



OPIS STANJA NA PODROČJU RAZISKOVALNIH PODATKOV V SLOVENIJI

Odprti podatki, predstavitev poročila predstavnikom ESFRI

JANEZ ŠTEBE, SANJA LUŽAR IN SONJA BEZJAK

ADP, Univerza v Ljubljani, 5. februar 2013

Program predstavitve

- Predstavitev udeležencev
- Projekt Odprti podatki
- Predstavitev po poglavjih:
 - načrtovanje, ustvarjanje podatkov, dodana vrednost in dokumentacija
 - digitalna hramba
 - iskanje, dostop in možnosti druge rabe podatkov
 - jasna postavitev odgovornosti za reševanje problemov
- Razprava

Projekt **Odprti podatki** 2011 – 2013- Priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji

Podlaga **„Načela in usmeritve OECD za dostop do javno financiranih raziskovalnih podatkov“**

- raziskovalni podatki iz raziskav, ki se izvajajo ob podpori iz javnih virov so javno dobro;
- potrebno je zagotoviti kolikor je mogoče odprt in prost dostop do raziskovalnih podatkov.

Koristi:

- Prihranek stroškov
- Preglednost, zaupanje v znanost se poveča
- Nova, bolj kakovostna odkritja ob večkratni analizi (...)

Prvi rezultati projekta:
Opis stanja na področju

Izvajamo niz predstavitev za seznanjanje ciljnih javnosti in pridobivanje kritičnih odzivov

USTVARJANJE PODATKOV

Sanja Lužar



Kaj so podatki?

- *„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.“*
- *„Mi nimamo podatkov, mi se ukvarjamo z viri.“*
- *„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.“*

Splošna opredelitev (*Principles and Guidelines* OECD):

“Raziskovalni podatki” so digitalno berljivi zapisi o dejstvih, ki predstavljajo osnovno podlago za znanstveno raziskovanje in ki v okviru znanstvene skupnosti veljajo kot ustrezno sredstvo za preverjanje veljavnosti raziskovalnih spoznanj.

Tipi podatkov

- 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,
- video,
- programska oprema,
- strojna oprema,
- aplikacije,
- spektri,
- senzogrami,
- lidarski podatki,
- herbariji,
- zbirke živali,
- vzorci tkiv.
- podatki pridobljeni z metodo zračnega laserskega skeniranja.

Podatki glede na metodo zbiranja

- podatki, ki jih zberejo raziskovalci sami,
- podatki narejeni z aparaturami,
- podatki, katerih osnova za generiranje so drugi viri npr. vzorci, kartoteke

Podatki iz drugih virov

- javne ustanove (SURS, ARSO, AJPES , Geodetska uprava RS ...)
- mednarodni podatkovni centri

Kakovost podatkov v fazi ustvarjanja

Kakovost podatkov - uporabnost podatkov za določen namen.

PROBLEM:

- Formalizirana in institucionalizirana pravila vs. raziskovalci prepuščeni lastni iznajdljivosti
- velike razlike v ravni poznavanja postopkov in meril za zagotavljanje kakovosti

REŠITEV:

- Deljenje izkušenj, dobrih praks in zgledov -> večja standardizacija -> večja kakovost -> večja možnost izkoriščanja
- strokovno usposabljanje za upravljanje s podatki

Izbor podatkov na podlagi kakovosti

- Kaj so dobri podatki?

Splošno soglasje v okviru znanstvene skupnosti.

- Hraniti zgolj kakovostne podatke, ni za vse sprejemljivo.

„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.“

- Ali je mogoče izbor na podlagi kakovosti podatkov formalizirati po analogiji z recenzijskim postopkom, ki velja za znanstvene članke?
Kdo bi ocenjeval in vrednotil kakovost podatkov?

REŠITVE:

- Prilagajanje tradicionalnih uredniških politik in recenzentskih postopkov področju vrednotenja podatkov
- Knjižničarji usposobljeni za vzdrževane formalnih kriterijev in omejitev

Dodana vrednost

- Čiščenje in transformacija
- Opremljanje podatkov z metapodatki

Stanje:

- uveljavljeni standardi,
- neformalni dogovori,
- Vse popolnoma prepuščeno raziskovalcu

Čiščenje, transformacija podatkov

- Običajno potrebna vsaj minimalna obdelava, obstajajo pa tudi podatki, ki so brez obdelave neuporabni.
- Pravila za čiščenje in transformacijo podatkov vsaj neformalno opredeljena kot uveljavljene prakse na posameznih inštitutih.

Končni proizvod so lahko podatki v različnih verzijah, ki so uporabni za različne raziskovalce.

Metapodatkovni standardi in dokumentacija

Podatki namenjeni za mednarodne zbirke –popolna dokumentacija nastajanja podatkov.

Področja, kjer še ni uveljavljenih standardov:

- Dublin Core, z dodatnimi polji,
- nestandarden xml zapis,
- interni pravilnik o organizaciji raziskovalnega dela, ki izhaja iz projektnega menedžmenta

Metapodatkovni standardi in dokumentacija

PROBLEM:

- Dokumentacija podatkov pogosto prepuščena samim raziskovalcem
- Dokumentiranje podatkov le kot dodatno delo

REŠITEV:

- Pomoč in svetovanje pri pripravi metapodatkov že v fazi ustvarjanja podatkov
- Svetovanje pri iskanju specializiranih storitev hrambe in dostopa do podatkov
- Usposabljanje knjižničarjev glede klasifikacij, tipov podatkovnih gradiv, in za podporo pri pripravi podatkov za objavo.

DOLGOTRAJNO DIGITALNO SHRANJEVANJE IN SKRBNIŠTVO

Kaj / kako / kje shranjevati?

Primer problemov in rešitev:

Problem:

Raziskovalni podatki: različni formati, velikosti, trajanje...

Rešitev:

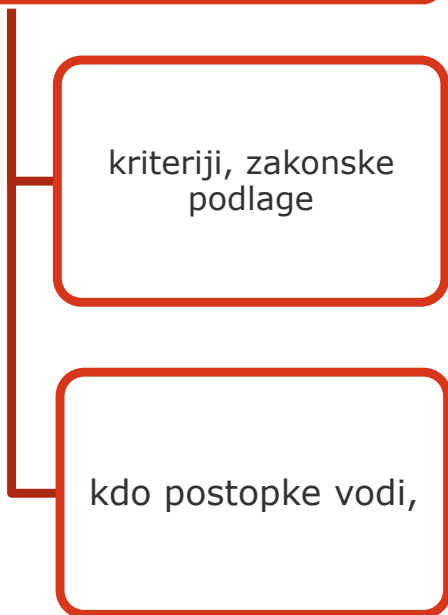
register formatov in prilagojenih postopkov shranjevanja...

Kdo bo prevzel nalogo? Vloga ostalih deležnikov?

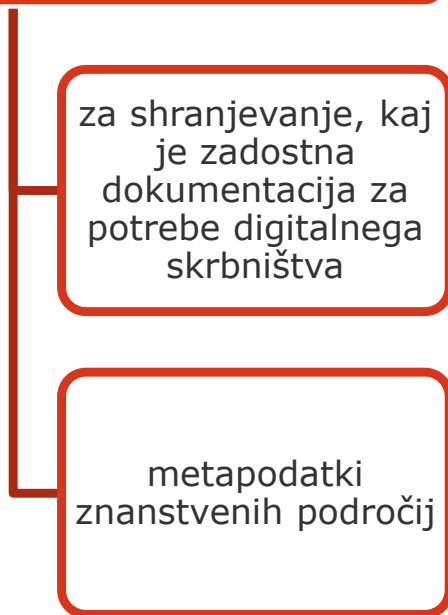
- Raziskovalci sami,
 - repozitoriji ustanov oz. splošni,
 - področni podatkovni centri,
 - nacionalne ustanove dediščine (NUK), pooblaščne ustanove (ARS)
-
- vloga področnih knjižnic in knjižnic ustanov, IT oddelki: podpora dejavnost strokovnjakov za digitalno skrbništvo?

Področja, ki jih je potrebno podrobneje opredeliti

Vrednotenje in izbor ter izločanje:



Metapodatki:



Preverjanje postopkov in organiziranosti hrambe:



Skupne potrebe na področju digitalne hrambe



ISKANJE, DOSTOP IN DRUGA RABA RAZISKOVALNIH PODATKOV

VLOGE IN NALOGE:

Ustvarjalci podatkov

uporabniki podatkov

Deliti ali ne?



Uporabiti ali ne?

Vprašanje, zadržki, dileme, strahovi, predsodki...



**nepoznavanje, pomanjkanje interesa, pomanjkanje časa,
nerazvita infrastruktura**

Kako?

vzpostavitev profesionalizirane podatkovne storitve

DOSTOP

V Sloveniji razvoj dveh hitrosti, velike razlike med področji:

1. Naravoslovci in družboslovci pogosteje vključeni v mednarodne podatkovne kroge:

Mednarodne podatkovne storitve:

- GenBank,
- UniProt,
- Protein Data Bank,
- Amadeus,
- ZACAT,
- ESS,
- Eurostat,

...

Slovenski repozitorij:

- Arhiv družboslovnih podatkov (ADP)

2. Na nacionalno zamejenih področjih se zametki podatkovnih storitev šele pojavljajo:

- Arzenal, Virtualna zakladnica nacionalne dediščine, ZRC SAZU
- DEDI, Enciklopedija naravne in kulturne dediščine na Slovenskem,
- SI-story, Zgodovina Slovenije,
- Slovensko leposlovje na spletu,
- SIDIH, Digitalna infrastruktura za humanistiko in umetnost,
- Inštitut za biomedicinsko informatiko (IBMI)
- Glasbenonarodopisni inštitut,
- ARKAS, Arheološki kataster Slovenije
- ...

OVIRE NA POTI K ODPRTEMU DOSTOPU:

Intervjuvanci so posebej izpostavili naslednje:

1. prisotnost **tekmovalnosti in konkurenčnosti** med raziskovalci,
2. slaba urejenost dostopa v primeru **državnih ustanov**,
3. problem nerešenega vprašanja **avtorstva**,
4. strah pred **izgubo nadzora** nad ‚lastnimi‘ podatki

1. TEKMOVALNOST IN KONKURENČNOST

Podatki so:

- neizkoriščen potencial
- osnova za prihodnje delo (objave, nova sredstva),
- last raziskovalca, ki jih je ustvaril

Več intervjuvancev je znanstveno okolje opisalo s pojmi: trg, konkurenca, tekma, tekmeč, lastništvo, monopol nad možnimi odkritji ...

Rešitev:

- deljenje podatkov in dostop do njih ni stvar raziskovalčeve osebne presoje, pač pa širšega konsenza znanstvene skupnosti in zahtev financerja (politike)
- reforma sistema vrednotenja znanstvenih prispevkov in financiranja projektov (pravično ovrednotiti prispevke in napore pri ustvarjanju in pripravi podatkov, ovrednotiti objave podatkov, tudi v netradicionalnih oblikah) (P-R 8)

2. DRŽAVNE USTANOVE SO SLAB ZGLED:

Državni podatki so:

- so težko dostopni,
- kdaj plačljivi

Intervjuvanci, ki so imeli izkušnje z državnimi uradi ali mednarodnimi organizacijami (npr. SURS, Svetovna banka), so bili posebej zadržani glede deljenja podatkov.

Rešitev:

- ureditev splošnih načel o dostopu do podatkov, ki jih ustvarja javni sektor (politike),
- del storitve in nalog evidentiranja, iskanja in dostopa prevzamejo in izvajajo področni podatkovni centri (infrastruktura in storitve) (P-R 9)

3. VPRAŠANJE AVTORSTVA:

Podatki:

- so ustvarjeni,
- predelani, obogateni, dopolnjeni (leksikoni, zemljevidi),
- osiroteli ...

Nekateri intervjuvanci so negotovi, si podatkov ne upajo deliti z drugimi, da ne bi bili v prekršku.

Rešitev:

- koordinirano reševanje širših vidikov avtorskih pravic,
- ustvariti pregledna navodila glede avtorskih pravic in licenc,
- priporočiti OA licence, ki ne omejujejo rabe podatkov bolj, kot je nujno (P-R 10)

4. STRAH PRED IZGUBO NADZORA NAD PODATKI

Podatke:

- je mogoče zlorabiti,
- lahko povzročijo škodo

Več intervjuvancev je previdnih, se bojijo odprtega dostopa, ponekod so za to strokovne podlage (npr. arheologija), drugod nepoznavanje rešitev.

Rešitev:

- urediti protokole glede dostopa (npr. registrirani uporabniki, etična komisija, ‚varna soba‘),
- vzpostaviti sistem nadzora nad uporabniki (P-R 13)

TOČKE ZA RAZPRAVO:

1. **MESTO DOSTOPA** (na obstoječi lokaciji, področni podatkovni c.)
2. **OMEJITEV DOSTOPA** (embargo, kategorije up., kontrola dostopa)
3. **OA LICENCE, VREDNOTENJE IN CITIRANJE PODATKOV**
4. **STANDARDI ZA METAPODATKE** (združljivost podatkov)

ODGOVORNOST V ZVEZI Z REŠEVANJEM PROBLEMOM

Potreba po jasni krovni politiki

- Citati
 - „Če smo vsi za, sem tudi jaz za“.
 - Toda: „da ne bodo neke trapaste direktive, ki jih potem ne bomo mogli izvrševati“.
 - Vloga jasnih politik pri ozaveščanju, spreminjanju navad.
- Sankcije
 - Pogodbena obveznost raziskovalnega projekta
- Spodbude
 - Dodatno delo ustrezno finančno ovrednoteno: obvezen Načrt ravnanja s podatki, ki dobi finančno pokritje v okviru projekta
 - Dosledno navajanje uporabljenih raziskovalnih podatkov v seznamu literature
 - Ustrezno vrednotenje pri osebnih bibliografijah

Infrastrukturna raziskovalna podporna storitev

- Zlasti pa za uresničevanje politik potrebujemo
 - Učinkovit sistem izvedbe storitev prevzema, hrambe in dostopa do podatkov
 - Ima zagotovilo dolgotrajnega delovanja
 - Ima ustrezno kadrovsko pokritje, možnost strokovnega usposabljanja
- Področni podatkovni centri/ Kriteriji za delovanje
 - Imajo možnosti pridobivanja znanja in to znanje širijo med uporabnike: ustvarjalce (svetovanje in pomoč pri pripravi in izvajanju NRRP) in uporabnike podatkov (aktivno spodbujanje druge rabe)
 - Imajo mednarodne povezave, so po potrebi nacionalno med sabo integrirani, hkrati pa razvijajo področno diferencirane rešitve
 - Imajo vzpostavljene javno objavljene kriterije in postopke za vrednotenje kakovosti podatkov in temu ustrezen izbor
 - So usposobljeni in imajo ustrezna orodja ter jasno določene postopke izvajanja storitev
 - Imajo mehanizme za zagotavljanje odziva na zahteve strok in okolja da prilagajajo svoje storitve