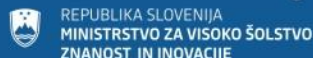


Raziskovalni podatki Provenienca v družboslovju in humanistiki

Andrej PANČUR, Inštitut za novejšo zgodovino

Ljubljana in online, Centralna tehniška knjižnica Univerze v Ljubljani,
Usposabljanje podatkovnih strokovnjakov, 14. – 17. 10. 2024



Različne definicije

- Raziskovalni podatki so definirani kot stvarni zapisi (numerični rezultati, besedilni zapisi, slikovno in zvočno gradivo), ki se uporabljajo kot primarni viri za namene znanstvenih raziskav in so v znanstveni skupnosti splošno sprejeti kot nujni za potrditev raziskovalnih izsledkov. Nabor raziskovalnih podatkov predstavlja sistematično, delno predstavitev raziskovalne tematike. Ta pojem ne vključuje: laboratorijskih dnevnikov, preliminarne analiz, osnutkov znanstvenih člankov, načrtov bodočih raziskav, strokovnih recenzij, komunikacije s kolegi in predmetov (npr. laboratorijskih vzorcev, bakterijskih sevov in testnih živali, kot so miši). (OECD Principles and Guidelines for Access to Research Data from Public Funding)
- Raziskovalni podatki so podatki, ki se uporabljajo kot primarni viri za podporo tehničnim ali znanstvenim raziskavam, akademskemu znanju ali umetniški dejavnosti, ki se uporabljajo kot dokazi v raziskovalnem procesu in/ali so v raziskovalni skupnosti splošno sprejeti kot nujni za potrjevanje raziskovalnih ugotovitev in rezultatov. Vse ostale digitalne in nedigitalne vsebine imajo potencial, da postanejo raziskovalni podatki. Raziskovalni podatki so lahko eksperimentalni podatki, podatki opazovanj, operativni podatki, podatki tretjih oseb, podatki javnega sektorja, podatki monitoringa, obdelani podatki ali ponovno uporabljeni podatki. (CODATA (Komisija za raziskovalne podatke pri Mednarodnem znanstvenem svetu))
- Podatki, pridobljeni z različnimi metodami, ki služijo za spoznavanje, preverjanje ali potrditev hipotez in izpeljavo zaključkov, ki so nastali ali bili obdelani v raziskavi. (Spoznav FAIR: priročnik o odprti znanosti v Sloveniji)

Raziskovalni podatki v družboslovju

Raziskovalne podatke lahko opredelimo glede na:

- Tipe raziskovalnih podatkov:
 - Viri podatkov: registri (npr. administrativni, zgodovinski, rezultati glasovanja, zdravstveni itd.), obstoječi raziskovalni podatki, populacijske skupine, komunikacije ipd.
 - Fizični formati podatkov: numerični, besedilni, slikovni, geoprostorski, avdio, video in programska oprema.
 - Kako so podatki ustvarjeni/zajeti: elektronski besedilni dokumenti, preglednice, laboratorijski zvezki, terenski zvezki in dnevniki, vprašalniki, prepisi in šifranti, avdio trakovi in video trakovi, fotografije in filmi, rezultati preiskav, primerki, vzorci, artefakti, diapozitivi, sheme baz podatkov, vsebina baz podatkov, modeli, algoritmi in skripte, poteki dela, standardni operativni postopki in protokoli, eksperimentalni rezultati, metapodatki in druge podatkovne datoteke, kot so npr. zapisi o pregledu literature in arhivi elektronske pošte.
 - Masivni podatki: 3 x V = Volume (Obseg: podatki so zelo veliki in da njihova obdelava zahteva veliko računalniško moč); Velocity (Hitrost: podatki se proizvajajo zaporedno in novi podatki nastajajo vsak trenutek); Variety (Raznolikost: podatki so nestrukturirani in neurejeni in zato niso pripravljeni za takojšnjo analizo).
- Formate raziskovalnih podatkov: Različni formati, v katerih so shranjene vrste podatkov (besedilni, numerični, multimedijski, strukturirani, programska koda itd.). Npr. statistični podatki so lahko shranjeni kot datotečni formati SPSS (*.sav) ali STATA, filmi kot *.mpg ali *.avi, strukturirani podatki kot *.xml ali v relacijski bazi podatkov MySQL in besedilne datoteke kot *.docx, *.pdf ali *.rtf.
- Velikost in kompleksnost: Velikost datotek je pomembna in prav tako zapletenost. Upravljanje razmeroma majhnega in preprostega nabora podatkov predstavlja drugačne izzive kot upravljanje velikih, kompleksnih podatkovnih datotek.
- Raziskovalne faze: Različne stopnje, skozi katere obdelujemo podatke: surovi, očiščeni, obdelani, analizirani podatki.

Kvantitativni podatki

- Pri kvantitativnem raziskovanju so zbrane informacije v numerični obliki. Kvantitativne raziskave se uporabljajo za kvantificiranje vedenja, stališč ali mnenj. Cilj kvantitativnih raziskav je pogosto preizkusiti ideje, navedene na začetku raziskave, oblikovati dejstva in odkriti vzorce.
- Podatki so izraženi v številkah, ki jih je mogoče oceniti s statističnimi analizami.
- Metode kvantitativnega zbiranja podatkov vključujejo različne oblike anket – spletne ankete, papirne ankete, mobilne ankete in ankete na kiosku, osebne intervjuje, telefonske intervjuje, prestreznike spletnih strani, spletne ankete, eksperimente in sistematična opazovanja. V večini primerov generalizira rezultate iz večje vzorčne populacije.

Kvalitativni podatki

- Kvalitativna raziskava je predvsem predhodna raziskava. Zbira informacije, ki niso v numerični obliki. Cilj kvalitativnih raziskav je pogosto razviti (nove) ideje in globlje razumevanje, ki ni dosegljivo s številčnimi rezultati.
- Podatki so izraženi v naravnem jeziku, pogosto besedilnem ali vizualnem.
- Kvalitativne metode zbiranja podatkov vključujejo fotografije, zvočne posnetke, video, nestrukturirane intervjuje, polstrukturirane intervjuje, odprte vprašalnike, dnevniške zapise, fokusne skupine (skupinske razprave), individualne intervjuje in nestrukturirana opazovanja. Velikost vzorca je običajno manjša od kvantitativnih vzorcev.

Katere vrste podatkov so dostopne:

- Makro podatki:
 - Agregirani:
 - O populaciji, o različnih skupinah, občinah, regijah in državah; običajno zgrajeni z združevanjem podatkov na nižjih enotah (npr. stopnja brezposelnosti, stopnja rodnosti).
 - Na sistemski stopnji:
 - Značilnosti enot višje ravni, kot sta država ali politični sistem, npr. volilni.
- Mezo podatki:
 - Se nanašajo na podatke o kolektivnih in družbenih akterjih, kot so podjetja, organizacije ali politične stranke.
- Mikro podatki:
 - Podatki posameznih enot (pogosto ljudi ali gospodinjev) pogosto iz anket, popisa ali administrativnih evidenc.

Vrste podatkov glede na časovno opredelitev:

- Presečno:
 - Ena časovna točka (posnetek časa)
 - Običajno informacije o več enotah in spremenljivkah
- Ponovljeno presečno:
 - Presečna raziskava ponovljena na novem vzorcu
 - Podatki na različnih vzorcih omogočajo analizo trendov
- Časovne vrste:
 - Niz podatkovnih točk v časovnem vrstnem redu (najpogosteje enakomerno razporejeni – npr. letna ponovitve)
 - Agregirani makro podatki so pogosto predstavljeni v časovnih točkah.
- Longitudinalno:
 - Sledenje istim enotam skozi čas, npr. študije o gospodinjskih panelih zbirajo informacije iz vzorca gospodinjestev v rednih "valovih" (primer tudi raziskava SHARE – 10 ponovitev)

Raziskovalni podatki v humanistiki?

- Zmeda glede definiranja raziskovalnih podatkov v humanistiki.
- “Raziskovalni podatki v humanistiki so med vsemi znanstvenimi disciplinami najrazličnejši, saj se skoraj vsi podatki o človeški dejavnosti lahko štejejo za raziskovalne podatke, vključno s časopisi, fotografijami, dnevniki, cerkvenimi zapisi, sodnimi spisi itd.”

Definiranje raziskovalnih podatkov v humanistiki

- Raziskovalni podatki v [digitalni] humanistiki so vsi viri in rezultati, ki se jih zbira, opisuje, vrednoti in/ali proizvaja v kontekstu raziskav v umetnosti in humanistiki in katere se lahko (dolgoročno) hrani v strojno berljivi obliki za namen arhiviranja, citiranosti in nadaljne uporabe. (DARIAH-DE)
- Ni možna enotna definicija.
- VSE, RAZEN ZNANSTVENIH PUBLIKACIJ: “definition of data encompasses all inputs and outputs of research that are not publications” ((ALLEA) Working Group on E-Humanities: Sustainable and FAIR Data Sharing in the Humanities)
- VSE: “build a bridge between SSH data and the EOSC, widening the concept of “research data” to include all of the types of digital research output linked to scholarly communication that are part of the research process” (OPERAS)

Raziskovalni podatki kot informacije javnega značaja

- 2022 novela ZDIJZ-G: implementirana Direktiva (EU) 2019/1024 Evropskega parlamenta in Sveta z dne 20. junija 2019 o odprtih podatkih in ponovni uporabi informacij avnega sektorja (t.i. Open Data Direktiva), ki nalaga državam članicam, da sprejmejo politike odprtega dostopa za podatke, pridobljene v javno financiranih raziskavah in zagotovijo, da takšne politike izvajajo vse raziskovalne organizacije in organizacije, ki raziskave financirajo.
- 3.a.(8) člen ZDIJZ-G: Izraz raziskovalni podatki pomeni podatke v digitalni obliki, razen znanstvenih publikacij, ki se zbirajo ali ustvarjajo v okviru znanstvenega raziskovanja in se uporabljajo kot dokaz v raziskovalnem procesu, ali pa je v raziskovalni skupnosti splošno sprejeto, da so potrebni za potrditev ugotovitev in rezultatov raziskav.
- 6.č(1): Organi zagotavljajo ponovno uporabo raziskovalnih podatkov s posredovanjem v svetovni splet z objavo v odprtih formatih, ki upoštevajo formalne odprte standarde, v strojno berljivi obliki, skupaj z metapodatki, razen kadar bi bilo to v nasprotju z izjemami od dostopa iz 5.a ali 6. člena tega zakona.
- 6.č(2): Ne glede na 3. točko šestega odstavka 6. člena tega zakona obveznost iz prejšnjega odstavka velja tudi za izvajalce javne službe na področju raziskovalne dejavnosti in izvajalce javne službe na področju izobraževalne dejavnosti, višje od srednješolske stopnje, glede raziskovalnih podatkov, ki so financirani iz javnih sredstev.

Zakaj in katere raziskovalne podatke je potrebno hraniti?

Upoštevamo dve perspektivi:

Znanstvena integriteta

Odprti dostop do znanstvenih publikacij:

2. člen Uredbe o izvajanju znanstvenoraziskovalnega dela v skladu z načeli odprte znanosti :

“zagotovijo odprti dostop do digitalnih različic znanstvenih publikacij in drugih rezultatov raziskav, ki so obravnavani v teh publikacijah ter so nujno potrebni za ponovitev raziskav

[ali za ponovno uporabo rezultatov raziskav v drugih raziskavah].”



Ponovna uporaba

Odprti dostop do raziskovalnih podatkov in drugih rezultatov raziskav v digitalni obliki:

4. in 5. člen Uredbe, (komentar k uredbi):

“Raziskovalci morajo v podatkovne repozitorije oddati tudi tiste /.../ podatke, ki niso vezani na nobeno objavljeno publikacijo, kadar jih zaradi svoje celovitosti in pomena za znanost lahko uporabijo drugi raziskovalci za izvajanje raziskav.”

Kaj so raziskovalni podatki?

- Vsi viri, ki se uporabljajo v raziskavah, razen znanstvenih publikacij:
 - Raziskovalni podatki so vse informacije, digitalne in nedigitalne, ustvarjene kot del znanstvenega procesa, na katerih temeljijo znanstveni zaključki.
 - Običajno ima vsak raziskovalec, ki izvaja izvirno raziskavo, raziskovalne podatke. Izvirna raziskava je raziskava, ki ne temelji izključno na povzetku, pregledu ali interpretaciji prejšnjih objav.
 - Tudi znanstvene publikacije kot raziskovalni podatki?
- V digitalni obliki:
 - Raziskovalni podatki v humanistiki so vsi viri in rezultati, ki se jih zbira, opisuje, vrednoti in/ali proizvaja v kontekstu humanističnih raziskav in katere se lahko hrani v strojno berljivi obliki za namen arhiviranja, citiranosti in ponovne uporabe.
 - Humanistika: večinoma uporaba kulturnih objektov kot so rokopisi, besedila, slike, posnetki ipd., ki se jih kot digitalne surogate nato lahko še dodatno obdela, vizualizira, označi, poveže in interpretira. GLAM ustanove: kulturna dediščina.

Raziskovalni rezultati:

- Raziskovalni podatki:
 - Kot pri družboslovju +
 - Objekti kulturne dediščine: rokopisi, tiskana besedila, slike, skulpture ipd.
 - Originalno digitalni ali (najbolj pogosto) digitalni nadomestki: digitalne reprodukcije materialnih objektov.
 - Datotečni formati (najbolj pogosti): TXT, CSV, XML, JPEG, TIFF, MP3, MP4.
 - Metapodatki (najbolj pogosti): Dublin Core, teiHeader, METS, MODS, LIDO, EAD, PREMIS, CIDOC CRM.
- Dodatni raziskovalni rezultati:
 - Raziskovalni podatki: Raziskovalci digitalne nadomestke praviloma še dodatno uredijo, prepišejo, označijo, povežejo, procesirajo, vizualizirajo in analizirajo. Najbolj običajna uporaba smernic TEI (Text Encoding Initiative).
 - Programska oprema: Raziskovalna programska oprema ter ustvarjanje prezentacijskih okolij in aplikacij, ki omogočajo interpretacijo podatkov, njihovo iskanje, filtriranje in povezovanje.
- Znanstvena in strokovna literatura, izvedbena dela:
 - o zgoraj naštetih raziskovalnih rezultatih (raziskovalnih podatkih, programski opremi).
 - o novih raziskovalnih spoznanjih, ki jih prinašajo novi podatki in metode.

Vrste raziskovalnih podatkov:

- **Neobdelani raziskovalni podatki:** so izvirni podatki, ki ste jih zbrali, vendar jih še niste obdelali ali analizirali. Na primer: zvočne datoteke, arhivi, opazovanja, zapiski s terena in podatki iz poskusov.
- **Obdelani raziskovalni podatki:** so podatki, ki ste jih digitalizirali, prevedli, prepisali, očistili, potrdili, preverili in/ali anonimizirali
- **Analizirani raziskovalni podatki:** so modeli, grafi, tabele, besedila ipd., ki ste jih ustvarili na podlagi neobdelanih in obdelanih podatkov in so namenjeni pomoči pri odkrivanju koristnih informacij, podajanju zaključkov in odločanju.

Raziskovalni podatki v raziskovalnem procesu

- Ker v humanistiki ni soglasja o tem, kaj so raziskovalni podatki v različnih humanističnih disciplinah, bi morala vsaka definicija upoštevati proces ustvarjanja podatkov v vsakokratni raziskovalni praksi.

Potek raziskovalnega dela

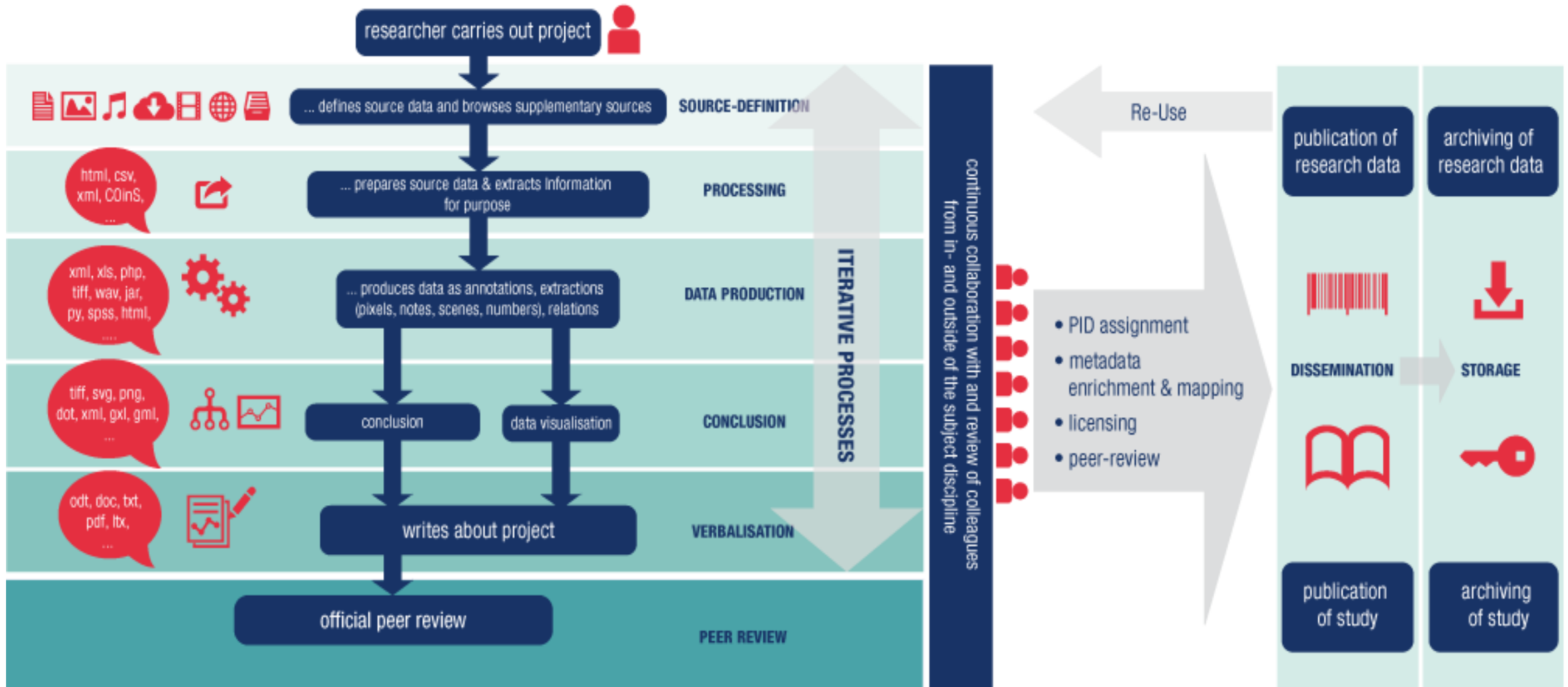
Analogen

- Načrtovanje:
 - razvoj raziskovalnega vprašanja,
 - opredelitev in oblikovanje poteka raziskave.
- Zbiranje:
 - določitev in izbor virov v arhivih, knjižnicah itd.
 - oblikovanje zbirke virov,
- Analiza:
 - določiti vzorce in strukture,
 - kontekstualizacija opažanj,
 - potrjevanje argumentov in hipotez.
- Objava.

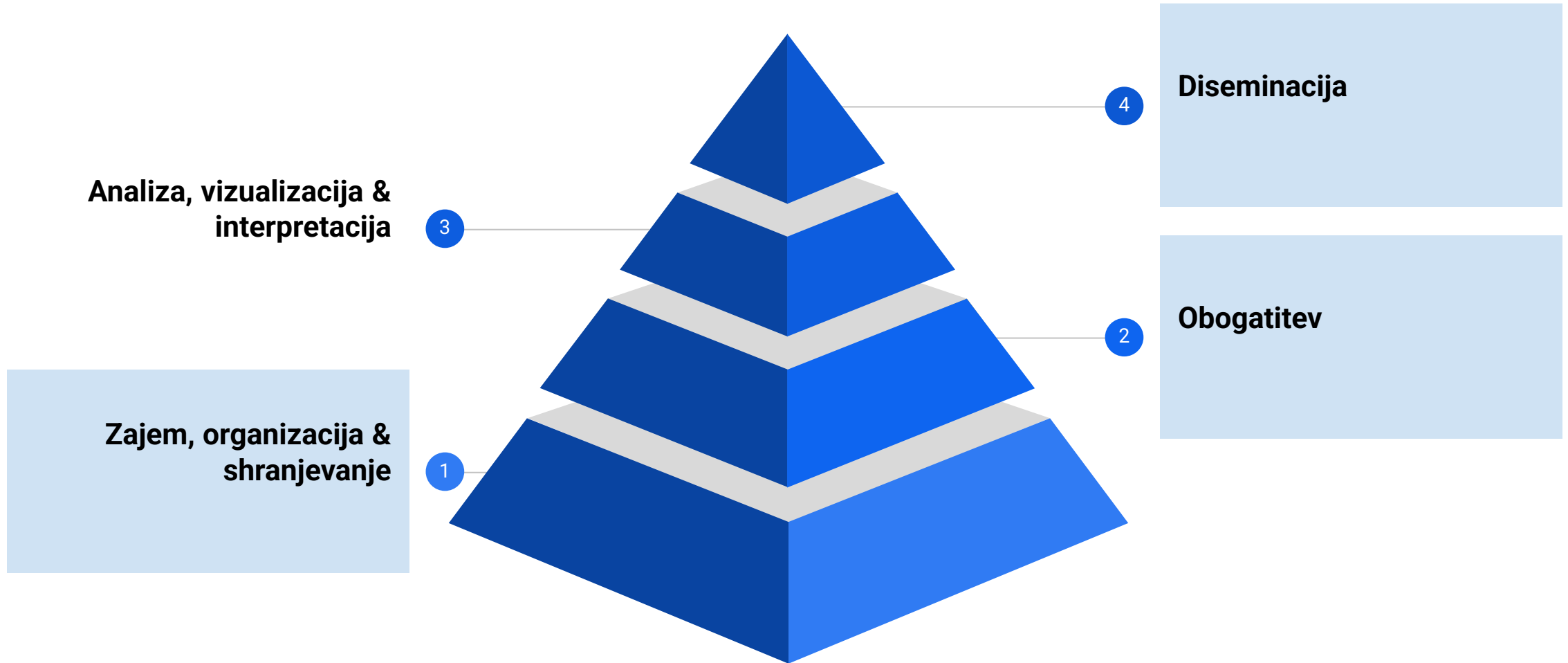
Digitalen

- Načrtovanje:
- Zbiranje:
- Digitalizacija:
 - digitalizacija dokumentov (v skladu s standardi za digitalizacijo),
 - zbiranje metapodatkov,
 - omogočiti dostop do digitalnih reprodukcij v skladu z načeli FAIR.
- Dopolnjevanje, obdelava in analiza:
 - ureditev metapodatkov,
 - prepis besedila,
 - dodati normirane podatke (kraje, osebe ipd.),
 - analiza z digitalnimi metodami in orodji.
- Digitalna objava.
- Hramba: v repozitorijih, v skladu z načeli FAIR.

Raziskovalni podatki v digitalni humanistiki v okviru življenjskega kroga raziskovalnih podatkov



Vse faze raziskovanja:



Primer arhivskih
virov:

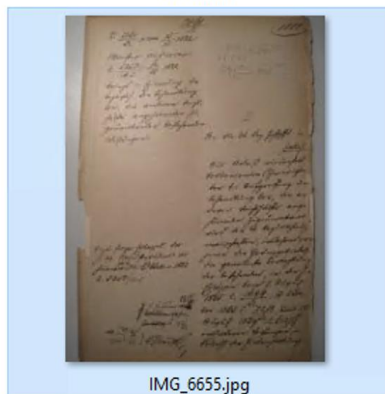
digitalizirani,

dodatni
metapodatki,

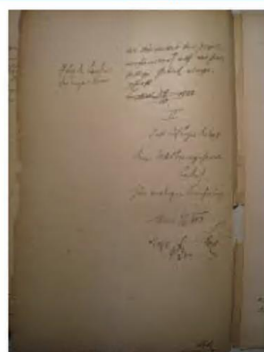
prepis



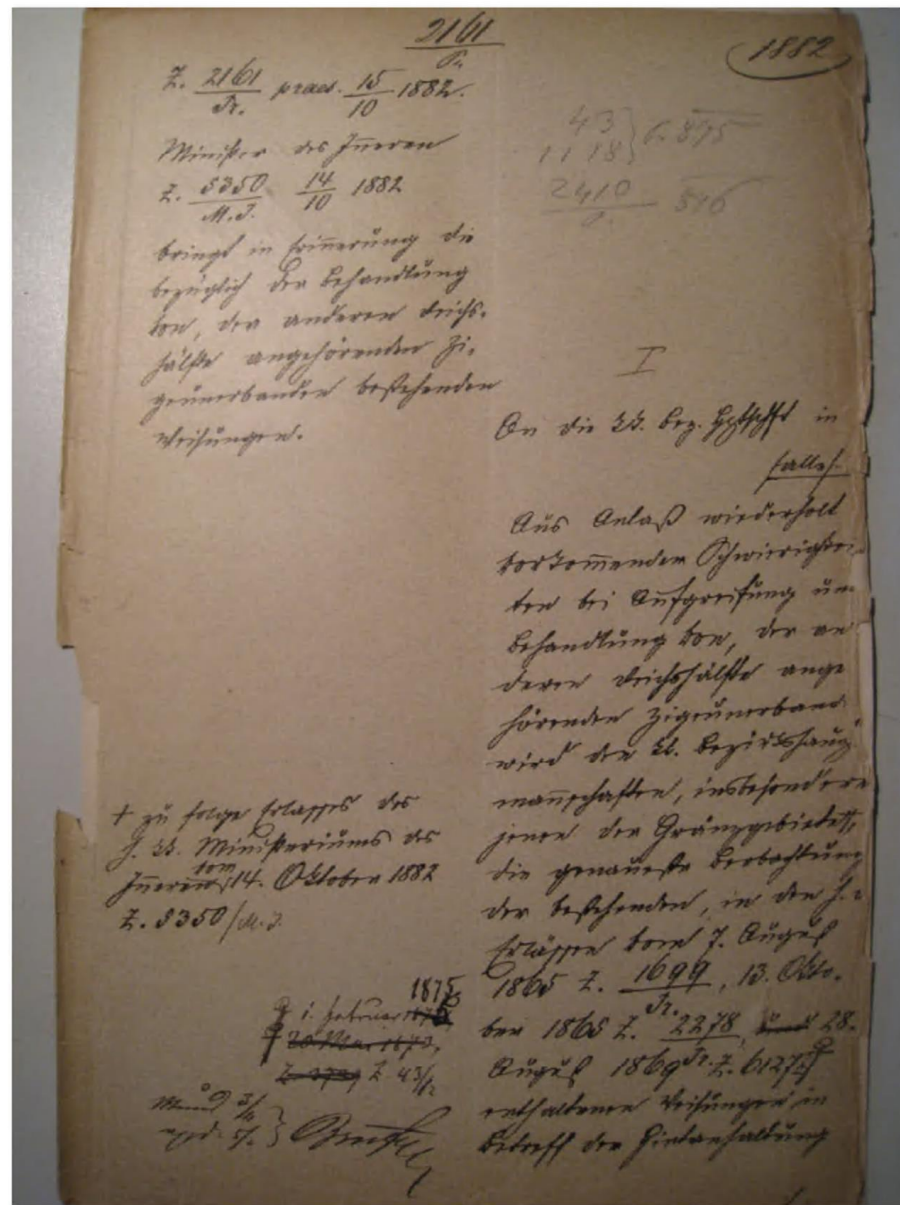
Doc1.doc



IMG_6655.jpg



IMG_6656.jpg



Projekt Slovenski literati in cesarska cenzura v dolgem 19. stoletju, zbirka Gledališka cenzura in uprizoritve
Dramatičnega društva v Ljubljani

C402																	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N			
1	Ime	Datum	Avtor	Naslov	Sprejeto	Zahtevek	Moralna	Cerkev	Državne in	Politika	Signatura	Uprizorjen	Faksimile	Opomba			
2	AS16/168/3047	14.2.1873	Adrien Decourcelle/Thiboust La	Na kosilu bom pri svoji materi	DA							15.11.1873					
3	AS16/168/885	17.4.1874	Salomon Hermann Mosenthal	Deborah	DA							19.4.1874					
4	AS16/168/48	17.1.1876	Ferdinand Raimund	Simit milijonar/ Zapravljujvec	DA							2.4.1877					
5	AS16/168/887	14.4.1876	Theodor Dunkel	Pariški potepuh	DA							30. 4. 1876					
6	AS16/168/2854	17.2.1876	H. Uhde	Ena se mora omožiti	Da							27. 11. 1876					
7	AS16/166/	12.3.1879	Jakob Alešovec	Nemški ne znajo	DA				DA			19. 12. 1885	DA	v mapi Theater			
8	AS16/168/43	4.1. 1887	Nikolaj Vasiljevič Gogolj	Revizor	DA							9. 1. 1887					
9	AS16/168/1100	27.4. 1887	Karl Görlitz	Popolna žena	DA							30. 4. 1887					
10	AS16/168a/1266	21.6. 1889	Josip Vošnjak	Svoji k svojim	DA	DA			DA	DA	C/DD709	29. 6. 1889	DA				
11	AS16/168/224	29.1.1889	Franz Nissel	Čarovnica pri jezeru	DA							27. 1. 1889					
12	AS16/168/563	9.3. 1889	August Kotzebue	Vrban Debeluhova ženitev	DA							17. 3. 1889					
13	AS16/168/1162	10.5. 1891	Josef Štolba	Testament	DA							NE					
14	AS16/168/2547	30. 9. 1891	Josef Štolba	Bratranec	DA							18. 10. 1891					
15	AS16/168/2593	30.10.1891	Georges Ohnet	Fužinar	DA							13. 2. 1892					
16	AS16/168/2596	3.10.1891	Gustav Moser	Knjižničar	DA							14. 10. 1894					
17	AS16/168/819	20.3. 1892	Henrik Ibsen	Nora	DA	DA						27. 3. 1892	DA				
18	AS16/168/3047	14.2.1893	Adrien Decourcelle/Thiboust La	Na kosilu bom pri svoji materi	DA							15. 4. 1877					
19	AS16/168/83176	23. 9. 1893	Albin Albini	Slednjič jo je vendarle pogodil	DA							NE					
20	AS16/168/885	17.4.1894	Salomon Hermann Mosenthal	Deborah	DA							19. 4. 1874					
21	AS16/168/2223	1.8.1894	Josef Freudenreich	Graničarji	DA							9. 10. 1894					
22	AS16/168/2229	1.8.1894	Giovanni Verga	Kmetски vitezi	DA							NE					
23	AS16/168/2230	1.8.1894	J.Böhm	Samo dve čašici	DA						DD/722	NE					
24	AS16/168/2231	1.8.1894	Friedrich Schiller	Viljem Tell	DA						C/DD669	NE					
25	AS16/168/2245	2.8.1894	Mihael Klappe	Rosenkranz in Güldenstern	DA							NE					
26	AS16/168/2246	2.8.1894	Neznani avtor	Amerikanski medved	DA							NE					
27	AS16/168/2247	2.8.1894	Émile Girardin	Svetohlinka	DA							NE					
28	AS16/168/2343	9.8.1894	Friedrich Wilhelm Hackländer	Monisieur de Blé	DA							NE					
29	AS16/168/2350	9.8.1894	Tihamér Almási	Po dveh letih	DA							NE					
30	AS16/168/2351	9. 8. 1894	Jean-François-Alfre Bayard/Van	Cunjar parižanski	DA							22. 12. 1894					
31	AS16/168/2431	18.8.1894	Avtor neznan	Oživela zakonska	DA							NE					
32	AS16/168/2432	18.8.1894	Emil Pohl	Sedem gavranov ali zakleti bra	DA							9. 10. 1900					

<<

Maskarada

>>

Naslov dela: Maskarada
Signatura ARS: AS/16/168/10603
Datum dokumenta: 22.11.1895
Avtor cenzurirane drame: Alexandre-Charles-Auguste Bisson
Cenzura: POGOJNO ODOBRENO
Področje cenzure - morala: DA
Datum uprizoritve: 23. 11. 1895

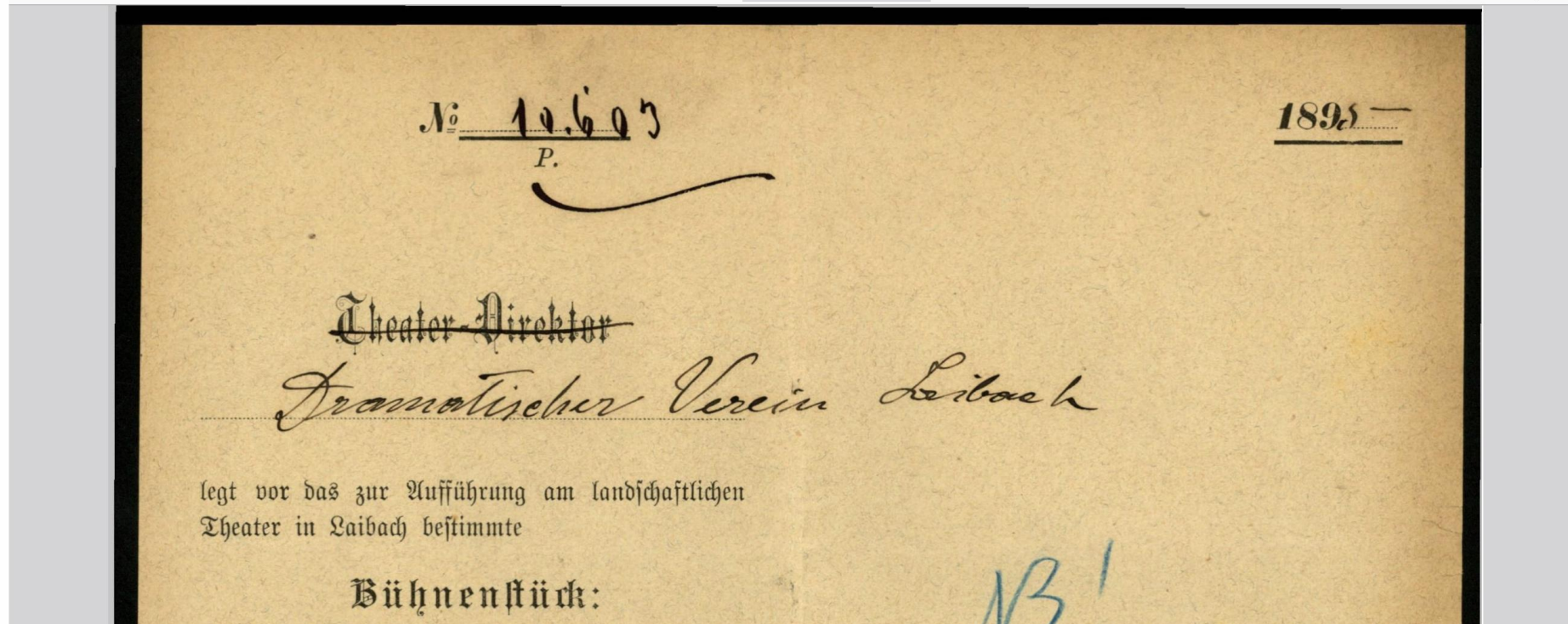


1

od 1



Samodejno ▾



Primer prepisa:

The Letters of Count Ignaz Maria von Attems-Heiligenkreuz to his Father (1734–1738)

https://dariah-si.github.io/Attems/attems.1.08.html

eZMono

Cover PageColophons ▼TOCFrontLetters ▼Studies ▼

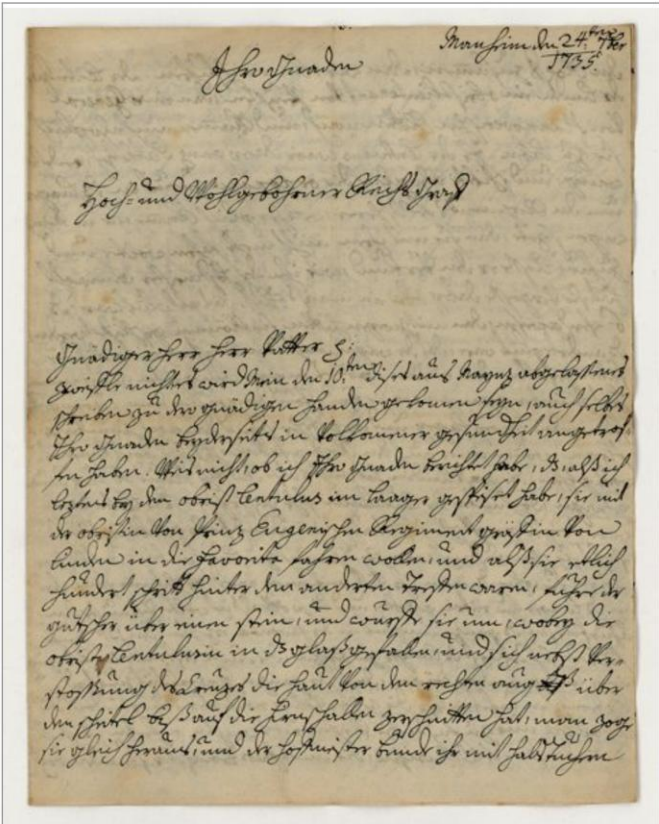
Your search text

<<8. Mannheim, 24. 9. 1735>>

den 24.^{ter} 7ber 1735.
Ihro Gnaden
Hoch- und Wohlgebohrner Reichs Graff

Gnädiger herr herr Vatter **etc.**

[1] Zweiffle nicht, es wird Mein den 10.^{ten} dises aus Maynz abgelassenes schreiben zu dero gnädigen handen gekommen seyn, auch selbes Ihro Gnaden beyderseits in Volkomener gesundheit angetroffen haben. Weis nicht, ob ich Ihro Gnaden berichtet habe, **das**, alß ich leztens bey den obrist **lentulus** im **Lager** gespeiset habe, sie mit der obristin von Prinz **Eugenischen** Regiment gräffin Von linden in die **favorita** fahren wollen, und alß sie etlich hundert schritt hinter dem anderten Treffen waren, fuhre der **Kutscher** über einen stein, und **warf** sie um, wobey die obrist **lentulus** in **das** glaß gefallen und sich nebst Verstoßung des Creuzes die haut Von dem rechten aug **biß** über den scheitel biß auf die **Hirnschale** zerschnitten hat, man zoge sie gleich heraus, und der hoffmeister bande ihr mit halsTuchern

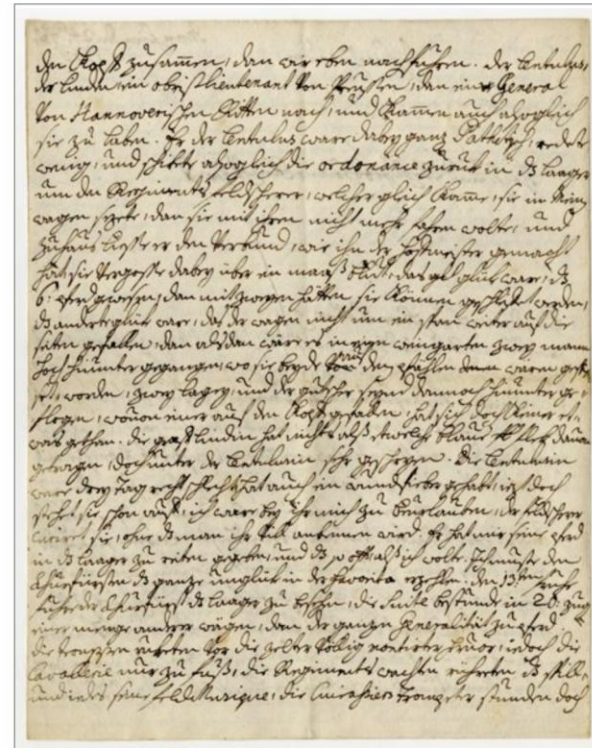


Arhiviranje prezentacije – spletne izdaje



stosßung des Creuzes die haut Von dem rechten aug **biß** über
den scheitel biß auf die **Hirnschale** zerschnitten hat, man zoge
sie gleich heraus, und der hoffmeister bande ihr mit halsTuchern

den Kopff zus**amm**en, dan wir eben nachfuhren. der *lentulus*,
der linden, ein obristlieutenant Von Preussen, dan eine General
Von Hannoverischen Ritten nach, und **Kamm**en auch alsogleich
sie zu laben. Er der *lentulus* ware dabey ganz *Pathetisch*, redete
wenig, und schikete alsogleich die *ordonance* **zurück** in das laager
um den Regiments feldscherer, welcher gleich **Kamme**, sie in Meinen
wagen **setzte**, dan sie mit ihren nicht mehr fahren wolte, und
zuhaus liesße er den **Verband**, wie ihn der hoffmeister gemacht
hat, sie Vergosse dabey über ein maaß blut, das **gl** glück ware, das
6 pferd gewesen, dan mit zweyen hätten sie Können **geschleift** werden,
das anderte glück ware, das der wagen nicht um ein **Spann(e)** weiter auf die
seiten gefallen, dan als dan wäre es in einen weingarten zwey mann
hoch hinunter gegangen, wo sie beyde **Von** auf den**en** pfählen **denen** wären **aufge-**
spießt worden, zwey **Lakaen** und der gutscher seynd dannoch hinunter ge-
flogen, wovon einer auf den Kopff gefallen, hat sich doch Keiner et-
was gethan. die graff lindin hat nichts alß etwelche blaue **#** fleck dauon
getragen, doch unter der *lentulusin* sehr geschryen. Die *lentulusin*
ware drey Tag recht schlecht, hat auch ein wundfieber gehabt, **jetzt** doch
stehet sie schon auff, ich ware bey ihr mich zu beurlauben, der feldscherer
curiret sie, ohne **das** man ihr Vill ankennen wird. Er hat mir seine pferd
in das laager zu reiten gegeben, und **das** so oft, alß ich wolte. Ich muste den
Churfürsten **das** ganze unglück in der *favorita* erzehlen. den 13.^{ten} **Früh**
fuhre der Churfürst **das** laager zu besehen, die *suite* bestunde in 20 zug
einer menge andrer wägen, dan der ganzen *Generalität* zu pferd.
die **Truppen rückten** Vor die zelter Völlig montirter hervor, jedoch die
Cavallerie nur zu fuß, die Regiments wachten rühreten **das** spill,
und jedes seine *feldmusique*, die *cuirassiers* Trompeter **standen** doch



Raziskovalni podatki:

```
<div type="letter" xml:id="attems.1.08" n="8">
  <head xml:id="attems-head-12">Mannheim, 24. 9. 1735</head>
  <pb xml:id="attems-dipl-pb-27" facs="#facs-08-1_re"/>
  <opener xml:id="attems-opener-8">
    <dateline>
      <placeName>
        <name type="place">Manheim</name>
      </placeName> den <date when="1735-09-24">24.<hi rend="superscript">ter</hi>
        <hi rend="italic">7ber</hi> 1735.</date>
    </dateline>
    <salute>
      <name type="person">Ihro Gnaden <lb/>Hoch- und Wohlgebohrner Reichs Graff </name>
    </salute>
    <lb/>
    <salute>
      <name type="person">Gnädiger herr herr Vatter <choice>
        <abbr>s:</abbr>
        <expan>etc.</expan>
      </choice>
    </name>
    </salute>
  </opener>
  <p xml:id="attems-p-18">Zweiffle nicht, es wird Mein den 10.<hi rend="superscript">
    >ten</hi> dises aus <name type="place">Maynz</name> abgelassenes <lb/>schreiben zu
    dero gnädigen handen gekomen seyn, auch selbes <lb/>
    <name type="person">Ihro Gnaden beyderseits</name> in Volkomener gesundheit
    angetrof-<lb/>fen haben. Weis nicht, ob ich <name type="person">Ihro Gnaden</name>
    berichtet habe, d<expan>a</expan>s, alß ich <lb/>leztens bey den <name type="person">
      >obrist <hi rend="italic">lentulus</hi>
    </name> im <choice>
      <orig>laager</orig>
      <reg>Lager</reg>
    </choice> gespeiset habe, sie mit <lb/>der <name type="person">obristin von Prinz <hi
      rend="italic">Eugeni</hi>schen Regiment gräffin Von <lb/>linden</name> in die
```


Kaj je provenienca raziskovalnih podatkov?

- Pojem provenienca oz. izvor raziskovalnih podatkov se nanaša na vse informacije o okoliščinah nastanka podatkov: zakaj in kako so bili podatki proizvedeni, kje, kdaj in kdo jih zbira.
- Metapodatki o poreklu podatkov segajo od človeku lahko berljivih do zelo tehničnih in običajno zahtevajo nekaj znanja o domeni za ustvarjanje.
- Metapodatki o izvoru podatkov omogočajo interpretacijo in ponovno uporabo podatkov; gradi zaupanje, verodostojnost in ponovljivost.
- Pomembno, ker:
 - uporabnik podatkov ni nujno tudi izdelovalec podatkov,
 - pomembno za določanje kakovosti, stopnje zaupanja v rezultate, ponovljivosti rezultatov in ponovne uporabe podatkov,
 - zagotavljajo verodostojnosti in zanesljivosti njihovih podatkov.

Postopek

- Osnovne vrste metapodatkov o provenienci: datum ustvarjanja, ustvarjalec, instrument ali uporabljena programska oprema, metode obdelave podatkov.
- Zajem in vzdrževanje metapodatkov o provenienci bi moralo biti običajen del procesov raziskovanja in upravljanja podatkov:
 - Zapisano vsaj v prostem besedilu: besedilna datoteka README, ki mdr. opisuje tudi uporabljene metode zbiranja in obdelave podatkov.
 - Na bolj strukturiran način z uporabo generičnih metapodatkovnih standardov (npr. Dublin Core). Uporaba metapodatkovnih standardov za posamezne discipline, kot je ISO 19115-2.
 - Uporaba zelo abstraktnih podatkovnih modelov, kot sta W3C Provenance Data Model (PROV-DM) in Provenance Ontology (PROV-O).

Dojemanje provenience v humanistiki

- Sprememba lastništva statičnega objekta: Objekti kulturne dediščine: Provenienca kot zabeležena zgodovina (sprememb) lastništva objektov kulturne dediščine. Provenienca kot zgodovina lastništva cenjenega predmeta, kot je umetniško delo. Provenienca zagotavlja dokumentirano zgodovino, ki lahko pomaga dokazati lastništvo, dodeliti delo znanemu umetniku in ugotoviti pristnost umetniškega dela.
- Transformacija raziskovalnih podatkov: Provenienca raziskovalnih podatkov sledi transformacijam v različico nabora podatkov, ki nato ustvari nadaljnje različice. Zapisi o poreklu podatkov lahko nato vključujejo informacije o avtorstvu, ime in različico programske opreme, uporabljene za izdelavo transformacij, ter opise samih transformacij.

Transformacija raziskovalnih podatkov na dokumentih

- Besedilni dokumenti: v digitalni humanistiki pogosta uporaba programa Git za kontrolo verzij. Uporaba platform GitHub in GitLab.
- Možnost natančnega beleženja revizij dokumenta je sestavni del TEI standarda za kodiranje besedil:
 - Zagotavlja bistvene informacije za upravljanje velikega števila datotek, ki se posodablajo, popravljajo ali drugače spreminjajo, kot tudi izjemno uporabno dokumentacijo za datoteke, ki se prenašajo od raziskovalca do raziskovalca ali sistema do sistema.
 - Brez beleženja sprememb lahko hitro pride do zamenjav različnih različic datotek ali pa se spregleda majhne, vendar pomembne spremembe, ki jih je v datoteki naredil neki prejšnji člen v distribucijski verigi.

```
<!-- ... --><revisionDesc>
  <change n="RCS:1.39" when="2007-08-08"
    who="#jwernimo.lrv">Changed <val>drama.verse</val>
      <gi>lg</gi>s to <gi>p</gi>s. <note>we have opened a discussion about the need for a new
        value for <att>type</att> of <gi>lg</gi>, <val>drama.free.verse</val>, in order to address
        the verse of Behn which is not in regular iambic pentameter. For the time being these
        instances are marked with a comment note until we are able to fully consider the best
            way
        to encode these instances.</note>
  </change>
  <change n="RCS:1.33" when="2007-06-28"
    who="#pcaton.xzc">Added <att>key</att> and <att>reg</att>
      to <gi>name</gi>s.</change>
  <change n="RCS:1.31" when="2006-12-04"
    who="#wgui.ner">Completed renovation. Validated.</change>
</revisionDesc>
```

Zapisniki sej parlamenta:

Od masivnih podatkov (big data) do pametnih podatkov (smart data): masivni pametni podatki (big smart data)

Masivni podatki

- metapodatki;
- nestrukturirani podatki;
- PDF:
 - dLib: kranjski deželni zbor (903)
 - Slstory: senat in skupščina Kraljevine SHS oz. Kraljevine Jugoslavije (830)
 - Slstory: Ljudska skupščina LRS (35)
 - Slstory: Skupščina SRS (2.159)
- HTML: Državni zbor RS (več kot 11.620).

Pametni masivni podatki

- metapodatki;
- delno strukturirani podatki;
- XML TEI;
- Slovenian parliamentary corpus (1990-2022) siParl 3.0:
 - 11.620 dokumentov
 - 1.134.287 govorov
 - 213.579.732 besed

Spletna stran Državnega zbora: primer zapisnika seje delovnega telesa



REPUBLIKA SLOVENIJA
DRŽAVNI ZBOR



| Poslanke in poslanci ▾ | Seje ▾ | Zakonodaja ▾ | Druge dejavnosti ▾ | O Državnem zboru ▾ | Informacijsko središče ▾ |

> Poslovnik Državnega zbora > Notranji red > Predlogi zakonov in aktov > Peticije > Fotogalerija > 30. obletnica Državnega zbora > Kontakt

ODBOR DRŽAVNEGA ZBORA REPUBLIKE SLOVENIJE ZA KMETIJSTVO, GOZDARSTVO IN PREHRANO

25. seja
(8. december 1998)

Sejo je vodil Franc Potočnik, predsednik odbora.
Seja se je pričela ob 18.14 uri.

PREDSEDNIK FRANC POTOČNIK: ... državnega zbora Republike Slovenije za kmetijstvo, gozdarstvo in prehrano. Na naši sredi posebej pozdravljam državnega sekretarja gospoda Mohoriča(?). Dnevni red, predlog dnevnega reda je: dopolnitev delne pobude za pogajanje sklenitev mednarodne pogodbe o članstvu evropske unije za področje delovne skupine osem, to je ribištvo in pa točka razno. Še preden preidemo na razpravo o dnevnem redu naj povem, da sta se za današnjo sejo upravičila gospod Jerovšek in gospod Čebulj. Da pa je po pooblastilu vodje poslanske skupine namesto Andreja Fabjana na seji prisoten gospod Štefan Klinc. Želi kdo o dnevnem redu razpravljati? V kolikor ne ga dajem na glasovanje.

Kdo je za?

Soglasno je dnevni red sprejet. Uvodno besedo dajem gospodu Mohoriču, državnemu sekretarju na ministrstvu za kmetijstvo, gozdarstvo in prehrano, ki je tudi odgovoren za področje, med drugim tudi za področje ribištva.

GOSPOD MOHORIČ: Hvala gospod predsednik. Na kratko bi želel predstaviti pogajalska izhodišča Republike Slovenije za področje številka 8, to je ribištvo. Slovensko ribištvo, ki zajema tako ribolov na teritorialnem morju, ribolov na odprtem morju, mari kulturo, gojenje školjk in gojenje rib. Potem je predelava, trženje,

v HTML formatu

```
2082 </table><div id="viewns_Z7_J9KAJKG10G5G80QTOKRJHB08M0_:form1:fieldSet1" style="margin-top:24px" class="fieldset"><h5 class='label'><span>Vsebina zapisa seje (v pregledu)</span></h5><div align="cent
2083 <p><b><font size="4" face="Arial CE">ODBOR DR&#381;AVNEGA ZBORA REPUBLIKE SLOVENIJE </font></b><br>
2084 <b><font size="4" face="Arial CE">ZA KMETIJSTVO, GOZDARSTVO IN PREHRANO</font></b></div><div align="center"><b><font size="4" face="Arial CE">25. seja</font></b></div><div align="center"><b><font s
2085 <br>
2086 <b><font size="4" face="Arial CE"> Sejo je vodil Franc Poto&#269;nik, predsednik odbora.</font></b><br>
2087 <b><font size="4" face="Arial CE"> Seja se je pri&#269;ela ob 18.14 uri.</font></b><br>
2088 <br>
2089 <br>
2090 <b><font size="4" face="Arial CE"> PREDSEDNIK FRANC POTO&#268;NIK:</font></b><font size="4" face="Arial CE"> ... dr&#382;avnega zbora Republike Slovenije za kmetijstvo, gozdarstvo in prehrano. Na
2091 <font size="4" face="Arial CE"> Kdo je za? </font><br>
2092 <font size="4" face="Arial CE"> Soglasno je dnevni red sprejet. Uvodno besedo dajem gospodu Mohori&#269;u, dr&#382;avnemu sekretarju na ministrstvu za kmetijstvo, gozdarstvo in prehrano, ki je tudi
2093 <br>
2094 <font size="4" face="Arial CE"> </font><b><font size="4" face="Arial CE">GOSPOD MOHORI&#268;: </font></b><font size="4" face="Arial CE">Hvala gospod predsednik. Na kratko bi &#382;elel predstaviti
2095 <font size="4" face="Arial CE"> Slovensko ribi&#353;tvo, ki zajema tako ribolov na teritorialnem morju, ribolov na odprtem morju, mari kulturo, gojenje &#353;koljk in gojenje rib. Potem je predelav
2096 <font size="4" face="Arial CE"> Evropa je razdelila po evropskih razdelitvah je to podro&#269;je razdelila na &#353;tiri poglavja z vidika zakonodaje in z vidika upravnih aktov. In sicer prvo pogl
2097 <font size="4" face="Arial CE"> Poleg tega bo potrebno postaviti vse in&#353;titucije, ki bodo zagotavljale izvajanje te zakonodaje, kakor tudi zagotoviti ustrezna finan&#269;na sredstva in pa kadri
2098 <br>
2099 <b><font size="4" face="Arial CE">2. TRAK: (TS)</font></b><br>
2100 <br>
2101 <b><font size="4" face="Arial CE"> </font></b><font size="4" face="Arial CE">(nadaljevanje) Kot veste imamo te kvote ima Unija z direktivo opredeljene kvote za severno morje, za Atlantik, Biskajsk
2102 <font size="4" face="Arial CE"> Zakonodaja, ki to pokriva je sedaj veljavni zakon o morskem ribi&#353;tvu. Potrebne statisti&#269;ne podatke pridobivamo iz naslova oziroma v okviru zakona o statisti
2103 <font size="4" face="Arial CE"> Zaradi racionalizacije nalog iz vidika izvajanja zakonodaje na podro&#269;ju morju, je v pripravi organizacija medresorskega nadzornega centra. Center, to je podobnc
2104 <br>
2105 <b><font size="4" face="Arial CE">3. TRAK: (TS)</font></b><br>
2106 <br>
2107 <b><font size="4" face="Arial CE"> </font></b><font size="4" face="Arial CE">(nadaljevanje) Ministrstvo za kmetijstvo in gozdarstvo je &#382;e pripravilo v vladi predlog sklepa, da se za te naloge
2108 <font size="4" face="Arial CE"> Potem so strokovni ukrepi in ukrepi dr&#382;avne pomo&#269;i. Mi v okviru prora&#269;una &#382;e izvajamo dolo&#269;ene strukturne ukrepe, ki so po vsebini primerljji
2109 <font size="4" face="Arial CE"> Delno so na&#353;i ukrepi primerljivi. Nekateri pa &#353;e niso in jih bo potrebno uskladiti. Na tem podro&#269;ju sem &#382;e enkrat omenil, ... (Ni razumljivo.) ..
2110 <br>
2111 <b><font size="4" face="Arial CE">4. TRAK: (DAG) </font></b><br>
2112 <br>
2113 <b><font size="4" face="Arial CE"> </font></b><font size="4" face="Arial CE">(Nadaljevanje) tudi Evropska unija. </font><br>
2114 <font size="4" face="Arial CE"> Ukrepi na podro&#269;ju strukturnih oziroma to podro&#269;je bi izvajala agencija za razvoj pode&#382;elja kot organ v sestavi ministrstva. Kajti vedno mora Evropa i
2115 <font size="4" face="Arial CE"> Mednarodni sporazumi. Mi imamo podpisan en mednarodni sporazum z Republiko Hrva&#353;ko, vendar se tega sporazuma ne izvaja. V glavnem ribi&#269;i posredujejo inform
2116 <font size="4" face="Arial CE"> Gospod predsednik, zaenkrat bi bila moja informacija v tem obsegu, &#269;e bodo vpra&#353;anja, bom pa posku&#353;al odgovoriti. </font><br>
2117 <br>
2118 <font size="4" face="Arial CE"> </font><b><font size="4" face="Arial CE">PREDSEDNIK FRANC POTO&#268;NIK: </font></b><font size="4" face="Arial CE">Hvala lepa za uvodno predstavitev problematike. Zc
2119 <br>
2120 <font size="4" face="Arial CE"> </font><b><font size="4" face="Arial CE">MARIO GASPARINI: </font></b><font size="4" face="Arial CE">&#268;isto na kratko. Na tistem na&#353;em ve&#269;jem igri&#353;
2121 <font size="4" face="Arial CE"> Ampak hotel sem dodati &#353;e nekaj drugega. S Hrva&#353;ko imamo sporazum, samo se ga ne izvaja zaradi povsem tehni&#269;nih zadev in tako naprej. </font><br>
2122 <font size="4" face="Arial CE"> </font><br>
2123 <b><font size="4" face="Arial CE">5. TRAK (M&#381;)</font></b><br>
2124 <br>
2125 <font size="4" face="Arial CE"> (Nadaljevanje) Bi pa zelo pomagalo, &#269;e bi mi sporazum s Hrva&#353;ko sprejeli, ta maloobmejni sporazum, ki smo ga imeli na 10. seji, pa na 12., pa na 13., pa na
2126 <br>
2127 <font size="4" face="Arial CE"> </font><b><font size="4" face="Arial CE">PREDSEDNIK FRANC POTO&#268;NIK:</font></b><font size="4" face="Arial CE"> Hvala. &#352;e &#382;eli kdo? Gospod Bo&#382;i&#269;
2128 <br>
2129 <font size="4" face="Arial CE"> </font><b><font size="4" face="Arial CE">IVAN BO&#381;I&#268;:</font></b><font size="4" face="Arial CE"> Samo majhno vpra&#353;anje, ker nisem ribi&#269;, ampak pon
2130 <br>
2131 <font size="4" face="Arial CE"> </font><b><font size="4" face="Arial CE">PREDSEDNIK FRANC POTO&#268;NIK:</font></b><font size="4" face="Arial CE"> Ja, prosim, gospod Ili&#269;.</font><br>
2132 <br>
```

pretvorba v XML format: TEI modul za dramska besedila

```
</profileDesc>
</teiHeader>
<text>
  <body>
    <div>
      <stage type="title">ODBOR DRŽAVNEGA ZBORA REPUBLIKE SLOVENIJE ZA KMETIJSTVO, GOZDARSTVO IN PREHRANO</stage>
      <stage type="session">25. seja</stage>
      <stage type="date">(8. december 1998)</stage>
      <stage type="chairman">Sejo je vodil Franc Potočnik, predsednik odbora.</stage>
      <stage type="time">Seja se je pričela ob 18.14 uri.</stage>
      <sp who="SDT2:PotočnikFranc">
        <speaker>PREDSEDNIK FRANC POTOČNIK:</speaker>
        <p><gap n="..." /> državnega zbora Republike Slovenije za kmetijstvo, gozdarstvo in prehrano. Na
naši sredi posebej pozdravljam državnega sekretarja gospoda Mohoriča<stage>(?)</stage>. Dnevni red,
predlog dnevnega reda je: dopolnitev delne pobude za pogajanje sklenitev mednarodne
pogodbe o članstvu evropske unije za področje delovne skupine osem, to je ribištvo in
pa točka razno. Še preden preidemo na razpravo o dnevnem redu naj povem, da sta se za
današnjo sejo upravičila gospod Jerovšek in gospod Čebulj. Da pa je po pooblastilu
vodje poslanske skupine namesto Andreja Fabjana na seji prisoten gospod Štefan Klinc.
Želi kdo o dnevnem redu razpravljati? V kolikor ne ga dajem na glasovanje.</p>
<p>Kdo je za?</p>
<p>Soglasno je dnevni red sprejet. Uvodno besedo dajem gospodu Mohoriču, državnemu
sekretarju na ministrstvu za kmetijstvo, gozdarstvo in prehrano, ki je tudi odgovoren
za področje, med drugim tudi za področje ribištva.</p>
      </sp>
      <sp who="SDT2:unknown-M">
        <speaker>GOSPOD MOHORIČ:</speaker>
        <p>Hvala gospod predsednik. Na kratko bi želel predstaviti pogajalska izhodišča
Republike Slovenije za področje številka 8, to je ribištvo.</p>
        <p>Slovensko ribištvo, ki zajema tako ribolov na teritorialnem morju, ribolov na
odprtem morju, mari kulturo, gojenje školjk in gojenje rib. Potem je predelava,
```

pretvorba v končni dokument TEI: del skupnega korpusa besedil (TEI modul za prepise govorov)

```
62      </tagsvec>
63    </encodingDesc>
64    <profileDesc>
65      <abstract>
66        <list type="agenda">
67          <item xml:id="OZKGIP-Redna-025-1998-12-08.toc-item0">
68            <title>Pred dnevnim redom</title>
69            <ptr target="#OZKGIP-Redna-025-1998-12-08.seg1"/>
70            <ptr target="#OZKGIP-Redna-025-1998-12-08.seg2"/>
71            <ptr target="#OZKGIP-Redna-025-1998-12-08.seg3"/>
72            <ptr target="#OZKGIP-Redna-025-1998-12-08.seg4"/>
73          </item>
74        </list>
75      </abstract>
76    </profileDesc>
77  </teiHeader>
78  <text>
79    <front>
80      <div type="preface">
81        <head>ODBOR DRŽAVNEGA ZBORA REPUBLIKE SLOVENIJE ZA KMETIJSTVO, GOZDARSTVO IN PREHRANO</head>
82        <head type="session">25. seja</head>
83        <note type="chairman">Sejo je vodil Franc Potočnik, predsednik odbora.</note>
84      </div>
85    </front>
86    <body>
87      <div>
88        <note type="time">Seja se je pričela ob 18.14 uri.</note>
89        <note type="speaker">PRESEDIK FRANC POTOČNIK:</note>
90        <u who="#PotočnikFranc"
91          xml:id="OZKGIP-Redna-025-1998-12-08.u1"
92          ana="#chair">
93          <seg xml:id="OZKGIP-Redna-025-1998-12-08.seg1">
94            <gap reason="inaudible"/> državnega zbora Republike Slovenije za kmetijstvo, gozdarstvo in prehrano. Na naši sredi posebej pozdravljam d
95          </seg>
96          <note?</note>
97          <seg xml:id="OZKGIP-Redna-025-1998-12-08.seg2">. Dnevni red, predlog dnevnega reda je: dopolnitev delne pobude za pogajanje sklenitev medna
98          </seg>
99          <seg xml:id="OZKGIP-Redna-025-1998-12-08.seg3">Kdo je za?</seg>
100         <seg xml:id="OZKGIP-Redna-025-1998-12-08.seg4">Soglasno je dnevni red sprejet. Uvodno besedo dajem gospodu Mohoriču, državnemu sekretarju n
101       </u>
102     </div>
103   </body>
104 </text>
105 </teiText>
106 </teiDocument>
```

lingvistično označeni govori korpusa (tokenizacija, lematizacija, oblikoslovno označevanje, imenske entitete)

```
131 <div type="preface">|
132 <head>ODBOR DRŽAVNEGA ZBORA REPUBLIKE SLOVENIJE ZA KMETIJSTVO, GOZDARSTVO IN PREHRANO</head>
133 <head type="session">25. seja</head>
134 <note type="chairman">Sejo je vodil Franc Potočnik, predsednik odbora.</note>
135 </div>
136 </front>
137 <body>
138 <div>
139 <note type="time">Seja se je pričela ob 18.14 uri.</note>
140 <note type="speaker">PRESEDNIK FRANC POTOČNIK:</note>
141 <u who="#PotočnikFranc" xml:id="OZKGIP-Redna-025-1998-12-08.u1" ana="#chair">
142 <seg xml:id="OZKGIP-Redna-025-1998-12-08.seg1">
143 <s xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1">
144 <gap reason="inaudible"/>
145 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t1" ana="mte:Agpmsg" msd="UposTag=ADJ|Case=Gen|Degree=Pos|Gender=Masc|Number=Sing" lemma="državen">državnega</w>
146 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t2" ana="mte:Ncmmsg" msd="UposTag=NOUN|Case=Gen|Gender=Masc|Number=Sing" lemma="zbor">zbor</w>
147 <name type="org">
148 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t3" ana="mte:Ncfsg" msd="UposTag=NOUN|Case=Gen|Gender=Fem|Number=Sing" lemma="republika">Republike</w>
149 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t4" ana="mte:Npfs" msd="UposTag=PROPNI|Case=Gen|Gender=Fem|Number=Sing" lemma="Slovenija">Slovenije</w>
150 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t5" ana="mte:Sa" msd="UposTag=ADP|Case=Acc" lemma="za">za</w>
151 <w join="right" xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t6" ana="mte:Ncnsa" msd="UposTag=NOUN|Case=Acc|Gender=Neut|Number=Sing" lemma="kmetijstvo">kmetijstvo</w>
152 <pc xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t7" ana="mte:Z" msd="UposTag=PUNCT">,</pc>
153 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t8" ana="mte:Ncnsa" msd="UposTag=NOUN|Case=Acc|Gender=Neut|Number=Sing" lemma="gozdarstvo">gozdarstvo</w>
154 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t9" ana="mte:Cc" msd="UposTag=CCONJ" lemma="in">in</w>
155 <w join="right" xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t10" ana="mte:Ncfsa" msd="UposTag=NOUN|Case=Acc|Gender=Fem|Number=Sing" lemma="prehrana">prehrano</w>
156 </name>
157 <pc xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s1.t11" ana="mte:Z" msd="UposTag=PUNCT">.</pc>
158 <linkGrp corresp="#OZKGIP-Redna-025-1998-12-08.seg1.s1" targFunc="head argument" type="UD-SYN">
159 <link ana="ud-syn:amod" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t2 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t1"/>
160 <link ana="ud-syn:root" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t2"/>
161 <link ana="ud-syn:nmod" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t2 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t3"/>
162 <link ana="ud-syn:nmod" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t3 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t4"/>
163 <link ana="ud-syn:case" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t6 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t5"/>
164 <link ana="ud-syn:nmod" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t3 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t6"/>
165 <link ana="ud-syn:punct" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t8 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t7"/>
166 <link ana="ud-syn:conj" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t6 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t8"/>
167 <link ana="ud-syn:cc" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t10 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t9"/>
168 <link ana="ud-syn:conj" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1.t6 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t10"/>
169 <link ana="ud-syn:root" target="#OZKGIP-Redna-025-1998-12-08.seg1.s1 #OZKGIP-Redna-025-1998-12-08.seg1.s1.t11"/>
170 </linkGrp>
171 </s>
172 <s xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s2">
173 <w xml:id="OZKGIP-Redna-025-1998-12-08.seg1.s2.t1" ana="mte:S1" msd="UposTag=ADP|Case=Loc" lemma="na">Na</w>
```

vse v skladu s pravili (validiranje podatkov): izmenljivost ali celo interoperabilnost (primer ParlaMint) podatkov

Parla-CLARIN

A TEI Schema for Corpora of Parliamentary Proceedings

v0.3

2022-05-03

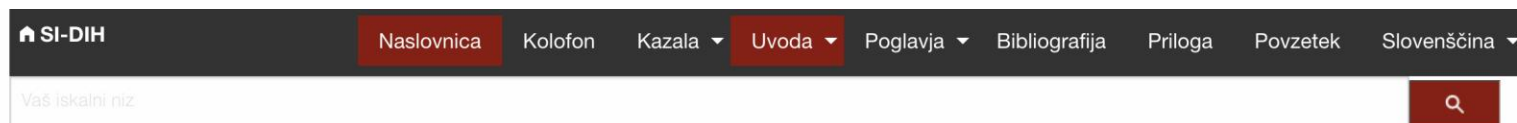
Table of contents

1. [Introduction](#)
 - 1.1. [Scope and purpose](#)
 - 1.2. [Background](#)
2. [General requirements](#)
 - 2.1. [Characters](#)
 - 2.2. [Documenting the encoding process](#)
 - 2.3. [Languages](#)
 - 2.4. [Identifiers and referencing](#)
 - 2.5. [Temporal information](#)
 - 2.6. [Files](#)
3. [Overall document structure](#)
 - 3.1. [Corpus structure](#)
 - 3.2. [Text divisions](#)
 - 3.3. [Document variants](#)
4. [Corpus metadata](#)
 - 4.1. [Typologies](#)
 - 4.2. [Speaker metadata](#)
 - 4.3. [Party metadata](#)
 - 4.4. [Relationships between people and parties](#)
5. [Transcriptions](#)
 - 5.1. [Utterances](#)
 - 5.2. [Transcriber comments](#)
 - 5.3. [Gaps](#)
 - 5.4. [Interrupted utterances](#)
 - 5.5. [Addressees, questions and answers](#)
 - 5.6. [Voting results](#)
6. [Linguistic annotation](#)
 - 6.1. [Basic word-level annotation](#)
 - 6.2. [Normalised and syntactic words](#)
 - 6.3. [Segmental annotation](#)
 - 6.4. [Linking annotation](#)
7. [Multimedia](#)
 - 7.1. [Speech and video](#)
 - 7.2. [Facsimile](#)
8. [Conversions](#)
 - 8.1. [Conversion from Akoma Ntoso](#)
 - 8.2. [Conversion to RDF](#)
9. [Acknowledgements](#)

Sledi lahko pretvorba v digitalno izdajo (HTML): brskanje, iskanje, relacije, vizualizacije ipd.

Možno pa še veliko, veliko več: analiza z različnimi programi in orodji, npr.:

- konkordančniki: specializirana orodja za analizo korpusov (npr. SketchEngine – vertikalni format)



- Orange (CONLL-U format): omogoča uporabo naprednih avtomatskih analitičnih tehnik – rudarjenje besedi

ZBIRKA
PARLAMENTARIA

4

Ajda Pretnar Žagar
Kristina Pahor de Maiti
Darja Fišer

What's on the agenda?

Topic modelling parliamentary debates
before and during the COVID-19 pandemic

Kaj je na dnevnem redu?

Tematsko modeliranje
parlamentarnih razprav pred
in med epidemijo
covid-19





Centralna tehniška knjižnica Univerze v Ljubljani
Central Technical Library at the University of Ljubljana
Trg republike 3, SI-1000 Ljubljana
Slovenija / *Slovenia*



Univerza v Ljubljani

