podlagi monopola nad podatki.<sup>13</sup> Do deljenja podatkov so raziskovalci veliko bolj odprti v tisti fazi življenjskega kroga, ko so prepričani, da so že vse obdelali in objavili, kar so imeli v načrtu. Iz tega sledi zahteva po reformi obstoječih sistemov točkovanj, ovrednotiti bo potrebno tudi objave v netradicionalnih oblikah, zlasti še odprto dostopne objave in podatkovne vire, ki naj bodo obravnavani enakovredno kot ostale znanstvene objave.

Podatki so osnova za objavo in kot taki nosijo potencial za prihodnje delo (nova sredstva, objave), so konkurenčna prednost. V tem smislu imajo za ustvarjalce pomembno vrednost.<sup>14</sup> Za opisovanje zadržkov ustvarjalci pogosto uporabljajo pojme, kot so trg, konkurenca, tekmece, tekma, kar je posledica prenosa tržnih modelov delovanja v znanstveno in strokovno sfero.<sup>15,16,17</sup>

Tekmovalnost se kaže v izraziti individualizaciji pri prijavi in izvajanju projektov. Intervjuvanci se zavedajo, da tržni modeli (pritisk denarja in časa) negativno vplivajo na znanstveno dejavnost in pričakujejo, da bodo bodoče politike prinesle rešitve, ki bodo izpostavile vrednote, lastne znanosti in motivirale raziskovalce k večjemu sodelovanju. Podatki so eden od elementov, ki vabijo k medpodročnemu izkoriščanju, pri tem pa je pomembno pravično ovrednotenje prispevkov in naporov pri ustvarjanju in pripravi podatkov.

### *Pomen velikega osebnega angažmaja raziskovalcev*

Intervjuvanci so odnos do deljenja podatkov pogosto opisovali emotivno in s subjektivnega vidika. Pogosto so razlagali, da ob pozivu k deljenju podatkov občutijo ljubosumje.<sup>18</sup> Ljubosumje so opisovali kot logičen odziv in posledico tega, da ustvarjanje podatkov ni ustrezno ovrednoteno. Ker so v proces vložili veliko osebnega truda, se ustvarjalci vedejo, kot bi bili lastniki podatkov. Če za podatke niso bili plačani, ne vidijo razloga za deljenje, saj so vanj vložili lastna sredstva (znanje, trud in čas). Kadar so vložili 'gromozanski osebni angažma', so prepričani, da jim tega nihče ne more poplačati.<sup>19</sup>

Intervjuvanec, ki je ustvarjalec in uporabnik podatkov, je opozoril na prisotnost občutka lastništva nad zbirkami, iz katerega izhaja ljubosumje in posledično omejevanje odprtega dostopa.<sup>20</sup> Pomembno vlogo igra preračunljivost v smislu lastnega vložka in povračila zanj.<sup>21</sup> Če so za ustvarjanje podatkov prejeli plačilo, se bolj zavedajo odgovornosti in dolžnosti deljenja. Zavedajo se, da so podatki končni produkt dela, za katerega so bili plačani.<sup>22</sup>

#### *Okvir 7.1: P-R 8*

##### *PROBLEM:*

Med intervjuvanci prevladuje mnenje, da so podatki sredstvo za cilj, to je morebitna objava. Podatki so dragoceno gradivo, ker nosijo potencial za publikacijo. So tudi valuta za menjavo in sodelovanje pri pisanju člankov, monografij.

Zaradi takega pojmovanja podatkov je razširjena težnja po monopolizaciji podatkov. Prevladuje ljubosumje, občutek, da je podatek lastnina in da ima raziskovalec, ki je zbiral podatke, monopol nad možnimi odkritji, ki bi sledili iz podatkov.

Kultura odprte izmenjave podatkov še ni vzpostavljena. Ni sistema, ki bi natančno določal vloge, odnose, odgovornosti, dolžnosti in pravice ustvarjalcev in uporabnikov. Zato je akterjem prepuščeno, da se v raziskovalnem postopku ravnajo po lastnih občutenjih in osebnih presojah.

##### *REŠITEV:*

Potrebno je poskrbeti za bolj jasno določene vloge in odgovornosti nosilcev. Za večjo pretočnost podatkov in večje medsebojno sodelovanje akterjev v življenjskem krogu podatkov bo potrebno upoštevati problem lastništva podatkov. Prepričanja, da so ustvarjalci in skrbniki podatkov zaradi vložene dela edini, ki morejo odločati o ravnanju z njimi, bo mogoče premostiti z ustreznim vrednotenjem podatkov.

Najbolj očiten ukrep, ki se ponuja v tem segmentu, je reforma sistema vrednotenja znanstvenih prispevkov in financiranja projektov, ki bo izpostavila zahtevo po dostopu do podatkov, ali povezanih s publikacijo ali pa kot samostojni produkt, ter s tem ustrezne zahteve po navajanju podatkovnih virov v primerih druge rabe. To naj gre v smeri, da se prizna avtorske pravice in se uredi ekvivalent nagrad za objavo podatkov po načelih, ki veljajo za objavo člankov.

### Državne ustanove so slab zgled

Več nezadovoljstva v zvezi s fazo iskanja, dostopa in uporabe podatkov so izkazali intervjuvanci s področja ekonomskih ved. Opozorili so na dve slabosti, ki otežujeta delovni proces. Ena je komercializacija podatkov v državnih ustanovah, druga so dolgotrajne procedure, ki ovirajo dostop. Tak sistem so navajali kot negativen zgled in kot argument za restriktivno vedenje, kadar so sami postavljeni v vlogo dajalca.<sup>23,24</sup>

K dodatnemu premisleku glede bodočih politik vodijo tudi opažanja o dosedanjih rešitvah s področja hrambe in dostopa do podatkov. Ustvarjalci, skrbniki in uporabniki so pri svojem delu pogosto omejeni z obstoječimi modeli v širšem okolju, na katere kot uporabniki nimajo neposrednega vpliva. Intervjuvanci so kot primer slabe prakse izpostavili tako državne (npr. SURS) kot tudi mednarodne ustanove (npr. OECD, Svetovna banka), ki zaračunavajo dostop do podatkov, čeprav so bili ti podatki že plačani iz javnih sredstev.<sup>25, 26, 27, 28, 29, 30</sup>

Takšne rešitve so slab zgled in krepijo zadržke raziskovalcev, kadar se sami odločajo o prostem dostopu do lastnih gradiv. Tisti, ki morajo plačati dostop do podatkov oziroma morajo podatke kupiti, so se negativno odzvali na pobudo, da bi z drugimi delili svoje podatke in so v vlogi dajalca podatkov zavzeli restriktivno stališče. To še ne pomeni, da podatkov ne bodo dali, vendar imajo negativno stališče do krovne ureditve, ki bi odprtost terjala le od raziskovalcev. Izpostavljen je bil očitek, da so državne ustanove slab zgled in da s svojimi praksami ne delujejo motivacijsko za raziskovalce.<sup>31</sup>

Po drugi strani pa so navajali tudi posamične primere dobrih praks iz tujine ter razlike med ameriškim in evropskim sistemom.<sup>32</sup> V ZDA imajo davkoplačevalci prost dostop do podatkov, ki so bili plačani z javnimi sredstvi in za dostop ni potrebno še enkrat plačati, saj že kot davkoplačevalci plačujejo infrastrukturno mrežo in hrambo. Intervjuvanci so opozorili tudi na 'luknjo' v sistemu: kjer bi morali pri slovenskih inštitucijah za dostop do podatkov opraviti določene postopke, lahko preko ameriških strani do teh istih podatkov prosto dostopajo. V tem smislu je slovenski sistem veliko bolj restriktiven in za uporabnike omejujoč.<sup>33</sup>

### Okvir 7.2: P-R 9

#### PROBLEM:

Intervjuvanci so na podlagi slabih domačih praks in na podlagi dobrih primerov iz tujine izrazili eksplicitna pričakovanja, da se uredijo splošna načela glede dostopa do podatkov, ki jih producira javni sektor.

#### REŠITEV:

Na več mestih smo izpostavili, da bi lahko del storitev in nalog evidentiranja, iskanja, in dostopa do podatkov javnega sektorja prevzeli področni akademski podatkovni arhivi, tako kot je primer UK DA, ki vodi storitve 'Varnega dostopa do zaupnih mikropodatkov'<sup>iv</sup>, ali pa 'podatkovno pribežališče' v ZDA, ki ga vzdržuje NORC.<sup>v</sup>

Na splošno lahko k dvigu pozitivne klime deljenja podatkov prispeva aktivnejše vključevanje države v iniciative 'odprtih državnih/javnih podatkov' (primerjaj Open government data).

### Vprašanje avtorstva

<sup>iv</sup> <http://securedata.data-archive.ac.uk/>, 7. 8. 2012

<sup>v</sup> <http://www.dataenclave.org/index.php/home/set/visitor>, 7. 8. 2012

Precej zadržkov in različnih stališč se je pojavilo v zvezi z vprašanjem avtorstva. En vidik zadeva avtorstvo v smislu priznanja zasluge za ustvarjalni napor, drugi kot oviro pred prostim deljenjem in dostopom do podatkov in del, kadar avtorske pravice to omejujejo.

Poleg dilem v zvezi z avtorskimi pravicami se pojavlja tudi vprašanje, kako sploh opredeliti avtorstvo, kajti iz tega izvirajo pravice in obveznosti v zvezi s hrambo, dostopom in drugo rabo. To vprašanje pride do izraza še zlasti pri uporabnikih, ki dodajo podatkom novo vrednost (npr. izdelava zemljevidov). Tak primer so tudi slovarji in druge zbirke. Kdo je po očiščenju, transformaciji lastnik podatkov in kaj lahko da v prosto rabo?<sup>34</sup> Ker se to področje pogosto navaja kot vir negotovosti ali pa tudi kot izgovor za to, da se omejuje dostop do določenih podatkov, bi bilo potrebno ustvariti pregledna navodila glede problematike avtorskih pravic in modelov licenc. Obstaja kar nekaj mednarodnih svetovalnih strani in projektov, ki pripravljajo orodja in podporo za zagotavljanje ustreznih citatov in referenc na vire, skupaj z njihovim ohranjanjem avtentičnosti glede na izvor (stalni identifikatorji in modeli citiranja, ki jih vsebujejo), ki tudi omogočajo sledljivost. Za podatke je priporočen poseben seznam variant OA licenc, ki ne omejujejo njihove rabe, bolj kot je nujno.<sup>vi</sup>

Vprašanje ravnanja z vsebinami, ki imajo različno avtorsko zaščito (filmi, umetniška dela, zgodovinski viri, glasba...), na posameznih področjih predstavlja resno oviro za strokovno delo, recimo na področju humanistike.<sup>35,36</sup> Podatke predstavljajo dokumenti ali dela, za katera ni jasno, ali še zapadejo pod omejitve avtorskih pravic, njihovi 'skrbniki' pa ne vedo, kaj sploh lahko delijo z drugimi in pod kakšnimi pogoji. Obstaja močan interes in pripravljenost za reševanje problemov, s tem da je potrebno preseči ozko razumevanje avtorstva, ki znanost zapira. Avtorske pravice so za to, da omogočajo javni dostop do 'kulturne dediščine' in ne da ga omejujejo. Intervjuvanec opozarja na prisotnost t. i. avtorskega napuha, ki ga vidi kot oviro na poti k odprtim podatkom. Povezuje jo z logiko ekonomskega profita, za katero meni, da se bo morala spremeniti, saj ne sodi v znanstveno etiko.<sup>37, 38</sup>

Vprašanje avtorstva je problematično tudi pri ustvarjanju mednarodnih zbirk, saj so različne države to vprašanje različno rešile. Eden od sogovornikov predlaga, da bi na mednarodni ravni ponudili zgolj metapodatke, dostop do podatkov pa bi vsaka država uredila skladno z lastnimi politikami.<sup>39</sup>

---

<sup>vi</sup> Avtorske pravice in licence: <http://www.openaccess.si/avtorske-pravice-in-licence/>, ter Viri in orodja za delo s podatki: <http://www.openaccess.si/viri-in-orodja-za-delo-s-podatki/>, 7. 8. 2012



**Okvir 7.3: P-R 10**

**PROBLEM:**

V zvezi z vprašanjem avtorstva je jasno izraženo eksplicitno povpraševanje po znanju, rešitvah, nasvetih, protokolih in dobrih praksah. Problem je najbolj jasno artikuliran na področjih, kjer se pojavljajo zametki storitev hrambe in izmenjav gradiv.

**REŠITEV:**

Eden od nujnih in logičnih korakov je koordinirano reševanje problema oziroma celega področja, ki obsega širše vidike avtorskih pravic v digitalnem okolju. Vprašanje avtorstva kot tako izkazuje resno potrebo po politikah, s katerimi bodo določena pravila in obveznosti ustvarjalcev podatkov, posledično pa bodo razrešene tudi dileme skrbnikov zbirk in zadržki v zvezi z drugo rabo.

Ena od teženj je v vse širšem pojavljanju t. i. podatkovnih revij, ki v sodelovanju z uveljavljenimi repozitoriji ali podatkovnimi arhivi, ki podatkom dodeljujejo stalne identifikatorje in zagotavljajo dolgoročno skrbništvo, pripravljajo recenzijsko pregledane članke o ustvarjenih podatkih.

**Strah pred izgubo nadzora**

Več intervjuvancev je izpostavilo strah, da bodo kot dajalci podatkov izgubili nadzor nad gradivom. Prepričani so tudi, da bo širši dostop povečal tveganje za zlorabe. Na določenih področjih je previdnost posledica strokovne podlage (npr. arheologija), na drugih je posledica odsotnosti praks in nepoznavanja preverjenih sistemov.<sup>40, 41</sup> Slednje bo potrebno seznaniti s pozitivnimi izkušnjami. Na primer, da se javnosti ponudijo le prečiščeni podatki, selekcionirani kot 'smetana' podatkovne produkcije.

Posamezna področja so regulirana tako, da so podatki dostopni le registriranim raziskovalcem, ki izrazijo svoj namen. Ena od možnosti je vloga oziroma prošnja, ki jo preveri in oceni za to določena oseba ali komisija.<sup>42</sup> Na občutljivih področjih, kot je denimo medicina, imajo etično komisijo, ki odloči, ali bodo vlagatelju odobrili dostop in pod kakšnimi pogoji.<sup>43, 44</sup> Dodatna rešitev je lahko tudi izjava oziroma pogodba, ki vlagatelja obveže, da z gradivom ravna pod natančno določenimi pogoji in je za kršitve lahko tudi kazensko odgovoren.

#### Okvir 7.4: P-R 11

**PROBLEM:**

Ustvarjalci in uporabniki so menjavam bolj naklonjeni in imajo manj zadržkov in strahov na tistih področjih, na katerih je ponovna raba podatkov ustaljena praksa. Drugod so skeptični in pri ravnanju zadržani oziroma odprtemu dostopu niso naklonjeni.

**REŠITEV:**

Pozitivne izkušnje lahko služijo kot primer dobre prakse na tistih področjih, na katerih imajo akterji manj izkušenj z drugo rabo podatkov, in lahko pomagajo tako na tehnološki ravni kot tudi na ravni ureditve razmerij med akterji (ustvarjalci, skrbniki, uporabniki).

Pri osveščanju naj večjo vlogo odigrajo osrednji nacionalni podatkovni centri. Vloga osrednjih nacionalnih centrov za ustvarjanje ali za dostop do določene vrste podatkov je lahko tudi svetovalna, lahko posredujejo in organizirajo delavnice za večšine uporabe določenih vrst podatkov, delijo dobre primere analiz in posredujejo orodja<sup>vii</sup> za analizo. Predvsem pa imajo urejene protokole in preverjene obrazce za zagotavljanje nadzora nad izročanjem in dostopom do podatkov.

#### Okvir 7.5: P-R 12

**PROBLEM:**

Med splošnimi zaključki lahko izpostavimo presenetljivo podobnost zaznave problemov in odgovornosti za kakovost in zagotavljanje dostopnosti v različnih strokah. Morda se nakazuje, da je v humanistiki ter bioloških vedah in medicini (life sciences) manj kompetitivnosti kot v naravoslovju in družboslovju in več poudarka na vrednosti priprave 'referenčnih' zbirk podatkov, ki koristijo celotni skupnosti. Pri prvih so problemi specifični, izpostavljam o odsotnost razlikovanja med referenčnimi zbirkami in popisi gradiv ter surovimi podatki v digitalni obliki posameznih najdišč, vrst, eksperimentov ipd.

**REŠITEV:**

Več pozornosti bi bilo treba usmeriti k nalogam, s katerimi bi zagotovili dostop do surovih gradiv, ki so lahko podlaga za nadaljnje raziskave, ter oblikovati ustrezne politike zajema gradiv in regulacije pravic dostopa do gradiv in storitev v okviru dostopa.

Potrebno bo nadaljnje delo pri identificiranju potencialov in nastavkov storitvenih dejavnosti na teh področjih, npr. na področju naravoslovja bi lahko del nalog prevzele znanstvene knjižnice ali podobni skupni servisi velikih inštitutov in univerz, npr. z oblikovanjem splošnega odprtodostopnega repozitorija objav in podatkov, pri oblikovanju zahtev in politik naj imajo večjo vlogo področne revije.

V družboslovju bi bilo potrebno preveriti sprejemljivost sodelovanja z že vzpostavljenimi storitvami Arhiva družboslovnih podatkov na sedaj manj zastopanih področjih, kot je ekonomija, psihologija ipd.

<sup>vii</sup> Glej Gesis, Leibniz Institute for Social Sciences: <http://www.gesis.org/en/services/data-analysis/official-microdata/microcensus/microdata-tools/>, 7. 8. 2012.

### **7.3 VREDNOTENJE PODATKOV**

Zadnji sklop vprašanj je namenjen vrednotenju podatkov. Ker smo vprašanje o vrednotenju vključili že v vseh predhodnih fazah, smo ga ob koncu zastavili le še po potrebi. Zanimalo nas je, ali se upošteva vrednotenje podatkov in ali je zagotovljeno v predhodnih fazah priprave podatkov, ali se pri vrednotenju upošteva pomen za druge.

Na področju razumevanja odgovornosti (pravic in dolžnosti) nas je zanimalo, kje bi bilo primerno mesto za vrednotenje podatkov: ob nastanku, shranjevanju, dostopu, objavi. Kakšen je pomen politik, vodil, standardov: ali obstajajo merila vrednotenja, kakšne so ovire, sankcije, finančne zahteve; in po drugi strani spodbude, nagrade. Ali se kakovost podatkov splača, ali je nagrajena?

Odsotnost skupnih politik in standardov iskanja, dostopa in druge rabe je vidna tudi na ravni povezovanja. Pomanjkljivi ali neobstoječi metapodatki namreč raziskovalce odvrta od tega, da bi izkoristili potencial obstoječega gradiva. V nekontroliranih ali vsaj neenotnih pogojih, kakor so značilni za številna področja ustvarjanja podatkov, potencialni uporabniki menijo, da je druga raba praktično nemogoča.<sup>45, 46</sup> Vendar so zadržki ali skepsa glede možnosti uporabe dostikrat pomanjkljivo utemeljeni in ne presegajo preizkusa ovrednotenja v okviru potrjenih načrtov DMP. Priročnik o zadržkih v zvezi z deljenjem podatkov, ki so pogosti med ustvarjalci podatkov, in odgovorih nanje, so pripravili v UK Data Archive (glej Tabela 7.1).

Med spodbudami za dostop do podatkov je tudi možnost izrabe v smislu povezovanja lastnih podatkov s podatki od drugih. Kakovost priprave podatkov se v tem pogledu vrednoti skozi uporabo ustreznih oznak vsebine in oblike podatkov, ki jo je zlasti v mednarodnem okolju včasih težko poenotiti.<sup>47</sup> Uporaba standardov in priporočil kljub vsemu lahko pomaga pri povezovanju.<sup>48</sup>

Povezovanje podatkov, primerjave in raba skupnih orodij (naprav, modelov...) prispeva k izkoriščanju potenciala, ki ga nosijo podatki. V več primerih se je to povezovanje izkazalo koristno ne le med tistimi, ki uporabljajo uveljavljene mednarodne zbirke,<sup>49</sup> pač pa tudi med domačimi raziskovalci. Tak primer so denimo humanistične vede, kjer se med seboj povezujejo arheologi, zgodovinarji, geografi. Izmenjava podatkov znižuje stroške, kar je še posebej pomembno, kadar so v igri drage metodologije (npr. fotografiranje iz zraka).<sup>50</sup>

K večji povezljivosti podatkov lahko prispevajo specializirana podatkovna središča, ki imajo razvit sistem hrambe in izkušnje z drugo rabo. Lahko so tudi pomoč v tem smislu, da izdelajo analize ali poročila.

**Tabela 7.1: Zakaj ne bi delili podatkov (zadržki in odgovori nanje)**

	<b>Zakaj ne bi delili podatkov?</b>	<b>Zakaj je vredno deliti podatke?</b>
1.	Moji podatki nikogar ne zanimajo.	Raziskovalci želijo imeti dostop do podatkov iz različnih študij, metodologij in področij. Zapiski nekega vrtnarja lahko postanejo pomemben vir za raziskovanje meteoroloških sprememb.
2.	Želim objaviti, preden podatke pokažem drugim.	Dostop do podatkov je mogoče urediti tako, da ima ustvarjalec podatkov prednost (npr. časovni embargo).
3.	Nimam časa ali denarja, da bi pripravil podatke za deljenje.	Upravljanje s podatki mora postati sestavni del raziskovalnega postopka. Načrtovati je potrebno pred začetkom raziskovanja, s čimer se prihrani čas in denar.
4.	Če bom respondente vprašal za dovoljenje za deljenje podatkov, ne bodo želeli sodelovati v raziskavi.	Respondentom je potrebno razložiti, kaj pomeni arhiviranje in deljenje podatkov. Običajno jih to ne skrbi tako, kot si mislijo raziskovalci. Če dovoljenja še nimate, ga je mogoče pridobiti tudi naknadno.
5.	Ukvarjam se z zelo občutljivo raziskavo. Podatki preprosto ne morejo biti dostopni drugim.	Najprej je treba vprašati respondente, če dovolijo deljenje. Druga pot je anonimizacija podatkov. Tretji način je embargo za določeno obdobje. Poleg tega je dostop do podatkov običajno omejen, npr. samo na raziskovalce.
6.	Imam kvantitativne podatke, povezovanje spremenljivk bo razkrilo identiteto udeležencev.	Kvantitativne podatke je mogoče anonimizirati s postopki anonimizacije, npr. zaokrožanje zgornjih kategorij distribucije, odstranitev spremenljivk, nadzorovan dostop do določenih spremenljivk (kot so poštna številke ali imena naselij).
7.	Imam avdiovizualne podatke in jih ne morem anonimizirati, zato teh podatkov ne morem deliti.	Vizualne podatke je mogoče anonimizirati z zameglitvijo obrazov in s popačenjem glasu. To je lahko zamudno in drago, zato je smiselno prej preveriti, če kljub vsemu dovolijo dostop, morda pod omejenimi pogoji.
8.	Obljubil sem, da bom po koncu raziskave podatke uničil.	Zakaj ste kaj takega obljubili? Vsekakor se je v prihodnje temu bolje izogniti. Kljub vsemu premislite, ali se obljuba oziroma zaveza nanaša na vse podatke.
9.	Podatke sem zbral z obljubo strogo zaupno.	V prihodnje se takim nepotrebim obljubam izognite. Z anonimizacijo je mogoče zaščititi istovetnost, popolne zaupnosti nikoli ne morete zagotoviti.
10.	Moja podatkovna zbirka in nastali prepis (transkript) sta v tujem jeziku.	To ne bi smel biti problem. Podatkovni arhiv sprejme tudi transkripte v tujem jeziku, čeprav so prevodi bolj zaželeni.
11.	Nemogoče je anonimizirati prepise, saj se bo izgubilo preveč pomembnih informacij.	Stopite v stik s podatkovnim arhivom in skušali vam bodo pomagati. Boljša rešitev je lahko omejen dostop.
12.	Moja podatkovna zbirka vključuje podatke, ki sem jih kupil, zato ne morejo biti javni.	Potrebno je ugotoviti, kdo je lastnik avtorskih pravic in pridobiti ustrezna dovoljenja. Raziskovalec mora biti seznanjen z licencami in pogoji rabe podatkov.
13.	Drugi raziskovalci ne bodo razumeli mojih podatkov – ali jih bodo uporabili za napačne namene.	Za razumevanje podatkov je treba ustvariti dobro dokumentacijo in zagotoviti informacije o raziskovalnem kontekstu.
14.	Podatki so zaščiteni kot intelektualna lastnina.	Potrebno je pridobiti dovoljenje od lastnika pravic. Najbolje v začetku raziskovalnega projekta, mogoče tudi za nazaj.

Vir: (Corti in drugi 2011, 11-12)

### *Katere podatke shranjevati*

Vprašanje prostega dostopa do podatkov odpira vrsto dodatnih vprašanj, ki segajo tudi na druga področja življenjskega cikla podatkov. Eno takih je, katere podatke ohraniti in katere zavreči. Ker politike hrambe še niso razvite, tovrstni premisleki zadevajo predvsem ustvarjalce podatkov in zbirke, posledično pa vplivajo tudi na možnost dostopa in druge rabe.

Ponekod velja prepričanje, da je pomembno ohraniti vse. Na posameznih področjih so raziskovalci prepričani, da je vsak podatek vreden hrambe, saj je bil vanj vložen nek trud.<sup>51,52,53</sup> V družboslovju in humanistiki, kjer se pojavi nenehno spreminjajo, vlada prepričanje, da je hramba na dolgi rok toliko bolj pomembna. A je problem, kako podatke ustrezno opremiti, pripraviti za drugo rabo in zagotoviti dostop. Gre tudi za problem t. i. 'big data', s katerim se soočamo v zadnjem obdobju vsesplošno razširjenega generiranja digitalnih vsebin, kjer je težko postaviti mejo, kaj še hraniti.<sup>viii</sup>

Večinoma se ustvarjalci podatkov in zbirke strinjajo, da ni smiselno hraniti vsega in se sprašujejo o tem, kje postaviti mejo. K takemu sklepu jih vodijo različna vprašanja in pomisleki, ti so v nadaljevanju na kratko predstavljeni.

Intervjuvanci so večkrat izpostavili predlog, da bi dostop do gradiv uredili glede na to, kdo je financer. Strinjajo se, da o dostopu do gradiv, ki so financirani z zasebnimi sredstvi, odloča plačnik, ki je tudi lastnik teh podatkov.<sup>54</sup> Seveda je tudi na teh področjih ob ustrezni ureditvi mogoče doseči dogovor o možnostih dostopa za raziskovalne namene, ki ga kot uravnotežen pristop omenjajo tudi načela OECD.<sup>ix</sup> Veliko več pomislekov so imeli s tem, kako opredeliti in določiti mejo dostopa do podatkov, plačanih iz javnih sredstev.<sup>55, 56, 57</sup>

Večinoma se torej strinjajo, da ne morejo biti dostopni vsi podatki. Nekateri menijo, da potem ne bi bilo več pregleda, kaj je zanesljiv podatek in kaj ni. Iz tega stališča izvira jasna potreba po storitvah, ki bi pomagale pri presoji, kaj je vredno hraniti in česa ne. Vprašanje je, ali je mogoče zagotoviti enoten, skupen servis, saj so področja raziskovanja in ustvarjanja podatkov raznolika, s tem pa tudi ovire in dileme za hrambo.

Za določena gradiva so zadostna rešitev splošni repozitoriji (npr. Dryad), za druga se kaže potreba po razvoju specializiranih, področnih podatkovnih središč. Zlasti na področjih, kjer se že kažejo zametki hrambe in izmenjav gradiv, je smotrno povezovanje in vzpostavljanje rešitev, ki bodo omogočile integracijo (npr. poenotenje registracije in pravic dostopa za uporabnike), kar bo dodatno spodbudilo izmenjavo in povečalo možnosti izkoristka obstoječega potenciala.

Strinjajo se glede tega, da je smotrno hraniti kakovostne podatke.<sup>58</sup> Kakovost pa ni edino merilo, po katerem bi raziskovalci presojali o hrambi. Nekateri podatki so vse preveč dragoceni za stroko in jih ne želijo deliti s širšo javnostjo, saj se, kot je bilo že večkrat omenjeno, bojijo zlorab.<sup>59</sup> Zavzemajo se,

<sup>viii</sup> Glej npr. JCDL 2012: Panel 1: Big Data Is Already Here, and It's Not Always What We Think: <http://jcdl2012.info/panels>, 7. 8. 2012

<sup>ix</sup> Glej Načelo E, OECD Principles and Guidelines for Access to Research Data from Public Funding (2009). Načelo E, Protection of intellectual property, pravi sledeče: Zaradi naraščanja javnega/zasebnega partnerstva v financiranju raziskovalne in druge dejavnosti, povezane s podatki, je potreba po uravnoteženih javno/zasebnih dogovorih, ki bodo omogočili širok dostop do raziskovalnih podatkov, kjer je to mogoče. Samo dejstvo, da je v podatkovne zbirke vključen tudi zasebni sektor, ne bi smel biti razlog za omejevanje dostopa do podatkov. Premisliti bi bilo potrebno ukrepe, ki spodbujajo nekomercialen dostop in rabo, in hkrati ščitijo komercialne interese, npr. zakasneli ali delni dostop do podatkov. Glej tudi: Knowledge Exchange: <http://www.knowledge-exchange.info/Default.aspx?ID=461> Glej Improving knowledge transfer between research institutions and industry across Europe: embracing open innovation. Implementing the Lisbon agenda (2007): [http://ec.europa.eu/invest-in-research/pdf/download\\_en/knowledge\\_transfe\\_07.pdf](http://ec.europa.eu/invest-in-research/pdf/download_en/knowledge_transfe_07.pdf), 7. 8. 2012

da bi bilo treba določiti nacionalne prioritete, za katere bi bila zagotovljena sredstva za vzdrževanje, hrambo in dostop. Kar se drugih gradiv tiče, so nekateri mnenja, da naj ostane v presoji posameznikov, kako bodo ravnali z njimi.<sup>60</sup>

Eden od predlogov je bil, naj se hranijo tisti podatki, po katerih je veliko povpraševanje. Pojavilo se je tudi mnenje, ki takšnemu razvrščanju podatkov ostro oporeka z utemeljitvijo, da privilegiranje gradiv vzpostavlja hierarhijo moči.<sup>61, 62, 63</sup> Zato je pri postavljanju prioritet bolj kot trenutno povpraševanje treba upoštevati dolgoročne strategije, nacionalne posebnosti in podobno. V bodočih politikah dostopa do podatkov mora biti ovrednoten pomen gradiv, v konsenzu z ustvarjalci z različnih področij pa je treba določiti, katere so pomembne na dolgi rok.

Bodoče politike morajo vključevati nacionalne prioritete gradiv, smernice in rešitve glede vzpostavljanja splošnih in področnih servisov. Usposobiti bo potrebno kader, ki bo po načelih knjižničarjev zagotavljal hrambo, vzdrževanje in dostop.<sup>64</sup>

**Okvir 7.6: P-R 13**

**PROBLEM:**

Raziskovalci, ki tako ali drugače rokujejo s podatki, nimajo enotnega mnenja, kaj bi bilo smiselno hraniti in kaj ne. Nimajo prave predstave o tem, kakšne naj bodo pravice in odgovornosti ustvarjalcev in uporabnikov v fazi iskanja, dostopa in hrambe raziskovalnih podatkov, plačanih iz javnih sredstev.

**REŠITEV:**

S politikami bo potrebno določiti tudi obveznosti in pravice ustvarjalcev podatkov in zbirk. Eden od predlogov je, da bi bodoče politike vključevale t. i. embargo oziroma časovno obdobje, znotraj katerega bi imel ustvarjalec 'naravno' pravico, da podatke 'izkorišča' za potrebe svojega dela, po preteku tega obdobja pa jih mora dati na voljo javnosti. S tem bi ustvarjalcu zagotovili določeno obdobje, ko bi lahko objavljal in izkoriščal potencial zbranega gradiva.

**Problem rabe računalniških tehnologij**

Področja, ki niso a priori vezana na digitalne podatke, se soočajo z dodatnimi, in sicer tehničnimi oziroma tehnološkimi ovirami. Na primer na področju slovenistike se ustvarjalci podatkov soočajo s predsodki, ki zadevajo digitalizacijo besedil.<sup>65</sup> Navajajo nezadostno infrastrukturno podporo pri rabi računalniške tehnologije (npr. dodaten prostor na strežnikih). Zadržki so povezani z različnim pogledom znotraj vede, ki raziskovalce deli na tradicionaliste, ki gradijo na natančnem branju in na druge, ki literarno znanost gradijo na podatkovnih zbirkah. Po ocenah intervjuvanca slednji nimajo veliko podpore.

Čeprav tega primera ne gre posploševati, pa vendarle kaže na to, da je potrebno dodatno pozornost namenjat vedam, ki so v tehničnem smislu v 'zaostanku' in jih je zato težko primerjati z drugimi disciplinami, v katerih je raba računalniških tehnologij in digitalnih podatkov že desetletja stalnica. Eden izmed intervjuvancev je izpostavil primer bioinformatike, kjer je prisotnost oziroma udeležba informatikov na tem področju dvignila ozaveščenost in povečala možnosti avtomatizacije.<sup>66</sup> Ena od sogovornic je izpostavila pomen prisotnosti mlajših sodelavcev, ki se šele uvajajo v delo in so pripravljeni na novosti, prav nasprotno starejši zaposleni velikokrat nasprotujejo novostim.<sup>67</sup>

**Okvir 7.7: P-R 14**

**PROBLEM:**

Neenaka stopnja razvitosti infrastrukture, rabe računalniških tehnologij in sodobnih pristopov, pomeni, da se raziskovalci soočajo z različnimi konflikti in težavami nerazumevanja.

**REŠITEV:**

Načelo pri tem naj bi bilo, da se nove rešitve in pristopi uvajajo postopno in inovativni pristopi dobijo priložnost na način spodbujanja in financiranja pilotskih projektov, iz katerih zrastejo bodoče storitvene dejavnosti dolgoročnejšega značaja.



## 7.5 CITATI

<sup>1</sup> Za določene tipe podatkov je to *[obstoj centrov, ki se ukvarjajo z elektronsko hrambo]* na našem področju že nekako urejeno, ravno te tri mednarodne banke. Za kakšne bolj specialne stvari pa ... Recimo, tipičen je naš center, kjer to ni glavna tehnika, ki jo raziskovalci uporabljajo pri svojem eksperimentalnem delu, ampak je to dodaten poskus, ki jim pomaga razložiti neko stvar, ki jo širše študirajo. Ne vidim prav smisla, da bi delali neko tako veliko podatkovno zbirko, kamor bi vse to nalagali ter potem do tega dostopali in analizirali. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>2</sup> Tukaj pri nas, kot jaz vidim, sta dva ekstrema. So mednarodni servisi, ki so zelo dobro organizirani in ustaljeni, vendar vzdrževanje tega veliko stane, tako da si niti ne zamišljamo, da bi kaj takega lahko bilo narejeno pri nas. To so na stotine milijonov podatkov v eni taki zbirki, GenBank, konkretno. Po drugi strani so pa ti poskusi oziroma podatki, ki jih generirajo raziskovalne skupine, mogoče tako zelo ozko specifični, da morda niso zanimivi za vse biologe v Sloveniji. V tem primeru je čisto dovolj, če za te podatke skrbi en tak center, kot je naš. Na mojem področju, biokemijskem, je tako. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>3</sup> *Ali vidite potrebo po tem, da se tudi na nacionalni ravni vlaga v to, da se to področje izpostavi, da se poznavanje te problematike razširi in podobno?* Kar se mojega področja tiče, mislim, da je v redu. Tako kot sem omenil, biokemija, molekulska biologija so discipline, ki so bile vedno navzven usmerjene. Morda je potreba pri kakšnih bolj lokalnih raziskavah, mogoče razširjenost raznih živali, rastlin. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>4</sup> Odvisno, kako je stvar urejena, ampak nek zapis o tem je bolj ali manj strukturiran. Določeni krogi tukaj pri nas, določene raziskave, ravno te genomske, se izrazito zavedajo problemov protokolov in točnega zapisa, ker so tudi zavezani temu, da jih že morajo objavljati. Medtem ko imajo drugi, bolj tradicionalni poskusi, to zapisano, ampak včasih je, včasih pa ni. (znanstvenik s področja biologije, NO6)

<sup>5</sup> Tukaj bi rekel, da politiko literature oblikujeta predvsem Matija Ogrin in Tomaž Erjavec. Oba intenzivno zastopata stališča, da morajo biti znanstveni podatki oziroma podatki, ki se uporabljajo v humanističnih vedah, prvič, javno dostopni, v kolikor je to le mogoče. Pri korpusih včasih ni, ker so narejeni iz zaščitenih virov, ki so pod 'copyrightom'. Kjer to ni mogoče, se najde pot, da je stvar vsaj deloma realno dostopna, se naredi recimo pod-korpus, ki ima iz vsake knjige samo nekaj stavkov in je še zmeraj relevanten ter se da z njim marsikaj delati, ampak je povsem javno dostopen. In drugič, da se naredi dostopne tudi novo nastale stvari, torej edicije, če se le da, v izvorni obliki, v našem primeru je to TEI, in pod čim bolj liberalno licenco. V praksi to pomeni, da so recimo znanstveno kritične publikacije, ki jih objavljajo na spletu, dostopne v surovem TEI formatu, in če hočeš iz tega natisniti knjigo, je vse, kar moraš narediti, xml transformacija v pdf. To velja za vse formate, kjer je to vsaj približno mogoče narediti. Hkrati nas to včasih omejuje pri sodelovanju s kakšnimi podjetji, ki delajo tudi na tem področju, ki si tega ne želijo, ampak se nam vsem zdi, da ta pot dolgoročno ne pelje nikamor. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>6</sup> Večinoma kompliciramo iz enega praznega pravnega strahu. Ta država je pravna tam, kjer ne bi smela biti. Namesto tega se moramo oprijeti tistih civilizacijskih spodbud, ki nam jih prinašajo mediji, ter pa gredo v smeri večjega odpiranja kulturne dediščine, ne pa njenega zapiranja, kar se trenutno dogaja. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)



---

<sup>7</sup> Če mi tekst popravimo in ga postavimo posebej tako, da ni več v okviru recimo Ljubljanskega zvona, ki je v javni lasti in kjer smo ga našli, potem nastane nekakšna vmesna pravna pozicija, kjer človek pravzaprav ne ve, kam sodi njegovo obnašanje: ali je robinhudovsko (ilegalno) ali pa je mogoče čisto nekaj normalnega. Zagovarjam pozicijo, ki je malo drugačna od večinskega prepričanja na Wikipediji in na Wikiviru, kjer vlada velika občutljivost glede avtorskih pravic. Sam sem bolj zarobinhudovstvo. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>8</sup> Odprti so [besedilni korpusi] za kogar koli, brez registracije. Ne gledamo, če boste iskali kakšne demografske vsebine ali pa kaj podobnega, nič vas ne nadziramo. (znanstvenik s področja jezikoslovja, NO13)

<sup>9</sup> Tudi na Wikiviru dajemo prednost objavi prvih izdaj, brez sprememb in brez posodabljanja. Torej so redakcijski principi enaki kot pri inštitutski ediciji. Ampak tam imajo vsega skupaj šest skupin besedil, medtem ko je na Wikiviru letni prirastek 4 milijone besed. Gre nam za čim širšo bazo, četudi besedila niso vedno najboljše redigirana. Važno je, da so dostopna, da so javno uporabna. Redakcija besedil na inštitutu pa je znanstveno početje, gre za zelo natančne komentarje, primerjavo variant in tako naprej. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>10</sup> Gre za opozicijo med dvema tipoma dela v stroki. Tradicionalno literarnovedno pisanje nima z velikimi podatkovnimi zbirkami nobenih opravkov in je kljub temu še vedno deležno določenega družbenega ugleda, saj kreira družbena pričakovanja v zvezi s slovensko književnostjo in jim ustreza. Drugo je pa to, čemur se reče literarna znanost in si jo predstavljamo samo na osnovi čim boljših, čim bolj natančno in popolno zbranih podatkov. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>11</sup> Pri nas v socialnem delu je v raziskovanju takšno, že precej uveljavljeno pravilo, čeprav tega ne kontroliramo, da damo zapise, prepise ali intervjuje nekih opazovanj, na vpogled respondentom, da povedo, ali je to res verno zapisano, kar smo se pogovarjali. Ker nimamo zelo velikih vzorcev, včasih naše raziskave temeljijo na študijah nekaj primerov. To je en vidik, drugi vidik nastopi kasneje, ko pravzaprav že naredimo analizo in dobimo ključne raziskovalne ugotovitve. Ponavadi, predno poročila predstavimo širši javnosti, s to skupino respondentov spet predebatiramo. Ne morem pa zagotoviti, da v vseh raziskovalnih skupinah, ki tukaj delujejo, oziroma v vseh projektih, to poteka točno tako. (znanstvenik s področja socialnega dela, NO7)

<sup>12</sup> Tudi sama imam to izkušnjo, da je tam, kjer sama zbiram podatke, tudi analiza lažja. Ko pa dobim intervjuje, ki jih je nekdo drug naredil ... Pri nas je marsikdaj tudi tako, ampak bolj za takšne raziskave, ki jih potem niti ne objavljamo širše, bolj za vpoglede v situacije, da včasih prosimo študente, da naredijo kakšne intervjuje. In tisto je včasih zelo skopo, ker tisti, ki je delal intervju, ni postavljajl podvprašanj, ni poskušal pravzaprav zares razumeti, kaj mu je oseba povedala in tam imamo težave. In zna biti, da je lahko to ovira. Nekih zelo poglobljenih dodatnih analiz potem najbrž iz the ... Ali pa je bolj odvisno od gradiva samega, kako popolno je bilo gradivo zares zbrano. (znanstvenik s področja socialnega dela, NO7)

<sup>13</sup> Zelo veliko razliko čutijo raziskovalci do tega, kar je že objavljeno in kar še ni objavljeno. Ko je enkrat objavljeno, je potem pač na voljo. Dokler ni objavljeno, bo zelo težko kdo od eksperimentalcev kaj dal od sebe za javno rabo. /.../ Tukaj morda jaz malo naivno gledam, ker gledam mogoče malo manj specialno. Jaz nimam nobenih podatkov, ki bi bili tako moji, da bi sedel na njih. Drugače splošno vzdušje in klima je pa dostikrat takšna, da "je to moje". Ni rečeno, da je vsem pogodu, da bi še nekdo

drug v teh podatkih našel nekaj, česar sam nisi videl. (znanstvenik s področja biologije, NO6)

<sup>14</sup> Mikropodatki sami veliko veljajo v znanstveni javnosti. Če imaš podatke, objavljaš. Saj ni noben problem pognat program, saj to vsi znamo. Vse sintakse so on-line. Me razumete, v čem je problem? (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>15</sup> Če sem jaz normalno financiran kot javni raziskovalni delavec in imam plačo zagotovljeno, da raziskujem in delam, potem ne vidim nobenega problema. Dokler smo pa mi volkovi in ovce ... Tudi sodelovanje. Tam, kjer smo prisiljeni, sodelujemo, če pa nismo prisiljeni, potem pa konkuriramo. 'Know how-a', ki ga imaš, nočeš dajati svojemu potencialnemu konkurentu, to je problem. Če so stvari normalne, če nimaš težav s preživetjem, ali boš na koncu meseca imel plačo, potem res ne vidim nobenega problema. Ampak nihče me še ni vprašal 'Mi daš te podatke?', pa bi jaz rekel OK, ne bom ti jih dal. Nihče me ni še tega vprašal eksplicitno. Sem pa jaz nekoga vprašal za eno bazo podatkov, ki je mi ne uporabljamo. Odgovor je bil: "Pojdi na Statistiko in izvoli." Je pa on dve leti to čistil, pripravljal, povezoval in tako naprej. Noče dati in ga razumem. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>16</sup> Znanost je izjemno naporen biznis, znanstveniki nimajo časa, ravno zaradi konkurence. In naložiti jim neko dodatno obvezo v smislu objave nekkih na pol pretiranih rezultatov, je zelo težko. Kar bi bilo mogoče narediti, je objaviti surove podatke. To, kar mašina 'izbruha'. Preden greš obdelovati tabele, je to mogoče objaviti. To je za ljudi, ki so s področja, ki imajo podobne mašine, ki vedo, kaj je uporabno. Ko pa ti podatke obdelaš, pride končna faza seveda kot dodatek k objavi v revijah. To se objavi v preprostih verzijah, v obliki že obdelanih tabel. Vmes je pa po mojem mnenju nerealistično pričakovati, da bo kdor koli dodaten čas vložil v to, da bi kakšne vmesne rezultate naredil čitljive. Ali tiste surove, ali pa tiste končne. (raziskovalec s področja naravoslovnih ved, NO3)

<sup>17</sup> Konkurenca v znanstvenih krogih je velika, da bi kdo dal še neobdelane podatke drugim na voljo, to se ne bo zgodilo. Vsaj kolikor jaz poznam stanje. Gotovo bo na delu tudi subjektivni faktor, mogoče ne bodo vsi videli samih prednosti v takšnem sistemu, mogoče je kdo bolj ljubosumen kot nekdo, kar se tiče njegovih lastnih dognanj. Po mojem mnenju je tukaj potrebno upoštevati tudi subjektivni dejavnik. (vodja knjižnice, NO18)

<sup>18</sup> Načeloma je pa tako, da so tukaj po teh protokolih vsi podatki na poti, da postanejo javni. Roko se drži nad njimi samo, dokler niso v celoti obdelani. *Obdelani v smislu čiščenja?* Ne, tudi analize. Analize v smislu, ali smo kaj našli. To je pa ljubosumje. Ni tukaj cilj, da damo surove podatke ven, pa nekdo, ki ni 14 let gradil detektorja, prej najde nek ... (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>19</sup> Mislim, da bi se moral cel sistem reformirati. Saj kakšnih deset let nazaj so imeli enak projekt in cilje pri Svetovni banki preko Global Development Network, financirat projekte, ki bi zbirali podatke, metapodatke, ki bi bili širše uporabni. Mislim, da dalje od Labour Force Survey-a niso prišli. Ker so raziskovalci enostavno povedali, da tudi če oni financirajo projekt in dobijo za nek projekt 10.000 evrov, je to še zmeraj majhen delež vsega napora, ki je v to vložen. Ker zbiranje podatkov, sploh iz podjetij, če to niso raziskave javnega mnenja, je izjemno težaški posel. Moraš imeti kontakt, da sploh prideš v podjetje, potem si moraš vzeti čas, da greš tja, kar pomeni, da na dan lahko narediš dve podjetji. Hitro lahko izračunamo, v kolikšnem času dobiš recimo 300 podjetij. In to ponavlja leto za letom, to je ogromen osebni angažma. Nekdo, ki to dela, dela zato, da bo to potem koristil, ne zato, da bo dal potem 'for free' na razpolago. In to je problem, ker tega noben ne daje javno na razpolago. Če pogledate po svetu, to nikjer ni javno dostopno. (Polona Domadenik, znanstvenica s področja

---

ekonomije)

<sup>20</sup> Se pravi ljubosumje, da bi kdo od zunaj posegel po podatkih, ki so bili zbrani z muko ali pa z določenimi finančnimi sredstvi, in jih izrabil za veliko sintezo. Se pravi, da bi po bližnjici prišel do rezultatov in pri tem prehitel tistega, ki se je na sintezo pripravljaval v potu svojega obraza. Zdelo se je nekako normalno, da je tisti, ki je podatke zbral, tudi edini njihov upravnik in da ne gredo v javno rabo. Ljubosumnost je bila velika. To naziranje se sedaj dejansko spreminja in k temu prispevam tudi sam: podatki naj bodo od samega začetka javno dostopni. Celo v fazi kreiranja, ko še niso popolni, naj bodo dostopni javnosti. Pa ne samo dostopni za uporabo, ampak tudi za dopolnjevanje in popravljanje s strani javnosti. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>21</sup> Ne vem, če so tukaj problem nagrade. Bolj je problem v tem, da ko se lotiš enega problema ugotoviš, da je vse, kar dobiš zastonj, v bistvu nič vredno. In če nisi ravno na kakšni prestižni univerzi, kot je Harvard, ki ima dostop do treh milijonov podjetniških podatkov, govorim za svoje področje, se lahko lepo pod nosom obrišeš za kar koli. Vidiš, da se moraš v bistvu sam ogromno angažirati. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>22</sup> Ja, majčkeno smo ljubosumni. To smo mi delali, za to smo porabili toliko in toliko ur in svojega dragocenega razuma in se nam zdi škoda to nekomu zastonj dati. Čeprav po drugi strani smo pa dobili plačano, to je le končni produkt projektov, ki so nam jih plačali. (znanstvenik s področja favnistike jamskih nevretenčarjev, NO22)

<sup>23</sup> Če želimo dobiti podatke od SURS-a, ki je plačan, da samo to dela, moramo plačati. Sedaj imam ravno en CRP, kjer delamo s podatki s področja trga dela. Vsi materialni stroški, pa upam, da bo zadostovalo, bodo šli za to, da plačamo dostop do mikro podatkov na SURS-u. In če se ne morejo zmeniti na državni ravni, da bi državni uslužbenci to počeli, ki so že tako ali tako plačani, ne vem, zakaj bi jaz to dala, ki porabim 150 % svojega prostega časa, da te stvari urejam in delam. Problem je v sistemu, sistem je narobe postavljen. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>24</sup> Sedaj skušamo dobiti podatke za upokojujence, zgodovino njihovih podatkov, koliko, kdaj so bili zaposleni, koliko je bila njihova plača, koliko prispevkov so plačali, kakšna so bila nadomestila, vse. Že eno leto se pregovarjamo, če bo Statistika dala vzorec, EMŠO, Pokojninskemu zavodu. Statistika hoče imeti vse podatke, na SPIZ-u pravijo "Nemogoče, tudi če vam damo vse naše podatke, vi iz tega ne boste mogli nič narediti, ker so to baze, ki so različne, ki so med seboj nepovezane." Mi jih znamo skupaj zložiti in nekaj odbrati. Smo potrebovali nekaj mesecev, da smo prišli do tega, da so jim dali samo EMŠO vzorca gospodinjstev, posameznikov, ki jih imamo v vzorcu, da so oni potem prilepili pokojninske podatke. Dali so samo za eno leto, da smo testirali. Po šestih mesecih smo rekli, da sedaj pa rabimo še druga leta in so rekli, da zdaj ta pogodba ne velja več. Sedaj bomo sklenili novo pogodbo, nam se bo pa projekt končal konec leta in še nimamo podatkov. Po letu in pol še nimamo podatkov. Lasje mi gredo pokonci, ko pomislim, kako in koliko časa traja, da prideš do teh podatkov. A podatki pa obstajajo, zagotovo. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>25</sup> Če hočeš dobiti nek evropski projekt, moraš imeti dobre objave. In kako boš prišel do dobre objave? Američanom zelo težko konkuriraš drugače, kot s tem da imaš neke super podatke. Tiste podatke obdelaš na standarden način in na ta način dobiš objavo. Me razumete, v čem je problem? To vsi delajo. In še najbolj v OECD-ju in na Svetovni banki. Veste, zato grejo meni kar lasje pokonci, ko slišim za ta dostop do podatkov, ker sama vem, kako je s tem. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>26</sup> Potem mi pa povejte, zakaj ne dajejo [OECD] najnovjših poročil on-line, da bi jih lahko vsi prebrali? Saj so financirani, vse to imajo plačano. Zakaj potem še prodajajo, povejte mi. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>27</sup> Sploh nisem še nikoli niti razmišljal, kako jih bo Statistika uporabila oziroma dovolila. Podatke, ki jih imajo oni, imajo pod kontrolo, podatke, ki jih imate vi, imajo tudi pod kontrolo. Ne vem, verjetno bi se to dalo dobiti. Če po pravici povem, sem že enega raziskovalca spraševal in je jasno rekel, da za živo glavo ne dobiš teh podatkov. Pojdi na Statistiko in vprašaj, pa jih boš dobil. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>28</sup> Če nekdo reče, da bi rabil recimo tole karto, mu jo pošljemo. Običajno mu je ne damo v osnovni verziji, ker je problem, da se lahko potem kaj poruši, ampak mu damo karto. Ne morem mu pa dati podatkov, lahko ga napotim, da jih dobi na Statističnem uradu. Osnovnih podatkov mu v tem primeru ne morem dati. (znanstvenik s področja geografije, NO8)

<sup>29</sup> Ne vem če OECD že sam ne krši svojega prvega načela s tem, da svojih publikacij ne dajejo on-line in da jih je treba naročevati. Podobno je bilo pri Svetovni banki, jaz sem bila tam kot raziskovalka en mesec. Oni so enako imeli vsa ta priporočila, zato jaz prav nič ne dam na njihova priporočila, jih sploh ne berem, ker mi gredo lasje pokonci. Kot sem že rekla ta GDN, ki je bil financiran s strani Svetovne banke, enako: dajmo ljudem denar, da bodo zbirali podatke in potem bomo dali mi te podatke on-line. Jaz sem takrat prišla na Svetovno banko kot raziskovalka, imela sem neke makedonske podatke, potrebovala sem samo eno dodatno leto, ker smo delali nek članek. In sem šla do človeka na Svetovni banki, ki je imel podatke točno za tisto leto, ki sem ga potrebovala, pa ni bilo nobene možnosti. Samo on jih je imel in za svojo lastno uporabo, zato, da je pisal reporte. To je en velik blef. Oni sami ne dajo podatkov. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>30</sup> On mora prosit za te podatke Statistiko in Statistika mu jih bo dala, neodvisno od mene. Jaz mu ne morem tega dovoliti, ker nisem lastnik. On mora prositi za te podatke in Statistika mu odobri. Ne jaz ne Inštitut nima nobene veze s tem. Podatki, ki so na razpolago na Statistiki, so na razpolago vsem, ne samo nam. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>31</sup> Če bodo državne ustanove same zgled v dostopu do podatkov, mislim, da bo tudi raziskovalce veliko lažje prepričati, da bodo dali podatke. Če meni kdo reče, če bom zbrala podatke, bom rekla: "Ne, hvala. Kar plačajte nekoga drugega in ko jih bo zbral, bom tudi jaz prišla ponje." (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>32</sup> Menda celo razpisujejo, niti ne projekte, neke takšne vrste dela, da če delaš na teh podatkih, je to dobro za državo, ker vedno malo drugače pogledaš na podatke. Kaj novega narediš in ti država celo plača za vsak članek, ki je narejen na teh podatkih. Pri nas pa moraš ti plačati, da sploh prideš do podatkov. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>33</sup> ARSO v tem primeru skrbi za to, da se shranijo podatki vseh meritev. Razlikujemo sinoptične in klimatološke meritve. Sinoptične meritve so le nekaj minut potem, ko so narejene, skozi posebne kanale za izmenjavo meteoroloških podatkov dostopne meteorologom kjerkoli po svetu. Če raziskovalci potrebujemo podatke, so v meteorologiji vsi izmerjeni podatki prosto dostopni. *To je samo za raziskovalce ali tudi za posameznika, če bi želel?* So razlike med meritvami in podatki, ki so rezultat numeričnih izračunov. Pri tem mislim predvsem na produkte numeričnih modelov za napovedovanje vremena. Prostega dostopa do teh podatkov v realnem času v Evropi zaenkrat

---

nasplošno ni, medtem ko so v ZDA tudi rezultati numeričnih modelov prosto dostopni takoj po nastanku. Ameriško sodišče je namreč odločilo, da davkoplačevalcu ki je preko davkov plačal za opazovanje, oziroma plačuje ARSO, da skrbi za mrežo opazovanj, shranjuje in obdeluje podatke, pripravlja numerične modele in njihove napovedi, ni potrebno še enkrat plačati za isti izdelek. V večini evropskih držav je pristop do podatkov bolj omejen. Posebej to velja za specialna opazovanja, zbrana med obdobji intenzivnih meritev. V Evropi se na spletnih straneh dobijo vremenske napovedi za posamezna mesta in slike oz. grafični prikazi podatkov. Odločitev Ameriške meteorološke službe NCEP v prvi polovici 90-ih let, da dajo prost dostop do 'golih' podatkov, torej izračunov modelov je omogočila številnim manj bogatim državam, posameznikom in privatnem sektorju, da razvijejo lastne napovedi vremena v realnem času in jih tako uporabljajo za opravljanje gospodarskih dejavnosti. Namreč, napovedi vremena za omejena območja rabijo začetne in stranske robne pogoje in prav te začetne in stranske pogoje so Američani naredili prosto dostopne. Zahvaljujoč tem podatkom so npr. študentje s Strojne fakultete par let nazaj pognali vremensko napoved, mislim, da je spletna stran provreme.si. Tudi mi na katedri v okviru Centra odličnosti Vesolje SI poganjamo vremensko napoved v realnem času, ki je dostopna na naši spletni strani, vendar mi glede tega sodelujemo tudi z ARSO-m. Ker evropske meteorološke agencije ponavadi dajejo produkte, na primer slike napovedi, na spletne strani, ne pa tudi podatkov, podjetnik iz teh slik ne more razviti dodatnih produktov. To je bistvena razlika od ameriškega pristopa. Moram pa omeniti, da se situacija v Evropi počasi spreminja proti večji odprtosti oz. prostemu dostopu do vseh produktov meteoroloških služb. Izjemno močna uporaba meteoroloških podatkov je na področjih gospodarstva, ki se ukvarjajo z vplivi onesnaženja in z izkoriščanjem obnovljivih virov energije, posebej vetra. (Nedjeljka Žagar, znanstvenica s področja meteorologije)

<sup>34</sup> To je pa posebno poglavje, mislim da, od vsega še najbolj pomembno. To je tako pomembno vprašanje, da bi morali po mojem mnenju že v projektu imeti kak razgovor s pravniki, ki se ukvarjajo prav s tem področjem. Ker avtorstvo mora biti zagotovljeno. (znanstvenik s področja geografije, NO8)

<sup>35</sup> Če pa gledam svojo analizo, bi pa jaz čim več mojega dala on-line, ampak si ne upam zaradi tega, ker ne vem, kje so tu uredniške in kje so moje avtorske pravice. Ker tukaj se to povezuje še s citiranostjo. Vedno več hočemo biti citirani in citiran si, če daješ stvari on-line. Da bi me čim več ljudi prebralo, sem naredila svojo stran, kamor sem dala svoje članke, sicer v angleščini, ker to mednarodni prostor bolj zanima, in čim več mojih zapisov, ker bi rada, da me ljudje opazijo. (Nina Vodopivec, raziskovalka s področja socialne antropologije)

<sup>36</sup> To bo ovira, čim več bo ljudi prihajalo z digitaliziranim gradivom, s fotografijami, s filmi, z oralno, ustno zgodovino, z glasbo ... To je zelo veliko različnih avtorskih pravic. Mi določenih gradiv sploh ne moremo dati on-line zato, ker se bojimo zaradi avtorskih pravic. (Nina Vodopivec, raziskovalka s področja socialne antropologije)

<sup>37</sup> Vedno več objavljamo na wiki straneh, kjer je vsako avtorstvo do vejice natančno dokumentirano. Kdor ima željo, da nastopa pod svojim imenom, to seveda lahko, vendar ne z namenom, da bi se proslavil kot urednik ali avtor, kot nekdo z zaslugami. Avtorsko prispevanje bi moralo biti samoumevno, ne pa da avtor preži, kdo bo njegove podatke uporabil ali pa celo zaračunava zanje. Avtorski napuh in želja, da se vsaka objava honorira, sta zunaj novega koncepta avtorstva na spletu. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>38</sup> Družba se zelo spreminja in vprašanje avtorskega napuha počasi izginja z dnevnega reda. Opažam, da mnogi ljudje, ki ogromno delajo, npr. prostovoljci pri wikijih, ne čutijo čisto nobene potrebe več,

da bi nastopali s svojim imenom in priimkom. Zadošča jim vzdevek, kljub temu da opravljajo družbeno pomembna infrastrukturna dela. Predstava o tem, da je avtorski ego tisto, kar sploh družbo naprej vodi, je v zadnjih letih zbledela, prav tako ekonomska logika, da mora biti zadaj profit. To se začuda dejansko dogaja in je treba podpirati. Spreminja se kulturna paradigma. Avtorstvu in avtorskim pravicam v prihodnje sploh ne bomo pripisovali tako velikega pomena, kot ga imajo danes. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>39</sup> *Ste določili nek minimalni nabor metapodatkov?* Tako, čeprav tu ni bilo predpisano, ker se zgodi, da kdaj nekdo ne sme dati imena izvajalca, ker so gor avtorske pravice, ena država ima takšne, druga malo drugačne predpise ... Opremimo z osnovnimi podatki, da lahko najdeš gradivo in po drugi strani, ko slišiš, da dobiš nekaj več informacij. Še zmerom pa je ideja ta, da se gre na posamezni arhiv, če želiš podrobnosti. (znanstvenik s področja zvočnega arhiviranja, NO21)

<sup>40</sup> Mene skrbi zloraba, ko gredo podatki enkrat iz hiše. Prvič, ali jih bo nekdo potem pravilno interpretiral. Drugič, ne morejo biti podatki na desetih mestih, ker morajo biti na enem, da se primerno ažurirajo. In tretjič, če ne govorimo o podatkih, ki jih dobimo od SURS-a, GURS-a in podobno, so to včasih stvari, ki jih naši znanstveniki zbirajo desetletja. On bo delal na tem še naslednjih deset let, ne more sedaj podatkov posredovati nekemu, ki ni vložil še nobenega dela v to, na koncu bo pa rezultate objavil pred njim. (znanstvenik s področja geografije, NO8)

<sup>41</sup> Seveda, saj raziskovalci imajo dostope. Ampak z njimi se mora prav tako podpisati neko pogodbo, kjer se zavežejo, da bodo zadevo uporabili izključno za določen namen. Nekje mora biti nek dokument, s katerim se zavežeš, kako boš podatke uporabljal. Zato jaz govorim o zlorabi, ker se bojim, da bi se potem to kar povprek uporabljalo, nenadzorovano. Nekdo, ki je podatke zbiral, mora vedeti, kaj se z njimi dogaja. (znanstvenik s področja geografije, NO8)

<sup>42</sup> Podatki iz zbirke ARKAS (Arheološki kataster Slovenije) so za zdaj dostopni le za strokovne namene. Interesenti si dostop prek spleta lahko zagotovijo na podlagi prošnje za dovoljenje, ki je praviloma odobrena vsem iz arheološke stroke in študentom arheologije, kakor tudi tistim, ki sodelujejo z nami bodisi v skupnih projektih ali tako, da nam posredujejo nove podatke o arheoloških najdiščih. Za tak restriktiven dostop smo se odločili zato, ker si arheologi ne želijo nestrokovnih in nenadzorovanih posegov raznih ljubiteljev in zbirateljev v arheološka najdišča ter iskanja arheoloških predmetov iz nekih osebnih nagibov in materialnih koristi, kar se dogaja. Za namene splošnega izobraževanja smo pripravili druge aplikacije, ki so odprte širši zainteresirani javnosti (npr. Interaktivna karta Slovenije z zbirkami ZRC SAZU, ARZENAL, ZBIVA) in ne vključujejo občutljivih podatkov npr. o natančni lokaciji najdišča. (raziskovalec s področja arheologije, NO14)

<sup>43</sup> Tudi za takšne raziskave, kjer boste samo uporabljali stvari, ki so shranjene, ponavadi potrebujete odobritev iz etične komisije. Pacient ponavadi podpiše nekaj zelo splošnega, kjer odobri, da se tisto kar bo ostalo, en del vzorca je uporabljen za diagnostiko, lahko uporablja tudi za druge raziskave. (znanstvenik s področja medicinskih vez in statistike, NO15)

<sup>44</sup> Tudi za retrospektivne študije potrebuje raziskovalec dovoljenje etične komisije. /.../ Jaz sem se veliko ukvarjala prav z retrospektivnimi analizami shranjenih tkiv in v tistem primeru na začetku niso vedeli točno, katere vrste dovoljenja potrebujejo od pacienta. Na začetku so mislili, da bodo morali za vsako raziskavo vprašati bolnika, če se strinja, da bodo tkivo uporabljali za ta in ta namen, za točno določeno raziskavo. Kar je seveda popolnoma neizvedljivo, če pomislite, da boste šli 20 let kasneje vprašat neko osebo, ki po možnosti niti ne bo več živa, pa dednike, če se strinjajo. In torej podpišejo

---

nekaj zelo generičnega. Za posamezne podatke pa ne vem točno, če sploh obstaja pravica bolnika, da noče biti vključen. (znanstvenik s področja medicinskih vez in statistike, NO15)

<sup>45</sup> Veliko je potem v interpretaciji. Vse kar tako skupaj potegniti, je včasih mogoče res težko. Zato bi mogoče bil nek servis, ki dela podatkovje na standardiziran, določen način. (znanstvenik s področja biologije, NO6)

<sup>46</sup> Težava, ki jo jaz tukaj vidim, je ravno pri metaanalizi, saj velja, kot da so enaki pogoji, ampak je grozno težko vso stvar skupaj sestaviti, ker te stvari ne gredo direktno skupaj. To ni lanski podatek o nekem stanju v družbi in letošnji, da bi to še vedno nekako približno skupaj pasalo, tudi če je metodologija malo drugačna. Tukaj so stvari takšne, da sta ponovljivost in dejanska primerljivost včasih zelo, zelo težki. /.../ Eno je standardizacija postopkov, drugo je pa standardizacija pogojev. Niti pogojev, v katerih delaš, ne moreš standardizirati. (znanstvenik s področje biologije, NO6)

<sup>47</sup> Tak sistem, ki bi omogočal povezovanje različnih podatkovnih zbirk, si je težko zamisliti, ker bi zahteval vrsto modifikacij obstoječih zbirk ter razne standarde. Razen tega je zaradi zelo različnih vsebin in specifik težko najti skupni imenovalac. (raziskovalec s področja arheologije, NO14)

<sup>48</sup> Včasih so skorajda površnosti, recimo med raziskavo ne zabeležijo spola živali in na koncu se izkaže, da imaš dve skupini, ki sta čisto različni. Obstaja možnost, da je to, vendar tega ne moreš preveriti, ker v samem toku stvar ni bila tako postavljena. Za določene stvari so taki servisi na tem genomskem, bioinformacijskem področju. Sekvenciranja enostavno ne more narediti kdor koli. Pošlješ material in dobiš nazaj datoteko s podatki z opisom, kaj so naredili in kako so do tega prišli. Sedaj se mi zdi, da je ena možnost ta delitev dela. (znanstvenik s področje biologije, NO6)

<sup>49</sup> Večinoma so [servisi za analize] plačljivi. Imajo toliko 'inputa', toliko stvari gre skozi, da stvari niso drage. Če hočeš narediti svoj laboratorij, dobiti kader in vse to narediti, stane ogromno denarja, potrebuješ pa to mogoče trikrat, štirikrat na leto. To je treba najeti. In ta delitev pri naravoslovju se mi zdi, da je lahko zelo ... Saj v družboslovju se verjetno tudi lahko najame neko agencijo, ki izvede anketo, verjetno ni treba, da imaš svoj kader. (znanstvenik s področje biologije, NO6)

<sup>50</sup> Tisto, kar mi uporabljamo, lahko vse delimo - GPRS izmerjene podatke, daljinsko zaznavanje, satelite, lidar ... Vsak, ki bo šel nekaj novega na tistem območju delat, bo najprej vprašal, če ti že to imaš, iz katerega leta imaš. Ker potem njemu ne bo treba novega naročiti in bo samo odkupil tvoje podatke, če jih bo potreboval v raziskovalni namen, pa še to ne bo treba oziroma za neko minimalno ceno. Škoda je zapravljati denar za to, da se trikrat posname isto območje, če je že nekaj narejeno. (raziskovalec s področja geodezije, NO19)

<sup>51</sup> *Bi lahko dali kakšne kriterije za to vrednotenje?* Težko je tako na pamet oceniti. Ampak škoda bi bilo čisto zavreči vrednost katerega koli gradiva, ki je zbrano. Sploh na ta način, ko dobimo potem tudi avtorizacijo, potrditev, da smo pravilno razumeli stvari, ki so jih ljudje povedali. *Se pravi, vso gradivo ima neko vrednost? Čisto v osnovi, kot dokument neke raziskave.* Kot osnovni podatki, ne zgolj kot raziskave. Vsaj neko zgodovinsko vrednost, če bi kdo želel za nazaj neke ponovne analize delati. Je pa res, da se ukvarjamo s področjem, kjer so spremembe dokaj hitre, ampak mislim, da je to povsod. Ne vem, težko bi postavila nek kriterij, da bi rekla, teh podatkov pa ne bi dali. (znanstvenik s področja socialnega dela, NO7)

<sup>52</sup> Malo večji zadržek imam do čisto tržnih raziskav. Čeprav se mi zdi, da bi, sploh pri nas specifično v

stroki socialnega dela, morala biti javna korist prevladujoč kriterij, ne glede na to kdo nam plača, in bi že zaradi tega morale biti stvari dostopne. (znanstvenik s področja socialnega dela, NO7)

<sup>53</sup> Meni se zdijo vsi teksti pomembni in je sprejemljivost ali nesprejemljivost podatkov odvisna samo od tega, ali so popolni ali nepopolni. Ne pa to, ali se nanašajo na kvalitetno ali nekvalitetno literaturo. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>54</sup> So tudi uporabniki, ki ne bi pustili, da je to dostopno, predvsem firme, s katerimi sodelujemo. Mi moramo podpisat 'Non disclosure agreement-e'. Recimo eni iz Italije prinesejo material, odnesejo, nič ne pustijo, kar ostane. (Gregor Anderluh, znanstvenik s področja biologije)

<sup>55</sup> *Kako vi trenutno delate, kaj delate s temi podatki, koliko ste jih pripravljene odpreti ali pa ne, pod kakšnimi pogoji?* Mi smo javni zavod. Odvisno je od tega, s kakšnega projekta smo stvari pridobili oziroma za kakšne namene smo jih delali. Največkrat so tako ali tako naročniki ministrstva ali pa gre za neke mednarodne projekte in to je neke vrste javni denar. Samo ne moremo kar reči, vse kar je država plačala, bo javno nedostopno. Kam bi pa prišli? Na koncu se ne bi več vedelo, kaj je še zanesljiv podatek in kaj ni zanesljiv podatek, če bi kar vsak nekaj predelal. Drugič, eno so surovi podatki, ko gremo na teren, poberemo lokacije, vpišemo itd. potem pa cela ekipa ljudi dela iz tega neke analize, katerih rezultat so običajno elaborati projektov ali znanstvene objave v knjigah. (znanstvenik s področja geografije, NO8)

<sup>56</sup> Tudi moje osebno mnenje je, da je to, kar so naredili Američani, povsem prav, posebej če je že enkrat plačano za te podatke. Seveda če gre za raziskovalne rezultate, je drugače. Jasno, če nekaj raziskujem in proces še vedno poteka, ne morem tega dati javno na razpolago, dokler ne dokončam raziskav in dokler jaz tega ne publiciram. Ampak ko so enkrat rezultati publicirani ali patentirani, mislim, da mora biti vse odprto in prosto dostopno vsem. (Nedjeljka Žagar, znanstvenica s področja meteorologije)

<sup>57</sup> Problem, na katerega bi naleteli, je ravno to, kdaj podatke narediti dostopne. Eni podatki so financirani iz javnih virov, drugi iz privatnih. Tukaj je zelo lahko narediti mejo. Pri javnih pa je pomembna časovnost, kdaj se pusti, koliko časa so podatki na nek način samo navedeni, ampak ne moreš do podrobnosti, več takšnih stopenj mora biti. Embargo recimo, ali je to konec raziskave, takrat je treba dati vse od sebe, si objavil ali pa nisi objavil, to je problem. Z objavljanjem je pa tako, ti članek oddaš in potem traja dve leti, da je objavljen, medtem pa raziskava že poteče. Jaz sem sicer naiven in še zmeraj mislim, da so ljudje pošteni, korektni in viteški, ampak se zgodi, da se pojavi celo slika ali članek, ali pa vsaj ideja, v obliki nekega drugega članka, ki je bil že dve leti nekje v referenčnem postopku. (znanstvenik s področja biologije, NO6)

<sup>58</sup> Jaz kot raziskovalka zagovarjam tudi kvaliteto podatkov. Moti me, ko danes na spletu najdeš vse, redkokdaj pa res kvalitetne stvari. Ko smo delali projekt DEDI, je bila najprej ideja računalničarjev, da naj vsak nekaj dodaja. Ampak mi se kot znanstvena ustanova ne moremo iti projekta, ki bo na koncu Wikipedija, ker moramo stati za svojimi trditvami. Smo imeli že uredniški odbor, ki je presojal, kaj je kvalitetno in kaj ni, pa še tu nismo našli skupnega jezika. Ne pa, da bo kar vsak dodajal. Saj so se na nas obračali ljudje, ker so hoteli, da kaj dodamo, ko se videli, kaj je DEDI. To je Enciklopedija naravne in kulturne dediščine na Slovenskem, ki ima opisane takšne osnovne objekte, sedaj pa jih počasi dodajamo. Ampak zapletlo se je ravno pri tem, ko smo rekli, da če mi pri tem vsebinsko zadaj stojimo, bil je še NUK, ZAP - Zavod za gradbeništvo, računalniške firme, potem moramo zagotavljat kvaliteto podatkov. (znanstvenik s področja geografije, NO8)



<sup>59</sup> Ampak arheološki podatki so pa zopet problem, ker oni ne morejo dati vseh podatkov. Običajno niti nočejo dajat lokacij, da ne bi nekdo prišel tja in začel izkopavati, da ne bi prišlo do ropanja in podobnih stvari. (znanstvenik s področja geografije, NO8)

<sup>60</sup> Smo rekli, da bi morala biti za vzdrževanje zagotovljena velika sredstva, ampak ali je res vse dovolj kvalitetno, da se spleča vzdrževati, ali je res vse tako nujno z nacionalnega vidika? Morali bi se dogovoriti, kaj so tisti nacionalno pomembni podatki in samo na tistih graditi in za tiste zagotavljati sredstva. Vsi ostali bi pa bili izmenljivi, kakor bi pač kdo ... Ker kadar imaš interes določene podatke najti, te bodo že usmerili do pravega človeka, ki ti jih bo potegnil iz predala. Ampak če bosta recimo v treh letih dva potrebovala določene podatke, potem ni smiselno vlagati v njih, potem je bolje, da si jih individualno poiščeta. Morali bi graditi na tistih nacionalno pomembnih podatkih, po katerih se veliko povprašuje in veliko dostopa, ne pa kar vse povprek. Ker potem bo zopet zmešnjava in ne boš vedel, ali je sedaj v redu, ali ni, ali so zadnji, ali niso. Pa tudi neizvedljivo bi bilo, ne nazadnje. (znanstvenik s področja geografije, NO8)

<sup>61</sup> Meni se zdi, da privilegiranje gradiva ni v redu, ker je vedno nekdo tisti, ki privilegira v okviru nekih svojih kriterijev in je to zelo težko za tako različne discipline, za tako različne oblike podatkov in upam, da za vedno več različnih institucij, zaenkrat govorimo mogoče o javnih zavodih, mogoče bodo v prihodnosti to različne privatne zbirke, ki jih ustanavljajo privatni raziskovalci ... Kako povezovat različne oblike institucij in kdo bo privilegiral to na podlagi tega v prostoru hierarhije moči, to se mi ne zdi v redu. (Nina Vodopivec, raziskovalka s področja socialne antropologije)

<sup>62</sup> Meni se to ne zdi slabo, slaba se mi zdi samo centralizacija. Sem zelo proti centralizaciji, ker se mi zdi, da je to lahko neka svojevrstna diktatura. Ampak drugače se mi zdi pa prav, da so ti prostori za tiste, ki se pač ne morejo drugače vključevati. (Nina Vodopivec, raziskovalka s področja socialne antropologije)

<sup>63</sup> Jaz sem takrat imela idejo, da je v bistvu dobro stvari centralizirati, ampak potem, ko vidiš, kako so različni kriteriji, kako so različne vrednote, kako so različne discipline, kako so različne oblike podatkov ... Se mi zdi, da je dobro, da stvari obstajajo centralizirano decentralizirane oziroma koordinirano decentralizirane. (Nina Vodopivec, raziskovalka s področja socialne antropologije)

<sup>64</sup> Mislim, da manjkajo metapodatki. *Manjkajo vsebinske oznake, ali kako?* Ja. Vsebinske oznake, neki ključni in tako naprej, nek 'tagging' bi bil že dovolj. Je delno narejeno, ampak ni tako, da bi imel to človek v kosteh, da mora to narediti. *Če je področje slabo narejeno ali pa pomanjkljivo, kaj bi bilo treba narediti ali pa kdo bi bil odgovoren za to, da bi to prevzel. Ali vidite koga v hiši? Ali vidite možnost?* Zavest raziskovalcev mogoče ni problematična. Problem je, da stvari niso tako enostavne, da ne bi imeli občutka, da je to neko dodatno delo, za katerega niso prepričani, če se spleča, če bo zato kaj boljše. Da bi imeli v hiši koga, pa trenutno zagotovo ne. Če bi hoteli to urejati, bi morali neko službo vpeljati. (znanstvenik s področja biologije, NO6)

<sup>65</sup> Očitno fakultetna oblast nima moči nad računalniško službo. Ta ne želi zapletati svoje eksistence in zato ostajajo zadeve nespremenjene. Na repozitorij diplomskih nalog čakamo že desetletje, za vzdrževanje katalogov in podobnih podatkovnih zbirk ni prave volje, kaj šele navdušenja. Enako je z oddelčnimi spletnimi stranmi, ki jih že četrto leto potrpežljivo čakamo in vmes improviziramo, kot se pač znajdemo, vsak po svoje. Stanje je tu precej žalostno in v ostri opoziciji s pogumnim novim svetom, ki ga reprezentirajo javna spletišča, ki jih vzdržujejo zavzete skupnosti prostovoljcev. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>66</sup> Bioinformatika, genomika ima tako imenovane antologije in tam se da napisati marsikaj. To je zato, ker so v tem delno udeleženi tudi računalničarji in informatiki, ki to vidijo na ta način, da se da to nekako avtomatično preoblikovati. Tukaj je ta zavest najbolj prisotna, drugje pa niti ne. (znanstvenik s področja biologije, NO6)

<sup>67</sup> Jaz vidim, da res potrebujem mlajše kolegice, prav zaradi tega, ker ste že na začetku rekli. Starejši kolegi so navajeni na svoj način dela in tudi če bomo nekaj vzpostavili, jih je ponavadi zelo težko prepričati o teh stvareh, pokazati, da so smiselne, kaj so koristi ... Pri mlajših se mi pa zdi, da pravzaprav šele vstopajo v to polje in je nekoliko lažje. Tudi za izobraževanja in za vse te stvari so še zelo odprti. Čeprav po drugi strani pa potem včasih naletimo na težavo, ko si nekdo misli, kaj mi bo sedaj tam en mladi pamet solil, da moram jaz tukaj neko stvar izpolnit. (znanstvenik s področja socialnega dela, NO7)

## 8. ODGOVORNOST V ZVEZI Z REŠEVANJEM PROBLEMOV

---

Ustvarjalci in uporabniki podatkov se zavedajo, da so del širše strokovne in znanstvene javnosti, znotraj katere delujejo. Hkrati pa se zavedajo tudi obveznosti in omejitev, ki jih zastavlja financer.<sup>1</sup>

### 8.1 POTREBA PO KROVNI POLITIKI

Intervjuvanci so izpostavili potrebo po skupni politiki, s katero bi fazo iskanja, dostopa in ponovne rabe podatkov uredili od zgoraj. Zavedajo se, da bodo od spodaj navzgor sami težko dosegli spremembe, zato je zaželen krovni rešitev, dosežena v sporazumu s predstavniki različnih področij.<sup>2,3,4</sup>

Ena od sogovornic, ki je bila na vodstvenem položaju, pravi, da je skušala vplivati na zaposlene in vpeljati nove prakse, vendar sama v kolektivu težko kaj spremeni. Izrazila je potrebo po zunanji podpori, ne le infrastrukturni, ampak tudi pri ozaveščanju. Kot je že bilo omenjeno na drugih mestih, so nekatera področja neozaveščena v smislu digitalizacije, pa tudi deljenja in ponovne rabe, še zlasti tista, na katerih ne rokujejo s standardiziranimi podatki.

Pri politikah je mogoče govoriti o spodbujevalnih ali kaznovalnih.<sup>5</sup> Intervjuvanci vidijo spodbude predvsem v smislu, da bi morale biti dodatno delo tudi dodatno ovrednoteno.<sup>6</sup> Kjer je potrebno, bi bilo nujno zagotoviti finančna sredstva, ki bi delo sploh omogočila. Poskrbeti pa bi bilo treba tudi za dosledno navajanje podatkov, ena od možnih rešitev je po sistemu citiranja, kot je uveljavljeno za (znanstvene) publikacije.<sup>7</sup> Kaznovalno pa v smislu, da je s pravili določeno, da mora ustvarjalec poskrbeti tudi za dostopnost podatkov, na primer, da je zahteva vključena že v projektno dokumentacijo in da ga k temu zavezuje pogodba.<sup>8</sup>

### 8.2 USTREZNO VREDNOTENJE PODATKOVNIH VIROV

Ustrezno vrednotenje raziskovalnega dela predstavlja pomemben del raziskovalnega postopka, katerega del so tudi raziskovalni podatki. Raziskovalci se bolj ali manj strinjajo, da denarne nagrade niso ključna motivacija za delo in razmišljajo o sistemu, ki bi spodbujal kakovostno delo in ga tudi nagrajeval.<sup>9</sup> Realen ukrep, ki je v tujini že uveljavljen, je obvezno citiranje podatkovnih virov. Obvezno citiranje podatkovnih virov ustvarjalce spodbuja, da svoje gradivo delijo z drugimi, po načelu, da bo ponovna raba koristila tudi njim.<sup>10,11</sup>

Na področjih, na katerih raziskovalcem primanjkuje objav, so do dajanja podatkov še bolj restriktivni. Jasno je bilo izraženo pričakovanje, da bodo podatke oddajali, če bodo ti ustrezno ovrednoteni s točkami za znanstveno uspešnost, kar je v njihovem interesu.<sup>12,13</sup> Primer dobre prakse s tega področja so revije, ki zahtevajo, da ustvarjalci oddajo tudi podatke, ti so ovrednoteni tako, da se štejejo kot avtorstvo, iz česar izhajajo točke, ki so kazalec znanstvene uspešnosti raziskovalca.

Z uvedbo obveznega citiranja podatkov oziroma podatkovnih virov bo vrednost priznana tudi celemu področju ustvarjanja podatkov in zbirk.<sup>14</sup> Predstavlja pozitivno spodbudo za delo, ki med raziskovalci velja za manjvredno in težaško in ki se ga številni sploh ne lotijo, če to ni neizbežno.<sup>15</sup>

Intervjuvanec je slikovito popisal problem, ko podatkovne zbirke nimajo statusa ekvivalenta znanstvene dejavnosti:

*Trenutno je stanje teh zbirk podobno statusu strokovnih publikacij, slovarju in kar je podobnega, tisto se pa v naših krogih relativno nizko vrednoti, ker preprosto nima statusa znanstvene dejavnosti, ampak strokovne dejavnosti in je takoj polovičko manj vredno. To seveda ni spodbudno. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)*

Problem je v neosveščenosti znanstvene skupnosti, katere predstavniki določajo merila za vrednotenje in neusposobljenost knjižničarjev, ki slabo poznajo posebnosti kategorije 'znanstvenih podatkov'. Medtem ko bo v znanstveni skupnosti potrebno sprožiti razprave, ki bodo dvignile ozaveščenost glede problema in rešitev, pa bo pri knjižničarjih potrebno vpeljati ustrezno usposabljanje.

Vrednotenje podatkovnih virov po načelu publiciranja člankov ustvarjalce spodbudi, da že v fazi ustvarjanja upoštevajo možnosti ponovne rabe, s tem pa povečajo potencial gradiva in sami stimulirajo drugo rabo, kar dodatno poveča plačilo za vloženo delo. Na kratko, obveznost citiranja lahko prispeva k dvigu kakovosti gradiv.<sup>16</sup>

Zgled dobre prakse na tem področju je mednarodna organizacija DATACITE,<sup>i</sup> ki obvezno citiranje spodbuja z naslednjimi utemeljitvami:

- obvezno citiranje poenostavlja pregled nad ponovno rabo in preverjanjem istovetnosti podatkov,
- mogoče je slediti vplivu podatkov,
- ustvari mehanizem za prepoznavanje in vrednotenje prispevka ustvarjalcev podatkov.

Večje priznavanje vrednosti podatkov v okviru vrednotenja znanstvenega ustvarjanja je eden od najpomembnejših ukrepov in kot tak mora biti vključen v bodoče politike. Le-te morajo obsegati ustrezno strokovno vrednotenje vložene dela in nastalega gradiva.

#### **Okvir 8.1: P-R 15**

**PROBLEM:**

V slovenskem raziskovalnem okolju je trenutno mogoče govoriti o delovanju neformalnega sistema, po katerem konkurenčne prednosti izvirajo iz monopoliziranja in restriktivnega ravnanja s podatkovnimi viri, iz česar sledi, da podatkovni viri ostajajo neizkoriščen potencial.

**REŠITEV:**

Potrebno bo vzpostaviti ustrezno vrednotenje, ki bo na sistemski ravni spodbudilo sodelovanje in nadomestilo obstoječi neformalni sistem.

<sup>i</sup> DataCite, <http://www.datacite.org/>, 7.8.2012

### 8.3 FINANČNA SREDSTVA

Intervjuvanci se strinjajo, da bo odprti dostop do podatkov mogoče uresničiti samo, če bo to eksplicitna zahteva financerja.<sup>17</sup> Zaželena je sistemska rešitev na ravni objavljenih politik, ki se potem izvajajo v obliki pogodbenih obveznosti, sprejetih pri prijavi v obliki načrtov ravnanja s podatki in pri njihovem izvajanju v teku projekta in ob zaključku. Intervjuvanci so eksplicitno izrazili, da je zahteva realna le, če bo plačnik zagotovil finančna sredstva.<sup>18</sup> Potreba po dodatnih finančnih sredstvih, ki bi omogočila zbiranje, ureditev, hrambo, vzdrževanje in dostop, je ena od pomembnejših zahtev za uresničenje akcijskega načrta projekta odprtih podatkov. Namreč, raziskovalci se večinoma soočajo z razmerami, ko nimajo sredstev za hrambo, kaj šele, da bi lahko poskrbeli za dostop do teh podatkov ali zbirk. Zato prevladuje stališče, da bi sredstva za ureditev in hrambo podatkov morala biti zagotovljena že v projektu. Zahtevo utemeljujejo s tem, da gre za dodatno strokovno znanje, ki ga je treba osvojiti in ga prenesti v prakso, za vse akterje pa pomeni tudi dodatno delo, ki ga z obstoječimi kadri in kapacitetami ne zmorejo opraviti.<sup>19, 20</sup>

Financiranje je tudi nasploh problem. Če je raziskovalno delo odvisno od tega, ali bodo dobili projektna sredstva ali ne, ne morejo kontinuirano zbirati določenih podatkov. Če ni projekta, ni denarja in se zbiranje ustavi. Zlasti je to problem pri podatkih, katerih kakovost je pogojena s posodabljanjem.<sup>21, 22</sup> Pri oblikovanju finančne sheme projektov je potrebno upoštevati tudi posebnosti posameznih področij, npr. nekateri potrebujejo več tehnične opreme, drugi več kadrov.<sup>23, 24, 25</sup>

Iz tega sledi potreba, da morajo bodoče politike vključevati jasno finančno shemo, ki bo že v prijavi projekta predvidela dodatna sredstva. Tako bodo prijavitelji že v pripravi na raziskavo presojali vrednost podatkov za ponovno rabo in ovrednotili stroške priprave za hrambo in dostop, s čimer se bo povečal izkoristek potenciala, ki ga nosijo gradiva. Vključevanje dodatnih sredstev je model Data Management Planning-a, ki ga spodbujajo sodobne raziskovalne politike. Neustrezno ureditev pri načrtovanju sredstev in poročanju o porabi, ki administrativno obremenjuje projekte danes, bo potrebno reformirati v smeri načrtovanja in porabe za projektno določene naloge, katerih sestavni del so tudi naloge skrbi za podatke.<sup>26</sup>

Poleg bodočih projektov je potrebno upoštevati tudi stanje obstoječih gradiv. Sredstva bo potrebno zagotoviti tudi za že zbrana gradiva, ki jih je potrebno zavarovati pred zastarevanjem, uničenjem... To pa je pogosto povezano z uvajanjem postopkov digitalne hrambe, ki zahtevajo denarna sredstva, čas in strokovno znanje.

### 8.4 DOBER SISTEM IZVEDBE: HRAMBE IN DOSTOPA DO PODATKOV

Bodoče politike morajo zagotavljati učinkovitejše sisteme, ki bodo omogočali hrambo in dostop do podatkov. Če je sistem dober, deluje spodbujevalno, saj olajša delo. V nasprotnem primeru pa kaznovalno, saj pomeni dodatno obremenitev za raziskovalce.

Pričakovanje novih rešitev je bilo še posebej močno izraženo na tistih raziskovalnih področjih, na katerih vključevanje v mednarodne repozitorije ni uveljavljena praksa. Tak primer so podatkovni viri, ki imajo nacionalno vrednost, zaradi nacionalnih posebnosti pa je vključevanje v mednarodne zbirke omejeno in oteženo.<sup>27, 28, 29, 30</sup> Intervjuvanci so izrazili potrebo in pričakovanje po dodatnih službah oziroma servisih, ki bodo imeli pregled nad podatkovnimi viri in tudi rešitve za posebnosti posameznih področij (etika, anonimnost, pravice, zaščita).

Na področjih, kjer ne poznajo sistemskih rešitev, vlada dodatno nezaupanje o kakovosti podatkov.<sup>31, 32,</sup>

<sup>33</sup> Iz tega izhaja potreba po vzpostavitvi sistema, ki bo presegal subjektivne ocene in merila kakovosti ustvarjenega in vpeljal standardizirane rešitve, v katerih kakovost podatkov na pregleden način jamčijo

licence ali drugi standardi kakovosti.<sup>ii</sup> Razviti bi bilo treba ekspertna znanja in kader, ki presoja o kakovosti.<sup>34, 35, 36, 37</sup> Zgled so lahko mednarodne zbirke,<sup>iii</sup> ki jih uporabljajo naravoslovci.<sup>iv</sup>

Ena od možnosti je skupen sistem, ki bi pokrival vse univerze in inštitute v državi ter skrbel za organizacijsko in vsebinsko raven.<sup>38, 39</sup> Tega pa bi ustrezno dopolnjevala področna podatkovna središča. Oblikovanje središč zahteva usposabljanje kadra, specializiranega za hrambo in dostop.

Raziskovalci so izrazili potrebo po razumni rešitvi, ki bo omogočala enostavno hrambo, spodbujala sodelovanje ter ustvarjalce servisirala v primerih dilem, vezanih na vprašanja hrambe in dostopa do podatkovnih virov.

#### **Okvir 8.2: P-R 16**

##### **PROBLEM:**

Že danes so omejitve velikokrat preveč omejujoče za kakovostno delo, zato so se raziskovalci v pogovoru največkrat strinjali, da od financerja pričakujejo razumno rešitev, ki ne bo povečala obsega administrativnih in drugih birokratskih del ter z dodatnimi obremenitvami zmanjševala kakovost dela. Izražena je bila tudi potreba po krovni politiki, ustreznem vrednotenju dodatnega dela ter učinkovitem sistemu hrambe, ki bi zagotavljal dostopnost podatkov. Raziskovalci imajo kvečjemu parcialen pregled nad problematiko, ki je podkrepjen z lastnimi izkušnjami. Velikokrat ne poznajo uveljavljenih in utečenih rešitev, ki jih poznajo akterji na drugih področjih ali v tujini.

##### **REŠITEV:**

Bodoče politike bodo morale poskrbeti za sistematizacijo podatkovnega upravljanja, to pomeni, da bo že v fazi priprave projekta potrebno vključiti in upoštevati pomen hrambe in dostopa do podatkov. Zato je smiselno, da zahteva po spremembi in predlogi pridejo od zgoraj navzdol. Izraženo je bilo pričakovanje, da zahteve postavi financer, vendar v dialogu z raziskovalci.

<sup>ii</sup> Primerjaj pravkar izdano publikacijo Allining national approach to digital preservation (ur. McGovern 2012)

<sup>iii</sup> Npr. že prej omenjene zbirke: GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), UniProt (<http://www.uniprot.org/>), Protein Data Bank (<http://www.pdb.org/>), Amadeus (<http://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/Amadeus.aspx>).

<sup>iv</sup> Zanimiv prispevek na to temo sta na IASIST konferenci predstavila Downs in Chen (2012).

**Okvir 8.3: P-R 17**

**PROBLEM:**

Številna področja se soočajo s težavami, ki izvirajo iz vzpostavljanja ali nadgradnje sistema podatkovnih storitev (družboslovje, humanistika ter posamezna področja naravoslovja). Ustvarjalci in uporabniki se ukvarjajo s tehničnimi, pravnimi, administrativnimi in finančnimi težavami. Določena vprašanja pa so povsem vsebinske narave in so kot taka vezana na konkretno stroko oziroma področje.

**REŠITEV:**

Ker gre za isto fazo razvoja in podobne težave, je smotrno razmišljati o povezanem in koordiniranem iskanju rešitev. Pogled od spodaj navzgor kaže, da so nekatere rešitve lahko skupne za vse akterje (npr. administrativne, pravne, finančne), druge bodo morale biti bolj ali manj diferencirane, usklajene in prilagojene področnim problematikam. Pomembno je, da so rešitve usklajene s hkratnim potekom spremenjenih norm in običajev znotraj posameznih disciplin in da upoštevajo posebne zahteve in veljavne pristope v disciplinah, tako da se zagotovi široka podpora njihovem uresničevanju. Zato je pomembno spremljati in posodabljati uveljavljene standarde na posameznih področjih in uvajati spremembe skladno z razvojem strok.

## 8.5 CITATI

---

<sup>1</sup> Ampak če bo pa politika tako zaostрила stvari, ali pa dala samo neke direktive, ki potem v praksi padejo na vsej fronti, pa nima smisla. Treba je vedeti, kakšno je dejansko stanje v državi, zato je zelo pomembno, da se pravilno opredeli, kakšne so stvari na posameznih inštitucijah, kakšno je stanje, kako gledajo. Zato, da ne bodo neke trapaste direktive, ki jih potem ne bomo mogli izvrševati. (znanstvenik s področja geografije, NO8)

<sup>2</sup> Če smo vsi za, potem sem tudi jaz za. Če pa eden ja, eden ne, potem nima smisla. *Se pravi, da je treba urediti na nivoju enih politik in da se obveznosti jasno določijo ter da se potem nadzoruje pravila igre in tako naprej.* Konec koncev bi rekel, da je to omejevanje. Človek verjetno ne želi dela, ki ga je vložil, nekemu dati. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>3</sup> Mi poskušamo pokazati pozitivne koristi tega, da so stvari na voljo: enkrat boš ti potreboval podatke od nekoga drugega, drugič bo on potreboval tvoje, in se mi zdi, da tudi preko tega mogoče vseeno malo premikamo stvari. Bi bilo pa najbrž zelo dobrodošlo. Ampak najprej se moramo pri nas dogovoriti o čisto načelnih stvareh dogovorit, kako to sploh speljati. Ali bo to nek pravilnik, ali bodo neka navodila, ker so tudi glede tega dileme. (znanstvenik s področja socialnega dela, NO7)

<sup>4</sup> Zbirke so načeloma na voljo vsakemu, ki je taksonomsko kvalificiran, in z zbirkami seveda tudi podatki o zbirki oziroma o vzorcih. Dostopno je vsakemu, domačemu ali pa tujemu. Bom vesel vsakega, ki bo pripravljen del te zbirke vzeti in obdelati, s tem da je pogoj, da bo to naredil v nekem razumnem času, da bo to objavil in da pove, da je to naša zbirka. Seveda so pa drugačne politike, tudi drugačni politiki, drugačni značaji. Nekateri držijo zase, da bodo sami naredili, prej ali slej, ali pa bo naredil nek Slovenec in tako naprej. In to je konec koncev težko sankcionirati. (znanstvenik s področja favnistike jamskih nevretenčarjev, NO22)

<sup>5</sup> *V osnovi me zanima, ali je ta ideja, da agencija za raziskovanje predpiše, da se to kot obvezno nekam deponira, pa da se naredi dostopno drugim, bi bila ta ideja smiselna, pozitivna?* Ja, meni se zdi to načeloma v redu. Menim, da če od nekoga nekaj zahtevaš, moraš povedati, kaj bo od tega imel. Lahko je to kaznovalno, lahko je pa nekako spodbujevalno. Kaznovalno vidim v tem smislu, da ne moreš prijaviti projekta, če nisi nečesa obljubil in zagotovil, toda to pomeni dodatno delo. Raziskovalci bodo to videli kot dodatno delo, kot po nepotrebnem naloženo delo in tukaj najbrž tega ni mogoče prijetno speljati. Spodbujevalno je lahko v smislu referenc. Se mi zdi, da bi bila potrebna zelo široka akcija, ki bi delovala kot korenček, zaradi katerega bi raziskovalci to raje naredili in dobro naredili. Če bi recimo ARRS nekako predpisoval, bi moral imeti dober sistem, ki bi omogočal izvedbo. Sistem, s katerim bi se to dalo narediti brez kakšnih posebnih težav ali posebnega, groznega dodatnega dela. (Andrej Blejec, znanstvenik s področja biologije)

<sup>6</sup> Zagotovo bodo problemi, ker pri nas je odpor do vsake nove stvari, vsakega novega papirja, novega dokumenta, ki ga uvedemo, da so stvari malo bolj pregledne in da imamo nek vpogled. /.../ *Že samo zaradi tega, ker so raziskovalci tako preobremenjeni?* Ja. Vsaka taka stvar, tudi če je mogoče na dolgi rok zelo koristna, se ti na prvi pogled zdi birokracija. In pri nas so zagotovo ti odpori. *Se pravi, bi bilo treba vzpostaviti tudi samo kulturo tega.* (znanstvenik s področja socialnega dela, NO7)

<sup>7</sup> Možno, da bi raziskovalci na začetku imeli zadržke, ker bi želeli prihraniti čim več in potem zelo pogosto iz istih podatkov pride do več objav, torej ne bi želeli, da drugi pridejo do njih. *V medicini je*



---

*torej precej prisoten občutek za lastništvo nad podatki, na način, da bi vsak rad čim več zase izkoristil.* Ja. Mislim, da je zelo odvisno od politike, ki jo izvajajo same revije in tudi viri financiranja, ker tudi to je pomembno. Če ima tisti, ki financira raziskave, svoje zahteve, potem postane to avtomatično. (znanstvenik s področja medicinskih vez in statistike, NO15)

<sup>8</sup> Jaz bi rekel tako, ne da bi tukaj morala biti spodbuda na tej strani, ne korenček ampak palica, za nekatere stvari čisto enostavno palica, ki bi zahtevala. /.../ Pač tako kot imaš po zakonu dostop do javnih podatkov. Ti nekdo zateži, da do nečesa prideš, tako bi bil lahko nek zakon o dostopu do javnih in znanstvenih podatkov in bi pravzaprav vsak lahko prišel do njih. Bi prebral v nekem članku, da je nekdo npr. strl nek nosilec in bi hotel videti surove podatke, da bi videl, na čem so delali. Se pravi, vse je nekje tam shranjeno.. (znanstvenik s področja gradbene informatike, NO20)

<sup>9</sup> Pri raziskovalcu je problem tudi, da ga boste težko motivirali z denarjem. Mi imamo problem v tem, ker je treba izpolnjevati habilitacijske pogoje. In sploh v ekonomiji in poslovnih vedah je tako, če imaš podatke, tudi če nimaš nekakšne dobre metodologije zadaj, lahko objaviš v SSCI reviji, včasih se hecamo, tudi če imaš samo povprečje. Resno, samo povprečja. Če bi to dal na širši dostop, ta možnost odpade. Ampak potem je tako. Zbiraš podatke in ko jih zbereš, moraš čim hitreje objaviti, ker če bodo podatki javni, pomeni, da bo že deset ljudi tam čakalo, da jih bodo dobili. /.../ V naravoslovnih znanostih imajo to drugače rešeno. Na ta način imajo laboratorije. Tisti, ki vodi laboratorij, je enostavno na vseh 'paper-jih' zraven, kar koli se dela v tistem laboratoriju. Če je za raziskovalce bolj pomembno, da se pišejo članki kot pa neka monetarna nagrada, je potem tisti, ki je vodja projekta lahko povsod zraven, kjer se uporabljajo ti podatki. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>10</sup> Predno se to sme objaviti kot uraden rezultat projekta, je veriga zelo kompleksna in jasno definirana. Dodaten problem, ki se ga začnejo raziskovalci zavedati šele, ko mislijo, da so kaj našli, pa je to, da se pri tako velikih projektih vsi zavedamo, da mora zelo veliko ljudi zelo trdo delati, da pridejo do rezultata, zelo veliko let in potem se bo nekemu nekje z malo sreče posrečilo, da bo prvi, ki bo v analizi zagledal rezultat. Nesprejemljivo bi bilo, da bi rezultate objavili z imenom tistega, ki bi jih prvi zagledal. Zato so tudi pravila, kako se to objavlja, in kontrola, da ne bi slučajno kaj pricurjalo. In potem se izkaže, da se ocenjuje ne sam raziskovalec, v bistvu se ne ocenjuje raziskovalno delo, ampak to trdo fizično delo - izdelava detektorja in nadzora nad software-jem in tako naprej. In s tem se poveča količnik prisotnosti na avtorstvu članka. (Jan Jona Javoršek, raziskovalec s področja eksperimentalne fizike osnovnih delcev)

<sup>11</sup> Po mojem je logično, da si vsi želimo, da drugi berejo naše članke in uporabljajo naše programe oz. raziskovalna orodja. Sicer so ljudje, ki morda mislijo, da sami še niso dovolj izkoristili vsega, kar se je dalo izkoristiti iz tega programa in potem mislijo, "Glej, jaz sem se trudil pet let razviti programsko orodje, če ga zdaj dam prosto dostopnega, bodo drugi to izkoristili bolje, kot sem jaz, in hitreje prišli do novih rezultatov in bodo v kratkem času publicirali več člankov kot jaz." Ampak to dolgoročno ni dober pristop. Rezultati raziskovalnega dela so dejansko koristni le, ko jih drugi uporabljajo. Če daste vaše rezultate prosto dostopne vsem, bodo vaše delo spoznali in vas citirali tudi ljudje, ki jih nikoli niste srečali in z njimi ne sodelujete na noben način. Pri nas je recimo pomemben kriterij za izvolitev v naziv število citatov. Sicer sem tudi sama imela izkušnje, da sem nekemu nekaj dala, potem je pa napisal članek in ne le da se mi ni zahvalil, celo vira ni citiral. Tako pač je v življenju, na koncu smo vsi ljudje razdeljeni po Gaussovi krivulji, neodvisno od tega ali smo znanstveniki, duhovniki, pravniki ali ostali. (Nedjeljka Žagar, znanstvenica s področja meteorologije)

---

<sup>12</sup> In tukaj jaz vidim en velik problem citiranja podatkov. V naravoslovnih znanosti je tako, da so podatki omenjeni v zahvali. Hvala temu in temu, da nam je dovolil uporabiti te in te podatke. Ampak to je samo v zahvali, ni pa evidentirano kot referenca. Jaz vidim to kot problem, morala bi biti referenca, na katere se nekdo sklicuje, kot da bi šlo za bibliografske vire, da bi se sklicevali na podatkovni vir na podoben način kot na referenco. /.../ In bi bil kot Science Citation Index, še Data Citation Index. (znanstvenik s področje biologije, NO6)

<sup>13</sup> Ponavadi vsi zbiramo, pa potem vsi tudi sodelujemo pri nekem članku. Če bi pa nekdo prišel do mene, pa bi hotel imeti te podatke, mu rečem, da lahko kaj skupaj narediva, podatkov pa ne dam. A me razumete? Ne tako, da bi jaz samo dala podatke, on pa bi naredil in mene zraven napisal. Veste, pri teh stvareh lahko pride tudi do zlorabe podatkov. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>14</sup> Ker to bi bil v bistvu kredit, ki bi štel. Pri teh naših raziskavah je tako, da je veliko terenskega dela. Eni ljudje so za teren, drugi niso za teren. In če je nekdo specialist za teren, bi lahko nabiral za pet raziskovalcev in delil stvari. Tu imamo takšne situacije. Druga stvar so laboratoriji, ki morajo biti tehnično opremljeni. Te tehnične opreme nima vsak in je predraga. Če bi lahko rekel, ti podatki so zbrani, pripravljeni, eksperiment so pa izvedli tam in tam, bi bila to referenca za tisti eksperimentalni laboratorij, ki dela toliko in toliko kvalitetno, bi to po moje spodbudilo tudi izmenjavo. (Andrej Blejec, znanstvenik s področja biologije)

<sup>15</sup> S tem se ljudje v glavnem ne želijo ukvarjati. Po moje je to malo karakterno pogojeno. Kdor rad doma pomiva, čisti, pospravlja, ta se bo rad lotil tudi postavljanja podatkovne zbirke. Kdo drug, ki se ima za velikega duha, za genija, se bo pa ukvarjal samo z velikimi sintezami. Drugi način je udobnejši, zato težko dobim med kolegi podporo za večja podatkovna infrastrukturna podjetja. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>16</sup> Recimo tudi to, kar ste takrat omenjali, da se sedaj šteje v COBISS-u. Pri nas imamo kar težave s tem, da imamo premalo točk. Jaz sem že omenila našim kolegom po tistem najinem pogovoru, ampak pri nas je potem tako, da vedno rabimo precej časa, da se premakne. *Saj to je povsod tako. Se pravi, da bi se to kot referenca upoštevalo v bibliografiji in bi potem to citirali drugi avtorji.* To bi bilo pri nas zelo pomembna spodbuda, ker imamo tukaj šibko točko. (znanstvenik s področja socialnega dela, NO7)

<sup>17</sup> Razen če bo neke vrsta prisila, da bo to v pogodbi o financiranju tudi eksplicitno pisalo, da je treba zagotoviti, da bodo podatki na določen način dostopni takoj in da jih je treba tudi ustrezno pripraviti. *Se vam zdi ta model realističen?* Ne, meni se ne zdi. Ker poznam naše raziskovalce in čim je nekaj takega, to že samo administriranje in tako. *Da bi sami videli korist od tega, pa ne verjamete?* Imam malo skepse do tega. Saj dostikrat slišimo, da je neka skupina objavila nek rezultat na enem področju in so neke določene zadeve naredili. In da je neka druga skupina enako zadevo preizkusila, če ne drugega ponovljivost takih nekih eksperimentov, in je prišla do takih in takih rezultatov. Tukaj je lahko vmes veliko zadev, da eni so delali pač na eni mašini, drugi na drugačni mašini in da so potem neka odstopanja določena. V vsakem primeru pa tisti, ki so na nekih mednarodnih kolaboracijah, tam pa to tako ali tako mora biti. Recimo kot smo omenili CERN, tam je na tisoče ljudi napisanih na nekem članku, to se pravi, da neka izmenjava podatkov že mora biti. Ker ena skupina dela nekaj, druga si mora te zadeve pač izmenjevat. Isto je na evropskih projektih, se mi zdi, da tam mora biti, ker je več partnerjev in ... (raziskovalec s področja naravoslovnih ved, NO3)

<sup>18</sup> Financerji bi morali direktno v okviru pogodbe zahtevati, če bo ta potreba, in če bo ta model in koncept tako, na kak način se protokolirajo izvorni podatki raziskav. Ali morajo biti dostopni na internetu, neki spletni strani, ali v knjižnici, ali v obliki nekega delovnega poročila, ali pa protokola meritev. Ampak to se mi zdi, ko ste ravnokar omenili, da bi bilo dodatno financirano, to bi bilo boljše sprejeto. Se mi zdi, ker pri nas raziskovalci, dobro tisto pisanje projektov, tisto še nekak, ker vedo, da za denar prosijo. Ampak ko projekt enkrat dobijo, ko so tista letna poročila, pa je na nek način vse to potem odveč. Ker ti zasleduješ ciljni rezultat in potem razmišljaš že naprej. Vse ostalo vidijo kot balast. (raziskovalec s področja naravoslovnih ved, NO3)

<sup>19</sup> Načrti seveda so, je pa odvisno od projektov, od financiranja. In glede na to, da je cela raziskovalna ura tako nizka in da je delež materialnih stroškov tako nizek, nam ponavadi ne ostane nič. Ko gremo zbirat podatke, nam ne ostane nič, da bi lahko na projektu zaposlili še kakšno tehnično pomoč, ali pa študentsko pomoč, ki bi nam to uredila. Saj tukaj gre potem na koncu samo za ureditev neke dokumentacije, kaj ti podatki pomenijo. V bistvu to ni kakšno hudo zahtevno delo. Drugače pa mislim, da bi bilo treba ob koncu projekta vsaj dva do tri tedne časa nameniti temu, da se podatki shranijo v ustrezni obliki, da lahko vsak, ki gre kasneje gledat te podatke tudi vidi, kaj je bilo. In mi ponavadi imamo tako, da ima tista oseba, ki je odgovorna za izpeljavo projekta, vprašalnike, ima podatke in jih seveda posreduje, saj ni problema, če kdo drug rabi te podatke. Je pa seveda vprašanje, v kakšni obliki je to vnešeno v baze in tako naprej. Manjka prevod tega, kaj pomeni številka v neki tabeli. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>20</sup> Sem za sistemske rešitve. Pri nas je v knjižnici največja težava to, da knjižničarke nimajo za to naraščajočo dejavnost, za popisovanje bibliografij ljudi, ki na inštituciji delajo, nobenega dodatka in delo komaj zmorejo. Naša knjižnica zato še zdaj nima vseh listkovnih katalogov v elektronski obliki. Tako je vedno zadrega, kadar pridemo s kakšnimi novimi zahtevami. Ta problem bi bilo treba rešiti v okviru promocije nove pismenosti na Slovenskem, izgradnje digitalne infrastrukture ali kakor koli drugače, saj v okviru rednega delovanja služb to preprosto ne gre. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>21</sup> Kdo bo zagotovil sredstva za to, da zadevo uredi. To je običajno stvar, ki je nekje po predalih, vprašanje v kakšnem stanju, ali je potrebna digitalizacija. Obdobje, ko so se podatki pri nas digitalizirali, ko so bila sredstva za te stvari, je mimo. Če nekdo individualno na tistem ne dela, potem to zastara, ostane nekje neobjavljeno, po predalih. Niti ne vemo, koliko je tega. /.../ Podatki, pravite, ki so nekje pri nekih sodelavcih v predalih, če jaz nimam denarja, da bi jim plačala za njihovo delo, od njih ne morem zahtevati, da to gradivo ustrezno uredijo. /.../ Najbolj so pri nas urejeni podatki, za katere imamo zagotovljeno sprotno financiranje. (znanstvenik s področja geografije, NO8)

<sup>22</sup> Naše pogodbe so vezane na projekte in če določeni inštituti nimajo projekta, morajo ljudi odpustiti. In s tem smo prisiljeni, da delamo točno tisto, za kar smo dobili projekt. Seveda ne dobimo zmeraj projekta, s katerim bi lahko zagotavljali ažuriranje nekih podatkov, ki smo jih delali pred desetimi leti, največkrat vse delamo zmeraj nekaj na novo. In to je naš največji problem, da ne moremo kontinuirano delati ene stvari, ker smo projektno vezani in pač delamo dve leti to, tri leta pa drugo. (znanstvenik s področja geografije, NO8)

<sup>23</sup> Sem pisal program in sem napisal, da to je to, kar potrebujemo, to je naš laboratorij pri družboslovju. Podatki in orodja so naš laboratorij, tega enostavno nihče noče sprejeti oziroma razumeti, samo nekateri to razumejo. Medtem ko kemiki, fiziki, matematiki lahko imajo laboratorij, pa imajo C in D ure za pokrivanje amortizacije in ne vem kaj še, ni noben problem. To je nekaj zelo

---

pomembnega, če nabavijo raziskovalno opremo. Mi pa sploh ne pridemo zraven, ker nimamo tako velikih zadev. Ampak za nas je razvoj enega orodja, recimo pa tudi za SHARE. Kdo nam to plača? Plača nam, kdor hoče imeti rezultate. Mi pa preko rezultatov razvijamo orodja. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>24</sup> Mislim, da vi še bolj ali pa enako razumete, nujnost razpolaganja uporabe podatkov. Ker to je za nas konec koncev laboratorij. Kar so za kemike epruvete, pa vse tiste njihove aparature, to je za nas to in to celo tako, da je treba razvit tako podatke in seveda potem orodja. To ni stvar, ki se jo kupi na trgu. Ko mi govori in sprašuje o nabavi opreme nad 50 tisoč evrov, jaz jim lahko povem pet orodij, ki so stala več kot 50 tisoč in kako jih čem nabavit. Jaz jih lahko samo razvijem ali pa kupim razvoj teh orodij. Tega sploh ni mogoče vpisati. Ali lahko vpišem? Ne morem. Tudi potem če gledate tip projekta, A,B,C,D. Mi ne moremo priti v B, tudi če smo ne vem kaj. Razen če bi imeli neko terensko delo, recimo razvoj tega orodja, to je A projekt, to ni nič. V bistvu mi to razvijemo v sobi. (Boris Majcen, znanstvenik s področja ekonomije)

<sup>25</sup> Jaz lahko rečem, da pri nas tega še ni. Tudi princip na Stefanu je nekako tak, da smo podporne dejavnosti. To je plačano iz projektnega denarja in zato ljudje, ki dobijo projekte, to vidijo kot strošek. Zato moramo paziti, da ne zapravljamo preveč. In v skladu s človeško naravo delamo samo tisto, kar hočejo od nas. Z drugimi besedami, mi se bomo v to spravili, ko bo to rabil bodisi financer bodisi raziskovalci. Do takrat nas to ne zanima. Ker delamo na nekem minimalnem osebju s toliko stroški, kolikor je potrebno za to, kar trenutno ljudje rabijo. (vodja znanstvenoinformacijskega centra, NO5)

<sup>26</sup> Treba bo tudi najti neke rešitve, ki bodo dobre za vse. Jaz osebno sem po dvajsetih ali pa devetnajstih letih, kar delam v tej raziskovalni dejavnosti, razočarana, predvsem zato, ker se mi zdi, da smo včasih imeli več časa za neko res konkretno delo, sedaj pa samo še administriramo. Res je, da imam sedaj kot pomočnica drugo funkcijo, kaj pa naj drugega pričakujem. Ampak že prej, ko sem bila na inštitutu polno zaposlena, nisem delala nič drugega, kot pisala poročila, finančna, taka in drugačna, za ta projekt, za drug projekt. Vsebinsko smo na projektu komaj kaj delali, poročali smo pa kar naprej in samo neke formularje izpolnjevali. Se mi zdi, da gredo stvari samo na slabše, namesto da bi denar šel v pravo smer. Dokazuješ za vsakih 30 centov in se pogajaš ali je bil ta DDV tak, pa odbitni, pa ni odbitni, pa ne vem kaj vse. Da ne govorim, da mi porabimo eno uro časa, zadaj je pa cela gora revizorjev, ki jih država plačuje, da potem vse te številke desetkrat pregledajo. In me skrbi, da bomo tudi pri teh stvareh šli v podobne, nespametne skrajnosti, temu se moramo izogniti. Ker ne nazadnje, mi smo raziskovalci, naš interes mora biti, da bodo stvari tako oblikovane, da nam bo čas prihranjen, ne pa, da bomo potem še ne vem kakšna poročila delali. (znanstvenik s področja geografije, NO8)

<sup>27</sup> Mi smo zelo nepovezani, kar je značilno za humanistiko v celoti. Osnovna pomanjkljivost humanistike je, da je nacionalno omejena. To ne velja samo za slovensko, ampak tudi za druge. Zato so nacionalne filologije označene kot konzervativne stroke, kar nam seveda ni prav in želimo našo podobo spremeniti. Sam opozarjam na podobne projekte v tujini v zadnjih 10 letih. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>28</sup> Eksotični jeziki, kot je slovenščina, so za tuje podatkovne zbirke interesantni samo zato, da se lahko pohvalijo s številom jezikov, prilagodijo se jim pa ne. Za zbirke slovenskih besedil moramo poskrbeti sami. Nihče drug tega ne bo počel, če ne bomo sami. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>29</sup> Tukaj imam jaz vseeno nekaj zadržkov povezanih s tem, kar sem že prej rekla, sploh zato, ker je

---

Slovenija tako majhna. Mi se marsikdaj ukvarjamo z življenjskimi zgodbami ali pa s situacijami posameznih primerov in v taki situaciji bi vedno moral biti nek postopek, nek protokol, kdaj je nujno, da so ti podatki predani, in kateri so zadržki, ko podatki ne bi smeli biti širše dostopni. (znanstvenik s področja socialnega dela, NO7)

<sup>30</sup> Pomagal sem, da so svoje besedilne zbirke dopolnili z našimi teksti, ampak to je bilo bolj simbolično sodelovanje. Pravo sodelovanje si predstavljam tako, da bi vsa orodja, ki jih na spletu uspešno uporabljamo za angleščino, lahko uporabili tudi na slovenskih besedilih. Tu pa se najdemo v zadregi, ker so orodja, ki omogočajo delo z besedili v zbirkah, ponavadi nacionalno specifična. Če že obvladajo preglase, pa č, ž, š-jev ne in ti pri č-ju besedo prelomijo na dvoje in naredijo potem napačne frekvence. (Miran Hladnik, znanstvenik s področja slovenistične literarne vede)

<sup>31</sup> Pri nas gre vedno na zaupanje. Saj veste, da so bili razni, tudi neki korejski znanstveniki, ki so klonirali ovce, a jih niso. Neka določena stopnja zaupanja mora biti vedno. V integriteto človeka, ki to dela in potem pa po njegovem slovesu v skupnosti ugotoviš, ali lahko dejansko podatkom zaupaš ali ne. Če je nekdo nov, pač ne moreš vedeti. Če je nekdo že uveljavljen raziskovalec, s svojim ugledom stoji za tem, kar je izmeril. (raziskovalec s področja naravoslovnih ved, NO3)

<sup>32</sup> Bolj ko greste v obdelane podatke, bolj pride ta problem, koliko zaupaš človeku, ki jih je obdeloval. (vodja znanstvenoinformacijskega centra, NO3)

<sup>33</sup> Torej vključili bomo vse, tukaj ni neke izbire. Ampak komu lahko zaupaš, komu ne? To je pa tako, preden daš nekemu material v obdelavo, se boš skušal prepričati, če ima za seboj neko kilometrino. Ali je že kaj podobnega delal in solidno naredil. To po literaturi ugotavljaš. (znanstvenik s področja favnistike jamskih nevretenčarjev, NO22)

<sup>34</sup> Veste, samo problem je v tem, če mi ne moremo do teh podatkov, pa moramo plačevati veliko denarja. Potem pa na koncu ugotoviš, da so slabi podatki zbrani. Kaj ti preostane drugega, kot da greš sam zbirat podatke. Potem pa te nekdo vpraša, ali boš dal te podatke. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>35</sup> Še nekaj drugega sem hotela reči, kar se mi zdi hud problem, kar se tiče odprtosti podatkov, zlasti za nekatere podatke, na primer že za intervjuje. Recimo v raznih dokumentacijskih centrih je zbranih veliko intervjujev, v Sloveniji pa vprašanje etike še ni dovolj dobro izdelano, pa tudi merila, kaj dajat 'on-line', kaj dajat v javni prostor. Ponavadi so ti intervjuji pridobljeni za lastno uporabo in ne gre za to, da bi transkripte dajali 'on-line'. V antropologiji zelo cenimo sugestivnost posameznega raziskovalca, zato ker podatkov ne dajemo v javnost, ampak zaupamo, da jih vsak posameznik zna interpretirati. (Nina Vodopivec, raziskovalka s področja socialne antropologije)

<sup>36</sup> Je pa zopet problem, kako zagotoviti, da v bistvu govoriš o istem fenomenu. Ker je tukaj toliko specifičnosti, sploh na družboslovnem področju. Lažje to naredijo naravoslovci in tehniki kot družboslovci. (Polona Domadenik, znanstvenica s področja ekonomije)

<sup>37</sup> To ste sedaj v bistvu dobro načeli. Jaz osebno malo dvomim o tem, da bi kdor koli imel interes dobivati podatke od nekoga drugega. Jaz bi imel precej dvomov vzeti od nekoga podatke. Naši raziskovalci, kakor vidim takole na inštitutu, bi obvezno arhiviranje ali pa pripravo podatkov za splošno dostopnost smatrali kot neko dodatno obligacijo in neko dodatno administriranje. Razen če bi bila na voljo, kot ste rekli, dodatna sredstva, mogoče mesec ali kaj takega, da je tudi plačano, da se ti

---

podatki pripravijo v tako obliko, da so širše dostopni. Zakaj jaz osebno dvomim? Ljudje so različni, eni so zelo hitri, delno tudi majčkeno površni, eni so pa zelo natančni in precizni. Če je odčitek recimo čas, to lahko tudi stroji avtomatsko naredijo, ampak če igra čas neko vlogo, tudi manjši časi, pa nekdo to na deset minut naredi, moral bi pa na minuto natančno, pa je temu zadevno potem točnost teh rezultatov plus minus 5 %. Ali pa npr. podobno glede temperature, ali kar koli drugega. In če ti dvomiš, ne veš, kakšen je tisti človek, če reče, ah, saj je tudi tako dobro, pa odčita na pet minut... Sem človek, ki zaupam sebi in sem zelo dvomljiv oziroma res bi mogel točno vedeti, kako je kdo to, a je res to čisto tako ... Skratka če bi bila zahteva, da mora biti v okviru nekega projekta vzpostavljen sistem javnega dostopa podatkov, za katerega moraš sam poskrbeti, da je ustrezno opremljen z vsemi ustreznimi komentarji in tako naprej, seveda to zahteva nek dodaten čas. Če jih moraš v taki obliki predstaviti, se mi zdi, da je, če je potem ustrezno finančno tudi ovrednoteno, da bi to bilo mogoče manj. Ker drugače ljudje potem smatrajo tudi neko dodatno administriranje, ki odkrito povedano, si težko predstavljam, da bi nekdo rekel, čakaj, kje bi jaz sedaj te izvirne podatke dobil od tega. (raziskovalec s področja naravoslovnih ved, NO3)

<sup>38</sup> Kar se mene tiče, si želim en nacionalni sistem, v katerega bi se kot eden od številnih drugih uporabnikov tudi mi vključili. Če je to mogoče, pa ne vem. V glavnem želim si en nacionalni sistem, ki bi pokrival vse univerze in vse inštitute v državi. Na ta način bi se nekako znebili, otresli strahov glede organizacije. Mene resno skrbi. (vodja knjižnice, NO18)

<sup>39</sup> Mislim, da če bi bil vzpostavljen en tak nacionalni sistem, bi verjetno moralo zraven sodelovati nekaj raziskovalcev iz vsakega področja. Vsaj recimo na nivoju naravoslovja, nek eminenten širok raziskovalec, ki bi bil sposoben ocenjevati te stvari. Mogoče po področjih, biotehnika, naravoslovje, ne vem, kaj še obstaja. Vsaj, če ne celo bolj podrobno, mogoče bi bilo še to preširoko. (vodja knjižnice, NO18)

# ZAKLJUČKI NACIONALNE ANALIZE POLOŽAJA ODPRTIH ZNANSTVENIH PODATKOV

---

Predstavitev analize pogojev in možnosti za vzpostavitev nacionalnega sistema zagotavljanja odprtega dostopa do raziskovalnih podatkov smo začeli s prikazom aktualnosti te tematike v strokovni in poljudni literaturi. O aktualnosti in novosti, inovativnem potencialu možnosti odprtega izkoriščanja raziskovalnih podatkov govorijo tudi že omenjeni izrazi kot je 'četrti paradigma' (ur. Hey in drugi 2009), 'digitalna humanistika', 'veliki podatki', ipd. Vendarle gre bolj za spremenjene predstave o nas samih in načinu, kako dojemamo vlogo podatkov v prostem dostopu v dobi interneta kot pa za povsem drugačen način ustvarjanja znanstvenih odkritij. Se pravi, da ne začenjamo z nič, ampak kot smo izpostavili v našem pristopu, iščemo in se opiramo na že obstoječe nastavke in dejavnosti, ki so v večini primerov tesno vpete v okolje ustvarjanja podatkov. Gre torej za to, da izboljšave in novosti v to okolje vtremo čim bolj nemoteče in na način, da bodo imeli vsi vpleteni dejansko nekaj pozitivnega od tega. Izboljšave, četudi niso revolucionarne, pa tako kot vsaka sprememba, potrebujejo obdobje prilagajanja in usposabljanja, preden bodo sprejete in zaživele.

Že v pristopu k projektu smo omenili, da naš namen ni zbrati seznam ukrepov, ki naj bi jih Ministrstvo uvedlo od zgoraj. Izhodišče, da že v prvi fazi projekta, katere rezultate tu predstavljamo, namenimo posebno pozornost ugotavljanju, kje in kako različni nosilci vlog v znanosti vidijo probleme in možne rešitve, izhaja iz strateške usmeritve, da je predvsem potrebno ustvariti možnosti in okoliščine, v katerih se bodo vsi vpleteni na svojih področjih, ki jih tudi najbolj poznajo, lotili ukvarjanja s problemi odprtega dostopa do podatkov. In sicer po korakih, od tega, da se problemov sploh začnejo zavedati, da najdejo primerne kapacitete in organizacijske možnosti, da se z njimi ukvarjajo, ter do tega da prispevajo svoje pobude tudi za izgradnjo institucionalnega okolja z ustreznimi politikami in načeli, ki spodbujajo in tudi zavezujejo k ravnanju v prid odprtega dostopa. Tako so tudi rešitve, ki smo jih izluščili iz te prve faze projekta, kljub temu da so oprte na študij literature in na obsežne ter poglobljene intervjuje, samo predlogi, dani v premislek in nadaljnjo dopolnitev vsem, ki jih zadevajo. Na nosilcih znotraj posameznih področij je, da se prepoznajo v okrepljenih vlogah v dejavnostih povezanih s podatki, da se usposobijo ter se vzpostavijo kot pomemben člen s svojo dejavnostjo znotraj posameznih disciplinarnih področij, katerih potrebe in možnosti so sposobni uvideti zaradi vsebinske bližine. Ministrstvo in drugi oblikovalci politik lahko dajejo svoje pobude, obenem pa je njihova vloga, da se odzivajo na izražene pobude z nivoja znanstvenih skupnosti in omogočijo organizacijske in institucionalne okvire za njihovo uresničevanje. Pri tem učenju seveda zelo pomagajo zgledi iz bolj razvitih okolij in meddisciplinarna z gledovanja, kadar gre za sorodne probleme, kar pa zopet terja prilagajanje in aktivno prevzemanje, se pravi, nek dodaten angažma. Bistveno pri tem je, da se vzpostavi okolje, v katerem bodo prevladale koristi tako za znanstveno skupnost v smislu napredovanja spoznanj, kot v smislu nagrad za vložene dodatne napore (ustrezne priprave podatkov) ter žrtve (npr. izgubljanje monopola nad ustvarjenimi podatki) za izboljšanje dostopa do raziskovalnih podatkov.

Glavna ovira, ki smo jo identificirali že iz pregleda sorodnih študij drugje, je lahko kapaciteta in dobro delovanje usposobljenih podpornih dejavnosti storitev, povezanih z organizacijo dostopa do podatkov. Pri nas je še v zametkih, v razvitih okoljih pa je že dobro vzpostavljen, sistem institucionalnih repozitorijev objav v obliki znanstvenih besedil, ki se vse bolj usmerjajo tudi v usposabljanje za

podporo bolj kompleksnim oblikam znanstvenih digitalnih objektov, med katere sodijo tudi raziskovalni podatki. Glede na to da segment institucionalnih digitalnih repozitorijev pri nas ni tako razvit, bo potrebno institucionalno podporno dejavnost za skrb za podatke razvijati in vzpostavljati drugje. Predvsem to lahko pomeni precejšnje spremembe v nalogah knjižničarjev in njihovo večjo vključenost v vse faze ustvarjanja podatkov ter poteka raziskovalnih projektov na inštitutih in univerzah. Pomembna je lahko tudi izgradnja centraliziranih storitev nad nivojem raziskovalnih ustanov (npr. na področju usposobljenosti in orodij za podporo digitalne hrambe). Ključno pa je vzpostavljanje in podporno delovanje specializiranih področnih storitev nacionalnih arhivov podatkov (v povezavi s sorodnimi evropskimi in svetovnimi). Zgled za slednje so nekatere že delujoče povezave na različnih področjih, od družboslovja in humanistike do naravoslovja in biologije. Obenem je pomembno, da se delujoči znanstveniki usposobijo v vlogah podatkovnih znanstvenikov, za kar je potrebna dodatna spodbuda za zagotavljanje enakopravnosti v napredovanju in ugledu.

### **USTVARJANJE PODATKOV**

Spoznanja, ki smo jih oblikovali s pregledom intervjujev, se začnejo z ugotovitvami, da so podatki lahko pojmovani zelo različno glede na kontekst uporabe, da jih lahko klasificiramo na različne načine glede na tipe ter da se tudi v fazi ustvarjanja generirajo zelo raznoliki formati in zapisi podatkov. Posebej bi izpostavili pomembnost razlikovanja med referenčnimi zbirkami, ki običajno nastajajo s kolektivnimi prispevki in predstavljajo celovit pregled na določenem področju, ter raznolikimi raziskovalnimi podatki, ki nastajajo znotraj raziskovalnih projektov v teku raziskovanja. Prav na slednje se pogosto pozablja in je zanje slabo poskrbljeno, zato jim je potrebno posvetiti enako mero pozornosti in jih ovrednotiti z vidika nadaljnje uporabnosti.

Pomembne so ugotovitve, da so za znanost zanimivi tudi podatki, ki nastajajo drugod, zlasti tisti, ki nastajajo v javnih ustanovah, bi lahko s svojo večjo dostopnostjo in možnostmi za dodelave in nadaljnje razširjanje v obogatenih oblikah, predstavljali zgled tudi za ožje znanstveno ustvarjanje. To sicer ni predmet tega projekta, je pa področje, ki je vsebinsko tesno povezano z njegovimi cilji, namreč vprašanje 'Odprtih javnih podatkov' (*Open government data*).

Enako je pomembno, da se pozornost posveti ne samo '*big data*', ampak prevladujočemu ustvarjanju malih, po obliki in vsebini lahko zelo kompleksnih vrst podatkov. Poskrbeti je potrebno za možnosti izmenjave dobrih praks med področji, obenem pa je jasno, da zelo uniformirane rešitve in robustni pristopi s področja velikih podatkov, ne bodo vselej uporabni za kompleksno problematiko malih podatkov. (Okvir 5.1: P-R1)

V vseh fazah ravnanja s podatki je vprašanje skrbi za kakovost eno od ključnih. Na tem področju je mogoče nadzor vzpostaviti s formaliziranimi protokoli, opredeljenimi v pisni obliki, ki pa jih, po izkušnjah intervjuvancev, v vsakdanjih raziskovalnih postopkih ne uporabljajo vedno čisto dosledno, četudi ponekod obstajajo in so raziskovalci z njimi seznanjeni. Tako da se to vprašanje navezuje tudi na širšo problematiko vrednotenja podatkov in načrtovanja, kakšne standarde kakovosti bo nekdo uporabil, tudi glede na to, ali bodo podatki kasneje na voljo v sistemu odprtega dostopa. Zagotovila glede kakovosti morajo biti v tem primeru še bolj pregledna, kot je običajno praksa. (Okvir 5.2: P-R 2)

Pomembno je tako tudi naslednje spoznanje (Okvir 5.3: P-R3). Četudi je v vzpostavljenih znanstvenih okoljih skrb za kakovost podatkov na visoki ravni, kar je povezano s ciljem zagotoviti kakovostne zaključke v analizi in objavi, je skrb za podatke in njihovo dokumentacijo (načrtovanje, varovanje, zagotavljanje ustreznega mesta hrambe in dostopa) področje, na katerem se bodo morali raziskovalci dodatno usposobiti, šele tako bo zagotovljena ustrezná kakovostna raven priprave podatkov za uporabo kot celote.



Vrednotenje in izbor podatkov je samo po sebi posebna naloga, ki ima po eni strani za podatke enak pomen kot vrednotenje publikacij skozi postopek recenziranja, se pravi, gre za ustrezno priznavanje znanstvenega prispevka, po drugi strani pa terja drugačne in prilagojene kriterije ter postopke. Tudi tu je pomembno usposobiti tako nosilce vlog, kot so založniki, knjižničarji in drugi predstavniki znanstvene skupnosti, kot prilagoditi politike financerjev znanosti. Med drugim je pomembno, da recenzenti prijav na raziskovalne projekte poznajo načela priprave načrtov za upravljanje s podatki (DMP) in da to znajo ustrezno ovrednotiti. Dodatno delo in predvideni napor pa morajo biti tudi ustrezno finančno in kadrovske podprti s strani financerja (Okvir 5.4: P-R 4, Okvir 5.5: P-R 5).

### *DODANA VREDNOST PODATKOV*

Zlasti problem spodbud in usposobljenosti za pripravo ustreznih metapodatkov, ki predstavljajo zgodovino nastanka in dokumentacijo značilnosti samih raziskovalnih podatkov, se je pokazal kot izrazito odvisen tudi od strokovne usposobljenosti in podpore ustreznih podatkovnih arhivov na posameznih področjih. V kolikor ni motivacije za izročanje podatkov v podatkovni center, potem ni niti znanja niti pripravljenosti, da se nekdo temu področju dodatno posveti (Okvir 6.1: P-R 6).

Področje digitalne hrambe je v svetu v zadnjem obdobju zelo napredovalo, tako v tehnološkem pogledu kot v pogledu temu prilagojenih postopkov, preverjenih s pomočjo standardov in pod nadzorom politik. Temu ustrezno se opazi zaostajanje pretežne večine 'nespecialistov' za to področje, kar delujoči znanstveniki običajno so. Izraženo je zavedanje, da je ta problem prisoten, hkrati pa so prevladujoče 'ad hoc' rešitve in nemoč, da se ta problem ustrezno uredi na način, da se zmanjša nevarnost izgubljanja in uničevanja podatkov. Tudi tu je vloga specialistov in centralizirane podpore nepogrešljiva, hkrati pa je potreben dodaten napor vseh ostalih (Okvir 6.2: P-R 7).

### *ODPRTI DOSTOP DO PODATKOV*

Zadnja faza v življenjskem krogu podatkov je možnost nadaljnje analize drugih raziskovalcev, ki je podprta z ustreznimi storitvami iskanja, registracije in dovoljenjem dostopa. Da so podatki dostopni drugim, pa jih mora biti ustvarjalec pripravljen dati na razpolago na ustrezno mesto, ki ureja omenjene storitve dostopa. Kljub nekaterim zametkom te dejavnosti je seveda to področje kritično podhranjeno zaradi prevladujoče kulture zapiranja in monopoliziranja podatkov. Seveda problem ni nekaj posebnega za naše okolje, tako da so tudi rešitve, o katerih smo razpravljali z našimi intervjuvanci, že ponujene in se postopoma uveljavljajo v drugih okoljih. Zlasti gre za vse bolj uveljavljen ukrep, da se odprtodostopni podatki vrednotijo v okviru bibliografij, da se obravnavajo enako kot druge znanstvene objave ter da se zagotovi tudi obveznost navajanja med viri v znanstvenih člankih, v katerih so podatki bili uporabljeni. Obenem se vzpostavlja tehnološko in s pomočjo trajnih identifikatorjev podprta infrastrukturna storitev zagotavljanja poenotenega naslavljanja podatkovnih virov, podobno kot z znanstvenimi publikacijami, ali z njimi povezano, ali v obliki podatkovnih revij, ali pa skozi robustno zasnovane področne podatkovne arhive.

Ob tem, ko se uvaja ustrezna tehnološka infrastruktura, pa je pomembno začeti s spremembami pri razumevanju vloge in odgovornosti vseh vpletenih, od ustvarjalcev, dejavnosti storitev, do uporabnikov in financerjev. Naši intervjuvanci pozivajo financerja, naj določi jasna in za vse zavezujoča pravila ter omogoči njihovo izpolnjevanje. Na ta način zagotavljajo, da bo problem nizke kulture deljenja podatkov, hitro rešen (Okvir 7.1: P-R 8).

Enako je tudi na že omenjenem področju dostopa do javnih podatkov potrebno zagotoviti večje vključevanje raziskovalnih potreb, skupaj tudi s potrebno posredovalno vlogo področnih podatkovnih arhivov, katerega zgled je lahko sporazum med ADP in SURS in skupno delo na pripravi podatkov za dostop preko 'varne sobe' in 'oddaljenega varnega dostopa'. (Okvir 7.2: P-R 9)

Vprašanja avtorstva, licenc in priznavanja pravic ter druge omejitve, ki izhajajo iz etičnih ali pravnih obvez v povezavi s podatki, so področje, ki se običajno omenja med ovirami za prosti dostop. Tudi to je področje, ki je ozko strokovno, na katerem pa lahko rešitve, ko se enkrat oblikujejo in opredelijo, pomagajo praktikom, ki ustvarjajo ali posredujejo podatke drugih. Tako je povezava in forum za izmenjavo dobrih praks med ponudniki storitev dostopa do raziskovalnih podatkov logična rešitev, prav tako tudi v povezavi z vprašanji zaupnosti podatkov in drugimi pravnimi problemi. (Okvir 7.3: P-R 10)

Svetovanje in usposabljanje tako ustvarjalcev kot uporabnikov sta pglavitni vlogi obstoječih, že delujočih ponudnikov storitev dostopa do raziskovalnih podatkov ali tistih, ki se šele vzpostavljajo v tej dejavnosti. Zelo pogosto so ustvarjalci podatkov izpostavili problem in skrb, da bodo podatki uporabljeni na neustrezen način, uporabniki pa, da je podatke težko razumeti in da so orodja za uporabo preveč zapletena. To je mogoče rešiti tudi s pomočjo večje izmenjave znanj in izkušenj v skupnosti, npr. na način srečevanja ustvarjalcev in uporabnikov določenih podatkov. Tudi ostali pomisleki in ovire, ki jih navajajo posamezni intervjuvanci in se pojavljajo v javnosti, se lahko rešijo skozi uveljavljen, delujoč sistem zagotavljanja izročanja podatkov in dostopa preko za to pooblaščenih arhivov podatkov, z uporabo vnaprej določenih postopkov in protokolov, npr. pri pridobivanju izrecnega pristanka za sodelovanje pri raziskavi, skupaj s postopki anonimizacije podatkov, kadar gre za podatke o ljudeh. Na drugih področjih so bolj v ospredju vprašanja licenc, pri katerih je prav tako mogoče predvideti modele, ki zagotavljajo morebitno izključno komercialno izkoriščanje, hkrati pa ne omejujejo znanstvene rabe. Poznavanje in promocija tovrstnih pristopov in modelov lahko pomembno pripomore tudi k širšemu sprejemanju idej odprtega dostopa do podatkov in njegovega uveljavljanja.

Ob koncu polstrukturiranega intervjuja smo intervjuvance povprašali, ali bi bili pripravljeni sodelovati v fokusnih skupinah ter na katerih področjih iz Načel OECD bi bili pripravljeni sodelovati v razpravah (tehnologija, standardi; digitalna hramba; organizacija, odgovornost, politike; avtorske in druge pravne omejitve, zasebnost, patenti; kultura izmenjave, etika; interoperabilnost, učinkovitost; kakovost podatkov; ocenjevanje storitev in nadzor nad izvajanjem; obveznosti načrtov ravnanja s podatki in njihov nadzor ter vrednotenje izpolnjevanja v okviru projektov).

Pričujoči projekt ima bistveno vlogo pri seznanjanju s koristmi in deljenju pozitivnih izkušen odprtega dostopa do podatkov. Skupaj s spletno stranjo *Openaccess.Si*, v okviru katere v poglavju *Odprti podatki* sodelujemo svojimi prispevki, predstavlja forum, ki je namenjen promociji načel in podpora pri njihovem uresničevanju. Pričakujemo odzive in nove prispevke tako sedaj kot tudi v nadaljevanju izvajanja projekta, saj je vzpostavitev tega sistema skupna naloga in odgovornost.

# LITERATURA

---

- ANDS. 2011. What is research data? Dostopno prek: <http://ands.org.au/guides/what-is-research-data.html>. (7. avgust 2012).
- ANDS Technical Working Group. 2007. *Towards the Australian Data Commons A proposal for an Australian National Data Service*. Canberra. Dostopno prek: <http://www.pfc.org.au/bin/view/Main/Data>. (7. avgust 2012).
- Anderson, Chris. 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *WIRED MAGAZINE*. Dostopno prek: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory). (7. avgust 2012).
- Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age*. 2009. Dostopno prek: <http://www8.nationalacademies.org/cp/projectview.aspx?key=48721>. (7. avgust 2012).
- Pravilnik o kazalcih in merilih znanstvene in strokovne uspešnosti*. ARRS. Dostopno prek: <http://www.rrs.gov.si/sl/akti/prav-znan-strok-uspesn-maj09.asp>. (7. avgust 2012).
- CESSDA. Dostopno prek: <http://www.cessda.org/>. (7. avgust 2012).
- DDI. Dostopno prek: <http://www.ddialliance.org/>. (7. avgust 2012).
- Digging into data*. Dostopno prek: <http://www.diggingintodata.org/>. (7. avgust 2012).
- UK Data Archive*. Dostopno prek: <http://data-archive.ac.uk/>. (7. avgust 2012).
- Arzberger Peter, Peter Schröder, Anne Beaulieu, Geoffrey C. Bowker, Kathleen Casey, Leif Laaksonen, David Moorman, Paul F. Uhler, Paul Wouters. 2004. Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal* 3: 135-152.
- Association of Research Libraries (ARL). 2007. *Agenda for Developing E-Science in Research Libraries*. Dostopno prek: [http://www.arl.org/bm~doc/ARL\\_EScience\\_final.pdf](http://www.arl.org/bm~doc/ARL_EScience_final.pdf). (7. avgust 2012).
- BRTF-SDPA. 2010. *Sustainable Economics for a Digital Planet: Ensuring Long Term Access to Digital Information*. Dostopno prek: <http://brtf.sdsc.edu/>. (7. avgust 2012).
- Berman, Francine. 2010. We Need a Research Data Census. *Communications of the ACM*. Dostopno prek: [http://www.cs.rpi.edu/~bermaf/Data\\_Census\\_CACM\\_10.pdf](http://www.cs.rpi.edu/~bermaf/Data_Census_CACM_10.pdf). (7. avgust 2012).
- Biemer, Paul P. in Lars Lyberg. 2003. *Introduction to survey quality*. Hoboken, New Jersey: J. Wiley.
- Borgman, Christine L. 2010. Research Data: Who will share what, with whom, when, and why? V: Beijing: China-North America Library Conference. Dostopno prek: <http://works.bepress.com/borgman/238/>. (7. avgust 2012).
- . 2011. The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*: 1-40. Dostopno prek: <http://ssrn.com/abstract=1869155> or <http://dx.doi.org/10.2139/ssrn.1869155>. (7. avgust 2012).

- CORDIS. e-Infrastructure. Dostopno prek: [http://cordis.europa.eu/fp7/ict/e-infrastructure/publications\\_en.html](http://cordis.europa.eu/fp7/ict/e-infrastructure/publications_en.html). (7. avgust 2012).
- Colledge, Michael in Jörgen Dalén. 2009. *ESS Handbook for Quality Reports*. Luxembourg. Dostopno prek: [http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQOR\\_FINAL.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQOR_FINAL.pdf). (7. avgust 2012).
- Corti, Louise, Veerle Van den Eynden, Libby Bishop in Bethany Morgan-Brett. 2011. *Managing and Sharing Data*. Essex: UK Data Archive. Dostopno prek: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>. (7. avgust 2012).
- Cragin, Melissa H. and Jacob R. Carlson in Michael Witt Carole L. Palmer. 2010. Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society* 368, Št. 1926: 4023-4038. Dostopno prek: <http://rsta.royalsocietypublishing.org/content/368/1926/4023>. (7. avgust 2012).
- Diane Harley, Sophia Krzys in C. Judson King. 2010. *Interim Report: Assessing the Future Landscape of Scholarly Communication: An In-depth Study of Faculty Needs and Ways of Meeting them*. Los Angeles, California. Dostopno prek: <http://escholarship.org/uc/item/15x7385g#page-3>. (7. avgust 2012).
- Downs, Robert R. in Robert S. Chen. 2012. Improving the Trustworthiness of an Interdisciplinary Scientific Data Archive. In: Washington, DC, ZDA. Dostopno prek: <http://www.iassist2012.org/indexfolder/program/index.php?show=session:L>. (7. avgust 2012).
- E-IRG. 2009. *E-IRG Report on data management. Data Management Task Force*. Dostopno prek: [http://www.e-irg.eu/images/stories/publ/task\\_force\\_reports/dmtfjointreport.pdf](http://www.e-irg.eu/images/stories/publ/task_force_reports/dmtfjointreport.pdf). (7. avgust 2012).
- ESF. 2008. *Shared Responsibilities in Sharing Research Data: Policies and Partnerships*. Dostopno prek: [http://www.dfg.de/download/pdf/foerderung/programme/lis/sharing\\_research\\_data\\_esf\\_dfg\\_0709.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/sharing_research_data_esf_dfg_0709.pdf). (7. avgust 2012).
- ESRC. 2010. *ESRC Research data policy*. Dostopno prek: <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>. (7. avgust 2012).
- Editorial. 2008. Community cleverness required. *Nature* 455, Nr. 7209: 1. Dostopno prek: <http://www.nature.com/nature/journal/v455/n7209/full/455001a.html>. (7. avgust 2012).
- . 2009. Data's shameful neglect. *Nature* 461, Št. 7261: 145. Dostopno prek: <http://www.nature.com/nature/journal/v461/n7261/full/461145a.html>. (7. avgust 2012).
- European Commission. 2010. *Riding the wave: How Europe can gain from the rising tide of scientific data*. Dostopno prek: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>. (7. avgust 2012).
- European Commission Research & Innovation. Policy Initiatives. Dostopno prek: <http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1301>. (7. avgust 2012).
- European Social Survey. 2009. *Round 5 Specification for Participating Countries*. London. Dostopno prek: [http://europeansocialsurvey.org/index.php?option=com\\_content&view=article&id=248:r5-project-specs-release&catid=22:news&Itemid=48](http://europeansocialsurvey.org/index.php?option=com_content&view=article&id=248:r5-project-specs-release&catid=22:news&Itemid=48) (7. avgust 2012).

- Feijen, Martin. 2011. *What Researchers want*. Dostopno prek: [http://www.surf.nl/nl/publicaties/documents/what\\_researchers\\_want.pdf](http://www.surf.nl/nl/publicaties/documents/what_researchers_want.pdf). (7. avgust 2012).
- Graaf, Maurits van der in Leo Waaijers. 2011. *A Surfboard for Riding the Wave*. Dostopno prek: <http://www.knowledge-exchange.info/Default.aspx?ID=469>. (7. avgust 2012).
- Green, Ann in Myron Gutmann. 2006. *Building Partnerships Among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives*. Dostopno prek: <http://deepblue.lib.umich.edu/handle/2027.42/41214>. (7. avgust 2012).
- Griffin, Tom, Brian Matthews, Alistair Mills, Sri Nagella, Arif Shaon, Michael Wilson in Erica Yang. 2012. Moving from a scientific data collection system to an open data repository. V: *The 7th International Conference on Open Repositories*. Dostopno prek: [https://www.conftool.net/or2012/index.php?page=browseSessions&form\\_session=2https://www.conftool.net/or2012/index.php?page=browseSessions&form\\_session=2](https://www.conftool.net/or2012/index.php?page=browseSessions&form_session=2https://www.conftool.net/or2012/index.php?page=browseSessions&form_session=2). (7. avgust 2012).
- Heather A Piwowar. 2011. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. *PLoS ONE* 6(7), Št. e18657. Dostopno prek: doi:10.1371/journal.pone.0018657. (7. avgust 2012).
- Henty, Margaret, Belinda Weaver, Stephanie Bradbury in Simon Porter. 2008. *Investigating Data Management Practices in Australian Universities*. Dostopno prek: [aprs.anu.edu.au/orca/investigating\\_data\\_management.pdf](http://aprs.anu.edu.au/orca/investigating_data_management.pdf). (7. avgust 2012).
- Hey, Tony, Stewart Tansley in Kristin Tolle, ed. 2009. *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft Research. Dostopno prek: [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_complete\\_lr.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf). (7. avgust 2012).
- ICPSR. ICPSR: Deposit Data & Findings. Dostopno prek: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/deposit/>. (7. avgust 2012).
- Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. 2009. *Harnessing the Power of Digital Data for Science and Society*. Dostopno prek: [http://www.nitrd.gov/About/Harnessing\\_Power\\_Web.pdf](http://www.nitrd.gov/About/Harnessing_Power_Web.pdf). (7. avgust 2012).
- Jones, Sarah. 2009. *A report on the range of policies required for and related to digital curation*. Glasgow, Velika Britanija. Dostopno prek: <http://hdl.handle.net/1842/3371>. (7. avgust 2012).
- Kaul, Inge, Isabelle Grunberg in Marc A. Stern, ed. 1999. *Defining Global Public Goods. Global Public Goods: International Cooperation in the 21st Century*. New York: Oxford University Press.
- Kevin Ashley. 2011. *Data Quality and Curation*. Dostopno prek: [http://www.grdi2020.eu/Pages/SelectedDocument.aspx?id\\_documento=201287b9-9f1c-4626-8b62-44cfe418707a](http://www.grdi2020.eu/Pages/SelectedDocument.aspx?id_documento=201287b9-9f1c-4626-8b62-44cfe418707a). (7. avgust 2012).
- Kuula Arja in Sami Borg. 2008. *Open Access to and Reuse of Research Data – The State of the Art in Finland*. Dostopno prek: [http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs07\\_OECD\\_en.pdf](http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs07_OECD_en.pdf) (7. avgust 2012).

- LeClere, Felicia B. 2006. Data Sharing for Demographic Research at ICPSR. *ICPSR Bulletin* 6, Št. 2: 3-8. Dostopno prek: <http://www.icpsr.umich.edu/files/ICPSR/org/publications/bulletin/2006-Q1.pdf>. (7. avgust 2012). (7. avgust 2012).
- Lossau, Norbert. 2012. An Overview of Research Infrastructures in Europe - and Recommendations to LIBER. *LIBER Quarterly* 21, Št. 3/4: 313-329. Dostopno prek: <http://liber.library.uu.nl/index.php/lq/article/view/8028>. (7. avgust 2012).
- Lyon, Liz. 2007. *Dealing with Data: Roles, Rights, Responsibilities and Relationships*. Dostopno prek: [http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing\\_with\\_data\\_report-final.pdf](http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf). (7. avgust 2012).
- McGovern, Nancy Y., ur. 2012. *Allining national approach to digital preservation*. Atlanta: Educopia Institut. Dostopno prek: [http://www.educopia.org/sites/default/files/ANADP\\_Educopia\\_2012.pdf](http://www.educopia.org/sites/default/files/ANADP_Educopia_2012.pdf). (7. avgust 2012).
- Ministrstvo za visoko šolstvo, znanost in tehnologijo. 2011. Načrt razvoja raziskovalnih infrastruktur 2011-2020. <http://www.arhiv.mvzt.gov.si/fileadmin/mvzt.gov.si/pageuploads/pdf/znanost/nacrt-RI.pdf>.
- Muñoz Palma, Patricia. 2011. *Developing a research data access policy in Chile*. Washington, DC. Dostopno prek: [http://sites.nationalacademies.org/PGA/biso/PGA\\_061353](http://sites.nationalacademies.org/PGA/biso/PGA_061353). (7. avgust 2012).
- National Research Council. 1999. *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington, DC: National Academy Press.
- National Science Board. 2005. *Long-Lived Digital Data Collections*. Dostopno prek: <http://www.nsf.gov/pubs/2005/nsb0540/>. (7. avgust 2012).
- OECD. 2007. *Principles and Guidelines for Access to Research Data from Public Funding*. OECD. Dostopno prek: <http://www.oecd.org/sti/scienceandtechnologypolicy/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>. (7. avgust 2012).
- PARSE.Insight. 2009. *INSIGHT into Issues of Permanent Access to the Records of Science in Europe*. Dostopno prek: [http://www.surf.nl/nl/publicaties/documents/what\\_researchers\\_want.pdf](http://www.surf.nl/nl/publicaties/documents/what_researchers_want.pdf). (7. avgust 2012).
- Parsons, Mark. 2010. Towards a 10-year vision for Global Data Infrastructures Workshop. Dostopno prek: <http://www.grdi2020.eu/Repository/FileScaricati/f43bfdb7-3dac-4cf6-a506-03b07d01b19f.pdf>. (7. avgust 2012).
- Research Information Network. 2008. *Stewardship of digital research data - principles and guidelines*. London. Dostopno prek: <http://www.rin.ac.uk/data-principles>. (7. avgust 2012).
- Research Information Network in British Library. 2010. *Patterns of information use and exchange: case studies of researchers in the life sciences*. Dostopno prek: <http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-informationuse->. (7. avgust 2012).
- Ruusalepp, Raivo. 2008. *A comparative study of international approaches to enabling the sharing of research data*. Dostopno prek: <http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/Data-Sharing-Report.pdf>. (7. avgust 2012).



- Sesink, Laurents, René van Horik in Henk Harmsen. 2008. *Quality guidelines for digital research data in the Netherlands*. Dostopno prek: [http://www.data-archive.ac.uk/media/57319/dsa\\_booklet.pdf](http://www.data-archive.ac.uk/media/57319/dsa_booklet.pdf). (7. avgust 2012).
- Soehner, Catherine, Catherine Steeves in Jennifer Ward. 2010. *E-Science and Data Support Services: A Study of ARL Member Institutions*. Dostopno prek: [http://www.arl.org/bm~doc/escience\\_report2010.pdf](http://www.arl.org/bm~doc/escience_report2010.pdf). (7. avgust 2012).
- Swan Alma in Sheridan Brown. 2008a. *To Share or not to Share: Publication and Quality Assurance of Research Data Outputs Report commissioned by the Research Information Network (RIN)*. Dostopno prek: <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf>. (7. avgust 2012).
- . 2008b. *The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs*. Dostopno prek: <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf>. (7. avgust 2012).
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff in Mike Frame. 2011. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6, Št. 6. Dostopno prek: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101>. (7. avgust 2012).
- The Consultative Committee for Space Data Systems. 2012. *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC, ZDA. Dostopno prek: <http://public.ccsds.org/publications/archive/650x0m2.pdf>. (7. avgust 2012).
- The Economist. 2010. Data, data everywhere. *The Economist*. Dostopno prek: <http://www.economist.com/node/15557443>. (7. avgust 2012).
- The Royal Society Science Policy Centre. 2012. *Science as an open enterprise*. Dostopno prek: [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf). (7. avgust 2012).
- Thessen Anne in David Patterson. 2011. Data issues in the life sciences. *ZooKeys* 150: 15-51. Dostopno prek: <http://www.pensoft.net/journals/zookeys/issue/150/>. (7. avgust 2012).
- UK Research Data Service. 2010. *Proposal and Business Plan for the Initial Pathfinder Development Phase*. Dostopno prek: <http://www.ukrds.ac.uk/resources/download/id/47>. (7. avgust 2012).
- Univerza v Oxfordu. Data Quality and Data Quality Assurance Policy. Dostopno prek: [http://www.admin.ox.ac.uk/pras/aboutus/data\\_quality/](http://www.admin.ox.ac.uk/pras/aboutus/data_quality/). (7. avgust 2012).
- Williford, Christa in Charles Henry. 2012. *One Culture. Computationally Intensive Research in the Humanities and Social Sciences*. Dostopno prek: <http://www.clir.org/pubs/reports/pub151/pub151.pdf>. (7. avgust 2012).
- Wynholds, Laura A., Jillian C. Wallis, Christine L. Borgman, Ashley Sands in Sharon Traweek. 2012. Data, data use, and scientific inquiry: Two case studies of data practices. *ACM Press*. Dostopno prek: <http://works.bepress.com/cgi/viewcontent.cgi?article=1282&context=borgman>. (7. avgust 2012).

Yoon, Ayoung in Helen Tibbo. 2011. Examination of Data Deposit Practices in Repositories with the OAIS Model. *IASSIST Quarterly* WINTER: 6-13. Dostopno prek: [http://www.iassistdata.org/downloads/iquol35\\_tibbo.pdf](http://www.iassistdata.org/downloads/iquol35_tibbo.pdf). (7. avgust 2012).

Štebe, Janez in Irena Vipavc Brvar. 2008. Vodič za pripravo družboslovnih podatkov za arhiviranje. Dostopno prek: [http://adp.fdv.uni-lj.si/za\\_uporabnike/delavnice/#vodic](http://adp.fdv.uni-lj.si/za_uporabnike/delavnice/#vodic). (7. avgust 2012).

---. 2011. Analiza stanja in perspektiva digitalne hrambe v Arhivu družboslovnih podatkov (ADP). *Knjižnica* 55, Št. 1: 57-85. Dostopno prek: [http://revija-knjiznica.zbds-zveza.si/Izvodi/K1101/Stebe\\_Vipavc-Brvar.pdf](http://revija-knjiznica.zbds-zveza.si/Izvodi/K1101/Stebe_Vipavc-Brvar.pdf). (7. avgust 2012).

Štebe, Janez, Jože Hudales in Boris Kragelj. 2011. Archiving and Re-using Qualitative and Qualitative Longitudinal Data in Slovenia. *IASSIST Quarterly* 2010, Št. Spring/Summer/Fall/Winter: 50-59. Dostopno prek: [http://iassistdata.org/downloads/iquol34\\_35\\_stebe.pdf](http://iassistdata.org/downloads/iquol34_35_stebe.pdf). (7. avgust 2012).

Vipavc Brvar, Irena in Jožica Klep. 2004. Primerjava metapodatkovnega standarda Arhiva družboslovnih podatkov (DDI) in standarda Statističnega urada RS (METIS). *Organizacija: revija za management, informatiko in kadre* 37, Št. 5: 296-304.



# PRILOGE

---

## Priloga 1: Prošnja za sodelovanje v projektu Odprti podatki



### Arhiv družboslovnih podatkov

Univerza v Ljubljani, Fakulteta za  
družbene vede  
Kardeljeva ploščad 5  
1000 Ljubljana  
tel: 01/5805-292  
arhiv.podatkov@fdv.uni-lj.si

Ljubljana, 25. 1. 2011

doc. dr. Janez Štebe, predstojnik ADP  
Kardeljeva ploščad 5  
1000 Ljubljana

Tel. 01/5805-292  
<http://www.adp.fdv.uni-lj.si/blog>  
[http://twitter.com/#!/Janez\\_Stebe](http://twitter.com/#!/Janez_Stebe)  
E-pošta: janez.stebe@fdv.uni-lj.si

### Zadeva: Prošnja za sodelovanje v projektu Odprti podatki

Spoštovani!

Arhiv družboslovnih podatkov za Ministrstvo za visoko šolstvo, znanost in tehnologijo v letih 2010-13 izvaja projekt CRP s šifro V5-1018 »Odprti podatki - Priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji«. Prosim vas za sodelovanje pri aktivnostih, ki jih bomo organizirali v okviru projekta za namenom zbiranja informacij o stanju na področju ustvarjanja in dostopa do raziskovalnih podatkov doma in v svetu. Z vašim sodelovanjem bomo v nadaljevanju iskali in preizkušali najbolj ustrezne rešitve, ki bodo ob zaključku projekta kot rezultat podane v predlogu akcijskega načrta. S svojim sodelovanjem nam boste pomagali priti do relevantnih rezultatov - takih, ki bodo ustrezali vašim željam in potrebam ter odsevali posebnosti vaše znanstvene discipline in institucionalnega okolja.

Odprt dostop do raziskovalnih podatkov je aktualen zaradi možnosti, ki jih ponujajo sodobne digitalne tehnologije in omrežja. Kot je navedeno v »Načelih in usmeritvah OECD za dostop do javno financiranih raziskovalnih podatkov«<sup>1</sup> (v nadaljevanju »Načela OECD«), bolj odprt dostop do podatkov povečuje raznolikost analiz in presoj, spodbuja nova raziskovanja, omogoča preizkušanje novih ali alternativnih hipotez in metod raziskovanja, proučevanje metod zbiranja podatkov in meritev, spodbuja izobraževanje novih raziskovalcev, omogoča raziskovanje prvotno nepredvidenih tem ter oblikovanje novih podatkovnih serij s kombiniranjem podatkov iz več virov. Odstranjene ovire pri dostopu do podatkov bodo torej zagotovile njihovo širšo uporabo, s tem izboljšale znanstveno odličnost rezultatov in njihovo uporabnost ter povečale donos javnih naložb v raziskovalno dejavnost.

K izpolnjevanju Načel OECD, ki predstavljajo izhodišče projekta, se je zavezala tudi Slovenija z nedavno včlanitvijo v OECD. Odprt dostop do podatkov po Načelih OECD pomeni preglednost nad viri raziskovalnih podatkov, ki so nastali v skladu z uveljavljenimi znanstvenimi merili kakovosti in uporabnosti za nadaljnje analize. Raziskovalne podatke je potrebno urediti in opremiti z vsemi potrebnimi dodatnimi informacijami (metapodatki), ki opredeljujejo avtorske, patentne, metodološke in vsebinske podrobnosti o nastanku podatkovnega vira in možnostih nadaljnje uporabe. Odstranjene ovire pri dostopu do podatkov bodo torej zagotovile njihovo širšo uporabo, s tem pa se bo izboljšala znanstvena odličnost in uporabnost rezultatov.

---

<sup>1</sup> OECD Principles and Guidelines for Access to Research Data from Public Funding.  
<http://www.oecd.org/dataoecd/9/61/38500813.pdf>

Vabimo Vas, da kot poznavalci svojega področja prispevate informacije, predloge in komentarje na različna izhodišča akcijskega načrta ter sodelujete v aktivnostih, ki jih bomo s tem namenom organizirali: v intervjujih, fokusnih skupinah in drugih dogodkih ter diskusijah, ki jih bomo vodili s pomočjo različnih spletnih orodij. Le tako bomo lahko zagotovili, da bo končni rezultat projekta – akcijski načrt z ukrepi za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji - realistično zasnovan, na podlagi obstoječih razmer na področju ustvarjanja podatkov, storitev njihovega shranjevanja in dostopa, ter dejanskih potreb po njihovem nadaljnjem izkoriščanju pri nas in na tujem.

Podrobnosti o poteku izvajanja projekta, pa tudi druge informacije iz področij, povezanih z vsebino projekta, so predstavljene na našem spletnem naslovu. Vabljeni ste, da čim bolj tekoče spremljate, komentirate in dodajate svoje vsebine tudi s pomočjo spletnih tehnologij (bloga, Twitterja, Facebooka, Connotee).

Za dodatna pojasnila smo vselej na voljo!

Za Vaše sodelovanje se Vam že vnaprej zahvaljujemo in Vas lepo pozdravljamo!

Vodja projekta Odprti podatki  
doc. dr. Janez Štebe

**PRILOGE:**

- Priporočilno pismo MVZT

## Priloga 2: Priporočilno pismo za sodelovanje v projektu Odprti podatki



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN TEHNOLOGIJO

Kotnikova 38, 1000 Ljubljana

T: 01 478 46 00  
F: 01 478 47 19  
E: gp.mvzt@gov.si  
www.mvzt.gov.si

Številka: 631-101/2010/2  
Datum: 17.12.2010

### Priporočilno pismo za sodelovanje pri projektu »Odprti podatki«

Arhiv družboslovnih podatkov (ADP) na Fakulteti za družbene vede, Univerza v Ljubljani, v okviru razpisa za CRP v letih 2010-13 po naročilu Ministrstva za visoko šolstvo, znanost in tehnologijo (MVZT) izvaja projekt »Odprti podatki - Priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji« (Šifra projekta: V5-1018). Republika Slovenija se je k ureditvi tega področja zavezala ob včlanitvi v organizacijo OECD. V ta namen vas prosimo, da sodelujete pri omenjenem projektu in na ta način prispevate k oblikovanju rezultatov, ki bodo upoštevali širše vidike učinkovite, kakovostne in sodobno zasnovane znanstvene dejavnosti.

Nosilci projekta, sodelavci ADP, vas bodo povabili k posameznim aktivnostim v različnih fazah poteka projekta, hkrati pa je predvideno tudi vključevanje predstavnikov MVZT. Vabimo vas torej, da skupaj z nami prispevate k uspehu projekta in ustvarjalno sodelujete v različnih fazah izvajanja projekta v svojih vlogah kot direktorji in vodje raziskovalnih ustanov, kot ustvarjalci podatkov, kot izvajalci podatkovnih storitev ali kot njihovi uporabniki.

S spoštovanjem,

Direktorat za znanost in tehnologijo  
Generalna direktorica

dr. Jana Kolar



Pripravil/a:  
Petra Tramte



## IZJAVA O STRINJANJU S POGOJI SODELOVANJA

Vaše sodelovanje pri intervjuju je prostovoljno. Intervju bo predvidoma trajal eno uro. Pridobljene podatke bomo uporabili za raziskovalne cilje projekta 'ODPRTI PODATKI' in jih shranili v podatkovnih arhivih, kjer bodo dostopni uporabnikom za druge raziskovalne ali izobraževalne namene. Osebni kontaktni podatki bodo iz tega izključeni in jih bomo uporabili samo za korespondenco v okviru tega projekta. Razen v kolikor izrecno ne želite, vaše ime ne bo navedeno v poročilih in drugih projektih gradivih v povezavi z izrečenim v okviru intervjuja.

Potrdite, ali izrecno želite, da je vaše ime navedeno v poročilih in drugih gradivih projekta, ki se navezujejo na izrečeno v okviru intervjuja.  DA  NE

Prosimo, da s podpisom potrdite, da se strinjate z zgoraj navedenimi pogoji sodelovanja.

\_\_\_\_\_  
Ime in priimek intervjuvanca

\_\_\_\_\_  
Podpis

\_\_\_\_\_  
Datum

Doc. dr. Janez Štebe  
Vodja projekta Odprti podatki

\_\_\_\_\_  
Podpis

\_\_\_\_\_  
Datum

ŠT. INT. \_\_\_\_\_



Datum: \_\_\_\_\_

ŠT. INT. \_\_\_\_\_

Kraj: \_\_\_\_\_

### OSNOVNI PODATKI O INTERVJUVANCU

Ime in priimek: \_\_\_\_\_

Organizacija, ki jo zastopa: \_\_\_\_\_

Znanstveno področje: \_\_\_\_\_

Položaj: \_\_\_\_\_

Izobrazba: \_\_\_\_\_

Leta dela: \_\_\_\_\_

Leto rojstva: \_\_\_\_\_

*Opreделите svojo sedanjo ali pričakovano vlogo v zvezi z raziskovalnimi podatki; poimenujte jo tako kot ste navajeni.*

Vloge (akterji) (nosilci pravic, dolžnosti, izvajajo naloge...)	Vloge v kateri nastopa	Lastno poimenovanje vloge
Ustvarjalci, raziskovalna enota, raziskovalni inštitut, podatkovni znanstvenik	<input type="checkbox"/>	
Storitve (skrb za podatke, katalogiziranje, opremljanje z metapodatki za uporabo, spravilo, povezovanje, specializirani področni podatkovni arhivi, mednarodni arhivi, digitalni institucionalni ali nacionalni repozitoriji, knjižnice)	<input type="checkbox"/>	
Uporabniki (drugi raziskovalci kot uporabniki, študentje, poučevanje)	<input type="checkbox"/>	
Vodstva in oblikovalci politik (ministrstvo, drugi financerji, vodstva institucij)	<input type="checkbox"/>	

*V katerih fazah življenjskega kroga podatkov nastopate ali se povezujete z drugimi; poimenujte jih tako kot ste navajeni.*

Faze življenjskega kroga podatkov	Faze v katerih sodeluje	Lastno poimenovanje aktivnosti
Načrtovanje	<input type="checkbox"/>	
Financiranje	<input type="checkbox"/>	
Ustvarjanje (zagotavljanje kakovosti, uporabnosti, evidentiranje nastanka)	<input type="checkbox"/>	
Dodana vrednost (metapodatki, transformacije, čiščenje)	<input type="checkbox"/>	
Izbor, vrednotenje kakovosti, uporabnosti	<input type="checkbox"/>	
Spravilo	<input type="checkbox"/>	
Analiza, objava rezultatov v raziskovalnem poročilu	<input type="checkbox"/>	
Dostop, iskanje (lociranje), <b>ponovna uporaba</b> podatkov drugih	<input type="checkbox"/>	

Faza: **USTVARJANJE PODATKOV**

**Opis stanja nastajanja podatkov, z njimi povezana orodja, postopki in navodila.**

(Opis nastanka podatkov in metapodatkov; kakšni podatki, kaj se z njimi dogaja; katera orodja, metapodatki, standardi, postopki so uporabljeni.)

**Problemi povezani z drugimi fazami življenjskega kroga (upoštevani ali niso upoštevani)**

(načrtovanje, kaj se bo zbiralo za kakšen namen; upoštevanje vrednotenja, izbora in zagotavljanje načrtovane kakovosti; upoštevanje zahtev hrambe; razreševanje problemov za zagotavljanje dostopa; vprašaj zakaj)

**Odgovornost (lastna in drugih) v zvezi z reševanjem problemov (primeri pomanjkljivosti in razlogi za nereševanje problemov)**

(kdo nastopa s kakšnimi dolžnostmi, pravicam, medsebojna pričakovanja in odvisnost nosilcev vlog; pomen politik, vodil, standardov pri opredeljevanju odgovornosti; ovire, sankcije in spodbude, nagrade);

**Identifikacija nosilcev dobrih praks, zgledov reševanja problemov.**

(Kdo so favoriti, zgledi, kje so že vzpostavljene za prenos znanja uporabne povezave doma in v tujini; uporaba orodij, dobrih praks in standardov, katerih? )

Zap. Št.	Kdo (poimenuj nosilce, njihove vloge, če to ni sam)	Kaj (kateri problem rešuje, katera sredstva uporablja, na katero področje se nanaša) *	Zunanja referenca

\*Sredstva, orodja, vodila

- Orodja (za ustvarjanje, za prikaz in analizo podatkov/ metapodatkov)
- Standardi, dobre prakse, navodila (glede kakovosti, podatkov, metapodatkov, hrambe)
- Politike, strategije, zakonodaja, načrti, projekti

(Opomba za analizo: namen je identificirati nosilce problemov ter nosilce zgledov njihovega reševanja; V nadaljnjih fazah iščemo rešitve na način preučevanja možnosti in pogojev za uveljavljanje zgledov rešitev).

Primeri vprašanj: USTVARJANJE PODATKOV

### Opis stanja

*Ali imate podatke: Kaj se s podatki zgodi, dogaja;*

*Kakšne podatke producirate in uporabljate pri svojem delu. (cf. Finsko poročilo)*

*Lahko pokažete konkretno datoteko in kako izgleda? Opišite.*

*Z vidika hrambe: Format (nekateri formati so pokriti, drugi ne). Uvrstite v razred tipa podatkov. Velikost (količina podatkov), kompleksnost.*

*ali so kakšna pravila glede tega? (poznavanje standardov, postopkov za zagotavljanje kakovosti;*

*Ugotavljanje podatkovnih potreb skladno z poslanstvom, ciljem projekta (Harnesing...)*

*Pregled uporabnosti obstoječih podatkov preden zbiraš nove (Harnesing...)*

*evidentiranje podatkov ob nastanku, dokumentiranje (cf. Parse: spremljajoči podatki o nastanku)*

*pregledovanje podatkov, kontrola kakovosti ob nastanku*

*pomen podatkov (izbor, vrednotenje za cilje projekta);*

*Povezave z objekti (metapodatki, standardi, publikacije)*

### Problemi

*Kateri od problemov, postopkov so pomembni, s kakšnim ciljem (Faze: shranjevanje, ponovna uporaba, objava, načrtovanje)*

*Pravni vidiki (avtorske pravice, zasebnost, etični vidiki)*

**Odgovornost**, pripravljenost prevzeti vlogo, odvisnost od pogojev (nagrada, sankcij), odgovornost drugih.

*Ali je ustvarjanje podatkov ustrezno vrednoteno (sodelovanje kot avtor pri objavi, posebej namenjena sredstva za skrb za podatke?)*

*Ustrezna kariera za podatkovnega znanstvenika, nagrade;*

*Ali so upoštevana pravila/zahteve sistema preverjanja kakovosti podatkov ob objavi: sankcija je zavrnjena objava; ali je zagotovljena kakovost ob nastanku*

*Preverjanje kakovosti ob shranjevanju: zavrnjen vstop v sistem hrambe;*

*Kdo skrbi za zagotavljanje pravnih vidikov (navodila, kdo jih je dolžan priskrbeti; etični kodeksi, etične komisije; skrb za izvajanje navodil).*

*Povezave z nosilci vlog (uporabnik, skrbnik); fazami*



Faza: **DODANA VREDNOST (METAPODATKI, TRANSFORMACIJE, ČIŠČENJE)**

**Opis stanja na področju dodane vrednosti podatkov, z njimi povezana orodja, postopki in navodila.**

(Opis spreminjanja podatkov in metapodatkov v fazi dodajanja vrednosti; kaj se s podatki in metapodatki dogaja; katera orodja, metapodatki, standardi, postopki so uporabljeni.)

**Problemi povezani z drugimi fazami življenjskega kroga (upoštevani ali niso upoštevani)**

(Za kakšne namene služijo postopki dodane vrednosti; upoštevanje načrtovane kakovosti; upoštevanje zahtev hrambe, ali je zagotovljena sledljivost postopkov in transformacij; razreševanje problemov za zagotavljanje dostopa, npr. lažja dostopnost, odstranjene ovire zaupnosti ipd.)

**Odgovornost (lastna in drugih) v zvezi z reševanjem problemov (primeri pomanjkljivosti in razlogi za nereševanje problemov)**

(kdo nastopa s kakšnimi dolžnostmi, pravicam, medsebojna pričakovanja in odvisnost nosilcev vlog; pomen politik, vodil, standardov pri opredeljevanju odgovornosti; ovire, sankcije in spodbude, nagrade);

**Identifikacija nosilcev dobrih praks, zgledov reševanja problemov.**

(Kdo so favoriti, zgledi, kje so že vzpostavljene za prenos znanja uporabne povezave doma in v tujini; uporaba orodij, dobrih praks in standardov, katerih? )

Zap. Št.	Kdo (poimenuj nosilce, njihove vloge, če to ni sam)	Kaj (kateri problem rešuje, katera sredstva uporablja, na katero področje se nanaša) *	Zunanja referenca

\*Sredstva, orodja, vodila

- Orodja (za ustvarjanje, za prikaz in analizo podatkov/ metapodatkov)
- Standardi, dobre prakse, navodila (glede kakovosti, podatkov, metapodatkov, hrambe)
- Politike, strategije, zakonodaja, načrti, projekti



Primeri vprašanj: DODANA VREDNOST (METAPODATKI, TRANSFORMACIJE, ČIŠČENJE)

### Opis stanja

*Ali surove podatke obdelate, da jim dodate vrednost?*

*Vrednotenje, uporabnost: ali so surovi podatki sploh uporabni (zase, za druge) preden jih opremimo, dopolnimo; kakšna znanja in orodja so potrebna za uporabo; katere dodatne informacije so potrebne (metapodatki, dokumentacija) o nastanku in dodajanju vrednosti;*

*Ali je možna standardizacija na tem področju, navodila, politike, metapodatki (jih pozna, so pomembne, ovire pri implementaciji)?*

*Kaj se dogaja s podatki od nastanka do točke spravila z vidika dodane vrednosti?*

### Problemi

*Čemu dodana vrednost koristi? (za namen spravila, iskanja, dokumentiranja, analize)*

**Odgovornost** Pripravljenost prevzeti vlogo, odgovornost, odvisnost od pogojev (nagrada, sankcij), odvisnost od objektov (obstoj načrtov, politik, strategij), odgovornost drugih (vloge, kdo bi lahko prevzel).

*Kdo opravlja, sam ali pooblaščen podatkovni center, repozitorij, ali specialist v okviru organizacije; centralni nacionalni digitalni arhiv; Mednarodni centri...*

*Ali je dodana vrednost nagrajena, izvajalec dobi povrnitev stroškov;*

*Ali so opredeljene obveznosti glede dodane vrednosti, metapodatkov (na nivoju organizacije, discipline, nacionalno); kdo bi moral opredeliti obveznosti;*

*Ali je obveznosti potrebno opredeliti za doseganje ciljev – enotna raziskovalna politika?*

### Identifikacija nosilcev dobrih praks

*Kje bi lahko dobil navodila, napotke, usposabljanje za delo na tem področju (področni centri za strokovno svetovanje, kdo je vodilni na tem področju, nastavki dejavnosti);*

Faza: **SPRAVILO, DOLGOROČNA RAZPOLOŽLJIVOST PODATKOV**

**Opis stanja na področju spravila podatkov, s skrbjo za podatke povezana orodja, postopki in navodila.**

(Opis korakov v fazi spravila; kaj se s podatki dogaja, kakšne oblike, formati podatkov nastopajo; kje se podatki hranijo; kakšna so znanja glede hrambe; katera orodja, metapodatki, standardi, postopki so uporabljeni – ali imajo načrte hrambe, poznajo OAIS.)

**Problemi povezani z drugimi fazami življenjskega kroga (upoštevani ali niso upoštevani)**

(Ali je shranjevanje za lastne potrebe ali tudi za potrebe drugih po koncu projekta; Prednosti hrambe na obstoječi lokaciji ali centralizirano po disciplinah; kaj je ciljna vsebina, ki se hrani; ali je opravljeno strokovno vrednotenje pomena podatkov – dolgotrajna vrednost za projekt in za druge, kaj se zgodi s podatki male vrednosti, ali so podatki 'enkratni' ali pa jih z lahkoto reproduciramo; upoštevanje načrtovane kakovosti – ali je zagotovljeno ohranjanje vrednosti podatkov – ali je razpoložljivost podatkov ogrožena zaradi odsotnosti hrambe; ali je v okviru sistema hrambe zagotovljena sledljivost izvora podatkov in postopkov v predhodnih fazah; razreševanje problemov povezave podatkov in metapodatkov, dokumentacije v paketu povezanih gradiv skozi proces hrambe)

**Odgovornost (lastna in drugih) v zvezi z reševanjem problemov (primeri pomanjkljivosti in razlogi za neresevanje problemov)**

(kdo nastopa s kakšnimi dolžnostmi, pravicam, medsebojna pričakovanja in odvisnost nosilcev vlog – kje bi bilo primerno mesto za odlaganje podatkov; pomen politik, vodil, standardov pri opredeljevanju odgovornosti – ali imamo merila in vrednotenje izpolnjevanja zahtev kakovostne hrambe; ovire, sankcije in spodbude, nagrade – kdo za hrambo plača, kdo skrbi za usposabljanje za hrambo; pravni problemi hrambe – prenos pravic, zaščita podatkov);

**Identifikacija nosilcev dobrih praks, zgledov reševanja problemov.**

(Kdo so favoriti, zgledi, kje so že vzpostavljene za prenos znanja uporabne povezave doma in v tujini; uporaba orodij, dobrih praks in standardov, katerih? )

Zap. Št.	Kdo (poimenuj nosilce, njihove vloge, če to ni sam)	Kaj (kateri problem rešuje, katera sredstva uporablja, na katero področje se nanaša) *	Zunanja referenca

Primeri vprašanj: SPRAVILO, DOLGOROČNA RAZPOLOŽLJIVOST PODATKOV

### Opis stanja

*Ali hranite podatke: Kaj se s podatki zgodi, dogaja;*

*Ali ima načrt in strategijo skrbi za podatke.*

*Kje hranijo proizvedeni svoje podatke (v mednarodnem/državnem/institucionalnem repozitoriju/na strežniku znotraj ustanove, kjer delajo/na svojem osebem računalniku...)?*

*Kakšna je skrb za podatke: kje, v kakšni obliki (formatu) se hranijo, verzije, načrt ravnanja s podatki*

*Kakšna je časovna potreba po hranjenju*

*Z vidika hrambe: Format (nekateri formati so pokriti, drugi ne). Uvrstite v razred tipa podatkov. Velikost (količina podatkov), kompleksnost.*

*Ali so kakšna pravila glede spravi in dolgotrajne hrambe tega? (poznavanje standardov, postopkov za zagotavljanje kakovosti;*

*Kakšna orodja uporablja ali načrtuje;*

### Problemi

*Evidentiranje in dokumentiranje spremljajočih podatkov o nastanku pomembni z vidika hrambe?*

*Pregledovanje podatkov, kontrola kakovosti ob vstopu v sistem hrambe*

*Pomen podatkov (izbor, vrednotenje za dolgotrajno razpoložljivost);*

*Nepomembni podatkovni viri: kaj z njimi? Bi tovrstne podatkovne vire dali v depo majhne vrednosti ali zavrgli? V povezavi z vrednotenjem;*

*Prenos podatkov v druge sisteme, povezani sistemi hrambe*

*Kaj so prednosti hranjenja podatkov na obstoječi lokaciji (priročnost/varnost/backup/enostavnost dostopa oz. prilagodljivost uporabe/nizka cena/kapaciteta hranjenja/tisto, kar je na razpolago/deljenje podatkov/samostojnost pri upravljanju oz. nadzoru) in slabosti (omejena kapaciteta/omejena varnost/omejena možnost backupa/zunanji težko dostopajo do podatkov/predraga cena storitev/ne ustreza potrebam/za back-up je potreben čas/omejena publiciteta)? Načrti za oblikovanje /pridruževanje / uporabo teh storitev?*

*Uničenje podatkov, primeri ravnanja? Koliko so podatki enkratni oz. se jih da ponovno pridobiti? Koliko dosega prag 'ogroženosti' uničenja;*

*Kje so med podatkovnimi viri praznine oz. do katerih podatkov ne morete dostopati (pomanjkljivosti kot je npr. obstoj pomembnega podatkovnega vira, ki je ogrožen, nedostopen, slabo dokumentiran...)? Kaj so grožnje shranjevanju: tehnologija, katastrofe, diskontinuiteta?*

*Kateri od problemov, postopkov so pomembni, s kakšnim ciljem (Faze: vrednotenje, ponovna uporaba za lastne namene, za namene drugih, objava, načrtovanje)*

*Tehnološki vidiki (orodja, postopki, standardi) - pomen*

*Uporaba specialnih metapodatkov za shranjevanje, standardi, navodila, ki jih upošteva.*

### Odgovornost

*Kdo je odgovoren? Kaj bi bilo treba narediti?*

*Kje bi jih lahko objavili oz shranili? (institucionalni repozitorij, svetovni repozitorij, revije, etc. ) Kam odlagati – kaj bi bilo primerno mesto...?.*

*Ali je shranjevanje podatkov ustrezno vrednoteno (posebej namenjena finančna sredstva?)*

*Ustrezna kariera za podatkovnega skrbnik, usposabljanje; kdo bi lahko za to nalogo usposabljal;*

*Preverjanje kakovosti ob shranjevanju, ali je opredeljeno, ali je nekdo pooblaščen?*

*Kdo skrbi za zagotavljanje pravnih vidikov (pravice glede spravila, prenos pravic, obveznosti, nadzor dostopa).*

Faza: **ISKANJE, DOSTOP (OBJAVA), PONOVNA UPORABA**

**Opis stanja na področju objave in dostopa do podatkov, z iskanjem, dostopom in ponovno uporabo povezana orodja, postopki in navodila.**

(Opišite komu, pod kakšnimi pogoji, kakšne oblike, formati in verzije podatkov so dostopni - kakšna je praksa glede omejevanja dostopa; kje se podatki hranijo; katera orodja, metapodatki, standardi, postopki so uporabljeni za iskanje in dostop; ali so podatki evidentirani, kako podrobno so opisani v katalogu.)

**Problemi povezani z drugimi fazami življenjskega kroga (upoštevani ali niso upoštevani)**

(Ali se dostopnost podatkov upošteva in zagotavlja v predhodnih fazah priprave podatkov; ali imajo podatki vrednost za druge (preverjanje objav, raziskovalni potencial za nova odkritja in objave, kombiniranje več podatkovnih virov med sabo); ali se za vrednost za ponovno uporabo vrednoti, kakšen je postopek, merila; ali ima uporabnik možnost oceniti vrednost in uporabnost podatkov; ali ima interes za podatke drugih, za kakšen namen; zagotovljena sledljivost izvora podatkov in postopkov v predhodnih fazah za zagotavljanja uporabnosti podatkov za druge, dostop do metapodatkov, dokumentacije; podatki in publikacije; )

**Odgovornost (lastna in drugih) v zvezi z reševanjem problemov (primeri pomanjkljivosti in razlogi za nereševanje problemov)**

(kdo nastopa s kakšnimi dolžnostmi, pravicam, medsebojna pričakovanja in odvisnost nosilcev vlog – kje bi bilo primerno mesto za dostop do podatkov; kdo bi lahko opravljal storitve zagotavljanja dostopa (knjižnice), katere funkcije bi moral izpolnjevati; pomen politik, vodil, standardov pri opredeljevanju odgovornosti – ali imamo merila kakovostne podatkovne objave, pozna in podpira strategije in politike za zagotavljanje dostopa; ovire, sankcije – nezmožnost lociranja podatkov, finančne zahteve; in spodbude, nagrade – kaj kot avtor pričakuje, citiranje, zabeležena objava; pravni problemi dostopa, avtorske pravice, etična občutljivost podatkov); pogled na obveznost dostopa do podatkov s strani naročnika (javno financirani projekti, zahteve EU), revije ipd.; kultura izmenjave, pogled glede omejevanja in pogojev dostopa; ali imaš kot interesent za podatke opredeljene pravice glede dostopa; odnos do OECD vodil (kdo in kako zagotoviti izpolnjevanje – nacionalno, ali vidi sebe, druge; samoumevnost ali neizvršljivost); model načrtovanja upravljanja podatkov);

**Identifikacija nosilcev dobrih praks, zgledov reševanja problemov.**

(Kdo so favoriti, zgledi, kje so že vzpostavljene za prenos znanja uporabne povezave doma in v tujini; uporaba orodij, dobrih praks in standardov, katerih? )

Zap. Št.	Kdo (poimenuj nosilce, njihove vloge, če to ni sam)	Kaj (kateri problem rešuje, katera sredstva uporablja, na katero področje se nanaša) *	Zunanja referenca



Primeri vprašanj: ISKANJE, DOSTOP (OBJAVA), PONOVNA UPORABA

### Opis stanja

*Ali je omogočeno lociranje, dostop in ponovna uporaba podatkov?*

*Ali omogoča dostop do podatkov. Komu, pod kakšnimi pogoji, kje, v kakšni obliki;*

*Podatki – publikacije (da objaviš v reviji in je zraven še povezava na podatke)*

*Koliko ljudi uporablja podatke (izbranega raziskovalca)? Jaz/ostali člani projektne skupine/omejeno število zunanjih uporabnikov (do 20)/veliko zunanjih uporabnikov (več kot 20)/drugi – meri pripravljenost, da daš na razpolago drugim/da uporabiš podatke drugih/da iščeš*

### Orodja in tehnologije za analizo, dostop

*Ali predstavljajo oviro z vidika dostopnosti podatkov?*

*Kje bi jih lahko objavili? (institucionalni repozitorij, svetovni repozitorij, revije, etc.)*

*Kam odlagati – kaj bi bilo primerno mesto spravila in dostopa...?*

*Ali ima načrt in strategijo objavljanja podatkov.*

*Kako hitro se producent/repozitorij/revija odzove na prošnjo za podatke? Ali se sploh odzovejo?*

### Problemi

*Ka so ovire dostopa (npr. nezmožnost lociranja podatkov; avtorske pravice; zasebnost; patent)?*

*Kateri od problemov, postopkov so pomembni, s kakšnim ciljem (Preizkusiti pomen posamezne faze za dostop: vrednotenje (recenzija, selekcija, zagotovilo znanstvene odličnosti, uporabnosti, ponovna uporaba za lastne namene, za namene drugih, načrtovanje objave...)*

*Tehnološki vidiki (orodja, postopki, standardi) – pomen za dostop*

### Odgovornost

*Kdo skrbi za zagotavljanje pravnih vidikov (pravice glede spravila, prenos pravic, obveznosti, nadzor dostopa).Povezave z nosilci vlog (uporabnik, skrbnik); fazami*

*Do katerih podatkov ne morete dostopati zaradi nerazpoložljivosti; kaj so ovire za dostop? Kdo je odgovoren? Kaj bi bilo treba narediti?*

*Ali je zagotavljanje dostopa nagrajeno/kaznovano (pravice uporabe do objave, konkurenca/ citiranje, priznanje avtorstva.*

*Ali je obvezno zagotoviti dostop (zahteva naročnika); Ali podpira obveznost zagotavljanja dostopa?*

*Kakšen pogled imate na obveznost izročanja podatkov – npr. v revijah;*

*Kultura? Splošna ocena koristi dostopa oz. kulture podatkovne izmenjave. Interdisciplinarne izmenjave (interes, navade)?*

*Ali so opredeljene pravice, jih uporabnik pozna, glede dostopa.*

*Pomembno z vidika pravic in dolžnosti; ali so stvari poljubne, določene...; Kakšna pa je dejanska praksa? Kakšen je prag dostopnosti (to je ločen vidik problema, tudi če so podatki shranjeni, so lahko dostopni ožjemu krogu);*

*Interes, da se citira...*

*Stališča glede odprtega dostopa (koristi, ali velja za njegovo delo, področje...= , zavedanje nevarnosti uničenja Percipirane ovire glede dostopa do podatkov...?*

### Koristi in spodbude

Politike (institucionalne, mednarodne, nacionalne) – raziskovalni projekti – podatki: ali se zahteva v okviru kakšnih slovenskih/tujih/mednarodnih projektov, da se poleg poročila shrani v javno dostopnem depozitoriju tudi podatke? Npr. FP7

Odnos imate možnost/želite objavljati/imate koristi?

Ali se vam šteje v dobro pri napredovanju, če objavljate podatke?

### Potrebe po ustreznih kadrih, finančnih in drugih sredstvih (Odgovornost):

Ali mislite, da bi za to moral dobiti denar? Koliko?

Javen/plačljiv dostop/kombinacija? (glej DRIVER)

Ali knjižnice/revije vidijo v hranjenju/dostopu do podatkov nišo za svojo dejavnost? V kolikšni meri bi jo izvajale na račun dosedanje?

OECD deklaracijo pozna, se mu zdi izvedljivo, kakšna sredstva, čigava odgovornost (etično, univerze same, financerji...) kdo določiti? Odvisnost od drugih (ocena moči, odzivnosti drugih akterjev na področju): kdo mislite, da mora rešiti problem odprtih podatkov? Sami/država/... Ali mora priti pobuda od spodaj ali od zgoraj?

Pomanjkanje poznavanja ustrezne tehnologije in znanja za njeno uporabo? (problemi, ovire pri dostopu; ali je potrebno poznavanje orodij; ali je treba kaj kompenzirati (npr. z obljubo soobjave...)

Ali je ideja dostopa po OECD samoumevna: kakšne so težave, bariere, pomisleki glede dostopa? Tu gre tudi za vidik izmenjave podatkov z vidika kulture.

Podpora nacionalni ureditvi tega področja? Poznavanje, ocena težav itd., izkušnje (ali je sam odklonil, dostop bil onemogočen)...: kdo računa na free riderje, kje so individualni interesi pretehtali nad skupno koristjo itd.

Faza objavjanja podatkov (motivacije, ovir; primerjaj pull push) v širšem smislu kultura izmenjave; problem citiranja in povezovanja podatkov in publikacij.

### Lastništvo podatkov:

Ali so avtorske pravice, intelektualna lastnina, patenti ipd. ovira za uveljavitev odprtih podatkov? Kako je s problemom zaupnosti podatkov? (meri pripravljenost na objavljanje z vidika ovir,;) PRAVICE

Katere politike, strategije poznate oz. delujejo kot spodbuda za objavljanje podatkov?

Vprašanje glede sprejemljivosti različnih politik: ali bi bilo sprejemljivo obvezno predajanje podatkov; Kaj pa model data management planning (DMP),

Zopet glede politik: centralizirano/ distribuirano: Kakšne so izhodiščne predstave, kako bi se lahko to reševalo? Vloga knjižnic, založnikov, medsebojna pričakovanja...?

Ali raziskovalci vidijo knjižnice kot partnerje pri hranjenju in dostopanju do podatkov?

### Od odkritja do dostopa

Katera znanja mora imeti uporabnik, da podatke lahko uporablja? Kako morajo biti shranjeni? V kakšnem formatu?

To vprašanje je povezano tudi s profilom podatkov: nekateri podatki so razumljivi sami po sebi, pri drugih rabiš razlago. Običajno potrebuješ za njihovo odpiranje določen program (npr. SPSS). Opremljeni morajo biti z meta podatki.

### Faza: IZBOR, VREDNOTENJE KAKOVOSTI, UPORABNOSTI

Pojasnilo: Vrednotenje je vključeno v večino predhodnih faz. Tukaj na enem mestu izpostavimo ta vprašanja z namenom, da preverimo, če smo izpustili kakšen pomemben vidik vrednotenja.

#### Opis stanja na področju vrednotenja podatkov, povezana orodja, postopki in navodila.

(Opišite pomen podatkov, njihovo enkratnost, uporabnost za lastno uporabo, za ponovno uporabo; katera orodja, metapodatki, standardi, postopki so uporabljeni za vrednotenje; ali je ocena vrednotenja kje označena, vidna.)

#### Problemi povezani z drugimi fazami življenjskega kroga (upoštevani ali niso upoštevani)

(Ali se vrednotenje podatkov upošteva in zagotavlja v predhodnih fazah priprave podatkov, ali se pri vrednotenju upošteva pomen za druge ; ali se ob objavi)

#### Odgovornost (lastna in drugih) v zvezi z reševanjem problemov (primeri pomanjkljivosti in razlogi za nereševanje problemov)

(kdo nastopa s kakšnimi dolžnostmi, pravicam, medsebojna pričakovanja in odvisnost nosilcev vlog – kje bi bilo primerno mesto za vrednotenje (ob nastanku, ob shranjevanju, ob dostopu, ob objavi); pomen politik, vodil, standardov – ali imamo merila vrednotenja, ovire, sankcije, finančne zahteve; in spodbude, nagrade – ali se kakovost podatkov izplača, je nagrajena?)

#### Identifikacija nosilcev dobrih praks, zgledov reševanja problemov.

(Kdo so favoriti, zgledi, kje so že vzpostavljene za prenos znanja uporabne povezave doma in v tujini; uporaba orodij, dobrih praks in standardov, katerih? )

Zap. Št.	Kdo (poimenuj nosilce, njihove vloge, če to ni sam)	Kaj (kateri problem rešuje, katera sredstva uporablja, na katero področje se nanaša) *	Zunanja referenca



Dodatna vprašanja: IZBOR, VREDNOTENJE KAKOVOSTI, UPORABNOSTI

### Opis stanja

*Ali se podatki ločijo na bolj pomembne in manj pomembne; Kako lahko jih reproduciramo, če se uničijo, izgubijo? (Ali imate opredeljene postopke za vrednotenje podatkov ob nastanku, ob sprejemu v hrambo, ob dajanju v dostop; Kakšni so postopki, kriteriji (recenziranje, ocenjevanje na podlagi česa?)*

*Kaj se dogaja s podatki od nastanka do točke spravila z vidika dodane vrednosti, ali podatki pridobivajo na vrednosti?*

*Ali so podatki dragoceni? (ponovljivi ali ne itd.; ali obstaja povpraševanje po podatkih, interes in možnosti ponovne uporabe – dostop do tujih podatkov)*

*Ali imate izoblikovana merila za ocenjevanje podatkov/bi morala biti? Koliko so disciplinarno specifična, koliko splošna?*

### Problemi

*Čemu vrednotenje koristi? (za namen spravila – selekcija najpomembnejših podatkov zaradi stroškov hrambe) ; iskanja, dokumentiranja, ponovne uporabe (rangiranje podatkov po pomenu za zagotovilo kakovosti pri ponovni uporabi; za ekvivalent znanstvene objave)*

*Ali vrednotenje obsega različne komponente, vključno s fazami in izpolnjevanje (opredeljenih) zahtev v posameznih fazah (nastanek, načrtovanje, uporaba standardov in meril; ustreznost dodane vrednosti, primernosti za hrambo, ustrezne opremljenosti z metapodatki...); ustreznost za ponovno uporabo (vrednost za uporabnike, za poučevanje, edinstvenost podatkov – praznina v znanju);*

### Odgovornost

*Kdo opravlja vrednotenje, sam ali pooblaščen podatkovni center, repozitorij, ali specialist v okviru organizacije; centralni nacionalni digitalni arhiv; Mednarodni centri...; znanstvena skupnost z opredeljenim postopkom, revija ob publikaciji.*

*Izvajalec vrednotenja dobi povrnitev stroškov;*

*Interes za vrednotenje (ali je ustrezno nagrajena izvedba, ali ima vrednotenje posledice za ugled, citiranost, in zato koristno),*

*Ali so opredeljeni kriteriji, politike vrednotenja (na nivoju organizacije, discipline, nacionalno);*

*Ali bi vrednotenje moralo biti preverjeno s statistikami uporabe (citiranje).*

### Identifikacija nosilcev dobrih praks, zgledov reševanja problemov.

*Ali so kakšne poznane prakse, zgledi; vključevanje v podatkovne revije; kdo bi ocenjeval podatke...*

*Kje bi lahko dobil navodila, napotke, usposabljanje za doseganje ciljev na tem področju (etični kodeksi in dobre prakse disciplin, organizacij, izdajateljev, podatkovnih centrov, repozitorijev);*

### *Sodelovanje v fokusnih skupinah*

Na katerih področjih iz Vodil OECD bi bil pripravljen sodelovati v razpravah (tehnologija, standardi; digitalna hramba; organizacija, odgovornost, politike; avtorske in druge pravne omejitve, zasebnost, patenti; kultura izmenjave, etika; interoperabilnost, učinkovitost; kakovost podatkov; ocenjevanje storitev in nadzor nad izvajanjem; obveznosti načrtov ravnanja s podatki in njihov nadzor ter vrednotenje izpolnjevanja v okviru projektov).

Spodnja tabela je namenjena identifikaciji nosilcev dobrih praks in zgledeov reševanja problemov. Če veste za kak takšen primer, ga prosimo navedite (nanaša se na vse faze življenjskega kroga podatkov)

Zap. Št.	Faza	Kdo (poimenuj nosilce, njihove vloge, če to ni sam)	Kaj (kateri problem rešuje, katera sredstva uporablja, na katero področje se nanaša)*	Zunanja referenca

ŠT. INT. \_\_\_\_\_