

REPOZITORIJ ROKOPISOV SLOVENSKEGA SLOVSTVA – ODDALJENA BLIŽINA?

Matija Ogrin

Inštitut za slovensko literaturo in literarne vede, ZRC SAZU, Ljubljana

Jan Jona Javoršek, Tomaž Erjavec

Institut "Jožef Stefan", Ljubljana

UDK 811.163.6'322-122:091"16/17"

Prispevek predstavi prve rezultate projekta NRSS – Neznani rokopisi slovenskega slovstva 17. in 18. stoletja: informacijsko-tehnološko podprta evidenca, znanstvene objave in analize. Razložimo kontekst in cilje projekta, zlasti glede na pomanjkljivo stanje raziskav s tega področja, in opišemo gradnike digitalnega repozitorija. Orisana sta digitalni zapis repozitorija v formatu XML, ki je skladen s smernicami iniciative za zapis besedil TEI, in programska platforma za iskanje po repozitoriju.

rokopisi slovenskega slovstva, iniciativa za zapis besedil TEI, Fedora Commons

The paper presents the first results of the NRSS project – Unknown Slovene literary manuscripts from the 17th and 18th centuries: informational-technological supported evidence, scientific publications and analyses. We present the context and goals of the project, in particular with regard to the lack of research in this field, and describe the building blocks of the digital repository. We sketch the digital encoding of the repository in XML, which is compliant with the Text Encoding Initiative Guidelines, and the software platform for searching the repository.

Slovene literary manuscripts, Text Encoding Initiative, Fedora Commons

1 Znani ali neznani slovenski rokopisi?

Raziskanost starejšega slovenskega slovstva v naši literarni zgodovini je razmeroma fragmentarna. S tem je misliti zlasti na dejstvo, da so določene osebnosti ali skupine ustvarjalcev znotraj obdobja podrobno in izčrpno obravnavane, nekateri njihovi sodobniki pa razmeroma orisno in obstransko. Določenim obdobjem je bila posvečena večja raziskovalna pozornost in je gradivo zanje podrobneje raziskano; nekatera druga obdobja pa so ostala slabše osvetljena – še zlasti barok in deloma razsvetljenstvo (tj. od srede 17. stol. do okrog 1830). Posledica tega in drugih dejavnikov je, da je neznanost število pomembnih ali vsaj slovstveno zanimivih besedil slovenskih piscev iz starejših obdobja ostalo neobjavljenih vse do danes. Čeprav so bila nekatera dela – a nikakor

ne vsa ta dela – deležna vsaj skopih omemb v literarnih zgodovinah, pa ta besedila prav gotovo niso vstopila v sistem literarne recepcije in komunikacije. Celotno avtorjem t. i. Matičine ZSS I – morda najboljše slovenske literarne zgodovine te dobe – se je zdela sprejemljiva prostodušna omemba, da se je iz dobe med Hrenovo in Schönlebnovo izdajo Evangelijev in listov »ohranilo blizu 50 slovenskih rokopisov, od kratkega zapisa do zajetnega zbornika, nastalih v najrazličnejših krajih naše domovine« (Legiša 1956: 281) in čeprav sta nekako dva ducata rokopisov tam kratko omenjena, skoraj nihče več ne ve, kateri rokopisi natanko to so in kje se danes nahajajo. V tem smislu ostajajo neizdana rokopisna besedila slovenskega slovstva 17. in 18. stoletja bela lisa slovenske kulturne zavesti, najsi gre za prevode

in priredbe ali celo za avtorska (pol)literarna besedila. Sistematično raziskovanje rokopisnih zbirk doma in v zamejstvu bi moglo še nemalo osvetliti baročno in razsvetljensko dobo slovenskega slovstva. To je bila izhodiščna hipoteza naše raziskave.

Najsi to hipotezo sprejemamo ali ne, vsekar je na dlani dejstvo, da v Sloveniji še ne obstaja sistematično zasnovan digitalni repozitorij rokopisov – še zlasti ne rokopisov slovenskega slovstva. Da bi to vrzel začeli zapolnjevati, smo kot nadaljevanje projekta eZISS zasnovali raziskavo rokopisov baročne in razsvetljenske dobe, deloma tudi zgodnjega 19. stoletja, napisanih v slovenskem jeziku. Raziskava naj bi bila izdatno tehnološko podprta in naj bi se končala s praktično delujočim modelom ali celo pilotno različico elektronskega registra rokopisov te dobe, kar je tudi v tehnološkem pogledu nemajhen izziv. Raziskavo smo poimenovali *Neznani rokopisi slovenskega slovstva 17. in 18. stoletja: informacijsko-tehnološko podprta evidenca, znanstvene objave in analize* (NRSS).

Raziskava NRSS je praktično koncipirana tako, da se rezultati posamičnih literarnozgodovinskih in historičnih raziskav znotraj nje stekajo v štiri ravni, ki se medsebojno dopolnjujejo in potekajo zvečine hkrati. Z njimi je moč strnjeno podati praktične cilje NRSS.

1. **Bibliografija NRSS.** Kot prvo pripravljalo delo je bilo treba izdelati bibliografski seznam (večine) znanstvenih objav, monografskih kakor analitičnih, v katerih so bili slovenski rokopisi 17. ali 18. stoletja doslej kakor koli obravnavani ali vsaj omenjeni.

2. **Indeks NRSS** – seznam evidentiranih rokopisov. Od najzgodnejših znanstvenih objav

v poročilih raznih muzejskih in zgodovinskih društev v 19. stoletju do novejšega časa so bili v strokovni periodiki in v monografijah omenjeni mnogi slovenski rokopisi. Izdelati je bilo torej treba indeks vseh slovenskih rokopisov 17. in 18. stoletja, ki so bili kdaj koli v pozitivni evidenci literarnih, narodopisnih in historičnih ved. Za mnoge med njimi danes (po 100 in več letih) ne vemo več, kje so shranjeni, mnogi so se tudi izgubili in so verjetno uničeni. Vsekakor je indeks omenjenih rokopisov eden od pomembnih rezultatov NRSS, saj nam posreduje ne le trenutno stanje raziskav na tem področju (ki ni zavirljivo),¹ ampak historični vpogled v to, kar je nekdanje že bilo v evidenci raznih ved, a se je zaradi dveh svetovnih vojn, emigracije, selitev, menjav lastnikov itn. do danes zbrisalo ali sploh izgubilo. To področje je presenetljivo veliko; indeks NRSS trenutno (avgust 2009) obsega 134 omenjenih rokopisnih enot. Poudarjamo, da so v indeksu zbrani podatki, pridobljeni iz strokovne literature, ne iz samih primarnih virov, saj nekateri od teh sploh niso dosegljivi.

3. Elektronski **register NRSS** je nedvomno osrednji cilj projekta. Obsegal bo *opise rokopisov*, narejene neposredno po samih primarnih virih, po možnosti skupaj z digitalnimi faksimili. Ideja registra izhaja iz dejstva, da so rokopisi baročne in razsvetljenske dobe raztreseni po številnih arhivih in knjižnicah; ker pa so arhivski registri največkrat jezikovno nevtralni, je uporabniku večkrat težko razbrati, ali je neka arhivalija napisana v latinščini, nemščini, italijanščini ali slovenščini – zato je potrebno težavno iskanje, ki otežuje raziskovanje. Register NRSS bo zato na enem mestu nudil podatke – podrobne opise rokopisov in

1 Po Jožetu Koruzi se je v okviru slovenske literarne zgodovine s problematiko slovenskih rokopisov baročne in razsvetljenske dobe komaj še kdo ukvarjal. Boljše je stanje v občem zgodovinopisju, kjer je bila razmera dobro raziskana vsaj uradovalna slovenika. To bo še posebej razvidno iz bibliografije NRSS, kjer bo moč strokovno literaturo tudi kronološko razvrstiti oz. pregledati. Pregled bo verjetno pokazal, da je bilo največ znanstvenih raziskav o rokopisih te dobe napisanih v času med vojnami, nato pa čedalje manj do časa okrog 1980, ko se (z menjavo generacij?) interes za to dobo v literarni zgodovini skorajda prekine. Pomenljivo je že dejstvo, da je bilo poleg Smolikove monografije o cerkveni pesmi edino temeljno delo, na katero se nam je bilo moč opreti, Kidričeva študija *Opombe k protireformacijski dobi*, napisana leta 1921 (Kidrič 1921).

njihove digitalne slike – medtem ko bodo originali kot taki seveda ostali tam, kjer trenutno so, pri svojih lastnikih.

V nadaljnje stopnje raziskovanja starejših slovenskih rokopisov vključujejo tudi analize in transkripcije manjšega korpusa izbranih besedil (četrti raven rezultatov) z namenom, da jih pripravimo za kritične izdaje. Toda osrednji rezultat raziskave bo nedvomno prav elektronski register NRSS. Prav z njim se bo pozitivno potrdila (ali ovrgla) izhodiščna hipoteza, da je več rokopisov slovenskega slovstva baročne in razsvetljske dobe ostalo večidel zunaj raziskovalne pozornosti, koliko teh rokopisov je, kakšni sploh so in kje se trenutno nahajajo.

2 Opis rokopisov

Glede na osrednje mesto, ki ga v celotni raziskavi zaseda načrtovani register, smo posebno pozornost posvetili metodi deskripcije rokopisov. Držali smo se v svetu sprejetih in ustaljenih metod analize rokopisnega gradiva, razvitih v pomožnih historičnih vedah, zlasti v kodikologiji, diplomatiki in arhivistiki. Sprejeli smo standardizirano metodo, kakor jo kodificira modul za opis rokopisov v Smernicah konzorcija Text Encoding Initiative (TEI 2007).² Vsak rokopis je bil ob ogledu *in situ* v skladu s to metodo podrobno opisan glede na štiri sklope podatkov, ki so vsi gnezdeni v krovni element <msDesc>. Strukturo teh podatkov predstavlja naslednja slika, pojasnjena v nadaljevanju.

```
<msDesc xml:id="ms_MOG_002">
  <msIdentifier>
    <country>Slovenija</country>
    <settlement>Ljubljana</settlement>
    <repository>Narodna in univerzitetna knjižnica</repository>
    <idno type="Nuk">Ms 1120/2</idno>
  </msIdentifier>
  <msContents>
    <msItemStruct class="#nrss_t_pridiga">
      <author xml:lang="sl">Rogerij Ljubljanski</author>
      <author xml:lang="la">Rogerius Labacensis</author>
      <title xml:lang="la" type="original">Symbolialis Siue: Concionum Catechisticarum
        Symbolum Apostolicum explicantium. Concio secunda. Dominica Secunda post
        Pentecosten.</title>
      <title xml:lang="sl" type="editorial">O veroizpovedi ali katehistične pridige z
        razlago apostolske vere. Druga pridiga. Druga nedelja po binškoših.</title>
      <incipit xml:lang="sl">De refsnižnu je tu, kar gouoril je Joannes ta S: Euangelifta
        od te lubefni Boŕhje, katero ima inu kasho Bug V: M: prutie temu reunimu zhloueftu,
        namezh: de On lubil je taiftu, prior dilexit nos, inu fizer poprei ueliku ...</incipit>
      <explicit>Spomnite, kai bo na dan te Sodbe sa enu vefsellie tem Vernim inu zhaloft tem
        Neuernim, kir takú perfhli bofte h'te Vezherij, kakor Amen.</explicit>
      <listBibl>
        <bibl copyOf="#bib_060" subtype="mentioned">
          <biblScope type="pp">128-130</biblScope>
          <note type="bibl">Omembe v zvezi z genezo Rogerijevega Parmarium empreum.</note>
        </bibl>
      </listBibl>
      <note type="regesta">Druga izmed desetih ohranjenih Rogerijevih katehističnih pridig.
        Besedilo je popolnoma izgotovljen čistopis, katehistična pridiga za drugo nedeljo po
        binškoših. Datirana je v Ljubljani 1716.</note>
      <textLang mainLang="sl">slovenščina</textLang>
    </msItemStruct>
  </msContents>
  <physDesc>
    <objectDesc form="quire">
      <supportDesc material="paper">
```

2 Razlogov za sprejetje te metode je več. S strani filoloških ved je razlog ta, da so Smernice TEI standard, v katerem se vsekakor steka in povezuje največja množina metod besedilne analize iz filoloških in historičnih ved, tako temeljnih kakor pomožnih, in da so te metode podane operativno in eksaktno. Z gledišča nadaljnje računalniške obdelave pa so Smernice TEI sprejete kot de facto standard za označevanje in strukturiranje besedil v zapisu XML, kar omogoča njihovo nadaljnjo strojno obdelavo, analizo in predstavitev.

1. Identifikator rokopisa <msIdentifier> vsebuje podatke o nahajališču oz. hranišču rokopisa, ustanovi ali lastniku, zbirki, signaturi. S temi najosnovnejšimi podatki je rokopis v fizičnem smislu »identificiran«, od tod ime tega sklopa.

2. Vsebinski sklop opisa <msContents> vsebuje bogato razčlenjene podatke o intelektualni vsebini rokopisa (ločeni od njegove materialne podstati), kjer so zbrane razpoložljive informacije o avtorju, naslov dela, zvrst, incipit, eksplicit, regista, oznaka jezika, predvsem pa členitev na morebitne dele, dodatke ipd.

3. Fizični opis <physDesc> vsebuje podatke o rokopisu kot materialnem artefaktu: o materialu, iz katerega je izdelan, in njegovih dimenzijah, o številu folijev, o rokah, ki so vanj pisale, ipd.

4. Zgodovina rokopisa <history> zajema sklop podatkov o kraju in času njegovega nastanka, o morebitnih prejšnjih lastnikih ali zbirkah.

5. Sklop dodatnih beležk <additional> vsebuje administrativne podatke o tem, kdo je opis rokopisa napravil, datum prvega opisa, ali se je k rokopisu pozneje vračal, zapis izpolnil s podrobnostmi itn.

Smernice TEI v tem primeru tudi omogočajo, da si urednik glede na potrebe svojega specifičnega dela in gradiva izbere bolj ali manj podrobno, granularno strukturo besedila. Lahko bi izbrali tudi manj zahteven sestav podatkov. Vendar smo se odločili za bolj zahteven sestav: tako nas shema XML konzorcija TEI, po kateri razčlenjevalnik validira našo datoteko XML, prisili, da uporabljamo vedno točno določene elemente opisa in v striktnem vrstnem redu. Ta prisila je še kako dragocena: zagotavlja nam, da bodo vsi opisi rokopisov strukturirani enako, kar pomeni, da ne bomo

mogli pozabiti nobenega podatka (ali pa bo element namenoma ostal prazen) in da bodo podatki vedno dosledno ločeni med seboj.

Tako razčlenjeni opis rokopisov naj bi bil podlaga za to, da bo po registru mogoče strojno iskati glede na razne kriterije, kakor so čas ali kraj nastanka, zvrst, avtor idr.

S tem naj bi register NRSS prispeval k čim boljši evidenci nad rokopisnim gradivom 17. in 18. stoletja v slovenščini. Register NRSS je torej specializiran tematski katalog rokopisov s faksimili v nastajanju. Pogoj, da bi lahko dejansko nastal, je uspešna združitev filološko-historične problematike s sodobnimi tehnologijami. Zgoraj je nakazano, kako je bila tehnologija elektronskega označevanja besedil konzorcija TEI aplicirana na opis rokopisov; toda za uspešno rešitev je potrebna še aplikacija orisane filološke problematike na izgradnjo podatkovnih zbirk in poizvedovalnih sistemov. Ta spoj ni enoznačen in enoplasten problem. Toliko manj, ker v Republiki Sloveniji še ni podobnega digitalnega repozitorija rokopisov, ki bi našemu delu utrl pot.³

3 Implementacija registra NRSS

Glede na to praznino smo se odločili, da bomo v okviru projekta implementirali register NRSS kot del digitalnega repozitorija dokumentov. Ta odločitev se je izkazala kot ustrezna, saj je bilo mogoče infrastrukturo digitalnega repozitorija uporabiti tudi za druge projekte, še več, pričakujemo tudi dodatne prednosti povezovanja podatkovnih virov iz različnih zbirk v istem repozitoriju.

Za implementacijo repozitorija smo izbrali platformo Fedora Commons (Lagoze: 2006), ki je sistemsko neodvisna platforma, napisana v jeziku Java in organizirana kot spletna aplikacija. Repozitorij dokumente obravnava kot

3 Edina nam znana elektronska zbirka, s katero je mogoča kakšna primerjava, je nastajajoča nova podatkovna zbirka Arhiva Republike Slovenije (<http://arsq.gov.si/Query/suchinfo.aspx>). Vendar je ta po eni strani neprimerno večja, saj je njen predmet arhivsko gradivo na ravni države Slovenije, po drugi strani pa je mnogo manj detaljna, saj je njen predmet predvsem evidenca fondov in zbirk kot takih, ne pa posameznih listin in rokopisov. Zato prava primerjava ni mogoča.

objekte, tipično zapisane v obliki datotek XML, tako da je vnos dokumentov, kakršni so dokumenti TEI XML, naraven in neposreden. Vsak dokument je sestavljen iz vrste podatkovnih tokov, ki so lahko interni (zapisani v obliki XML) ali eksterni, npr. slikovni v primeru faksimila ali binarni v primeru skenograma v formatu PDF. Vsak dokument vsebuje tudi kontrolne podatkovne tokove za sledenje spremembam, za zapis metapodatkov (v standardnem metapodatkovnem zapisu Dublin Core (ANSI/NISO Z39.85) ter za zapis medsebojnih odnosov med dokumenti. Platforma vsebuje vrsto spletnih storitev, ki omogočajo poizvedovanje po repozitoriju: poizvedovanje po metapodatkih, po medsebojnih odnosih med dokumenti ter po strukturi in podatkovnih tokovih posameznega dokumenta. Ta infrastruktura je razširljiva, v primeru implementacije registra NRSS smo npr. dodali še elemente za kompleksno poizvedovanje po celotnem besedilu, kjer smo uporabili sistem SOLR in knjižnico Lucene (Hatcher: 2004), ki omogočata poizvedovanje neposredno po ključnih besedah iz podatkov v formatu TEI ter poizvedovanje po celotnem besedilu (full text search). Poleg tega platforma Fedora Commons omogoča razširitev posameznih dokumentov z novimi storitvami, največkrat namenjenimi za predstavitev dokumenta ali zbirke dokumentov v obliki spletne strani.

Gradivo registra NRSS smo za vnos v digitalni repozitorij s skriptnim jezikom XSLT za transformacijo dokumentov XML razbili v vrsto manjših dokumentov, ki smo jih vnesli v repozitorij. Posamezni dokumenti so zbrani v zbirke dokumentov, iz katerih je mogoče sestaviti izvorne datoteke. Poleg tega imajo posamezni dokumenti definirane medsebojne odnose, ki delujejo kakor kazalke, ter so na osnovi podatkov v zapisu TEI samodejno opremljeni z metapodatki v formatu Dublin Core (Liddy 2002). Dokumenti imajo definirane tudi lastne storitve za predstavitev v obliki spletne strani (implementirane kot transformacije zapisa

XML TEI v HTML s pomočjo skriptnega jezika XSLT).

Vsak takšen dokument je tako mogoče prikazati kot posamezno spletno stran, posamezne zbirke pa kot spletne indekse dokumentov. Tako lahko uporabnik pregleda seznam vseh zapisov v treh zbirkah: zbirki bibliografije NRSS, zbirki evidentiranih rokopisov v indeksu NRSS ter zbirki opisanih rokopisov (v nekaterih primerih tudi transkribiranih in opremljenih s faksimilnimi posnetki). Vsak seznam uporabniku ponuja povezave na posamezne zapise, pri čemer si jih lahko ogleda kot spletno stran ali kot izvorni zapis v formatu XML TEI. Še več, medsebojni odnosi so prikazani kot spletne povezave, tako da so npr. vsi evidentirani rokopisi, omenjeni v določeni enoti bibliografije, neposredno dostopni preko povezav, če pa je bil dejanski rokopis najden in obdelan, je dostopna tudi povezava na zbirko opisanih rokopisov.

Uporabnik lahko dostopa do podatkov tudi preko poizvedovalnega sistema, ki omogoča iskanje po vsem besedilu dokumentov ali po posameznih ključnih besedah, npr. imenu avtorja strokovne študije, imenu avtorja rokopisa, podatkih o času nastanka, fizičnih podatkih o rokopisu ipd. Ta poizvedovalni sistem, ki je deloma še v razvoju, bo postal dragoceno orodje za nadaljnje raziskovalno delo na področju evidentiranja in obdelave starejših slovenskih rokopisov, saj bo bistveno olajšal iskanje znanih podatkov ter tako tudi doslej zahtevne naloge, kakor so identificiranje najdenega rokopisa ali primerjava različnih transkripcij. Poleg tega sistem omogoča dostop do predstavitve podatkov v obliki spletne strani, kar poveča dostopnost gradiva, do izvorne oblike, ki je primernejša za strokovno in strojno obdelavo, ter do standardno zapisanih metapodatkov o gradivu, kar omogoča integracijo v druge poizvedovalne sisteme ter povezavo med različnimi repozitoriji, npr. s sistemom Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH, prim. Ward: 2004), ki

ga podpira platforma Fedora Commons, ter naprednimi sistemi, npr. OAI-P2P (Ahlborn: 2002).

4 Sklep

Prispevek je predstavil hipotezo, cilje in konceptualno zasnovo projekta NRSS – *Neznani rokopisi slovenskega slovstva 17. in 18. stoletja: informacijsko-tehnološko podprta evidenca, znanstvene objave in analize*. Osrednji cilj je izdelava pilotnega elektronskega repozitorija z opisi in digitalnimi faksimili rokopisov, prvega te vrste v Sloveniji. Predstavljena je arhitektura repozitorija, predvsem njegov digitalni zapis, ki je skladen s smernicami iniciative za zapis besedil TEI ter tehnološko platformo za iskanje po repozitoriju, osnovano na programju Fedora Commons. Repozitorij je trenutno v testni fazi, saj se na eni strani še zbirajo opisi rokopisov, na drugi pa še razvijamo masko za prikaz in iskanje po repozitoriju. Register NRSS naj bi na enem mestu nudil zbrane in podrobne podatke o vseh pomembnejših rokopisih v slovenščini, ohranjenih iz 17. in 18. stoletja, s tem pa tudi spodbudil nove raziskave te dobe.

Literatura

- BENJAMIN, Ahlborn, NEJDL, Wolfgang, SIBERSKI, Wolf, 2002: OAI-P2P: A Peer-to-Peer Network for Open Archives, v *Workshop on Distributed Computing Architectures for Digital Libraries - ICPP2002. Dublin Core Metadata Element Set, Version 1.1, a DCMI Recommendation, 2009*: <<http://dublincore.org/documents/2008/01/14/dces/>>. (Dostop 19. 8. 2009.)
- HATCHER, Erik, GOSPODNETIC, Otis, 2004: *Lucene in Action*. New York: Manning Publications.
- KIDRIČ, France, 1921: *Opombe k protiref. (katol.) dobi v zgodovini slovenskega pismenstva*. ČJKZ III 1921/22. 71–133.
- LAGOZE, Carl, PAYETTE, Sandy, SHIN, Edwin, WILPER, Chris, 2006: Fedora: an architecture for complex objects and their relationships, *International Journal on Digital Libraries* 6/2. 124–138.
- LEGIŠA, Lino 1956: *Zgodovina slovenskega slovstva I. Do začetkov romantike*. Ljubljana: Slovenska matica.

- LIDDY, Elizabeth D., ALLEN, Eileen, HARWELL, Sarah, CORIERI, Susan, YILMAZEL, Ozgur, OZGENCIL, Ercan N., DIEKEMA, Anne, MCCRACKEN, Nancy, SILVERSTEIN, Joanne, SUTTON, Joanne, 2002: Automatic metadata generation & evaluation. V *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. New York: ACM Press. 401–402.
- MILLER, Eric, BRICKLEY, Dan, 2001: Expressing Simple Dublin Core in RDF/XML, *Dublin Core Metadata Initiative Proposed Recommendation*.
- TEI Consortium, 2007: TEI P5: *Guidelines for Electronic Text Encoding and Interchange*. Version 1.4.1. Last updated on July 1st 2009. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (Dostop 19. 8. 2009.)
- WARD, Jewel, 2004: *Unqualified Dublin Core usage in OAI-PMH data providers*. OCLC Systems & Services 20/1. 40–47.