



ARHIV DRUŽBOSLOVNIH PODATKOV

Analiziraj podatke! Deli raziskavo! Prispevaj k znanosti!

RAZVOJ SMERNIC ZA PREDAJO IN ARHIVIRANJE KVALITATIVNIH PODATKOV V ARHIVU DRUŽBOSLOVNIH PODATKOV

Maja Dolinar, Janez Štebe in Sonja Bezjak

Konferenca "Jezikovne tehnologije in digitalna humanistika", 20. 9.–21. 9. 2018, Fakulteta za elektrotehniko, Univerza v Ljubljani

O ADP

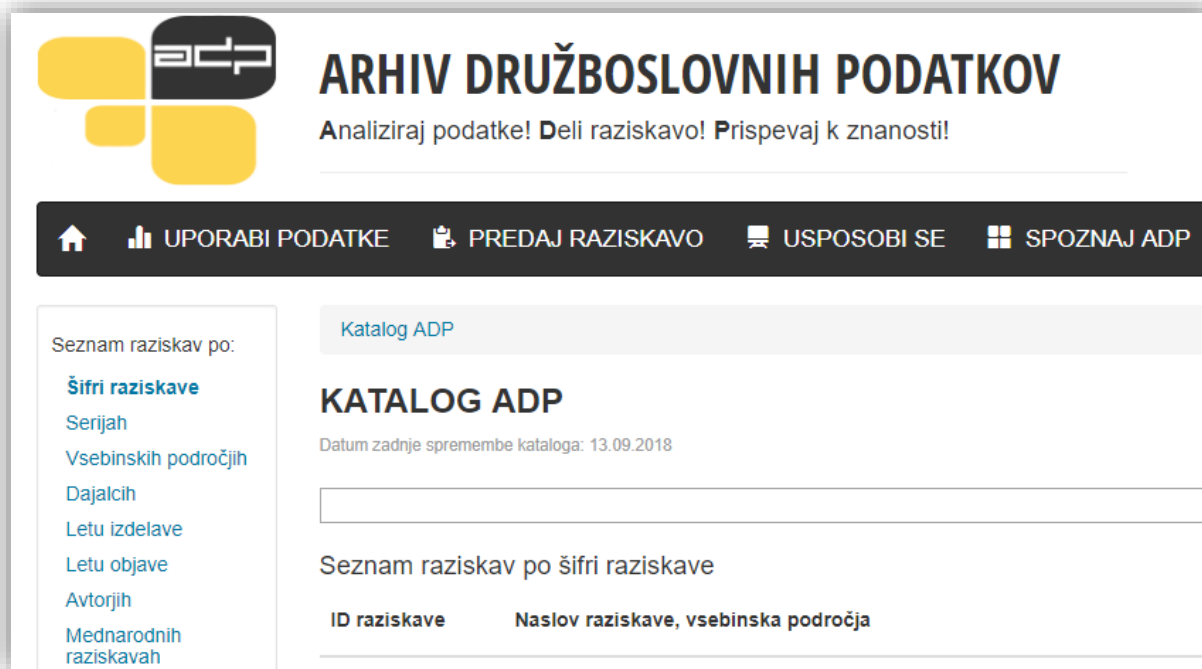


- 1997
- **nacionalno podatkovno središče za družboslovje**
- dajalci podatkov iz vseh 4 univerz, zasebnih raziskovalnih centrov, SURS ...
- preko 600 družboslovnih raziskav, najstarejša iz leta 1968 (malo kvalitativnih)
- cca. 700 uporabnikov letno (90 % izobraževanje, 10 % znanstveno-raziskovalni)
- član CESSDA ERIC
- različni mednarodni projekti (DwB, Foster, SEEDS, SaW, SERISS ...)
- 2018 – pridobljen osnovni certifikat zaupanja vrednih arhivov *CoreTrustSeal*



Namen

- vse več pobud v skupnosti uporabnikov glede ustvarjanja in uporabe različnih vrst in formatov kvalitativnih podatkov
- v Katalog ADP že vključenih nekaj kvalitativnih raziskav



ARHIV DRUŽBOSLOVNIH PODATKOV

Analiziraj podatke! Deli raziskavo! Prispevaj k znanosti!

UPORABI PODATKE PREDAJ RAZISKAVO USPOSOBI SE SPOZNAJ ADP

Katalog ADP

KATALOG ADP

Datum zadnje spremembe kataloga: 13.09.2018

Seznam raziskav po šifri raziskave

ID raziskave	Naslov raziskave, vsebinska področja
--------------	--------------------------------------

<https://www.adp.fdv.uni-lj.si/>

- predstavitev smernic za prevzem in arhiviranje kvalitativnih raziskav

Sprejem kvalitativnih raziskav

ADP dolgotrajno hrani raziskovalne podatke, zanimive za družboslovne analize, s poudarkom na problemih, povezanih s slovensko družbo.

- kriteriji: potencial nadaljnje uporabe podatkov ter skladnost s strategijo odprtega dostopa do podatkov

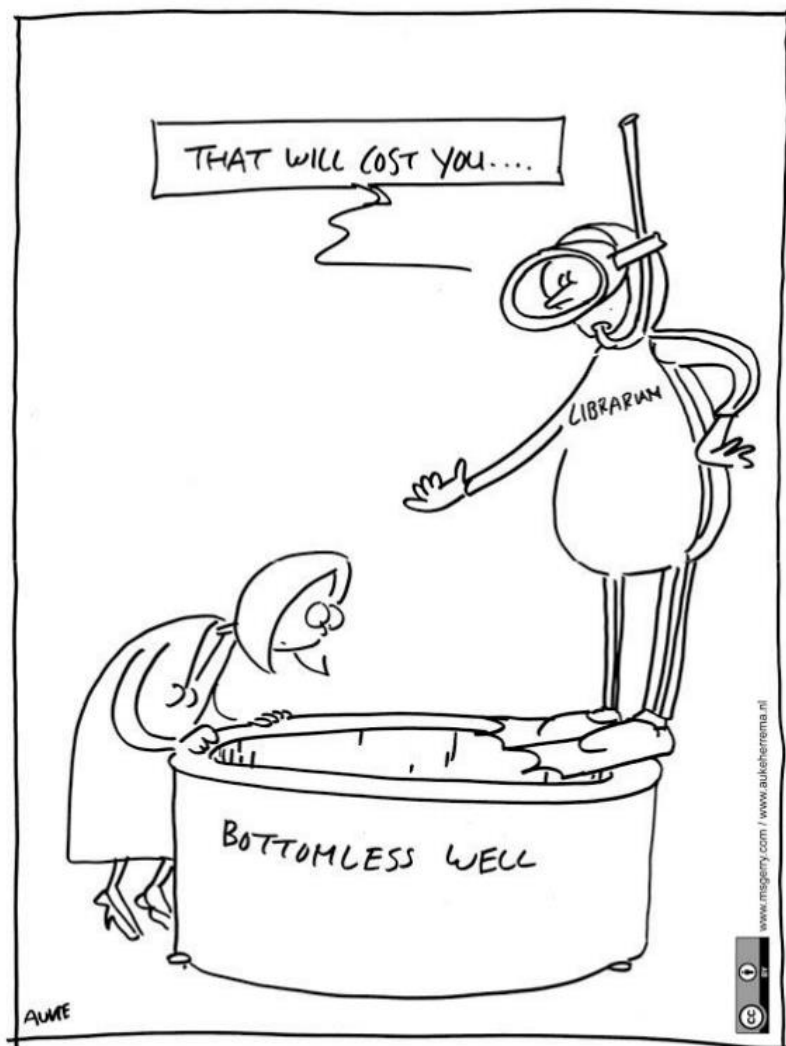
Za ADP za dolgotrajno arhiviranje zanimive 4 skupine kvalitativnih podatkov:

- besedilni podatki
- mirujoča slika
- avdio gradivo
- avdiovizualno gradivo



INFORMATION TYPES

Sprejem kvalitativnih raziskav



RECOVERY OF DATA

Vir: Research data alliance meeting 2014

Merila ADP za sprejem podatkov:

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

Prioriteta in presoja pomena z vidika stroškov dolgotrajne hrambe in ocene prihodnje uporabnosti podatkov.

Prevzem in arhiviranje

- 1) Raziskovalec preveri, ali podatki ustrezajo merilom za sprejem v arhiv.
- 2) Predlog za predajo lahko pošlje preko obrazca »Evidentiraj raziskavo«.
- 3) ADP opravi intervju z raziskovalcem, kjer se preuči ustreznosti sprejema in sprejme odločitve o predaji.
- 4) Raziskovalec izpolni »**Izjavo o izročitvi**«, v kateri podrobneje opiše predana gradiva in pogoje dostopa.
- 5) Raziskovalec izpolni »**Obrazec za opis raziskave**«.
- 6) Raziskovalec **pripravi ustrezno urejene in dokumentirane podatke ter ostalo gradivo** za predajo v Arhiv
 - „**Priporočila za urejanje podatkovnih datotek kvalitativnih raziskav**“ (2018)
 - podatki in celotna dokumentacija morajo biti v digitalni obliki
 - digitalizacija!



LOSS OF DATA

Več na: www.adp.fdv.uni-lj.si/deli/postopek

Vir: Research data alliance meeting 2014



Prevzem in arhiviranje

**ANONIMIZACIJA - že pred
predajo v arhiv! A ne vedno –
odvisno od vsebine raziskave!**

Zaupnost podatkov v ADP:

1. osnovni anonimizacijski postopki in odstranitev osebnih informacij
 2. omejen dostop do raziskave
 3. pogodbe z uporabniki, ki določajo posebne pogoje uporabe
- informirana soglasja udeležencev za sodelovanje v raziskavi → pogoj za prevzem gradiv v ADP

© MARK ANDERSON

WWW.ANDERSTOONS.COM



"Before I write my name on the board, I'll need to know how you're planning to use that data."

Prevzem in arhiviranje

Anonimizacija podatkov:

- utemeljena, dokler ne zmanjša analitične vrednosti podatkov!

Primeri:

- odstranitev vseh direktnih identifikatorjev v transkriptih,
- zameglitev obrazov na slikah,
- zameglitev uporabniških imen v objavah na družbenih medijih,
- itd.

Anonimizacija predanih podatkov je delo in odgovornost dajalca!



Priprava podatkov in dokumentacije

Naslov datoteke	
Opis	
URL povezava*	
Dolžina posnetka**	
Avtor/ji	
Dajalec	
Organizacija	
Datum nastanka	
Kraj nastanka	
Obdobje zbiranja	
Vsebinsko področje	CESSDA Vsebinsko področje
Vrsta podatka	Besedilni podatki
Jezik datoteke	
Primarni vir	
Licenca	

- dokumentacija mora omogočati novemu uporabniku razumevanje in možnost primerne uporabe originalnih podatkov.
- izpolnitev obrazca „**Opis raziskave**“
- ključen je podroben opis podatkovnih datotek!
 - obrazec „**Opis podatkov**“

* v primeru podatkovne datoteke, ki vsebuje podatke iz spletne strani, bloga

** v primeru podatkovne datoteke, ki vsebuje zvočne ali video posnetke

Obdelava in objava

- Skrben pregled predanih gradiv znotraj ADP
- Arhiviranje raziskave
 - Dokument „**Priporočila za prevzem kvalitativnih raziskav (ADP)**“ (2018): vključuje vrsto kvalitativnih podatkov (tipi formatov podatkovnih datotek) in priporočila za pripravo podatkov in metapodatkov za podatkovne arhiviste

Vrsta podatka	Prevzem Priporočeni formati	Prevzem Sprejemljivi formati	Formati za arhivsko hrambo	Formati za dostop
BESEDILNI PODATKI	.rtf ASCII (.txt) Xml text according to an appropriate Document Type Definition (DTD) or schema	HTML	PDF/A W3C XML ODF HTML WARC	PDF/A HTML
MIRUJOČA SLIKA	TIFF (.tif)	JPEG (.jpeg, .jpg, -jp2) GIF, TIFF, RAW, PSD, BMP, PNG, PDF/A, PDF	TIFF, JPEG, PNG, JPEG2000, SVG 2D v1.1 WRC, DWG 3D	TIFF, JPEG, JPEG2000, PNG, GeoTIFF, Primerno za prenos: JPEG, EPS, GIF

Dostop in uporaba podatkov

Cilj: prosto dostopne raziskave končnim uporabnikom Kataloga ADP → ideja odprte znanosti

Ob predaji dajalec podpiše „**Izjavo o izročitvi**“, ki določa pogoje dostopa do raziskave:

- podatki in dokumentacija uporabnikom na voljo pod pogoji *Creative Commons 4.0* licenc



Priznanje avtorstva

CC BY



Priznanje avtorstva-Nekomercialno

CC BY-NC

3 vrste dostopa v ADP:

- 1. Prosti dostop** (ni potrebno registrirati, uporaba omejena z zakonodajo, etičnimi pravili in avtorskimi pravicami)
- 2. Standarden dostop** (registracija, analiza in prenos podatkov, v celoti anonimizirani podatki – *PublicUseFile*)
- 3. Dostop pod posebnimi pogoji** (potrebno dovoljenje izvirnih avtorjev ker: o podatki morda niso v celoti anonimizirani – *ScientificUseFile*)
 - datoteka je pod embargom
 - datoteka dostopna le naročniku in izvirnim avtorjem.

Zaključek

- Razširitev vloge ADP ključna za ohranjanje in izmenjavo dragocenih kvalitativnih podatkov
- Izzivi v prihodnje:
 - nadaljnje dopolnjevanje postopkov, pravil in meril – prilagajanje različnim vrstam podatkov
 - razvoj načinov nadzorovanega omejenega dostopa do raziskav
 - razvoj spletnega orodja za prikaz in analizo kvalitativnih podatkov



Kontakt



*Hvala za vašo pozornost!
Vprašanja?*

Univerza v Ljubljani

Fakulteta za družbene vede

Arhiv družboslovnih podatkov

Kardeljeva ploščad 5

1000 Ljubljana



www.adp.fdv.uni-lj.si



arhiv.podatkov@fdv.uni-lj.si



Arhiv.Druzboslovnih.Podatkov



@ArhivPodatkov

Univerza v Ljubljani



CESSDA ERIC



seriss SYNERGIES FOR EUROPE'S
RESEARCH INFRASTRUCTURES
IN THE SOCIAL SCIENCES