

PROBLEMI DIGITALIZACIJE SLOVENSKE LITERARNE IN KULTURNE DEDIŠČINE

Marijan Dovič

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

UDK 821.163.6:316.7(497.4):004.55

Prispevek z infrastrukturno-tehnoloških, organizacijskih, deloma pa tudi pravnih vidikov obravnava ključne probleme digitalizacije slovenske slovstvene zapuščine v okviru namišljenega idealnega portala »Kulturna dediščina Slovenije«.

slovenska literatura, kulturna dediščina, digitalizacija, hipertekst, elektronske izdaje

The paper discusses key issues relating to the digitalisation of the Slovene written legacy as a part of the imagined ideal web portal »Cultural heritage of Slovenia« from infrastructural, technological, organisational and legal points of view.

Slovene literature, cultural heritage, digitalisation, hypertext, electronic editions

Danes ne more biti več nikakršnega dvoma o tem, da digitalizacija in spletno objavljanje slovenske slovstvene dediščine predstavljata prvovrsten *infrastrukturni izziv* sodobne slovenistike. Zahteva po celoviti in kakovostni elektronski predstavitvi slovenske kulturne in umetnostne dediščine, znotraj te pa seveda tudi literarne oziroma v širšem smislu slovstvene, se zdi vse bolj uresničljiva in legitimna. Mogoče je reči, da z razvojem informacijskih tehnologij pravzaprav dobiva značaj imperativa, ene izmed *prioritet* humanističnih disciplin.¹ Na svetovnem spletu je trenutno dostopnih veliko vsebin, ki bi jih lahko uvrstili v rubriko »slovenska kulturna dediščina«, težko pa bi rekli, da kot celota delujejo zares usklajeno, premišljeno, celovito ali pregledno.

»Kulturna dediščina Slovenije:« zamišljeni idealni portal

Preden se lotimo najrazličnejših načelnih in praktičnih zadreg, ki jih digitalizacija nujno prinaša s seboj – in tako skušamo odgovoriti na vprašanje *kako* – se za trenutek pomudimo pri vprašanju *kaj*. V ta namen si zamislimo (za zdaj še imaginarni) povsem brezplačen in javno dostopen spletni portal, imenovan recimo *Kulturna dediščina Slovenije*. Zamislimo si, da bi se na takšnem portalu lahko sprehajali po vseh ohranjenih Prešernovih rokopisih in tiskih (tudi apokrifnih), ki bi nam bili dostopni v obliki faksimilov in kakovostnih prepisov, da bi po potrebi sami generirali vzporedne postavitve zelenih variant (bodisi rokopisnih ali tiskanih), iskali konkordance, izvajali zahtevne empirične tekstne raziskave, brskali po drobovju natančnega znanstvenokritičnega

1 Tekstno kodiranje in digitalizacija sta od statusa »pomožnih« disciplin danes napredovali med osrednja vprašanja sodobne humanistike (Juvan 2006, Burnard 2005).

uredniškega aparata, si zraven ogledali še likovne upodobitve Prešerna ali motivov iz njegovih pesmi ter prisluhnili njihovim uglasbitvam – od entuziastičnih zborovskih priredb 19. stoletja do Kreslinove Vrbe. Sledeč Prešernovim prepesnitvam ljudskih pesmi, na primer Lepe Vide, bi se nato pogreznili v nelinearno, rizomatsko strukturirano omrežje prepletajočih se motivov in s tem slovenske ljudske pesmi, sledili različnim zapisom tekstovnih variant in melodij ter zraven poslušali zvočne posnetke njihovih izvedb s »terena«. Zatem si zamislimo, da bi pri ljudskem motivu »/.../ šel sem čez gmajnico«, ki ga Iacobus Gallus citira v znamenitem motetu *Preparate corda vestra*, preklopili v muzikološko bazo in brskali po celotnem Gallusovem opusu – in sicer po kompletnih rokopisih in originalnih tiskih iz 16. stoletja, ki bi jih lahko primerjali z različnimi zgodovinskimi transkripcijami Gallusovih maš, motetov in madrigalov, izvajalci pa bi si po potrebi iztislili optimalno izvedbeno verzijo in si nato pri interpretaciji pomagali z zvočnimi posnetki dostopnih izvedb, ki bi jih ravn tako dobili na portalu. Nadalje si zamislimo, da bi naš obisk nadaljevali v smiselno organizirani spletni likovni galeriji, kjer bi krmarili med visoko ločljivimi skenogrami vseh Groharjevih del (ne glede na njihovo fizično nahajališče), ki bi jih spremljal kakovosten kritični aparat z vsemi umetnostnozgodovinskimi metapodatki; ali pa bi si v virtualnih tridimenzionalnih prikazih od vseh strani ogledovali slovenske kiparske in arhitekturne mojstrovine in se na primer sprehodili po notranjosti Plečnikove cerkve srca Jezusovega v praški četrti Jiřího z Poděbrad.

Zastaviti takšen cilj bi bilo še pred nekaj leti prenapeto in fantastično. Zamisel je sicer še vedno nadvse smela, kajti na poti do njene uredništvne stojijo številne ovire in nedorečenosti,

ki jih je mogoče v grobem razdeliti na *ekonomske* (financiranje), *pravne* (razreševanje avtorske pravice avtorjev in dedičev, sorodne pravice itn.), *infrastrukturno-tehnološke* (kodiranje, označevalni standardi, hranjenje, konkretne digitalne rešitve) in *politično-organizacijske* (skrbništvo, avtoriteta/ nadzor). Kot običajno je problem tudi ali celo predvsem »v glavah«: narava zamisli namreč zahteva metarefleksiven paradigmatični zasuk, modifikacijo utrjenih praks posameznih humanističnih disciplin in ne nazadnje poglobljeno interdisciplinarno sodelovanje med humanistiko in informacijsko tehnologijo.

Slovenska slovstvena dediščina in njena digitalizacija

Vsaka izmed naštetih točk je tako kompleksna, da se je bo v tem okviru mogoče dotakniti le shematično in v zoženi perspektivi, torej z vidika slovstvene dediščine.² Bežen pregled današnjega stanja digitalizacije slovenskega leposlovja v zbirkah, kot so Wikipedija oz. Wikivir, Zbirka slovenskih leposlovnih besedil, Digitalna knjižnica Slovenije, virtualna knjigarna Beseda, zbirka znanstvenokritičnih e-izdaj pri ISSLV ZRC SAZU, Intratekst, besedilna korpusa Nova beseda in Fida, Google Books ter Internet Archive, kaže na prvi pogled zelo pestro podobo. Toda v nadaljevanju se bo izkazalo, da s temi dosežki še ne moremo biti povsem zadovoljni. Njihove pomanjkljivosti in parcialnosti morajo nujno postati predmet »infrastrukturne« refleksije. Kako zastaviti nadaljnje korake, da se bodo rezultati prizadevanj – sicer vse hvale vrednih – učinkoviteje povezovali v smiselno celoto? Preden se lotim posameznih problemskih točk, velja povedati, da podpiram štiri temeljna načela, ki so podlaga zamisli o Digitalni knjižnici Scaife (Scaife Digital Library).³

² Že problematika avtorskih pravic v zgornjem namišljenem »sprehodu« bi zadostovala za samostojno razpravo. Dela slovenskih slikarjev na primer niso brezplačno dostopna na spletu, četudi so njihovi avtorji pod rušo že več kot 70 let.

³ Projekt SDL temelji na dolgoletnih izkušnjah odlične zbirke Perseus (gl. Blackwell, Crane 2009).

Prvič, objavljena besedila morajo biti strokovno recenzirana (»peer reviewed«); drugič, morajo biti ustrezno formatirana in kodirana (XML/TEI); tretjič, imeti morajo trajen in stabilen dolgoročen repozitorij, neodvisen oziroma ločen od proizvajalca vsebine; in četrtič, morajo biti brezplačno dostopna javni rabi prek odprte licence (CC). V nadaljevanju se bo nekoliko natančneje pokazalo, za kaj gre pri omenjenih načelih.

Tipi digitalnih izdaj

Po začetni evforični digitalizaciji strani kot »bitmap« sličic in neoznačenemu prepisovanju tekstov v nezanesljivih formatih, kot so word, html ali pdf, se je »digitalna humanistika« odločno orientirala v smeri zahtevnejših strukturiranih in označenih zapisov. Čeprav še vedno nastajajo povsem uporabne zbirke preprostejšega tipa, je danes jasno, da mora resna digitalizacija upoštevati določene standarde. Na spletu seveda obstaja široka paleta različnih tipov in kategorij izdaj. Večina »slovenike« je pri nas dostopne v nestrukturirani obliki,⁴ imamo pa tudi nekaj vzorčnih primerov kompleksnih edicij v zbirki *Elektronske znanstvenokritične izdaje slovenskega slovstva*.⁵ Celo pri najbolj dodelanih tujih izdajah še ni razrešena dilema, ali je idealna organizacijska oblika *besedilna zbirka* ali kompleksen *hipermedijski arhiv*, ki vključuje tudi nebesedilna gradiva.

Standardi kodiranja

V morju potencialov kiberprostora bi bilo utopično pričakovati, da se bodo v elektronskih izdajah vzpostavili enotni načini strukturiranja, navigacije ali oblikovanja. A če so zbirke na prvi pogled res zelo različne, vzorčni tuji primeri, kot so Rossetti Archive, William Blake Archive, Perseus Digital Library in mnogi drugi, vsaj glede *kodiranja* ne puščajo veliko dvoma: izdelavo elektronskih izdaj je treba opreti na odprte mednarodne standarde elektronskega kodiranja, ki zagotavljajo ne le raznovrstno uporabnost označenih besedil, ampak tudi njihovo neodvisnost od tehnoloških sprememb, trajnost in izmenljivost.⁶ Besedilo naj bi bilo torej zapisano v jeziku XML (eXtensible Markup Language; gre za izvedenko predhodnega SGML) in obenem strukturirano po priporočilih TEI (Text Encoding Initiative, www.tei-c.org). Takšna besedila imajo bogato semantično strukturo, ki je računalniku razumljiva in manipulabilna.⁷ Vseh dimenzij uporabnosti tako označenih besedil se verjetno niti še ne zavedamo, četudi je že to, kar je mogoče danes, naravnost osupljivo.⁸

Avtoritarni strokovni nadzor ali kolektivni entuziazem?

Uspeh projektov v okviru Wikimedie sproža evforične napovedi o demokratizaciji participacije v svetu digitalnih izdaj. Kljub temu bi vsaj glede rubrike »dediščina« veljalo upoštevati

- 4 Najpreprostejša je povsem nemanipulabilna virtualna knjigarna Beseda, ki ponuja zaklenjene transkripcije knjig v formatu pdf. Zbirka torej ne izkorišča potenciala e-izdaj, je pa zato organizirana pregledno in uporabniku prijazno, kar še zdaleč ni nepomembno.
- 5 Izdaja jih Inštitut za slovensko literaturo in literarne vede ZRC SAZU v sodelovanju z Odsekom za tehnologije znanja Inštituta Jožef Stefan. Tak način objavljanja je visoko zahteven in zamuden, seveda pa posebej primeren za časovno oddaljene in kompleksne primere, kot so npr. Brižinski spomeniki (gl. Ogrin 2005, Erjavec 2005, Dović 2005).
- 6 Mnoge zbirke so bile v XML konvertirane naknadno, med drugim tudi zaradi težav s komercialnimi softverskimi standardi – na primer zbirka češke poezije (*Česká elektronická knhthovna*). Tudi sedanje neoznačene zbirke bodo verjetno prej ali slej konvertirane.
- 7 Opozoriti velja, da je bila shema XML deležna tudi kritik, češ da je preveč linearna in ni kompatibilna z avtopoetičnimi sistemi (gl. McGann, *Marking Texts in Many Dimensions*). Kljub temu za zdaj nima primerljive alternative.
- 8 Gl. Blackwell, Crane 2009.

nekaj hierarhije iz Gutenbergove dobe. Vsekakor morajo *elite* posameznih strok temeljito premisliti lastno tradicijo tekstne kritike oz. ekdotike in njena vodila artikulirati v novem kontekstu.⁹ Povsem naivno pa je od še tako navdušenih amaterskih digitalizatorjev pričakovati premišljene strukturne in organizacijske rešitve.¹⁰ Vsekakor je treba nove oblike »skupinskega« dela, ki jih omogoča kiberprostor, učinkovito integrirati v posamezne faze nastajanja izdaj, predvsem prepisovanje in korigiranje.¹¹

Avtorske pravice

Čprav še vedno nastajajo komercialni projekti, ki uporabniku zaračunavajo dostop, postaja uporaba vse številnejših besedilnih baz brezplačna in neomejena – oziroma omejena kvečjemu v okviru priljubljene licence CC (Creative commons, <http://creativecommons.org/>), ki je zasnovana z namenom, da razširi prostor javne rabe. Podobne težnje se usmerjajo tudi »nazaj«: krepí se gibanje za korekcijo obstoječe avtorske zakonodaje, ki v zahodnih državah po mnenju mnogih pretirano oz. predolgo ščiti avtorstvo.¹²

Vsebinske zamejitve in prioritete

Načelna neomejenost kiberprostora in inherentna nedokončanost spletnih izdaj – izdaje so lahko velikanske, kontekstualno zelo

goste in odprte za nenehno popravljanje in dopolnjevanje – predstavljata veliko zadrego pri vsebinskem zamejevanju in določanju prioritete. Problem dokončanja, nekoč zarezan z »oddajo v tisk«, je pri brezplačno dostopnih e-izdajah še bolj pereč; pri plačljivih je bilo namreč treba delo končati že zato, da je lahko šlo v prodajo.¹³ S problemom, koliko konteksta vključiti, kje začeti, kdaj (če sploh) izjaviti »končano je«, se soočajo vsi načrtovalci spletnih izdaj, rešitve in utemeljitve pa so zelo različne.¹⁴

Sklep

Digitalizacija (svetovne) kulturne dediščine je epohalen projekt. Vseh njegovih dimenzij ta hip gotovo še ni mogoče ugledati. Novi medij pomeni mnogo več kot zgolj nevtralno »ohranitev« dokumentov, pa vendar je ena od njegovih bistvenih potez ravno ta, da z njim na poseben način arhiviramo celotno epoho (ali kar dve): epoho rokopisne in tiskane dobe, epoho, ko je prenos znanja v družbi zaznamovala sekvenca pravokotnih strani, organiziranih v kodeks oziroma pozneje v natisnjeno knjigo. Izkušnje z digitalnim hipertekstom so pokazale, da konkretni materialni parametri tiskane kulture, dotlej nekaj neopaznega in samoumevnega, niso tako nepomembni; izkazalo se je, da je med platnice uvezen niz strani prinašal tudi stabilen, jasno zamejen in tihi

9 Gl. Ogrin 2005 in Dolinar 2005. Naloga strokovnih elit je med drugim gotovo tudi, da razrešijo vprašanje standardizacije navajanja digitalnih virov.

10 Celó Jerome McGann, ki poudarja nehierarhično in nelinearno naravo spletnih izdaj, priznava, da stran – četudi namenjena »nomadskemu« anarhičnemu bralcu – vendarle potrebuje arhitekta, ki načrtuje optimalno gibanje skozi kiberprostor (McGann, *The Rationale of Hypertext*). Podobno kot avtor v sodobni spletni literaturi ostaja nepričakovano močna figura in prihaja na dan celo tam, kjer ga ne bi pričakoval (gl. Hartling 2009), potrebujejo digitalne izdaje (avtorskega) načrtovalca, za lažjo navigacijo pa so jim pogosto dodana »linearna« pomagala (kronologije, biografije, bibliografije ipd.).

11 Primera odlične prakse sta Gutenbergov Distributed Proofreaders, sprva zasnovan kot podpora projekta Gutenberg, ali slovenski Wikivir.

12 O razlogih za to piše Lessig v *Svobodni kulturi*.

13 Gl. Sewell 2009, Kirschenbaum 2009 ter Brown idr. 2009.

14 Ob primeru slovenske kulturne dediščine se je tako treba odločiti, ali je meja slovenski jezik ali morda slovensko ozemlje; ali je treba enakovredno vključiti drugojezične literature na tleh današnje Slovenije, emigrantsko, zamejsko literaturo. Ker z napredovanjem tiskane kulture količina besedil postaja vse manj obvladljiva, se kot zelo smiselna kaže možnost začeti kronološko pri najstarejših besedilih (rokopisih), ki predstavljajo poseben edicijski in hermenevtični izziv.

tekst, ki je vednost organiziral kategorizacijsko, linearno in hierarhično.¹⁵ Virtualizacija gradiv je torej na novo osvetlila dileme, ki sta jih premlevala že Diderot ali Fichte – namreč o razmejitvi »dela« kot duhovne (virtualne) entitete od njegovih akcidenčnih, materialnih pojavnih oblik.¹⁶

Hipertekstne izdaje v nobenem primeru niso samo zaslonske verzije tiskanih¹⁷ – ontološki parametri tu pač ne ostajajo nedotaknjeni. Pa vendar je njihova dragocenost med drugimi tudi ta, da lahko epoho tiska z njeno linearno strukturo vred postavijo v muzej, jo ponudijo na ogled takšno, kot je nekoč bila, z vsemi njenimi aporijami vred. Ta akt muzealizacije pa seveda lahko pomeni tudi simbolično slovo od nje in usmeritev k izzivom novih, asociativnih struktur podatkovno gostega in »gladkega« prostora, v katerem ni več sledu o materialnosti, linearnosti in enosmerni komunikaciji tiskane dobe.

Literatura

- BLACKWELL, Christopher; CRANE, Gregory, 2009: Conclusion: Cyberinfrastructure, the Scaife Digital Library and Classics in a Digital age. *Digital Humanities Quarterly* 3/1: <<http://www.digitalhumanities.org/dhq/>>. (Dostop 24. 6. 2009.)
- BROWN, Susan, CLEMENTS, Patricia, GRUNDY, Isobel; RUECKER, Stan; ANTONIUK, Jeffery; BALAZS, Sharon, 2009: Published Yet Never Done: The Tension Between Projection and Completion in Digital Humanities Research. *Digital Humanities Quarterly* 3/2: <<http://www.digitalhumanities.org/dhq/>>. (Dostop 24. 6. 2009.)
- BURNARD, Lou, 2005: Encoding standards for the electronic edition. Ogrin, Matija (ur.): *Znanstvene izdaje in elektronski medij*. Ljubljana: ZRC SAZU. 25–42.
- CHARTIER, Roger, 2008: *Pisanje in brisanje*. Ljubljana: Studia Humanitatis.
- DOLINAR, Darko: Znanstvene izdaje in literarna veda. Ogrin, Matija (ur.): *Znanstvene izdaje in elektronski medij*. Ljubljana: ZRC SAZU. 155–166.
- DOVIČ, Marijan, 2005: Priprava elektronske znanstvenokritične izdaje pesmi Antona Podbevška. Ogrin, Matija (ur.): *Znanstvene izdaje in elektronski medij*. Ljubljana: ZRC SAZU. 101–116.
- ERJAVEC, Tomaž, 2005: Elektronske znanstvenokritične izdaje slovenskega slovstva: standardi in izzivi. Ogrin, Matija (ur.): *Znanstvene izdaje in elektronski medij*. Ljubljana: ZRC SAZU. 51–70.
- HARTLING, Florian, 2009: The digital author? Authorship in the digital era. *Primerjalna književnost* (pos. št., v tisku).
- HLADNIK, Miran: *Za elektronsko knjigo*: <<http://lit.ijs.si/zaeknjigo.html>>. (Dostop 24. 6. 2009.)
- JUVAN, Marko, 2006: *Literarna veda v rekonstrukciji*. Ljubljana: LUD Literatura.
- KIRSCHENBAUM, Matthew G., 2009: Done: Finishing Projects in the Digital Humanities: *Digital Humanities Quarterly* 3/2: <<http://www.digitalhumanities.org/dhq/>>. (Dostop 24. 6. 2009.)
- LANDOW, George P.: Material Object, Virtual Spaces: <<http://www.19.bbk.ac.uk/issue6/digital%20forum/landowapril2008.pdf>>. (Dostop 24. 6. 2009.)
- LESSIG, Lawrence: *Svobodna kultura*. Ljubljana: Krtina, 2005.
- MCGANN, Jerome: The Rationale of HyperText: <<http://jefferson.village.virginia.edu/~jjm2f/rationalale.htm>>. (Dostop 24. 6. 2009.)
- MCGANN, Jerome: Marking Texts of Many Dimensions: <<http://jefferson.village.virginia.edu/~jjm2f/blackwell.htm>>. (Dostop 24. 6. 2009.)
- OGRIN, Matija, 2005: Uvod. O znanstvenih izdajah in digitalni humanistiki. Ogrin, Matija (ur.): *Znanstvene izdaje in elektronski medij*. Ljubljana: ZRC SAZU. 7–21.
- STOICHEFF, Peter; TAYLOR, Andrew, 2004: Architecture, Ideologies, and Materials of the Page. Stoicheff, Peter; Taylor, Andrew (ur.): *The Future of the Page. Studies in Book and Print Culture*. Toronto: University of Toronto Press. 3–26.
- SEWELL, David, 2009: It's For Sale, So It Must Be Finished: Digital Projects in the Scholarly Publishing World. *Digital Humanities Quarterly* 3/2: <<http://www.digitalhumanities.org/dhq/>>. (Dostop 24. 6. 2009.)
- 15 Gl. Stoicheff, Taylor 2009. Ni presenetljivo, da so se najbolj pronicljive analize prehoda iz tekstne v hipertekstno dobo (McGann, Landow ...) porajale ravno iz izkušnji pri pripravi digitalnih arhivov.
- 16 Gre za debato, ki jo je sprožil problem prenašanja lastninskih pravic v zvezi z delom (gl. Chartier 2008: 247–270).
- 17 V e-okolju vsekakor izgublja smisel iskanje »ultimativne« variante teksta za vsako ceno, saj je mogoče nehierarhično predstaviti vse variante, proces nastajanja itd. (gl. npr. Podbevškove zbrane pesmi v e-ZISS). Sam tekst se torej izkaže kot dinamičen, nefiksiran, »ohlapien« pojav (Juvan 2006: 94–99). Whitmanov arhiv naravnost odlično prikaže fluidno, »procesno« naravo besedila.

Tuje spletne izdaje in arhivi

(Dostop 2. 7. 2009.)

The Complete Writings and Pictures of Dante Gabriel Rossetti. A Hypermedia Archive: <<http://www.rossettiarchive.org/>>.

Česká elektronická knihovna - Poezie 19. a počátku 20. století: <<http://www.ceska-poezie.cz/cek/>>.

Distributed proofreaders: <<http://www.pgdp.net/c/>>.

Google books: <<http://books.google.com/books>>.

Gutenberg Project: <<http://www.gutenberg.org>>.

Internet archive: <<http://www.archive.org/index>>.

IntraText Digital Library: <<http://www.intratext.com/SLV/>>. *Manuscriptorium:* <<http://www.manuscriptorium.com>>.

Orlando: Women's Writing in the British Isles from the Beginnings to the Present: <<http://www.cambridge.org/online/orlandoonline/>>.

Perseus Digital Library: <<http://www.perseus.tufts.edu/hopper/>>.

The Walt Whitman Archive: <<http://www.whitmanarchive.org/>>.

The William Blake Archive: <<http://www.blakearchive.org/blake/>>.

Slovenski spletni viri (Dostop 2. 7. 2009.)

Beseda. Virtualna slovenska knjigarna: <<http://www.omnibus.se/beseda/>>.

Digitalna knjižnica Slovenije: <<http://www.dlib.si/dlib.asp>>.

Elektronske znanstvenokritične izdaje slovenskega slovesstva (e-ZISS): <<http://nl.ijs.si/e-zrc/>>.

Nova beseda: <http://bos.zrc-sazu.si/s_nova_beseda.html>.

Wikivir: Slovenska leposlovna klasika: <http://sl.wikisource.org/wiki/Wikivir:Slovenska_leposlovna_klasika>.

Zbirka slovenskih leposlovnih besedil: <<http://lit.ijs.si/leposl.html>>.