

# Atviri bibliografiniai duomenys – pirmas žingsnis susietų atvirų duomenų link

Patrick DANOWSKI

Austrijos mokslo ir technologijų institutas (IST), Klosterneuburg, Austrija, el. p. [patrick.danowski@ist.ac.at](mailto:patrick.danowski@ist.ac.at)

*Vis daugiau bibliotekų įsitraukia į semantinio saityno projektus. Tačiau paprastai nesvarstomas duomenų licencijų klausimas arba toks svarstymas nukeliamas į projekto pabaigą. Šiame straipsnyje nagrinėjama, kodėl semantinio saityno kontekste licencijų klausimas yra toks svarbus ir turėtų būti vienu iš pagrindinių semantinio saityno projektų aspektų. Jame taip pat aptariama, kodėl viešasis tinklas yra vienas iš sprendimų, visiškai atitinkančių specifinius semantinio saityno reikalavimus ir užtikrinančių projektų tęstinumui būtiną daugkartinį semantinių bibliotekų duomenų naudojimą.*

*Reikšminiai žodžiai:* semantinis saitynas; atviri bibliografiniai duomenys.

Straipsnyje aptarsime, kodėl bet kuris semantinių duomenų projektas turėtų prasidėti nuo nuosavybės teisių aptarimo. Tai padėtų įgyvendinti pradinį semantinio projekto etapą, pavyzdžiui, skelbti nesemantinius, neapdorotus duomenis, siekiant gauti atsiliepimų apie juos, ir dalyvauti svarstymuose.

## Motyvacija

Mūsų nuomone, bibliotekos, norėdamos sėkmingai dalyvauti semantiniame saityne, turi įveikti tris pakopas:

1. Bibliotekų duomenys turi būti pateikiami plačiai pripažintu (ne tik bibliotekų bendruomenės) semantiniu formatu.

2. Bibliotekų duomenis turi būti įmanoma naudoti daug kartų.

3. Bibliotekų duomenys turi būti naudojami kituose projektuose.

Šiame straipsnyje pirmasis, t. y. formatų, klausimas neliečiamas. Pagrindinis dėmesys skiriamas kitiems dviem aspektams. Pirmiausia reikėtų atsakyti į klausimą, kodėl licencijos yra ypač svarbus dalykas. Mano asmenine nuomone, apie semantinius duomenis ir jų techninius aspektus kalbėti prasminga tik tuo atveju, jei siekiama, kad bibliotekų duomenys būtų atviri viešam daugkartiniam naudojimui. „Uždaram“, tik tarp bibliotekų vykstančiam, komunikavimui sukurti sė-

kmingai funkcionuojantys keitimosi duomenimis formatai. Didelis semantinio saityno privalumas yra tai, kad šiuos duomenis gali suprasti ir kiti. Jei svarbiausias semantinio tapsmo siekis yra viešas daugkartinis duomenų naudojimas, pirmutinis mūsų uždavinys būtų nustatyti savo požiūrį į tai. Pirmiausia tai formatų klausimas, tačiau yra ir kitas, daug svarbesnis, – ką bus leidžiama daryti su šiais duomenimis.

## Atvirosios prieigos judėjimo kontekstas

Šiandienos bibliotekų duomenys paslėpti duomenų bazėse, į kurių duomenų rinkinius dažnai nėra jokių giliųjų saitų. Duomenys yra giliai tinkle, izoliuoti. Net jeigu bibliotekoms ir būdinga tradicija atverti savo duomenis, dažniausiai tai susiejama su užmokesčiu, kuris paprastai apibūdinamas kaip „mažas“. Europoje duomenų bazės (visa duomenų bazė arba tik kokia nors svarbi jos dalis) saugomos autorių teisių įstatymų. Šie įstatymai neapima duomenų rinkinių. Viena iš semantinio saityno problemų yra nykstančios ribos tarp duomenų bazių ir duomenų rinkinių. Tai apsunkina diskusijas apie duomenų autorių teises.

Tuo pačiu metu bibliotekos užsiima atvirosios prieigos lobizmu. Berlyno deklaracijoje teigiama: „Mes suprantaime atvirąją prieigą kaip išsamų mokslo bendruomenės aprobuotą žmonijos žinių ir kultūros paveldo šaltinį“<sup>1</sup>.

<sup>1</sup> Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 22.10.2003  
<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

Pastaruoju metu bibliotekos pasisako už atvirąją prieigą prie mokslo duomenų. Bibliotekų duomenys priklauso kultūros paveldui ir jie traktuotini kaip mokslo duomenys. Todėl bibliotekų duomenų atvėrimas laikytinas natūraliu dalyku. Tim Berners-Lee savo pasisakyme TED konferencijoje iškėlė šūkį „Atviri duomenys dabar!“<sup>2</sup> Tai reikėtų suprasti kaip kvietimą jau dabar skelbti bibliotekų duomenis, nesileidžiant į formalias diskusijas, tačiau tuo tikslu pirmiausia būtina apsispręsti dėl licencijų.

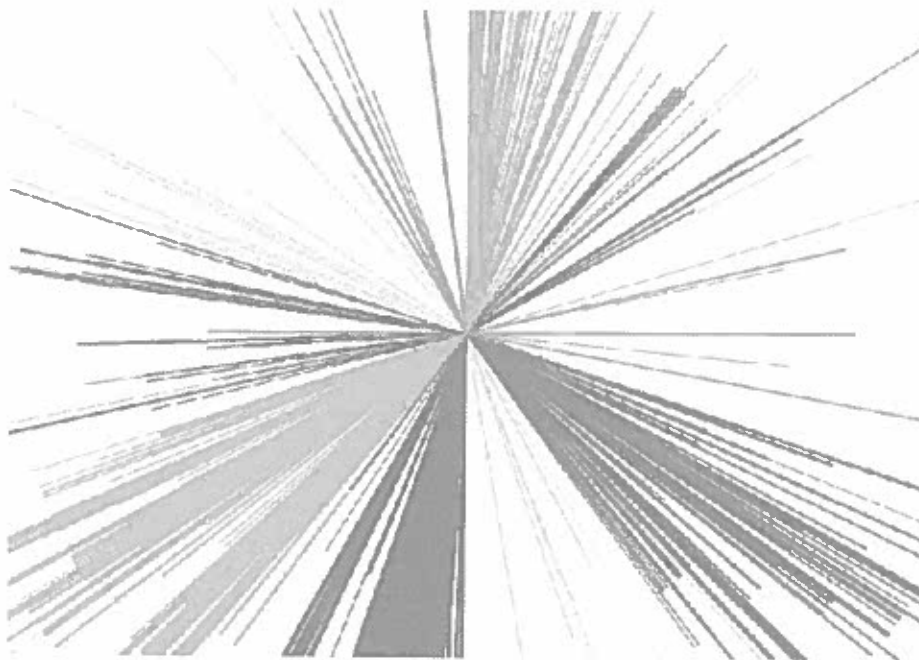
## Licencijos pasirinkimas

Bibliotekoms siekiant atverti savo duomenis klasikinės autorių teisės tikrai nėra geriausias pasirinkimas. Sunkumų sukelia ir *Creative Commons* licencijų taikymas. Bibliotekų srityje labai populiarios nekomercinės sutartys, tačiau šių sutarčių taikymas gali sukelti keblumų. Pagal tokias sutartis viešinamų duomenų negalės naudoti (arba pateikti kartu su kitais duomenimis) komercinės paieškos sistemos. Nėra taip paprasta nustatyti, kuri programinė priemonė yra komercinė, o kuri – ne. Ar galite įsivaizduoti, kad dabartiniu metu *Google* nebūtų leidžiama indeksuoti jūsų tinklalapio? Nemanau, kad toks kelias būtų mums priimtinas.

*Creative Commons* gali kelti sunkumų ir dėl duomenų konteksto, nes pagal šią licenciją paisoma išskirtinai kūrėjo teisių. Tai gali atrodyti puikus pasirinkimas – tekstų ir vaizdų atveju visada turi būti nurodomas autorius, tačiau duomenys grindžiami faktais ir kūrėjo vaidmuo čia nėra toks svarbus<sup>3</sup>. Taipogi vieno ištekliaus duomenys gali būti sujungiami su kitų išteklių duomenų rinkiniais. Šiuo atveju reikėtų nurodyti visus panaudotus duomenų rinkinius, ir juos išvardijantys sąrašai gali būti ilgesni už patį duomenų paieškos rezultatų rinkinį.

Yra ir kita priežastis, dėl kurios prieš pradėdant projektą būtina apsispręsti dėl tinkamos licencijos pasirinkimo: tai darys įtaką visai projekto eigai. Jeigu kartais projekto vykdymo metu nuspręstumėte pakeisti licenciją, tai gali išaugti į atskirą projektą (tai rodo Vikipedijos partitės, kai pereiti nuo GFDL prie šiuolaikiškesnės *Creative Commons* prireikė ne vienerių metų)<sup>4</sup>.

Optimali išeitis būtų duomenims netaikyti jokių autorių teisių arba skelbti juos viešosios nuosavybės teisėmis. Viešoji nuosavybė kaip licencija negalioja Europoje, nes čia neįmanoma, kad duomenims nustotų galioti visos teisės. Šios situacijos sprendimas jau rastas: norint atsisakyti autorių teisių, galima pasinaudoti teisių atsisakymo



Dan Brickley *UDC Star* („UDK žvaigždės“) (<http://www.flickr.com/photos/danbri/4326955233/sizes/o/>) „cern.ch/bookdata 1-ojo skerspjūvio vaizdas, kurio kiekvienas segmentas atitinka pagrindinius UDK dešimtinius skyrius“

<sup>2</sup> Berners-Lee, Tim: Tim Berners-Lee on the next Web, TED Talk, Februar 2009  
[http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html)

<sup>3</sup> Nebent būtų svarstomas pasitikėjimo klausimas, bet toks svarstymas būtų kito straipsnio tema.

<sup>4</sup> Walsh, Josh: Wikimedia community approves license migration 21.05.2009  
<http://blog.wikimedia.org/2009/wikimedia-community-approves-license-migration/>

priemonėmis, pvz., CCO<sup>5</sup>, pareiškiant, kad atsisakoma visų teisių ir savininkas dėl jų neturės pretenzijų.

### Judėjimas už atvirus bibliografinius duomenis

2010 m. sausio mėn. CERN biblioteka paskelbė savo bibliografinius duomenis<sup>6</sup> MARCXML formatu pagal *Public Domain Dedication and Licence*<sup>7</sup> ir CCO aiškiai pasisakydama už atvirąją prieigą. Šiuo pavyzdžiu pasekė kitos bibliotekos, pvz., Gento universiteto biblioteka<sup>8</sup> ir daugelis Kelno (Vokietija) bibliotekų, priklausančių Šiaurės Reino-Vestfalijos bibliotekų paslaugų centrui<sup>9</sup>. Šiomet į šį judėjimą ketina įsitraukti daugiau bibliotekų. Dalyvaudamos šiame judėjime bibliotekos turi galimybę pareikšti, kad jos rimtai suinteresuotos savo produktams suteikti atvirąją prieigą. CERN paskelbus duomenis, Dan Brickley vizualiai atvaizdavo, kaip CERN taiko UDK klasifikaciją. Tai leido lengvai nustatyti galimas duomenų paklaidas.

Bibliotekų duomenų atvirumas (ne tik bibliografinių, bet ir klasifikacijos sistemų bei autoritetinių duomenų) teikia joms galimybę atlikti svarbų vaidmenį

semantiniame saityne. Čia tiktų sakyti: „Atviroji prieiga prie bibliotekų duomenų tiesia kelius prieigai prie žinių“<sup>10</sup>. Bet koks prieigos ribojimas neleistų įvairiems projektams naudoti mūsų duomenis – taip sumažėtų šio duomenų šaltinio vertė. Mums reiktų kiek įmanoma plačiau leisti naudotis savo duomenimis ir juos teikti visiems projektams, ypač komerciniams, taip sukuriant turtingą viešosios prieigos sritį. Kaip teigiama Pasaulinės konferencijos, skirtos informacinei visuomenei, (*World Summit on the Information Society – WSIS*) dokumente:

„Turtinga viešosios prieigos sritis – tai esminis informacinės visuomenės plėtros veiksnys, teikiantis įvairiapusę naudą: tai ir išsilavinusi visuomenė, naujos darbo vietos, inovacijos, verslo galimybės, mokslo pažanga“<sup>11</sup> (WSIS Declaration of Principles, Section 26 12.12.2003).

*Iš anglų kalbos vertė T. Auškalnis*

Straipsnis parengtas pagal pranešimą, skaitytą 2010 m. Geteborge (Šveicarija) vykusioje 76-ojoje IFLA konferencijoje.

<sup>5</sup> Apie CCO: <http://creativecommons.org/choose/zero>

<sup>6</sup> <http://www.cern.ch/bookdata>

<sup>7</sup> ODC Public Domain Dedication and Licence (PDDL) Text: <http://www.opendatacommons.org/licenses/pddl/1-0/>

<sup>8</sup> Ghent University Library Exports: <http://lib.ugent.be/info/en/exports.shtml>

<sup>9</sup> Kelno universiteto ir viešoji biblioteka (USB), Kelno medijų meno akademijos biblioteka, Reino-Pfalzo bibliotekų centrui priklausanti Kelno taikomųjų mokslų universiteto biblioteka, Kelno viešoji biblioteka ir Vokietijos sporto universiteto biblioteka.

<sup>10</sup> IFLA prezidentės Ellen Tise (2008–2010) tema: Libraries Driving Access to Knowledge.

<sup>11</sup> <http://www.itu.int/wsis/docs/geneva/official/dop.html>