

# LNB atvērtie dati pētniecībai

Mārīte Apenīte, Uldis Bojārs, Ansis Garda,  
Anita Goldberga, Maira Kreislere,  
Anita Rašmane, Elza Ungure

Šo pētījumu ir finansējusi Latvijas Zinātnes padome, projekts “Latvijas atmiņas institūciju dati digitālajā telpā: vienojot kultūras mantojumu”, projekta Nr. lzp-2019/1-0365.

**Atslēgvārdi:** saistītie dati, atvērtie dati, LNB atvērto datu portāls, LNB ontoloģiju serviss, nacionālais tēzauris, nacionālā bibliogrāfija, oficiālā izdevējdarbības statistika

## Ievads

Kultūras mantojuma institūcijas – galerijas, bibliotēkas, arhīvi un muzeji – dokumentē, kontekstualizē, saglabā un indeksē kultūru un kolektīvo atmiņu (Pessach 2008). Tādējādi šīs institūcijas nodrošina pagātnes pieredzes un radošo izpausmju pārnesi uz tagadni un nākotni (Byrne 2014). Mūsdienās atmiņas institūcijām jārēķinās ar paradigmas maiņu tajā, kā zināšanas un dati tiek saglabāti un izplatīti, publiskajā diskursā arvien lielāku uzsvāru liekot uz digitālu brīvpiekļuves datu un saistīto atvērto datu<sup>1</sup> (*linked open data*) nozīmi, kas izpaužas tādos starptautiskos projektos kā *European Open Science Cloud* un no tā atvasinātajos projektos – *HIRMEOS*, *OPERAS* un *OpenAIRE*. Salīdzinājumā ar vēsturiskajām alternatīvām tieši atvērti, identificējami, strukturēti un saistīti mašīnlasāmi dati tiek uzverti kā labāk pieejami, atkārtoti lietojami un neizmeļami, tādēļ tiek uzskatīts, ka tiem ir lielāks potenciāls tikt izmantotiem dažādos pētījumos, uzlabot zinātnisko atklātību, kā arī veicināt sadarbību starp dažādām pusēm (Zhang u. c. 2011).

Eiropas līmenī augstas kvalitātes datu pieejamības nodrošināšanu radošiem un inovatīviem risinājumiem paredz Eiropas datu stratēģija<sup>2</sup>. Līdzīgi kā Eiropas datu stratēģijā, arī Latvijas attīstības stratēģijā līdz 2030. gadam (Latvija 2030) kā viens no rīcības virzieniem noteikta Latvijas kultūras kapitāla saglabāšana un attīstība, paplašinot Latvijas digitālo kultūras telpu, “piedāvājot precīzus un objektīvus informācijas avotus, kā arī kompetentus komentārus un analīzi” (Saeima 2010: 18). Tādēļ arī LNB 2019.–2023. gada attīstības stratēģija nosaka, ka “datus veidojam ne vien ērtai un efektīvai piekļuvei, bet arī pētniecībai. Nacionālas nozīmes datus bagātināsīm atkārtotai izmantošanai un tīmeklī publicēsīm saistīto datu formā, lai LNB veidotie dati būtu savienojami ar citu institūciju datiem, veidojot lietotājiem draudzīgas un pilnīgas piekļuves formas” (LNB 2019), kā arī lai veicinātu starpnozaru zināšanu jēgpilnu savietošānu un atkārtotu izmantošanu.

- 1 Saistītie dati ir principu un tehnoloģiju kopums (Berners-Lee 2006; Bizer et al. 2011) strukturētu datu publicēšanai tīmeklī. Saistītie dati ir veids, kā tīmeklī publicēt mašīnlasāmus datus par dažādiem objektiem (piemēram, bibliotēkas kolekcijas darbiem), to īpašībām un savstarpējām saitēm. Uz objektiem var atsaukties, izmantojot unikālus URI (vienota resursu identifikators). Atvērtie dati ir dati, kuriem jebkurš lietotājs var brīvi piekļūt, tos brīvi izmantot un izplatīt tālāk. Atvērto saistīto datu veidošanā saistīto datu principi tiek lietoti datu publicēšanai brīvpieejā.
- 2 <https://eur-lex.europa.eu/legal-content/LV/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>

LNB misija nepārtraukti attīstīt pārbaudītu, ticamu, mūsdienu standartiem atbilstošu zināšanu bāzi, izmantojot atbilstošus datu standartus un apstrādes rīkus, un tās piekļuvu pētniekiem, studentiem un plašākai sabiedrībai, izriet gan no *Latvija 2030* ietvertā rīcības virziena, gan LNB attīstības stratēģijas. Tomēr vienlaikus LNB kā atmiņas institūcija apzinās, ka tās glabātais kultūras mantojums ir tikai daļa no kopējā nacionālā kultūras mantojuma un lietotāju vajadzību labākai apmierināšanai nepieciešama sadarbība ar citām atmiņas institūcijām – kultūras mantojuma glabātājām.

Kultūras informāciju sistēmu centra (KISC) mērķis ir nodrošināt kultūras mantojuma institūcijas ar nepieciešamajiem informācijas tehnoloģiju resursiem un sistēmām katrā nozarē. Tas nozīmē, ka arhīvos, bibliotēkās, muzejos u. c. atmiņas institūcijās katrā nozarē tiek izmantotas līdzīgas vai vienas un tās pašas zināšanu organizācijas formas un datu vērtības tiek interpretētas līdzīgi. Tomēr vienlaikus tas nozīmē, ka kopumā Latvijas atmiņas institūcijas izmanto dažādus, savstarpēji ne vienmēr saderīgus datu un zināšanu organizācijas modeļus, kas apgrūtina resursu integrāciju vienotās digitālās platformās un kolekcijās. Digitālā vide rada jaunas sadarbības un semantiskās kontekstualizācijas iespējas, kas Latvijā līdz šim nav izpētītas un ieviestas atmiņu institūciju sadarbībā, kā tas, piemēram, veiksmīgi noticis Somijā u. c. Eiropas un pasaules valstīs<sup>3</sup>.

Tādēļ viens no Latvijas atmiņas institūciju, t. sk. LNB, mērķiem ir attīstīt semantiskās sadarbības potenciālu datu jomā un veicināt dažādu institūciju nevienmērīgu datu kopu sadarbību digitālajā telpā, ļaujot sasaistīt šo informāciju ar ārējiem avotiem un nodrošinot viedu, intuitīvu meklēšanas funkcionalitāti, kas var palīdzēt pētniekiem, citām ieinteresētajām pusēm (piemēram, izdevējiem, žurnālistiem utt.) un plašākai sabiedrībai (skat. Gruber 2008).

Paredzot uzdevumu uzlabot LNB veidoto datu kopu un vārdnīcu kvalitāti un radot iespējas tās padarīt par lietotājam maksimāli draudzīgiem datiem, t. sk. saistīto datu kopām un ontoloģijām, LNB 2020. gadā uzsāka projektu “Latvijas atmiņas institūciju dati digitālajā telpā: vienojot kultūras mantojumu” (projekta Nr. lzp-2019/1-0365). Tā ietvaros 2021. gadā tika izveidots LNB Atvērto datu portāls<sup>4</sup>, kurā publicētas četras uzticamu, pārbaudītu un strukturētu datu kopas.

Saistīto datu kolekcija “Rainis un Aspazija” satur atvērtus un strukturētus datus, kas kolekcijas pētniekus aizved uz konkrētiem faktiem par Raini un Aspaziju. Nacionālā tēzaura datu kopas, kas ir publicētas LNB ontoloģiju servīsā, atkārtotai izmantošanai padara pieejamu informāciju par tematiskajiem priekšmetiem, ģeogrāfiskajiem nosaukumiem un formas/žanra terminiem. Nacionālās bibliogrāfijas datu kopa satur informāciju par izvēlētajā laika periodā (2017–2020) Latvijā publicētajām grāmatām, to autoriem u. c. saistīto informāciju. Izdevējdarbības statistikas datu kopā ir dati par izdevējdarbību Latvijā, sākot no 1991. gada. Dati publicēti neapkopotā veidā, neuzrādot ziņas par izdevumu tirāžu.

3 Daži jau publicētu atmiņas institūciju atvērto saistīto datu kopu piemēri ir Britu nacionālā bibliogrāfija, ASV Kongresa bibliotēkas saistīto datu serviss, vairākas Somijas Nacionālās bibliotēkas datu kopas (Ontoloģiju serviss *Finto*, Nacionālā bibliogrāfija un citas) kā arī un VIAF projekts (Hickey and Toves 2014).

4 <https://dati.lnb.lv/>

Turpmākajās nodaļās iepazīstinām ar datu kopu saturu, datu kopu veidošanas specifiku un ar to saistītajiem izaicinājumiem un nākotnes perspektīvām, kas, iespējams, būs aktuālas arī citām atmiņas institūcijām, veidojot nacionālajā un starptautiskajā digitālajā telpā sadarbspējīgus datus.

## Saistīto datu kolekcija “Rainis un Aspazija” (RunA)

Saistīto datu kolekcija “Rainis un Aspazija” (RunA)<sup>5</sup> sāka veidot 2014. gadā un iecerēta kā vienots piekļuves punkts abu dzejnieku literārajam un epistulārajam mantojumam, kā arī visdažādākajiem ar viņu radošo darbību un dzīvi saistīto rakstisko materiālu, foto, audio, kinohroniku u. c. digitalizētajiem resursiem. Šobrīd kolekcijā ir iekļauti tādi satura veidi kā Raiņa un Aspazijas darbu pirmizdevumi (tajā skaitā darbi, kuri publicēti tikai periodikā), šo darbu literārās anotācijas, dzejnieku savstarpējā sarakste (pirmslaulību periods un Šveices trimdas periods), fotogrāfijas, arhīvu dokumenti u. c. materiāli.

RunA kolekcijas satura daudzpusību, dažādību un plašo aptvērumu nebūtu iespējams nodrošināt vienas institūcijas robežās, jo tajā iekļautie materiāli fiziski glabājas dažādās kultūras mantojuma institūcijās. Kolekcija ir kopdarbības rezultāts, kas tapis, iesaistoties Latvijas Nacionālā arhīva, Latvijas Nacionālās bibliotēkas, Latvijas Universitātes Literatūras, folkloras un mākslas institūta, Memoriālo muzeju apvienības un Rakstniecības un mūzikas muzeja speciālistiem.

Digitālo objektu skaits RunA kolekcijā 2021. oktobrī ir vairāk nekā 1300 objektu visdažādākajās digitālajās formās un izmantojamības līmeņos (t. sk. 83 literārie darbi un 395 vēstules un to anotētie teksti). Kolekcija satur arī audio un video datnes, arhīva dokumentu digitalizētos attēlus, prezentāciju datnes un citus satura veidus. Tekstveida objekts “vēstule” kolekcijā iekļauts gan kā vēstules vai atklātnes lapušu attēlu grupa, gan kā anotēts mašīnlasāms teksts; oriģināldarbs – kā digitalizēta grāmata un strukturēti bibliogrāfiskie dati. Kolekcijā šobrīd nav iekļauti Raiņa un Aspazijas tulkoto darbu teksti, izņemot Johana Volfganga fon Gētes “Faustu”. Turpmāk paredzēts iekļaut arī anotētus Raiņa dienasgrāmatas tekstus.

RunA kolekciju no citām digitālajām kolekcijām un digitālo resursu repozitorijiem atšķir datu organizācijas metode – tā ir saistīto datu kolekcija, kas papildus digitalizētajiem objektiem satur identificēto entīšu (personu, organizāciju, vietu, datējumu, notikumu u. c.) tīklu, kas sasaista šos objektus, piešķir tiem nozīmi, vietu un laiku, piesaista datus no citiem publiski pieejamiem datu avotiem, ļaujot analizēt un vizualizēt Raiņa un Aspazijas darbības sociālo vidi un biogrāfiskos faktus. RunA šim nolūkam tiek izmantoti arī bibliogrāfiskie metadati.

5 <https://runa.lnb.lv/en/>



1. attēls. Digitālo objektu un entīšu meklēšanas saskarne.

Daļa no kolekcijas tekstveida materiāliem – darbu literārās anotācijas un Raiņa un Aspazijas sarakste – ir manuāli anotēti, tekstā atzīmējot dažādu entīšu (personas, vietas, datumi u. c.) pieminējumus. Šī darbība tiek veikta īpašā teksta anotēšanas rīkā (Bojārs u. c. 2018), un anotācijas ir tās, kas veido tīklu ar saitēm starp teksta dokumentiem un tajos pieminētajām entītēm. Šobrīd, anotējot RunA kolekcijas materiālu saturiskajai apstrādei, tiek izmantotas deviņas entīšu klases (persona, institūcija, darbs, vieta, ēkas un celtnes, notikums, laiks, jēdziens, lieta), kurām tieši RunA materiālu apstrādei atbilst 78 specializētas anotāciju klases. Anotēšanas rīkā ir iespējams ne tikai atzīmēt entīšu pieminējumus, veidojot vienkāršas anotācijas un saglabājot atbilstošās entītes datus īpašā entīšu datubāzē, bet arī veidot saliktās anotācijas (vienkāršo anotāciju loģiskus “teikumus”), kā arī iezīmēt dokumenta struktūru ar strukturālajām anotācijām (Bojārs u. c. 2018).

Anotācijas veido datu kopu, kas ļauj jēgpilni saistīt gan RunA kolekcijas objektus un entītes, gan tekstā minētās entītes ar ārējos tīmekļa informācijas avotos atrodamiem datiem. RunA kolekcijā tiek publicēti atvērti, strukturēti dati par kolekcijas objektiem, entītēm un saitēm to starpā (piemēram, saite no dzejnieku darba literārās anotācijas uz vēstulē pieminētu entīti).

Balstoties uz datiem par kolekcijas objektiem, entītēm un saitēm to starpā, tiek veidoti RunA datu tīklojumi. Datu tīklojums vizuāli attēlo kolekcijas objektu un entīšu savstarpējo saistību. Kolekcijā ir divu veidu tīklojumi – objektu tīklojums un entīšu tīklojums. *Objektu tīklojums* ir datu vizualizācija (grafs), kurā divas virsotnes (kolekcijas objekti) ir savstarpēji saistītas, ja abos dokumentos ir pieminēta viena un tā pati entīte. Entīšu tīklojums ir vizualizācija, kurā divas virsotnes (entītes) ir savstarpēji saistītas, ja abas entītes ir pieminētas vienā un tajā pašā anotētajā dokumentā.

Anotējot kolekcijas tekstveida materiālus, tika izveidoti vairāk nekā 14 tūkstoši anotāciju, kas atsaucas uz 2600 identificētām entītēm. Publiski pieejamajā RunA kolekcijas portālā šo bagātīgo informāciju iespējams atrast un izmantot tikpat vienkārši kā digitālos objektus. Lai to izdarītu, tikai jāpārvelk virtuāls meklēšanas “slēdzītis” no “Objekti” uz “Entītes”:

Wikidata

**Mošanāsforma**  
Zana Pjēra Beranžē, 1780-1857

**Klase**  
Person

**Jaunās formas**

- **LV** Zana Pjēra Beranžē, 1780-1857 **arvārdi**
- **LV** Pjēra Zana Beranžē, 1780-1857
- **LV** Pjēra-Zana de Beranžē, 1780-1857 **arvārdi**
- **LV** Pjēra-Zana de Beranžē, 1780-1857 **arvārdi**
- **LV** Pjēra-Zana de Beranžē, 1780-1857 **arvārdi**
- **LV** Pjēra-Zana de Beranžē, 1780-1857 **arvārdi**

**Ārējo identifikāciju saraksts**

- **VIAF** ALPFA ALUJ <https://viaf.org/viaf/1223994>
- **VIAF** NMF <http://www.viaf.org/viaf/1223994>
- **VIAF** ISNI <https://www.isni.org/en/0000000120956640>
- **VIAF** Nekropols <https://nekropols.lv/ta/persona/1223994>
- **VIAF** EN Wikipedia [https://en.wikipedia.org/wiki/Pierre-Jean\\_de\\_Beranger](https://en.wikipedia.org/wiki/Pierre-Jean_de_Beranger)
- **VIAF** RU Wikipedia [https://ru.wikipedia.org/wiki/Пьер-Жан\\_де\\_Беранже](https://ru.wikipedia.org/wiki/Пьер-Жан_де_Беранже)
- **VIAF** DE Wikipedia [https://de.wikipedia.org/wiki/Pierre-Jean\\_de\\_Beranger](https://de.wikipedia.org/wiki/Pierre-Jean_de_Beranger)
- **VIAF** Wikidata <https://www.wikidata.org/wiki/Q1223994>
- **VIAF** VikiMāksla [https://commons.wikimedia.org/wiki/Category:Pierre-Jean\\_de\\_Beranger](https://commons.wikimedia.org/wiki/Category:Pierre-Jean_de_Beranger)
- **VIAF** ADP
- **VIAF** WorldCat Identity <https://www.worldcat.org/identities/ocn001209564>

**Pielikumi**  
Frānču dzejnieks.

Latviešu Wikipedia - Zana Pjēra Beranžē

**Zana Pjēra Beranžē, 1780-1857**

**Ārējo identifikāciju saraksts**

- **VIAF** ALPFA ALUJ <https://viaf.org/viaf/1223994>
- **VIAF** NMF <http://www.viaf.org/viaf/1223994>
- **VIAF** ISNI <https://www.isni.org/en/0000000120956640>
- **VIAF** Nekropols <https://nekropols.lv/ta/persona/1223994>
- **VIAF** EN Wikipedia [https://en.wikipedia.org/wiki/Pierre-Jean\\_de\\_Beranger](https://en.wikipedia.org/wiki/Pierre-Jean_de_Beranger)
- **VIAF** RU Wikipedia [https://ru.wikipedia.org/wiki/Пьер-Жан\\_де\\_Беранже](https://ru.wikipedia.org/wiki/Пьер-Жан_де_Беранже)
- **VIAF** DE Wikipedia [https://de.wikipedia.org/wiki/Pierre-Jean\\_de\\_Beranger](https://de.wikipedia.org/wiki/Pierre-Jean_de_Beranger)
- **VIAF** Wikidata <https://www.wikidata.org/wiki/Q1223994>
- **VIAF** VikiMāksla [https://commons.wikimedia.org/wiki/Category:Pierre-Jean\\_de\\_Beranger](https://commons.wikimedia.org/wiki/Category:Pierre-Jean_de_Beranger)
- **VIAF** ADP
- **VIAF** WorldCat Identity <https://www.worldcat.org/identities/ocn001209564>

**Sauksnis**

Arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →

arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →

arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →  
 arhivā kopā ar identifikāciju →

**Pierakstojumi (2)**

№	Titrs	Apraksts	Identifikācija dati
1	Arhivā kopā ar identifikāciju →	Arhivā kopā ar identifikāciju →	Arhivā kopā ar identifikāciju →
2	Arhivā kopā ar identifikāciju →	Arhivā kopā ar identifikāciju →	Arhivā kopā ar identifikāciju →

2. attēls. Franču dzejnieka Žana Pjēra Beranžē entītes lapa anotatora entīšu datubāzē.

3. attēls. Franču dzejnieka Žana Pjēra Beranžē entītes lapa kolekcijā RunA.

Piemēram, 2. attēlā ir redzami dati par franču dzejnieku Žanu Pjēru Beranžē, kas anotatora entīšu datubāzē glabājas teksta un hipersaišu formā.

RunA portālā šie dati tiek parādīti, informāciju bagātinot gan ar datiem no Vikipēdijas, gan informāciju, kuros RunA kolekcijas teksta dokumentos šī entīte pieminēta. Portālā iespējams piekļūt atvērtajiem saistītajiem datiem par šo entīti un tās datu tīklojumam.

Vairums kolekcijas satura ir brīvi pieejams atkārtotai lietošanai un ir izmantojams kā atvērtie dati. Mašīnlasāmi dati par kolekcijas objektiem, entītēm un saitēm to starpā ir publicēti atvērto saistīto datu formā (Bojārs u. c. 2020). Šie dati ir pieejami gan RunA kolekcijas vietnē, kur tie ir publicēti atbilstoši saistīto datu principiem (Berners-Lee 2006), gan apkopotā formā LNB atvērto datu portālā. No kolekcijas tīmekļa lapām objektu un entīšu datiem var piekļūt, spiežot pogas “Turtle RDF” un “RDF/XML” (vieglāk uztverams ir Turtle RDF formāts) vai pievienojot lapas adresei paplašinājumu atbilstošā formātā (.ttl vai .rdf) – piemēram, <https://runa.lnb.lv/60316.ttl>. LNB atvērto datu portālā ir publicētas divas šīs kolekcijas datu kopas, kas satur apkopotus kolekcijas datus Turtle RDF un RDF/XML formā.

## Nacionālais tēzauris

Datu un zināšanu organizācijas un pārvaldības politikas pārmaiņas notiek ne tikai bibliotēku katalogu vidē pasaulē, bet arī LNB Nacionālā tēzaura autoritatīvajos datos, jo ir būtiski nodrošināt vienotu kontrolēto terminu lietojumu ne vien bibliogrāfiskajos ierakstos bibliotēku elektroniskajos katalogos un dažādās datubāzēs, bet veicināt arī datu atkārtotu izmantojamību un savietojamību ar citām starptautiskajām kontrolētajām vārdnīcām tīmekļa vidē.

Latvijas Nacionālā tēzaura pamatu veido LNB Priekšmetu saraksts (LNBPS), kura izveidē kopš 2000. gada tiek lietots ASV Kongresa bibliotēkas Priekšmetu saraksts (*Library of Congress Subject Headings* – LCSH<sup>6</sup>), kas tiek adaptēts atbilstoši latviešu valodas un zinātņu nozaru terminoloģijas prasībām. LNBPS tiek veidots kā kontrolēta un strukturēta vārdnīca ar tēzauram līdzīgu struktūru, kurā jēdzieni tiek attēloti ar terminiem, starp kuriem ir hierarhiskās, ekvivalentās un asociatīvās attiecības, un tas tiek izmantots universāla satura indeksēšanai ar kontrolētiem terminiem, tādējādi nodrošinot dažādu informācijas resursu tematisko aptvērumu, piekļuvu un informācijas izguvi.

Lai nodrošinātu priekšmetu datu kvalitātes uzturēšanu, piekļuves un informācijas izguves precizitāti, piedāvājot lietotājiem izvēlēties starp plašākiem, šaurākiem, asociatīviem vai līdzīgas nozīmes terminiem, kontrolējot to sinonīmiju, kvazisinonīmiju un homonīmiju, LNB tiek veidota un uzturēta Autoritatīvā datubāze, kas ietver Nacionālā tēzaura autoritatīvos datus (tematiskie, ģeogrāfiskie nosaukumi un formas/žanra priekšmeti) un bibliotēku datu

6 <https://id.loc.gov/authorities/subjects.html>

atveidošanā ieviešamā jaunā starptautiskā standarta RDA (*Resource Description and Access*)<sup>7</sup> entīšu (personas; dzimtas/ģimenes; institūcijas; pasākumu nosaukumi; darbu nosaukumi) autoritatīvos datus.

Kopš 2004. gada tiek veidoti ģeogrāfisko nosaukumu un tematisko priekšmetu autoritatīvie ieraksti, kas nodrošina ne vien šo priekšmetu pamatformu kontrolētu lietošanu, bet arī bilingvālu (latviešu un angļu) un daudzvalodu priekšmetisko piekļuvi un informācijas izguvi atsevišķu nozaru termiņiem vai konkrētu objektu nosaukumiem kā tematiskajiem priekšmetiem vai ģeogrāfisko nosaukumu priekšmetiem.

2014. gadā uzsāka arī pakāpeniska formas/žanra priekšmetu veidošana, izmantojot FAST (*Faceted Application of Subject Terminology*)<sup>8</sup> un LCGFT (*Library of Congress Genre/Form Terms*)<sup>9</sup>. Atsevišķos gadījumos, ja nav ekvivalenta priekšmeta ne FAST, ne LCGFT, tad LNB tiek izmantoti arī Čehijas Nacionālās bibliotēkas ECZENAS autoritatīvie ieraksti, vai tiek veidoti jauni lokālie priekšmeti atbilstoši nacionālās valodas terminradei, kultūras vajadzībām un zinātņu nozaru terminoloģijai. Kopš 2018. gada Latvijas vietu nosaukumiem pievieno arī koordinātas no Latvijas Vietvārdu datubāzes.

Latvijas Nacionālā tēzaura kontrolētie priekšmeti tiek izmantoti LNB daudzveidīgā krājuma satura atspoguļošanai un priekšmetiskās piekļuves nodrošināšanai elektroniskajā katalogā<sup>10</sup>, speciālo krājumu kolekcijās, Nacionālajā analītikas datubāzē un Digitālās bibliotēkas kolekcijās<sup>11</sup>. Tie nodrošina kontrolētu tematisko piekļuvi un datu izguvi arī Valsts nozīmes kopkataloga bibliotēkās un publisko bibliotēku katalogos.

Lai arī LNBPS, kas veido Nacionālā tēzaura pamatu, lietošana ir nodrošinājusi vienotu LNB un citu bibliotēku datu pārvaldības un zināšanu organizēšanas politiku jau 21 gadu, tomēr mūsdienu informācijas tehnoloģiju attīstība, dažādu informācijas resursu plašā un daudzveidīgā pieejamība interneta vidē un kultūras mantojuma digitalizācija veicina ne vien datu apjoma palielināšanos, bet maina arī lietotāju informācijas meklēšanas un lietošanas paradumus un vajadzības.

Zināšanu organizācijas pārmaiņu nepieciešamību bibliotēku vidē ietekmē arī saistītie dati, kuri dod iespēju datus publicēt tīmeklī, piešķirot to ierakstiem unikālus URI identifikatorus<sup>12</sup> un veidojot saites ierakstu starpā. Tomēr bibliotēku datu veidošanas formāta MARC21 iespējas neatbilst saistīto datu veidošanas principiem un nenodrošina saistīto datu veidošanu

7 <https://www.rdtoolkit.org/about>

8 <https://www.oclc.org/research/areas/data-science/fast.html>

9 <https://id.loc.gov/authorities/genreForms.html>

10 <https://lnb.lv/>

11 <https://lnb.lv/lv/digitala-biblioteka>

12 URI (vienotais resursu identifikators) ir identifikatoru veids, kas ļauj viennozīmīgi norādīt uz informācijas resursiem. HTTP URI izmantošana resursu identifikācijai ir viens no saistīto datu principiem.



un eksponēšanu lietotājiem, tāpēc autoritatīvajiem datiem jāpievieno hipersaites, lai nodrošinātu veidu, kā no MARC ieraksta piekļūt tīmekļa resursiem.

2019. gadā tika sāta Nacionālā tēzaura jaunveidoto un iepriekšveidoto priekšmetu autoritatīvo ierakstu bagātināšana ar hipersaitēm “uz autoritatīviem ārējiem resursiem (identifikatoru sistēmām un enciklopēdijām)” (LNB 2021: 17), tādējādi nodrošinot LNB autoritatīvo datu iekļaušanu “tīmeklī esošu strukturētu enciklopēdisku datu vidē” (LNB 2021: 17).

2020. gadā automatizēti tika atlasītas LNB tematisko priekšmetu angļu formu sakritības ar LCSH. Tika izskatīti trīs atlasīto ierakstu saraksti – tiešās atbilstības (ieraksti ar vienu LCSH identifikatoru), dublikāti (ieraksti ar vairākiem LCSH identifikatoriem) un kļūdu gadījumi (neatbilst neviens LCSH identifikators). Pēc atlasīto datu izskatīšanas un kļūdu novēršanas LCSH URI identifikatori tika automatizēti pievienoti LNB tematisko priekšmetu autoritatīvajos ierakstos. 2021. gadā līdzīgi tika sāta arī FAST identifikatoru pievienošana LNB tematisko priekšmetu autoritatīvajiem ierakstiem.

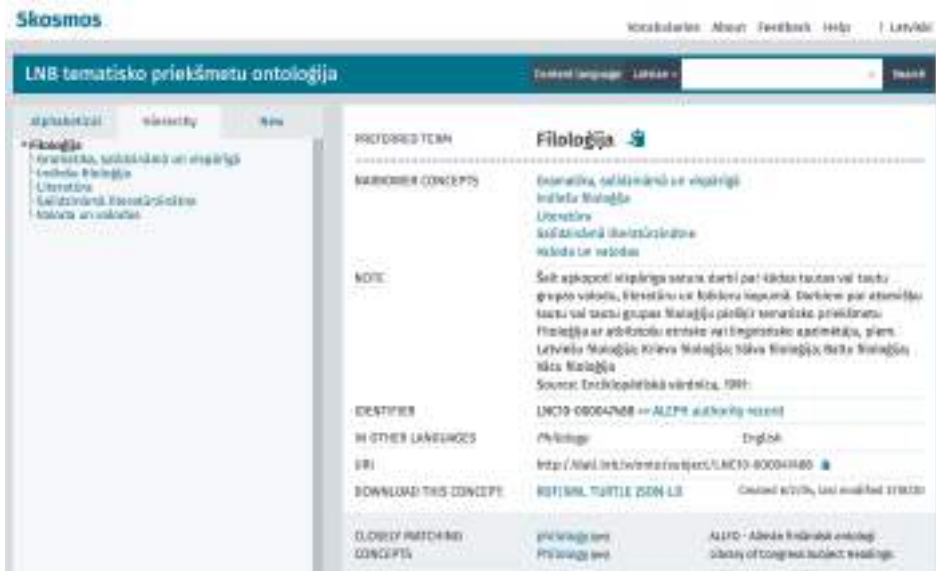
Saistot Nacionālā tēzaura autoritatīvos ierakstus ar Kongresa bibliotēkas saistītajiem datiem (*LC Linked Data*), kas ietver saites ar *FAST Linked Data*, *Wikidata* un *General Finnish Ontology YSO*, Nacionālā tēzaura autoritatīvie ieraksti var tikt savstarpēji saistīti arī ar citām atvērto saistīto datu kopām tīmeklī.

Turpinās arī jauno Nacionālā tēzaura autoritatīvo ierakstu bagātināšana ne vien ar LCSH un FAST identifikatoriem, bet tiek pievienotas hipersaites arī ar LCNAF (*Library of Congress Name Authority File*), LCGFT, VIAF, *Wikidata*, NKMP (Nacionālā kultūras mantojuma pārvalde), Latvijas Vietvārdu datubāzes un Nacionālās enciklopēdijas identifikatoriem. Pakāpeniski redakcijas procesā ar ārējo resursu hipersaitēm tiek bagātināti arī iepriekšveidotie autoritatīvie ieraksti.

Lai veicinātu Nacionālā tēzaura autoritatīvo datu atkalizmantošanu un sadarbību ar citu institūciju datiem un veidotu lietotājiem jaunas piekļuves formas, 2021. gadā tika arī sāts jauns daudzpakāpju process autoritatīvo datu pielāgošanai tīmekļa tehnoloģijām un publicēšanai atvērto saistīto datu formā, un LNB tika izveidots ontoloģiju serviss<sup>13</sup> (Apenīte, Bojārs 2021).

Ontoloģiju kā zināšanu pārvaldības modeļa izstrāde ir salīdzinoši jauna prakse kultūras un publiskajā sektorā. Daudzviet to potenciāls vēl nav pilnībā apzināts un izmantots, taču, piemēram, ir atzīts un ieviests Somijā, kur kopš 2011. gada tiek izmantots Somijas tēzauru un ontoloģiju serviss (*Finto*). Praktiski šajā pārvaldības modelī zināšanas, ar kurām līdz šim nav bijis iespējams apmainīties digitāli, ir jāattēlo tā, lai automatizēta sistēma tās saprastu, izmantotu un nodotu šo zināšanu kopumu to lietotājiem – cilvēkiem – saprotamā veidā (Milić u. c. 2018).

13 <https://dati.lnb.lv/onto/lv/>



4. attēls. Jēdziena 'filoloģija' lapa LNB ontoloģiju servisā, Skosmos vidē.

LNB ontoloģiju servisā ietilpst trīs datu kopas: vispārīgo un individuālo jēdzienu ontoloģija (*NLL subject authority ontology*), ģeogrāfisko nosaukumu ontoloģija (*NLL geo ontology*) un formas/žanra ontoloģija (*NLL genre/form ontology*). Kopumā tajās ir 55 740 jēdzieni.

Tematisko priekšmetu ontoloģija (*NLL subject authority ontology*) ietver tematiskos priekšmetus: abstraktus jēdzienus, dažādu nozaru terminus, svētku nosaukumus, arhitektoniskos objektus ar vai bez nosaukuma: arkas, baznīcas, klosterus katedrāles u. tml. reliģiskas celtnes, muižas un pilis, pieminekļus, memoriālos ansambļus, konkrētu ēku nosaukumus u. c., kas raksturo dažādu informācijas resursu saturu.

Ģeogrāfisko nosaukumu ontoloģijā (*NLL geo ontology*) ietverti administratīvie (valstis, apgabali, novadi, rajoni, provinces, pilsētas, pagasti, ciemi) un neadministratīvie vietu nosaukumi (planētas, kontinenti, okeāni, jūras, upes, salas, kalni, meži, kultūrvēsturiski novadi, neadministratīvas pilsētu daļas, ostas, dzelzceļa stacijas, dzelzceļa līnijas, izzudušas pilsētas, arheoloģisko izrakumu vietas, kapsētas, nacionālie parki un dabas liegumi, parki, dārzi u. tml. vietu nosaukumi).

Formas/žanra ontoloģija (*NLL genre/form ontology*) ietver priekšmetus, kas raksturo dažādu informācijas resursu formu vai žanru, piem., biogrāfijas, enciklopēdijas, katalogi, terminu vārdnīcas, dzeja, proza, latviešu proza, angļu stāsti u. tml. formas vai žanra priekšmeti.

Lai izveidotu ontoloģiju servisa datu kopas, Latvijas Nacionālā tēzaura autoritatīvie ieraksti tika pārveidoti atvērto saistīto datu formā un publicēti tīmeklī, izmantojot *Skosmos*<sup>14</sup>, kas ir atvērta pirmkoda tīmekļa rīks SKOS aprakstīto kontrolēto vārdnīcu un tēzauru publicēšanai, pārlūkošanai, meklēšanai un rediģēšanai (skat. 4. attēlu).

SKOS<sup>15</sup> (*Simple Knowledge Organization System – Vienkārša zināšanu organizēšanas sistēma*) izstrādātās specifikācijas un standarti (Miles, Bechhofer 2009) atbalsta zināšanu organizēšanas sistēmu, kā tēzauru, klasifikācijas shēmu, priekšmetu sarakstu un taksonomiju, izmantošanu semantiskajā tīmeklī un nodrošina priekšmetu autoritatīvo ierakstu pārveidi atvērto saistīto datu formā. SKOS nodrošina zināšanu organizēšanas sistēmu attēlošanas standarta veidu, informācijas kodēšanā izmantojot Resursu apraksta ietvaru (RDF), lai šo informāciju savstarpēji savietojami nodotu starp datoru lietojumprogrammām.

Sākotnēji bibliotēku datu standarta MARC21 formā pieejamo Nacionālā tēzaura autoritatīvo ierakstu pārveidošana un publicēšana atvērto saistīto datu formā ir daudzpakāpju process, kas ietver:

1. MARC21 autoritatīvo ierakstu bagātināšanu – ierakstiem tiek pievienoti unikāli identifikatori, ierakstu sistēmas numuri, lai norādītu, ka ieraksts satur saites ar citiem ierakstiem datu kopā, kā arī visiem pamatjēdzieniem tiek pievienots valodu koda apzīmējums *lv*;
2. MARC21 autoritatīvo ierakstu konvertēšanu SKOS jēdzienos, izmantojot *mc2skos* rīku<sup>16</sup>, kas MARC/XML datus pārveido SKOS RDF datu formā. Līdzīgi kā LCSH, arī Nacionālā tēzaura autoritatīvo ierakstu hipersaišu norādei ar citām ārējo resursu datu kopām tiek izmantota īpašība *skos:closeMatch*;
3. SKOS datu kopu publicēšanu atvērto saistīto datu formā, izmantojot *Skosmos*, kas nodrošina daudzvalodu lietotāju saskarni jēdzienu pārlūkošanai un meklēšanai, kā arī hierarhijas vizualizēšanai. Katram ontoloģijas jēdzienam ir savs unikāls URI identifikators, ko var izmantot, lai viennozīmīgi atsauktos uz šo resursu, un visi dati ir pieejami kā atvērtie saistītie dati.

Darbs LNB ontoloģiju servisa datu kopu bagātināšanā un uzlabošanā tiks turpināts ar šādiem procesiem: 1) datu bagātināšana ar valodu kodiem un citu ārējo resursu identifikatoriem; 2) regulāra datu atjaunināšana ar jaunajiem vai rediģētajiem Nacionālā tēzaura autoritatīvajiem ierakstiem; 3) jaunizveidoto datu publicēšana.

Nākotnē LNB ontoloģiju servisa datu kopas, kas publicētas atvērto saistīto datu formā, varētu tikt izmantotas bibliotēku informācijas saistīšanai ar Latvijas muzeju un arhīvu datiem, kultūras mantojuma objektu aprakstā izmantojot ontoloģiju servisa objektu URI

14 <http://skosmos.org/>

15 <https://www.w3.org/TR/skos-reference/>

16 <https://github.com/scriptotek/mc2skos>

identifikatorus. Tās varētu saistīt arī ar citu valstu atmiņas institūciju datiem un starptautiskajām kontrolētajām vārdnīcām tīmekļa vidē.

2022. gadā tika sākta arī jaunveidoto Nacionālā tēzaura tematisko priekšmetu autoritatīvo ierakstu bagātināšana ar Somijas Finto ontoloģijas YSO (*General Finnish Ontology*) somu un angļu valodas jēdzieniem (ja nav ekvivalenta jēdziena LCSH) un identifikatoriem, kā arī atlasītas LCSH un YSO angļu jēdzienu sakritības, lai veiktu arī iepriekšveidoto autoritatīvo datu automatizēto papildināšanu ar YSO identifikatoriem. 2022. gadā tiks sākta arī ģeogrāfiskā nosaukuma un formas/žanra priekšmetu automatizēta bagātināšana ar Kongresa bibliotēkas (LCNAF, LCSH un LCGFT) un FAST identifikatoriem. 2022. gada rudenī tiek plānots arī sākt Notikuma nosaukuma priekšmetu autoritatīvo ierakstu veidošanu un to papildināšanu ar ārējo resursu jēdzieniem un hipersaitēm.

Nacionālā tēzaura autoritatīvo datu bagātināšana ar jauniem priekšmetiem un ar dažādu ārējo resursu jēdzieniem un identifikatoriem ne vien veicina ilgtspējīgu tēzaura attīstību, bet arī šo datu sadarbību un atkārtotu izmantošanu atvērto saistīto datu formā.

## Nacionālā bibliogrāfija

Nacionālā bibliogrāfija ir informācijas kopums par fiziskā un digitālā formā Latvijā publicētiem izdevumiem, kas ir nacionālā kultūras mantojuma sastāvdaļa, un ar to saistīta struktūrēta, sistematizēta informācija un autoritatīvie dati – standartizētas, strukturētas, pārbaudītas un uzticamas ziņas par personām, institūcijām, vietām, notikumiem, pasākumiem, darbiem, priekšmetiem un citiem kontrolētu klasifikatoru vārdiskiem apzīmējumiem<sup>17</sup>. LNB kā nacionālās bibliogrāfijas centrs ir atbildīga par Latvijā radīto publikāciju autoritatīvu bibliogrāfisko datu nodrošināšanu un pieejamību tīmeklī. Ar mērķi palielināt datu atvērto pieejamību un izmantošanu gan citu valstu nacionālās bibliogrāfijas centriem, gan bibliotēkām un citām kopienām, t. sk. kultūras mantojuma institūcijām un izdevējiem, LNB publicē aktuālu starptautisku standartu prasībām atbilstošus, pārbaudītus un iespējami savlaicīgus nacionālās bibliogrāfijas datus. Veidojot nacionālās bibliogrāfijas datus, LNB izvērtē katra datu elementa atbilstību dažādu lietojumu vajadzībām un identificē saistīto datu pakalpojuma sniegšanai trūkstošos komponentus ne tikai aktuālajos, bet arī vēsturiskajos datos.

LNB atvērto datu portālā publicēto nacionālās bibliogrāfijas datu kopu veido strukturēti dati par saskaņā ar Obligāto eksemplāru likumu saņemtām Latvijā publicētām grāmatām, kas ir apjoma ziņā vislielākā nacionālās bibliogrāfijas daļa un kas tiek pakāpeniski rediģēta atbilstoši RDA standartam. Papildus minimālajām bibliogrāfiskajām ziņām, kas nodrošina konkrētā izdevuma atrašanu, identificēšanu, izvēli un atlasi, ierakstā tiek iekļauts detalizēts izdevumu tapšanā iesaistīto personu un institūciju uzskaitījums, ar priekšmetiem atklāts izdevumu

17 <https://likumi.lv/ta/id/62905-par-latvijas-nacionalo-biblioteku>

saturs, kas nodrošina tematisko meklēšanu, norādīti grāmatu paralēlie un oriģinālie nosaukumi tulkojumiem, izdevēji, tipogrāfijas, kā arī ziņas par līdzautoriem – tulkotājiem, ilustratoriem, sastādītājiem, priekšvārdu un pēcvārdu autoriem, fotogrāfiem, redaktoriem u. tml. Izvēlētajā nacionālās bibliogrāfijas datu kopā ietvertie dati par 2017.–2020. gadu ir labi strukturēti, kvalitatīvi un homogēni, jo pirms datu kopas publicēšanas atvērto datu portālā tie sagatavoti atbilstoši starptautiskajam datu veidošanas standartam RDA.

Portālā publicētā datu kopa satur informāciju par aptuveni 11,3 tūkstošiem grāmatu, un apjoms tiks pakāpeniski paplašināts gan ar jaunākajiem nacionālās bibliogrāfijas grāmatu daļas datiem (respektīvi, tuvākajā laikā tiks pievienoti dati arī par 2021. gadā LNB saņemtajām un reģistrētajām Latvijā publicētajām grāmatām), gan ar nacionālās bibliogrāfijas grāmatu daļas vēsturiskajiem datiem. Nacionālās bibliogrāfijas grāmatu daļas vēsturiskās datu kopas apjoms ir vairāk nekā 250 000 bibliogrāfisko ierakstu MARC21 formātā, un tas aptver laika periodu kopš pirmās zināmās latviešu valodā publicētās grāmatas 16. gadsimtā. Lai sagatavotu bibliogrāfiskos datus publicēšanai LNB atvērto datu portālā atbilstoši RDA un saistīto datu principiem, nacionālās bibliogrāfijas grāmatu daļas vēsturiskie dati tiek rediģēti un bagātināti, izmantojot automatizētus un manuālus datu labošanas paņēmienus. Publicētā datu kopa šobrīd satur tikai bibliogrāfiskos datus bez atbilstošajiem autoritatīvajiem datiem, kurus nākotnē paredzēts publicēt atsevišķā datu kopā.

Ilgāk nekā 20 gadus LNB sagatavo datus MARC21 formātā, kurā ir svarīgs datu elementu izkārtojums ierakstā noteiktā secībā, izmantojot īpaši kodētus laukus, apakšlaukus un indikatorus, ko veido tikai speciālistiem saprotams ciparu un zīmju kopums un ko spēj “nolasīt” slēgtas bibliotēku sistēmas. Saistīto datu standartos, t. sk. RDA, ir būtisks katrs atsevišķais informācijas apgalvojums, kuru var izmantot atkārtoti pilnīgi citā kontekstā neatkarīgi no pilna grāmatas ieraksta MARC formātā. Bibliotēku datu veidošana saskaņā ar RDA saistīto datu formā nākamās paaudzes bibliotēku datu veidošanas sistēmās balstītos uz entīšu savstarpējo attiecību un hierarhijas modeli, izmantojot autoritatīvos datus un RDA atsauces (*RDA Reference*) datus (dažādas entītes un to attiecību apzīmētāji).

Nemot vērā pasaules pieredzi un Starptautiskās bibliotēku asociāciju un institūciju federācijas (*International Federation of Library Associations and Institutions*) IFLA Bibliogrāfijas pastāvīgās komitejas sagatavotās vadlīnijas “Nacionālo bibliogrāfiju kopējā prakse elektroniskajā laikmetā” (*Common Practices for National Bibliographies in the Electronic Age*)<sup>18</sup>, LNB datu portālā nacionālās bibliogrāfijas dati ir publicēti (1) sākotnējā formā, bibliotēku datu standartā MARC21<sup>19</sup> un (2) pārveidoti saistīto datu formā atbilstoši BIBFRAME datu modelim un noformēti NTriples un RDF/XML datu formātos<sup>20</sup>. BIBFRAME

18 [https://www.ifla.org/wp-content/uploads/2019/05/assets/bibliography/common\\_practices\\_for\\_national\\_bibliographies\\_2021-01.pdf](https://www.ifla.org/wp-content/uploads/2019/05/assets/bibliography/common_practices_for_national_bibliographies_2021-01.pdf)

19 [https://dati.lnb.lv/files/nat\\_bibliography-2017\\_2020-marc.zip](https://dati.lnb.lv/files/nat_bibliography-2017_2020-marc.zip)

20 [https://dati.lnb.lv/files/nat\\_bibliography-2017\\_2020-bibframe-ntriples.zip](https://dati.lnb.lv/files/nat_bibliography-2017_2020-bibframe-ntriples.zip) un [https://dati.lnb.lv/files/nat\\_bibliography-2017\\_2020-bibframe-rdfxml.zip](https://dati.lnb.lv/files/nat_bibliography-2017_2020-bibframe-rdfxml.zip)

datu modelis ir paredzēts bibliogrāfiskās informācijas izteikšanai saistīto datu formā<sup>21</sup>. MARC21 datu pārveidošana BIBFRAME formā veikta, izmantojot *marc2bibframe2* atvērtā pirmkoda rīku<sup>22</sup>.

Publicējot nacionālo bibliogrāfiju tīmeklī saistīto datu formā, tiek veicināta šīs informācijas tālāka izmantošana un nodrošināta iespēja šos datus saistīt ar citām datu kopām, t. sk. ar datiem dažādos digitalizācijas projektos, piemēram, *Europeana* vai LNB digitalizēto grāmatu portālā. Tos varētu izmantot arī citos ar kultūras mantojumu un izdevējdarbību saistītos risinājumos, t. sk. nododot datus mantisko tiesību kolektīvā pārvaldījuma organizācijai LATREPRO. LATREPRO nodrošina, lai autori un izdevēji saņemtu taisnīgu atlīdzību par viņu darbu reprogrāfisko reproducēšanu Latvijā – saskaņā ar Autortiesību likuma 35. pantu atlīdzību par darbu kopēšanu var saņemt jebkurš autors un izdevējs, kura radītais darbs ir piegādāts LNB kā obligātais eksemplārs<sup>23</sup>. Šobrīd datus par iepriekšējā gadā publicētajām grāmatām LNB sniedz MARC/XML formātā. Attīstoties izdevēju organizāciju platformām un rīkiem, izdevējus un autorus pārstāvošā organizācija tos varētu saņemt modernākā, saistīto datu formā, līdz ar to iegūstot efektīvākas datu izmantošanas iespējas.

## Datu kopa pētniecībai par izdevējdarbību Latvijā

LNB ir viena no statistikas iestādēm, kas iesaistīta Oficiālās statistikas programmas<sup>24</sup> izpildē. Programmā LNB atbildībā ir arī statistikas nodrošināšana par izdevējdarbību Latvijā. Izdevējdarbības statistika tiek apkopota, pamatojoties uz Obligāto eksemplāru likumu<sup>25</sup>, Ministru kabineta noteikumiem Nr. 281<sup>26</sup>, Statistikas likuma 4. pantu<sup>27</sup> un balstoties uz LNB krājumā saņemtajiem izdevumu obligātajiem eksemplāriem (turpmāk – OE)<sup>28</sup>. Statistika par izdevējdarbību Latvijā tiek veidota un publicēta saskaņā ar Statistikas likumu<sup>29</sup>, Eiropas Statistikas prakses kodeksu<sup>30</sup> un Eiropas Parlamenta un Padomes Regulu (EK)

21 <https://www.loc.gov/bibframe/docs/bibframe2-model.html>

22 <https://github.com/lcnetdev/marc2bibframe2>

23 <https://likumi.lv/ta/id/5138-autortiesibu-likums>

24 <https://www.csb.gov.lv/lv/dokumenti/oficialas-statistikas-sistema/ofici%C4%81las-statistikas-programma>

25 <https://likumi.lv/doc.php?id=136682>

26 <https://likumi.lv/ta/id/291156>

27 <https://likumi.lv/ta/id/274749-statistikas-likums>

28 <https://lnb.lv/lv/izdevejiem/obligatie-eksemplari>

29 <https://likumi.lv/ta/id/274749-statistikas-likums>

30 <https://ec.europa.eu/eurostat/documents/4031688/9332152/KS-02-18-142-LV-N.pdf/82ba14e8-a49e-4215-a9ad-8f3f0b8b1b64?t=1540214583000>

Nr. 223/2009<sup>31</sup>. Atšķirībā no Oficiālās statistikas portālā un ikgadējā krājumā “Latvijas izdevējdarbības statistika” (iepriekš – “Latvijas prese”) publicētajiem apkopotajiem datiem, kas raksturo izdevējdarbības nozari, datu kopā pētniecībai par izdevējdarbību Latvijā LNB ievāktie un XLSX formātā sagatavotie dati<sup>32</sup> ir pieejami neapkopotā veidā un primāri paredzēti pētniekiem, kuri datu kopu var izmantot, veidojot konkrētiem pētnieciskajiem mērķiem atbilstošus apkopojumus.

Datu kopa ar neapkopotajiem izdevējdarbības statistikas datiem pirmo reizi tika publicēta 2019. gadā, ietverot datus par grāmatām un brošūrām no 1995. gada un citiem izdevumu veidiem no 2017. gada. Projektā datu kopa tika papildināta ar datiem par grāmatām un brošūrām 1991.–1994. gadā, digitalizējot iepriekš tikai kartīšu formātā pieejamos statistikas datus izdevumu nosaukumu līmenī. Dati datu kopā tiek atjaunoti un publicēti reizi gadā līdz 31. maijam. Datu kopa satur bibliogrāfiskus metadatus, bet nesniedz informāciju, piemēram, par pārdošanas rādītājiem un kanāliem, izdevumu patēriņu, izdevniecību darbinieku skaitu, apgrozījumu u. tml., kas, lai gan starptautiskajā praksē tradicionāli tiek izmantoti izdevējdarbības nozares raksturošanai, šobrīd LNB kā statistikas iestādei nav pieejami.

Datu kopa ietver datus par visiem Latvijā reģistrētu juridisku personu un Latvijā dzīvojošu fizisku personu kalendārā gada laikā izdotajiem (sagatavotajiem un publiskotajiem) iespaiddarbiem, elektroniskiem izdevumiem (kasešu, diskešu, disku u. tml. formātā), tiesšaisnes izdevumiem un nekonvencionālajiem jeb “pelēkajiem” izdevumiem (izdevumiem, kas satur vispārpieejamu informāciju un tiek izplatīti bez maksas, piemēram, konferenču materiāli, valsts pasūtītu pētījumu rezultātu ziņojumi, deponēti zinātniskie darbi u. tml.), kuru OE saņēmusi LNB neatkarīgi no tā, vai izdevumiem ir vai nav piešķirts ISBN, ISMN vai ISSN. Izdevumu veidi, uz kuriem attiecas Obligāto eksemplāru likums un kurus izdevējiem ir pienākums piegādāt LNB, ir definēti Obligāto eksemplāru likuma pielikumā<sup>33</sup>. Datu kopa satur datus par grāmatām un brošūrām, Masu informācijas līdzekļu reģistrā (turpmāk – MIL) ierakstītiem žurnāliem un laikrakstiem, citiem periodiskajiem izdevumiem, kuri atbilst laikraksta vai žurnāla pazīmēm, bet nav reģistrēti MIL, attēlizdevumiem, kartogrāfiskajiem materiāliem, nošu izdevumiem, promocijas darbu kopsavilkumiem, sīkiešpieddarbiem un e-grāmatām.

Datu kopā tiek ietverti arī dati par grāmatām, kuru OE piegādāti ar viena kalendārā gada novēlošanas – pēc tam, kad pabeigta iepriekšējā pārskata perioda datu apkopošana. Par citiem novēloti piegādātiem izdevumu veidiem dati netiek apkopoti. Šobrīd datu kopā ietvertie dati par e-grāmatām nav uzskatāmi par pilnīgiem un reprezentatīviem, jo normatīvajos aktos, kas regulē OE piegādi, pastāv nepilnības, kā dēļ daļa izdevēju, t. sk. lielākie e-grāmatu izdevēji, LNB nepiegādā e-grāmatu OE, līdz ar to statistikā nav atspoguļota liela daļa e-grāmatu.

31 <https://eur-lex.europa.eu/legal-content/LV/TXT/PDF/?uri=CELEX:02009R0223-20150608&from=LV>

32 <https://lnb.lv/lv/izdevējdarbības-statistika/datu-masivs-petniecibai-par-izdevējdarbību-latvija>

33 <https://likumi.lv/doc.php?id=136682>

Oficiālās statistikas programmā veidotā izdevējdarbības statistika ietver arī ziņas par tirāžu, kas iegūtas, balstoties uz izdevēju iesniegtajiem pārskatiem par izdevējdarbību<sup>34</sup>, kuros tiek sniegta ziņa par kalendārājā gadā izdoto iespiesto izdevumu tirāžu, kā arī jebkura izdevuma papildtirāžu. Šīs izdevēju sniegtās informācijas konfidencialitāti aizsargā Statistikas likuma 7. panta otrās daļas 8. punkts, kas statistikas iestādei nosaka pienākumu nodrošināt statistisko konfidencialitāti, 17. pants, kas nosaka datu apstrādes kārtību un to aizsardzības prasības, kā arī 19. panta pirmā daļa, kas regulē privātpersonu vai valsts institūciju identificējošu statistikas datu izplatīšanu. Tādēļ, lai nodrošinātu statistisko konfidencialitāti un neizpaustu ierobežotas pieejamības informāciju, datu kopā nav iekļauti un publiski nav pieejami dati par iespiesto izdevumu tirāžu izdevumu nosaukumu līmenī.

LNB ilgstoši strādā, lai nodrošinātu izdevējdarbības statistikas datu kvalitāti, kā arī saskaņotību un salīdzināmību gan Latvijas, gan starptautiskā mērogā. Tomēr gan datu konvertēšanas rezultātā, laika gaitā attīstoties informācijas tehnoloģijām un mainoties statistikas sagatavošanā lietotajai programmatūrai, gan ieviešot korekcijas datu atveides metodoloģijā, datu masīvā ietvertajos datos iespējams paralēls dažādu pieraksta formu lietojums, kam pētniekiem atsevišķos gadījumos jāpievērš uzmanība, kā arī nebūtiskas kļūdas, kas novērojamas, salīdzinot datus krājumos “Latvijas prese” un “Latvijas izdevējdarbības statistika” un Latvijas izdevējdarbības pētniecības datu kopā. Lai uzlabotu statistikas metodoloģiju, ieviestu metožu efektivitāti un diskutētu par statistikas veidotājiem un lietotājiem aktuālajiem jautājumiem, LNB sadarbojas ar zinātniekiem, pētniekiem un statistikas veidotājiem un lietotājiem no Latvijas un ārzemēm, piemēram Ziemeļvalstu tīklošanās grupu (*Nordic Networking Group*) un Eiropas Izdevējdarbības studiju asociāciju (*European Publishing Studies Association EuroPub*). Lai uzlabotu nacionālo datu atbilstību starptautiskiem standartiem, kas nodrošina datu salīdzināmību starptautiskā līmenī, šobrīd norit darbs pie Latvijas izdevējdarbības statistikas sagatavošanas metodoloģijas saskaņošanas ar ISO 9707 “Informācija un dokumentācija – Statistika par grāmatu, laikrakstu, periodisko izdevumu un elektronisko izdevumu izdošanu un izplatīšanu”.

Neapkopotu datu publicēšana atvērto datu kopas formātā valsts pārvaldē ir starptautiski atzīta labā prakse, jo nodrošina, ka dati pētniekiem un citiem interesentiem ir brīvi pieejami, atkārtoti izmantojami un lietojami atbilstoši individuālajām vajadzībām (Milić u. c. 2018). Tomēr, līdzīgi kā citu valstu pārvaldes iestāžu gadījumā (turpat: 55–56, 58–60), arī LNB vēl atrodas semantisko tehnoloģiju apgūšanas sākumposmā un tikai apzina iespējas padarīt izdevējdarbības datu lietotāju pieredzi “gudrāku” un bagātīgāku, atklājot arvien plašākas datu lietojuma un informācijas izguves iespējas, kā arī pilnveidojot statistikas izstrādi un padarot efektīvāku datu vākšanu, sagatavošanu un izplatīšanu.

Šobrīd projektā LNB ekspertu darba grupa strādā pie rekomendāciju sagatavošanas, lai uzlabotu izdevējdarbības statistikas datu kvalitāti, tos bagātinātu ar datiem no citām LNB datu kopām, kā arī potenciāli pilnveidotu efektivitāti iekšējai darba plūsmai statistikas datu

34 <https://lnb.lv/lv/izdevejiem/izdevejdarbibas-statistika#Kurpieejamaparskataveidlapa>



sagatavošanā. Izdevējdarbības statistikas dati tiek veidoti LNB portālā izdevējiem<sup>35,36</sup>, kurā mūsdienīgā veidā nodrošināta iespēja iesniegt pārskatu par izdevējdarbību, informāciju par izdevumu tirzāžām, kā arī saņemt starptautiskos standartnumurus (ISBN, ISMN, ISSN) un UDK. Līdzās statistikai, balstoties uz datiem, kas saņemti, izdevējiem pieprasot kādu no starptautiskajiem standartnumuriem vai UDK, LNB portālā izdevējiem top minimālie bibliogrāfiskie apraksti, kas tiek eksportēti uz Valsts nozīmes bibliotēku kopkatalogu, vienlaikus izdevumu ierakstos LNB portālā izdevējiem fiksējot kopkataloga bibliogrāfisko ierakstu sistēmas numuru jeb identifikatoru. Līdz ar to darba grupa secinājusi, ka statistikas datus iespējams padarīt kvalitatīvākus, t. sk. tos bagātinot ar autoritatīvajiem un citiem kontrolētiem datiem. Darba grupa turpina darbu, izstrādājot rekomendācijas datu bagātināšanai un apzinot gan datu, gan tehnoloģisko risinājumu aspektus.

## Nobeigums

Rakstā apskatītas četras Latvijas Nacionālās bibliotēkas datu kopas, kuras ir publicētas projektā, – saistīto datu kolekcija “Rainis un Aspazija”, Nacionālā tēzaura atvērte dati ontoloģijas formā, Nacionālās bibliogrāfijas atvērte dati un Latvijas izdevējdarbības statistikas dati. Publicējot minēto informāciju atvērto un saistīto datu formā, LNB radusi iespēju šo datu tālākai izmantošanai, t. sk. digitālajās humanitārajās zinātnēs. LNB plāno paplašināt LNB atvērto datu portālu ar vēl divām datu kopām, iekļaujot tajā aptuveni 5000 kultūras mantojuma jomā izmantotu terminu no RDA standarta, kā arī LNB autoritatīvo datu kopu.

Projektā “Latvijas atmiņas institūciju dati digitālajā telpā: vienojot kultūras mantojumu” izvirzītais uzdevums uzlabot LNB veidoto datu kopu un vārdnīcu kvalitāti un pārveidot tās par atkārtoti izmantojamām atvērto datu kopām, t. sk. saistīto datu kopām un ontoloģijām, ir izpildīts līdz ar četru atvērto datu kopu publicēšanu. Tomēr šim un vēl ielānotajām atvērto datu kopām vēl ir “jāiztur” praktiskā atkārtotā lietojuma pārbaude, t. i., izmantošana pētnieku vidē un citos praktiskos risinājumos (līdzīgi kā minētā kultūras mantojuma platforma). Ņemot vērā, ka Atvērto datu portāls tika atklāts īsi pirms šī raksta tapšanas, portālā jau publicēto un nākotnē publicējamo datu kopu funkcionalitāte un izmantošanas iespējas ir tālākas izpētes objekts. Tomēr nesen LNB ontoloģiju serviss ir publicēts Latvijas atvērto datu portālā<sup>37</sup>. Tāpat ontoloģija un atbilstošu entīšu informācija no pārējām LNB portāla datu kopām tiek integrēta Latvijas vienotajā kultūras mantojuma platformā (ERAF projekts “Kultūras mantojuma satura digitalizācija” (2. kārtā)). Šīs norises apstiprina, ka pārdomātam, pieredzē un reālu vajadzību izpētē balstītam semantisko jēdzienu tīklam ir potenciāls tikt veiksmīgi pārnestam tehnoloģiskā risinājumā, ko iespējams lietot starpinstitucionālā līmenī integrētas datu pārvaldības un zināšanu pārneses nodrošināšanai.

35 <https://izdevejiem.lnb.lv/>

36 <https://lnb.lv/lv/izdevejiem/izdeveju-portals>

37 <https://data.gov.lv/dati/lv/dataset/zanru-vietu-un-prieksmetu-tezauru-dati>

Paplašinot atvērto un atvērto saistīto datu apjomu, LNB pilda vairākus uzdevumus digitālās pētniecības atbalstam, t. i., piesaista pārbaudītiem datiem kontekstu no dažādiem avotiem, saista starpnozaru informāciju semantiskā līmenī un paplašina zināšanu bāzi. Tīmeklī publicēti atvērtie saistītie dati veicina demokrātiskāku piekļuvi zināšanām, tai pašā laikā nodrošinot saistīti ar pirmavotiem un veicinot piekļuvi tiem. Projektā radītais atvērto datu portāls ir arī solis publisko datu kopīgošanas virzienā. Tā veidotāji apzinās, ka datu savietojamībai, kvalitātei un noteiktai struktūrai ir būtiska nozīme datu plašākā izmantošanā, t. sk. pētniecībā un izglītībā, no kā labumu iespējams gūt visai sabiedrībai.

- Apenīte, Mārīte, Bojārs, Uldis (2021). National Library of Latvia Subject Headings as Linked Open Data. Verborgh, Ruben u. c. (ed.). *The Semantic Web: ESWC 2021 Satellite Events*. ESWC 2021. Print ISBN 978-3-030-80417-6; pp 33–37. Online ISBN 978-3-030-80418-3. Lecture Notes in Computer Science, vol 12739. Springer, Cham.
- Berners-Lee, Tim (2006). *Linked data*. Pieejams: <http://www.w3.org/DesignIssues/LinkedData.html> [skatīts 27.10.2021.].
- Bizer, Christian, Heath, Tom, Berners-Lee, Tim (2011). Linked data: The story so far. *Semantic services, interoperability and web applications: emerging concepts*. IGI global, pp. 205–227.
- Bojārs, Uldis, Rašmane, Anita, Žogla, Artūrs, Bāliņa, Signe, Salna, Edgars (2018). Semantic Annotation Tool for Cultural Heritage Content. *Baltic Journal of Modern Computing*, No. 6(4), pp. 449–463.
- Bojārs, Uldis, Rašmane, Anita, Goldberga, Anita (2020). Linked data for digital humanities scholars and researchers: “Rainis and Aspazija” (RunA) collection. *5th Conference Digital Humanities in the Nordic Countries, DHN 2020*. CEUR Workshop Proceedings, No. 2865, pp. 185–194.
- Byrne, Alex (2014). Memory institutions shaping the past, present and future. *The Australian library journal*, No. 64(4), pp. 1–11.
- Gruber, Tom (2008). Ontology. Liu, Ling, Özsu, Tamer M. (eds.). *Encyclopedia of Database Systems*. Springer-Verlag.
- Hickey, Thomas B., Toves, Jenny A. (2014). Managing ambiguity in VIAF. *D-lib Magazine*, No, 20(7/8).
- Sacima of the Republic of Latvia (2010). *Sustainable Development Strategy of Latvia until 2030*. Available: [https://pkc.gov.lv/sites/default/files/inline-files/LIAS\\_2030\\_en\\_1.pdf](https://pkc.gov.lv/sites/default/files/inline-files/LIAS_2030_en_1.pdf) [accessed 27.10.2021.].
- Latvijas Nacionālā bibliotēka (2019). *Latvijas Nacionālās bibliotēkas attīstības stratēģija 2019.–2023. gadam*. Pieejams: [https://www.lnb.lv/sites/default/files/media/par-lnb/institucija/LNB\\_strategija\\_LV.pdf](https://www.lnb.lv/sites/default/files/media/par-lnb/institucija/LNB_strategija_LV.pdf) [skatīts 27.10.2021.].
- Latvijas Nacionālā bibliotēka (2021). *Rekomendācijas Atsauču datu semantiskās sadarbības veicināšanai starp Latvijas kultūras mantojuma institūcijām*. DOI: <https://doi.org/10.52197/IBIV6839>. Pieejams: <https://dom.lndb.lv/data/obj/file/30193606.pdf> [skatīts 27.10.2021.].
- Miles, Alistair, Bechhofer, Sean (2009). *SKOS simple knowledge organization system reference*. W3C recommendation. Available: <https://www.w3.org/TR/skos-reference/> [accessed 27.10.2021.].
- Milić, Petar, Veljković, Nataša, Stojmenov, Leonid (2017). Semantic Technologies in e-government: Toward Openness and Transparency. Bolívar, Manuel Pedro Rodríguez (ed.). *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues*. Cham: Springer, pp. 55–66.
- Pessach, Guy (2008). *Networked memory institutions: Social remembering, privatization and its discontents*. Cardozo Arts & Ent.
- Zhang, Junsong, Zhao, Wu, Xie, Gang, Chen, Hong (2011). Ontology based Knowledge Management System and Application. *Procedia Engineering*, No. 15, pp. 1021–1029.

# Open Research Data of the National Library of Latvia

Mārīte Apenīte, Uldis Bojārs, Ansis Garda,  
Anita Goldberga, Maira Kreislere,  
Anita Rašmane, Elza Ungure

Keywords: linked data, open data, NLL open data portal, NLL ontology service, national thesaurus, national bibliography, official publishing statistics

This paper describes open and linked open datasets of the National Library of Latvia (NLL) that are available to researchers. This includes data from the linked digital collection *Rainis and Aspazija*, the National Thesaurus and National Bibliography datasets and the dataset of Latvia's publishing statistics. These datasets are published on NLL's open data portal <https://dati.lnb.lv/> and are available for research and reuse.

The datasets have been published as part of the project *Latvian memory institution data in the digital space: connecting cultural heritage* (No. lzp-2019/1-0365), which is funded by the Fundamental and Applied Research Programme of the Latvian Council of Science.

The paper describes the methods used for the dataset analysis, processing and transformation as well as the necessary data improvements. It also looks at the possibilities to use these datasets for cooperation with related fields (e.g. archives and museums) and at the plans for improvements and additions to the datasets and the open data portal.