

Slovenski biografski leksikon v prostem pristopu: med izdelavo in nadgradnjo

Petra Vide Ogrin*
Jan Jona Javoršek**
Mojca Mlinar Strgar*
Tomaž Erjavec**

*Slovenska akademija znanosti in umetnosti
** Institut Jožef Stefan

Vsebina

- SBL (1925-1991) – nekaj podatkov
- Smernice in označevalna shema TEI
- Digitalizacija SBL → spletni SBL
- Tehnologija: proizvodovalni sistem, repozitorij
- Prihodnost: SBL → Slovenska biografija

SBL (1925-1991)

- 15 zvezkov + kazalo
- 5047 gesel (85 rodbinskih), preko 5100 oseb, 16 mio znakov
- Preko 350 avtorjev, uglednih znanstvenikov več generacij
- Biografska metoda: primarni viri; J. Munda: „ves čas ista: opisati življenje in delo ljudi, ki so znanosti, kulturi in narodovemu razvoju sploh kaj tehtnega prispevali od daljnih do današnjih dni (...) zanesljivo podati bistvena biografska dejstva, važnejšo literaturo in jedrnato oceno, kjer je že možna.“
- Prvi zvezki nedostopni – odločitev za spletno izdajo

Temeljna vodila pri pripravi spletne izdaje SBL:

- Prost dostop do podatkov SBL
 - Uporaba in nadgradnja odprtokodnega programja
 - Uporaba standardov
 - Izmenljivost in trajnost podatkov
- Ustrezna: označevalna shema TEI in repozitorijski sistem Fedora Commons

Smernice TEI

- TEI – Text Encoding Initiative: začetek 1987, konzorcij strokovnjakov s področij humanistike, družboslovja in računalništva, prve smernice za strukturo označevanje el. besedil
- 2007 – TEI P5, *de facto* standard na podr. humanistike in družboslovja, za vse vrste besedil in podatkov, dobro dokumentirana, prilagodljiva, izmenljiva
- DLF, 1998, LC: *TEI in XML v digitalnih knjižnicah*; 2006 zadnji in izpopolnjen dokument, 5 stopenj označevanja dokumentov
- Zakaj TEI za SBL: ... + ker omogoča uporabo iste označevalne sheme v metapodatkih in celotnem besedilu

Digitalizacija SBL

- Skeniranje tiskanega SBL → grafični format
- OCR → besedilni format
- Ročna redakcija
- Dokument v formatu OpenOffice: prosto dostopen urejevalnik el. besedila + integriran pretvorbeni program (XSLT) → format XML/TEI P5, preprosta struktura
- Podrobno strukturno označevanje podatkov: avtomatsko izluščanje biografskih podatkov + obsežno ročno označevanje upošteva modul smernic TEI P5 za označevanje biografskih in prozopografskih podatkov.

Skeletni prikaz dokumenta v XML/TEI

```
<?xml version="1.0" encoding="utf-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader> <!-- vsebina --> </teiHeader>
  <text>
    <front> <!-- vsebina --> </front>
    <body> <!-- vsebina --> </body>
    <back> <!-- vsebina --> </back>
  </text>
</TEI>
```

- Strukturno označevanje podatkov znotraj <body> – dva namena:
 - Zahtevno iskanje
 - Izpis na zaslonu

Začetek članka iz SBL

Z bistvenimi biografskimi dejstvi na samem začetku:

- Variante priimka in imena
- Poklici in dejavnosti
- Datum in kraj rojstva
- Datum in kraj smrti
- (oz. obdobje delovanja, kadar datumi niso znani)

→ Te smo avtomatsko izluščili (Perl) + podrobno ročno označevanje

Pohlin Marko o., diskalceat (med svetom Anton), sprožitelj slovenskega književnega preroda, r. 13. apr. 1735 v šempet. predmestju Lj. gostilničarju Janezu Juriju P.-u iz Kamnika (gl. čl. Pohlin), u. 4. febr. 1801 za podagro v diskalc. samostanu v Mariabrunnu pri Hütteldorfu zap. od Dunaja. Vsiljuje se misel, da so v Lj. imeli z njim sitnosti, preden je prišel z 12. l., morda po posredovanju p. Bernarda P. (gl. čl.), v franč. gimn. v Novem mestu, kjer je dovršil 1747/48 drugi, 1748/49 tretji razred. Učitelj mu je bil 1747/48 p. Romualdus Laurin, 1748/49 p. Ildephonsus Karner, a med njegovimi součenci so bili tudi Hrvatje. Napredoval je dobro (1747/48: studii boni, bonorum morum; 1748/49: capacitatis et profectus praestantis, morum commendabilium), a se vrnil v Lj., kjer je bil morda že kot četrtošolec 1749/50 sprejet v jezuit-


```

<person xml:id="sbs443704" n="sbl02207">
  <sex value="1" />
  <persName>
    <roleName type="eccl">oče</roleName>
    <forename type="monastic">Marko</forename>
    <surname>Pohlin</surname>
    <roleName type="operosorum">operoz</roleName>
  </persName>
  <persName>
    <forename>Anton</forename>
    <surname>Pohlin</surname>
  </persName>
  <persName type="pseudo">
    <name>Peter Kumras</name>
  </persName>
  <persName>
    <roleName type="eccl">pater</roleName>
    <name type="monastic">Markus</name>
  </persName>
  <persName type="operosorum">
    <name>Novus</name>
  </persName>
  <note>sprožitelj slovenskega književnega preroda</note>
  <occupation>duhovnik</occupation>
  <occupation>redovnik</occupation>
  <occupation>diskalceat</occupation>
  <birth>
    <date when="1735-04-13">13. apr. 1735</date>
    <placeName>
      <settlement>Ljubljana</settlement>
    </placeName>
  </birth>
  <death>
    <date when="1801-02-04">4. febr. 1801</date>
    <placeName>
      <settlement>Dunaj</settlement>
      <district>Mariabrunn</district>
      <settlement xml:lang="ger">Wien</settlement>
      <district xml:lang="ger">Mariabrunn</district>
      <country>Avstrija</country>
    </placeName>
  </death>
</person>

```

1925

SLOVENSKI BIOGRAFSKI LEKSIKON

1991

Oseba:

Spol/tip:

?

Poklic:

Letnice:

do

rojstva / smrti

Kraj:

+

↕

Avtor:

Omembe:

Način:

točno

Vez:

ALI

Briši polja

Išči

- [Predgovor](#)
- [Kolofon](#)

Seznam: [A](#) [B](#) [C](#) [Č](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [Š](#) [T](#) [U](#) [V](#) [W](#) [Y](#) [Z](#) [Ž](#)

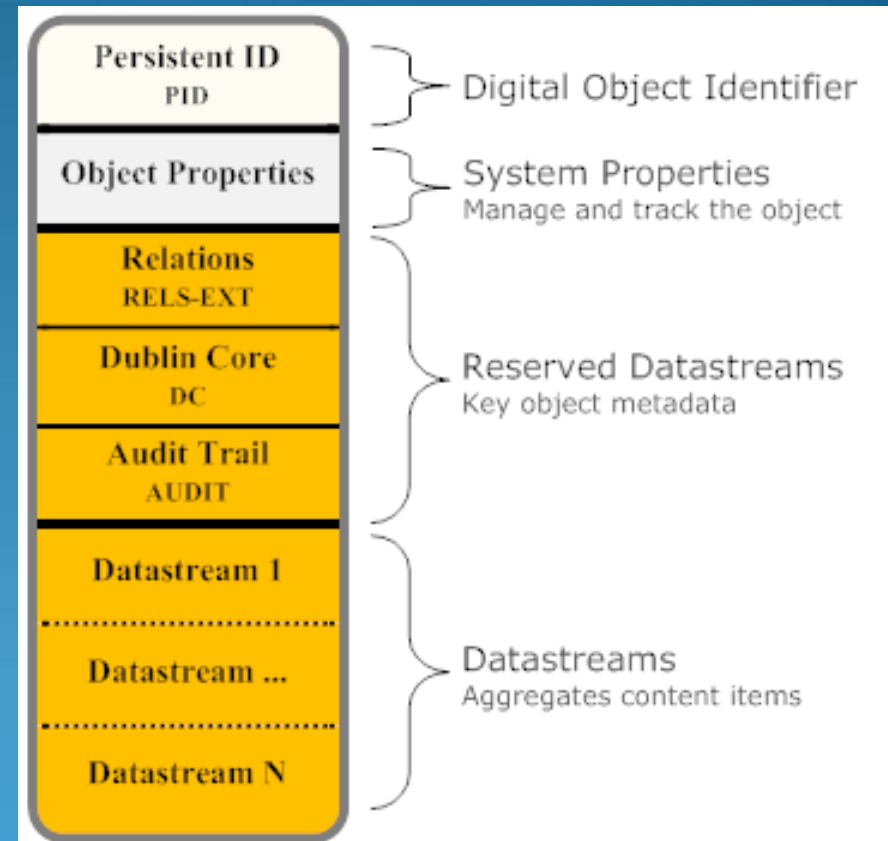
[Naključna oseba](#)[Seznam rodbin](#)[Rojeni na današnji dan](#)[Umrli na današnji dan](#)

Fedora

- **Standardi:** XML, spletne storitve (SOAP, REST), metapodatki (Dublin Core, OAI MH), XSLT, RDF itd.
- **Model digitalnih objektov v XML**
- **Hramba v datotekah**
(podatkovne baze v ozadju)
- **Replikacija**
(modeli so neodvisni od implementacije)

Fedorini objekti

- Preprosto: objekt = dokument
- Kompleksno: zbirke, medsebojni odnosi
- objektno-orientiran
 - podatkovni tokovi
 - diseminatorji
- Odnosi:
 - sistemski/objektni
 - Splošni (zbirke, odvisnost)



SBL v zbirki

- TEI ostaja izviren zapis
- Vsak članek lasten dokument
- Odnosi za zbirko SBL, črke, zvezke
- Poizvedovanja in transformacija XSLT: prikaz članka, abecedni sezname ipd.
- Iskalni vmesnik: Lucene SoLR + transformacije XSLT

Iskalni vmesnik

- Iskalni sistem kot vtič (Lucene SoLR)
- Poizvedovalni jezik Lucene
- Preprosta in kompleksna poizvedovanja v iskalnem oknu
- Napredna iskalna maska za preprosta kompleksna poizvedovanja

1925

SLOVENSKI BIOGRAFSKI LEKSIKON

1991

Oseba:

Spol/tip:



Poklic:

Letnice:

do

rojstva / smrti



Kraj:

+



Avtor:

Omembe:

Način:

točno



Vez:

ALI



Briši polja

Išči

- [Predgovor](#)
- [Kolofon](#)

Seznam: [A](#) [B](#) [C](#) [Č](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [Š](#) [T](#) [U](#) [V](#) [W](#) [Y](#) [Z](#) [Ž](#)

[Naključna oseba](#)[Seznam rodbin](#)[Rojeni na današnji dan](#)[Umrli na današnji dan](#)

Skrito poizvedovanje

- Vgrajene povezave, ki uporabljajo iskalnik:
 - Na današnji dan
 - Sodobniki
 - Generacija
 - Članki avtorja
- Možnosti:
 - Omembe krajev in drugih lastnih imen, poklicev, letnic, datumov ipd.

Načrti – Slovenska biografija

Slovenska biografija

podatkovna zbirka enotnih biografskih podatkov



Cilji

- enoten vzdrževan register osebnih podatkov
- dostopnost in uporabnost obstoječih virov
- enotna platforma za pripravo, razvoj in objavo biografskih podatkov in obdelav
- združenje nacionalnih in regionalnih virov in publikacij
- javno dostopen servis za strokovno in široko javnost

Prihodnost

- skupni razvoj platforme s sorodnimi projekti:
 - eZMono: Elektronske znanstvene monografije
 - NRSS: Neznani rokopisi slovenskega slovstva
 - Iskalni vmesnik za eZISS: Elektronske znanstvenokritične izdaje slovenskega slovstva
- skupni vmesnik knjižnice in iskalnika

<http://nl.ijs.si/fedora/sbl/>
(beta)