



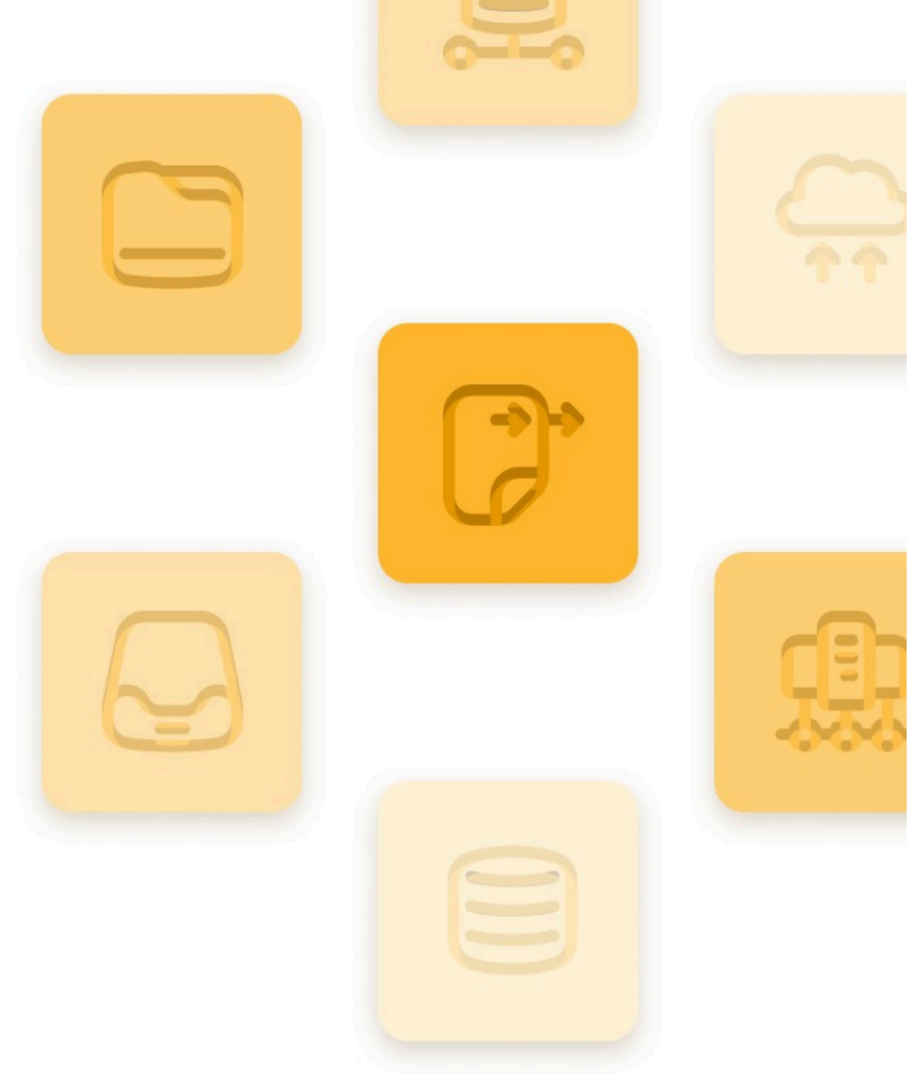
Finansē  
Eiropas Savienība  
NextGenerationEU

2027

Nacionālais  
attīstības plāns

# Pētniecības datu metadatu izveide: **kvalitāte, standarti, dokumentācija**

2026. gada 5. maijā



# Projekta finansējums

Materiāls izstrādāts projekta “Atbalsts atvērtās zinātnes ieviešanai praksē, kā arī izveidoti risinājumi zinātnes datu koplietošanai un dalībai ES atvērtajā zinātnes mākonī” ietvaros (ANM projekta Nr. 2.1.3.1.i) ar Eiropas Savienības Atvēršanas fonda un Latvijas valsts finansiālo atbalstu



Finansē  
Eiropas Savienība  
NextGenerationEU



Nacionālais  
attīstības plāns



Latvijas  
Biozinātņu un tehnoloģiju  
universitāte



RĪGAS TEHNISKĀ  
UNIVERSITĀTE



LATVIJAS  
UNIVERSITĀTE

RSU



VPC



Finansē  
Eiropas Savienība  
NextGenerationEU



Nacionālais  
attīstības plāns

# 3. sesija · FAIR

I + R — Saistītie dati, vārdnīcas, RDF

Eduards Skvireckis · 2026. gada 5. maijā · 13:15–14:45

## Šodienas plāns

Laiks	Sesija	Tēma
9:00–10:30	1	F — Atrodami: identifikatori, metadatu standarti
10:45–12:15	2	A + I — Pieejami un sadarbspējīgi
<b>13:15–14:45</b>	<b>3</b>	<b>I + R — Saistītie dati, vārdnīcas, RDF</b>
15:00–16:30	4	R praksē — Licences, izcelsme, DPP, SPARQL

Septiņi tematiskie bloki · četras aptaujas

## FAIR karte — kur mēs sākam

<b>F1</b> pastāvīgs ID	<b>F2</b> bagātīgi metadati	<b>F3</b> ID metadatos	<b>F4</b> reģistrēts
<b>A1</b> protokols	<b>A1.1</b> atvērts protokols	<b>A1.2</b> autorizācija protokolā	<b>A2</b> nezūdoši metadati
<b>I1</b> formālā valoda	<b>I2</b> FAIR vārdnīcas	<b>I3</b> jēgpilnas attiecības	
<b>R1</b> bagātīgs apraksts	<b>R1.1</b> licence	<b>R1.2</b> izcelsme	<b>R1.3</b> nozaru standarti

Šajā sesijā — **I2**, **I3** pilnībā un **R1.3** **pa pusei** praktiskajā darbā ar atslēgvārdiem.



Fināns  
Eiropas Savienībā  
NextGenerationEU



2027  
Nacionālais  
atlabas plāns

# I+R ievads

Semantiskais tīmeklis un saistītie dati

## Aptauja 1 — RDF un saistīto datu iepriekšējā pieredze

*Cik ļoti pazīstams jums ir RDF vai "saistīto datu" (Linked Data) jēdziens?*

- (a) Esmu lietojis RDF / SPARQL / Turtle praksē
- (b) Esmu dzirdējis terminus, bet sīkāk neesmu pētījis
- (c) Dzirdu pirmoreiz

## Kas ir semantiskais tīmeklis (I+R)

**Semantiskais tīmeklis** — tīmekļa paplašinājums, kurā **datiem ir jēga**, ko var saprast gan cilvēki, gan mašīnas. Berners-Lee, Hendler, Lassila, 2001.

- **Dokumentu tīmeklis** · cilvēki lasa lapas; saites starp dokumentiem
- **Datu tīmeklis** · mašīnas seko URI; saites starp *resursiem* (autoriem, vietām, jēdzieniem)
- **Saistītie dati** (*Linked Data*) · datu tīmekļa praktiskā realizācija — Berners-Lee, 2006

## 5-zvaigžņu kāpnes — kur ir *dv.dataverse.lv*? (I3)

- ★ — datu kopa pieejama jebkurā formātā ar atvērtu licenci
- ★★ — strukturēta (Excel, ne PDF)
- ★★★ — atvērts formāts (CSV, ne XLS)
- ★★★★★ — **RDF + URI** · *dv.dataverse.lv* šeit (schema.org JSON-LD; ORCID URI autoriem)
- ★★★★★ — saistīts ar citu RDF datu URI · *trūkst pussolis: vārdnīcu URI atslēgvārdu laukā*

**Atruna:** piecas zvaigznes ir publicēšanas mērogs *publicētāja pusē*. **Tās nav mērķis; tā ir kāpne, ko publicētāji iet pakāpeniski.**

*Berners-Lee, T. (2010). Linked Data — Design Issues.*

# RDF un trijnieki

Modelis, kas jau strādā JSON-LD `@context`-ā

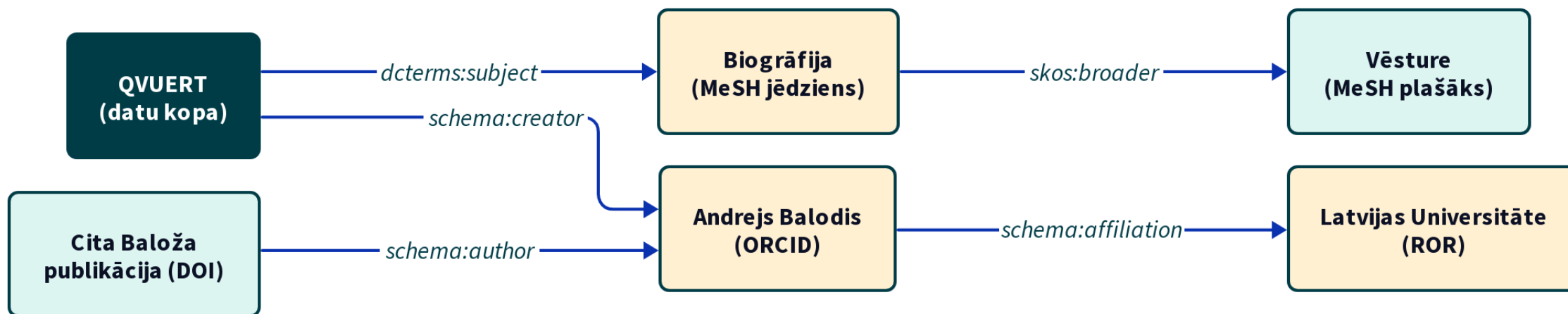
## Trijnieks, URI, tukšais mezgls (1)

```
<https://doi.org/10.71782/DATA/QVUERT> schema:creator <https://orcid.org/0000-0002-0688-9213> .  
      ↑                ↑                ↑  
subjekts          predikāts          objekts
```

- **Trijnieks** — trīs daļas: subjekts, predikāts, objekts
- **URI** — Uniform Resource Identifier; visas trīs trijnieka daļas var būt URI → durvis uz saistītajiem datiem
- **Tukšais mezgls** ( `blank node` ) — lokāls anonīms resurss bez URI (piem., autora afiliācija kā teksts bez ROR)

*Trīs jēdzieni. Tas ir RDF datu modelis. Tagad — kā to pierakstīt?*

## Mazs grafs — kā trijnieki savienojas (I+R)



- Katrs **trijnieks** ir maza durvis · vienam objektam vairāki predikāti
- Vairāki trijnieki ar kopēju subjektu vai objektu = **grafs**
- Subjekts vienā trijniekā kļūst par objektu citā · *saistīto datu būtība*

Tas pats QVUERT, ko mēs lasījām vienā trijniekā · grafs paplašinās ar katru jaunu predikātu.

## Piemērs: Turtle 5 rindas (11)

```
@prefix schema: <http://schema.org/> .  
  
<https://doi.org/10.71782/DATA/QVUERT>  
  a schema:Dataset ;  
  schema:name "Biogrāfisko interviju transkripcijas..." ;  
  schema:creator [ schema:name "Andrejs Balodis" ;  
                  schema:sameAs <https://orcid.org/0000-0002-0688-9213> ] .
```

- `@prefix` · saīsinājuma deklarācija — kā JSON-LD `@context`
- `a schema:Dataset` · tips (`a = rdf:type`); semikols beigās = "tas pats subjekts, vēl viens predikāts"
- Iekavas `[ ... ]` · tukšais mezgls — viens autors, divi predikāti par viņu
- Punkts beigās · ieraksta beigas

## RDF sintakses — viens modelis, vairākas pieraksta formas (I1)

Sintakse	Kur sastopaties	Cilvēkam lasāms
<b>JSON-LD</b>	dv.dataverse.lv eksports · schema.org <code>@context</code> · API atbildes	Vidēji
<b>Turtle</b>	RDF dokumenti · SPARQL piemēri · SKOS vārdnīcas	Visvairāk
<b>RDFa</b>	HTML lapā iegulti strukturēti dati (RDFa Lite) · Drupal, GoodRelations vēsturiski	Iegulsts atribūtos
<b>RDF/XML</b>	Vēsturiskais formāts · daļa no DataCite eksportiem	Vismazāk

**Sintakse, ne vārdnīca.** Visas pieraksta tos pašus trijniekus. Jūsu metadatu eksports izvēlas formātu — saturs paliek tas pats grafs.

# Vārdnīcas, ontoloģijas un to publicēšana

SKOS jēdzienu shēmas · OWL ontoloģijas · RDFa iegultie dati

## Vārdnīcas vs ontoloģijas — kur kura ir trijniekā (I2)

Vārdnīca → trijnieka *subjekti / objekti*

Kas ir resurss.

Saraksts ar jēdzieniem, autoriem, vietām, valodām — katram URI un apzīmējums.

Piemēri: MeSH, ORCID, GEMET, ROR

Ontoloģija → trijnieka *predikāti + klases + aksiomas*

Kā resursi viens ar otru saistās.

Predikātu un klašu definīcijas, ar formālu semantiku ( `subclassOf` , `inverseOf` , `functionalProperty` ).

Piemēri: RDFS, OWL, schema.org (daļēji), FOAF, DCAT

Wilkinson 2016 I2 prasa, lai **abi** — gan vārdnīcas, gan ontoloģijas — sekotu FAIR principiem. (*W3C nozīmē ontoloģijas arī ir vārdnīcas; šajā kursā lietojam praktisko atšķirību: vārdnīcas → vērtības, ontoloģijas → predikāti+klases.*) Šajā kursā fokuss uz vārdnīcām (vērtību līmenis); ontoloģijas — turpmākajām mācībām.

## 5 vārdnīcu tipi (12)

Tips	Definīcija (īsi)	Disciplīnas piemērs
<b>Vārdnīca</b> ( <i>controlled vocabulary</i> )	Apzināts terminu / vērtību / jēdzienu saraksts	DataCite <code>resourceType</code> (Dataset, Image, Text...)
<b>Tēzauris</b>	Vārdnīca + hierarhija (plašāks/šaurāks) + sinonīmi	UNESCO tēzauris · MeSH (medicīna)
<b>Taksonomija</b>	Stingri hierarhiska klasifikācija (parent-child)	Linnaeus nomenklatūra · failu sistēma
<b>Klasifikācijas shēma</b>	Iepriekš noteikta kategoriju shēma ar kodiem	Dewey Decimal · CPV (publiskie iepirkumi)
<b>Ontoloģija</b>	Vārdnīca + hierarhija + saites + loģiskās aksiomas	Gene Ontology · FOAF

## Kā tās izsaka mašīnlasāmi — SKOS, OWL, RDFa (I2)

Slānis	Kam paredzēts	Piemēri
<b>SKOS</b>	Jēdzienu shēmas — vārdnīcas, tēzauri, taksonomijas, <i>daļa</i> klasifikāciju	MeSH, GEMET, EuroVoc, AGROVOC
<b>OWL</b>	Ontoloģijas — predikāti, klases, formāla semantika	Gene Ontology, FOAF, schema.org daļēji
<b>RDFa</b>	RDF, iegulsts HTML lapās — meklētājiem un agregētājiem	Drupal CMS, GoodRelations, schema.org <b>Dataset</b> (RDFa Lite)

Ontoloģijas **netiek** izteiktas SKOS — tām ir OWL. SKOS ir jēdzienu apzīmējumu publicēšanai. Klasifikācijas shēmas tipiski paliek savās dzimtajās formās (Dewey, CPV); SKOS var kalpot kā **savietojuma karte** starp tām, ne kā oriģinālā publicēšanas forma.

## SKOS — jēdzienu shēmu publicēšanas slānis (I2)

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix mesh: <http://id.nlm.nih.gov/mesh/> .
@prefix wd:   <http://www.wikidata.org/entity/> .

mesh:D003920 a skos:Concept ;
  skos:prefLabel "Diabetes Mellitus, Type 2"@en ;
  skos:altLabel  "Type 2 Diabetes"@en, "T2DM"@en ;
  skos:broader   mesh:D003924 ;           # plašāks: Diabetes Mellitus
  skos:exactMatch wd:Q3025883 .         # Wikidata
```

- `prefLabel` jebkurā valodā · `altLabel` sinonīmi · `broader` hierarhijā
- `skos:exactMatch` — **savietojuma karte** ; sistēmas runā savā starpā

SKOS ir balstīts uz RDF; jēdzieni, apzīmējumi, hierarhijas un saites kā trijnieku grafs. **Vārdnīcām un tēzauriem** — ne ontoloģijām (tām OWL).

## Vārdnīcu reģistri — kur sākt meklēt (12)

### BARTOC

Vārdnīcu metareģistrs  
Visu tipu (KOS) reģistrs  
bartoc.org

### LOV

Linked Open Vocabularies  
RDF vārdnīcas  
lov.linkeddata.es

### LOD Cloud

Saistīto datu kopu karte  
Vizuāla pārskatu diagramma  
lod-cloud.net

### FAIRsharing

Standarti + datubāzes + politikas  
Daudznozaru  
fairsharing.org

### OBO Foundry

Dzīvības zinātņu ontoloģijas  
Kūrēta bibliotēka  
obofoundry.org

### OntoPortal family

BioPortal · AgroPortal · EcoPortal · MatPortal  
Disciplīnu portāli  
ontoportal.org

***0 latviski publicētu vārdnīcu LOV. Daudzas (GEMET, EuroVoc, AGROVOC) ir tulkotas — lietojiet angļu URI, saņemiet LV apzīmējumu klientā.***

## Aptauja 2 — vārdnīcu terminoloģijas atpazīstamība

*Kurus no šiem terminiem jūs jau lietojat darbā?*

- (a) Kontrolēta vārdnīca
- (b) Tēzauris
- (c) Taksonomija
- (d) Klasifikācijas shēma
- (e) Ontoloģija
- (f) Neviens

# I2 un I3 — pielietojums un R1.3 atšķirība

Vārdnīcas, attiecības un standarti

## I2 — vārdnīcas, kas seko FAIR (I2)

### Wilkinson 2016 (EN)

I2. (meta)data use vocabularies that follow FAIR principles.

- Lai izteiktu vērtības (autori, atslēgvārdi, vietas, valodas), lieto **kontrolētās vārdnīcas**
- F-UJI tests: [FsF-I2-01M](#)

### LV

I2. (Meta)dati lieto vārdnīcas, kas seko FAIR principiem.

*Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018.*

**Rekursija: FAIR dati lieto FAIR vārdnīcas.** Bet vārdnīca pati ir dati. Tāpēc **FAIR vārdnīca pati seko FAIR principiem** — atrodama, pieejama, sadarbspējīga, atkārtoti izmantojama.

## I3 — jēgpilnas attiecības ar citiem datiem / resursiem (I3)

### Wilkinson 2016 (EN)

I3. (meta)data include qualified references to other (meta)data.

### LV (kursa kanoniskā forma)

I3. (Meta)dati ietver jēgpilnas attiecības ar citiem (meta)datiem.

- Saites starp resursiem ir **kontrolētas** — ne vienkārši URLvai rakstu zīmju virknes
- Predikāti: `schema:sameAs`, `dcterms:isPartOf`, `dcterms:references`, `dcterms:isVersionOf`, `prov:wasDerivedFrom`
- Atšķirība no I1 (formāla valoda) — I3 ir formālā valodā izteiktas `_attiecības_` starp resursiem
- F-UJI tests: `FsF-I3-01M`

## R1.3 — kopienas standarti (*R1.3 ieskats*)

### Wilkinson 2016 (EN)

**R1.3.** (meta)data meet domain-relevant community standards.

### LV (kursa kanoniskā forma)

**R1.3.** (Meta)dati atbilst nozarei būtiskiem kopienas standartiem.

- Viss metadatu **ieraksts** atbilst kopienas-pieņemta standarta shēmai
- Disciplīnu piemēri: **DDI** (sociālās zin.) · **EML** (ekoloģija) · **MixS** (ģenētika) · **Croissant** (ML datu kopas)
- F-UJI tests: **FsF-R1.3-01M**

*R1.3 pilns izvērsums — 4. sesijā. Šeit minam definīciju, lai šobrīd nodalītu no I2.*

## I2 vs R1.3 — vērtību vārdnīca vs kopienas shēma (I2 + R1.3)

### I2 = vērtības līmenis (vārdnīca)

Vārdnīcas **konkrētiem laukiem**:

- autors → ORCID URI
- atslēgvārds → MeSH jēdziens
- vieta → Geonames

### R1.3 = ieraksta līmenis (~ontoloģija)

Visa shēma **kopienas standartam**:

- sociālo zinātņu kopiena → DDI
- ekoloģija → EML
- ģenētika → MixS

**Atšķirīgs tvērums; atšķirīgi F-UJI testi** (`FsF-I2-01M` vs `FsF-R1.3-01M`). Vienam ierakstam I2 var būt izpildīts, R1.3 — ne, un otrādi.

## I2 + I3 dubultloma (I2 + I3)

### I2 = vārdnīca

MeSH definē jēdzienu `D003920` (*Diabetes Mellitus, Type 2*). MeSH ir I2 resurss.

**Tas, ko jūs minat.**

### I3 = kvalificēta saite

`schema.org creator.sameAs https://orcid.org/0000-0002-0688-9213` — saite ir kvalificēta ar `sameAs` predikātu.

**Tas, kā jūs to saistāt.**

**Dubultloma:** `creator.sameAs` ar ORCID URL ir **I2 un I3 vienlaikus**. ORCID ir vārdnīca (I2); `sameAs` ir kvalificēta saite (I3).

## Vārdnīcu piemēri (12)

### MeSH

Medicīna

LV: daļējs

[id.nlm.nih.gov/mesh/](http://id.nlm.nih.gov/mesh/)

### GEMET

Vide

LV: pilns

[eionet.europa.eu/gemet/](http://eionet.europa.eu/gemet/)

### EuroVoc

ES politika

LV: pilns (24 val.)

[eur-lex.europa.eu/browse/eurovoc.html](http://eur-lex.europa.eu/browse/eurovoc.html)

### AGROVOC

Lauksaimniecība

LV: daļējs (FAO)

[fao.org/agrovoc/](http://fao.org/agrovoc/)

### ORCID

Autori (kā vārdnīca)

URI balstīta

[orcid.org/](http://orcid.org/)

### ROR

Iestādes (kā vārdnīca)

URI balstīta

[ror.org/](http://ror.org/)

Papildus piemēri: **Wikidata** (atvērta strukturēta zināšanu bāze — sk. nākamo slaidu)

## Wikidata — atvērta strukturēta zināšanu bāze (I2)

**Wikidata** — Wikimedia Foundation kopienas projekts (kopš 2012). Brīva, daudzvalodu strukturēto datu bāze; visi dati publicēti ar **CC0** licenci — Wikidata Q-ID var citēt jebkurš metadatu standarts bez juridiskām barjerām.

### Pamatafakti

- Vairāk nekā 110 milj. resursu · katram **Q-identifikators** (piem., `q42` = Douglas Adams)
- ~12 000 īpašību · katrai **P-identifikators** (piem., `p31` = *instance of*)
- 300+ valodas · LV apzīmējumi miljoniem resursu
- Bez maksas SPARQL galapunkts: `query.wikidata.org`

### Ietekme datu ekosistēmā

- Strukturēto datu avots **Wikipēdijai** (300+ valodu izdevumiem; infokastes)
- Lietota **GLAM** iestādēs · Europeana · DBpedia · OpenStreetMap
- **Universāls savienotājs** starp ORCID, ROR, MeSH, GeoNames, VIAF, Getty

Datu kuratoram: Wikidata Q-ID darbojas kā **savietojuma karte** starp disciplīnu vārdnīcām — viens jēdziens vienā vietā, ar saitēm uz visām vārdnīcām, kurās tas ir nosaukts.

## Aptauja 3 — disciplīnas kalibrācija

*Kuru disciplīnu jūs reprezentējat?*

- (a) Humanitārās zinātnes (literatūra, vēsture, filozofija, mākslas)
- (b) Sociālās zinātnes (socioloģija, psiholoģija, ekonomika, politoloģija)
- (c) Dzīvības zinātnes (medicīna, bioloģija, farmācija)
- (d) Vide / ekoloģija / Zemes zinātnes
- (e) Lauksaimniecība / pārtika
- (f) Cita disciplīna
- (g) Strādāju ar vairākām disciplīnām

# CVOC mehānisms (*pēcsesijas izpētei*)

Kā Dataverse savieno metadatu laukus ar vārdnīcu API

Šo bloku detalizēti nepārrunājam šajā sesijā — slaidi paliek pēcsesijas izpētei. Galvenā ziņa: **CVOC ir gatavs iespējošanai uz [dv.dataverse.lv](https://dv.dataverse.lv); konfigurācija ir 30 minūšu darbs**. Ja jūsu disciplīnai vajag konkrētu vārdnīcu — runājiet ar VPC datu kuratoriem.

## :CVocConf — konfigurācijas iestatījums (I2 + I3)

```
[
  {
    "field-name": "author",
    "term-uri-field": "authorIdentifier",
    "js-url": "https://gdcc.github.io/dataverse-external-vocab-support/scripts/orcid.js",
    "headers": { "Accept": "application/json" },
    "managed-fields": {
      "authorName": "schema:name",
      "authorAffiliation": "schema:affiliation"
    }
  },
  {
    "field-name": "keyword",
    "js-url": "https://gdcc.github.io/dataverse-external-vocab-support/scripts/ontoportal.js",
    "headers": { "Accept": "application/json" }
  }
]
```

- `field-name` · kuram metadatu laukam piesaistās · `term-uri-field` · kur saglabā URI
- `js-url` · skripts, ko Dataverse pievieno deponēšanas formai · `managed-fields` · automātiski aizpildāmie pavadošie lauki
- **API: 1 POST** uz `/api/admin/settings/:CVocConf` — *30 minūšu darbs administratora pusē*

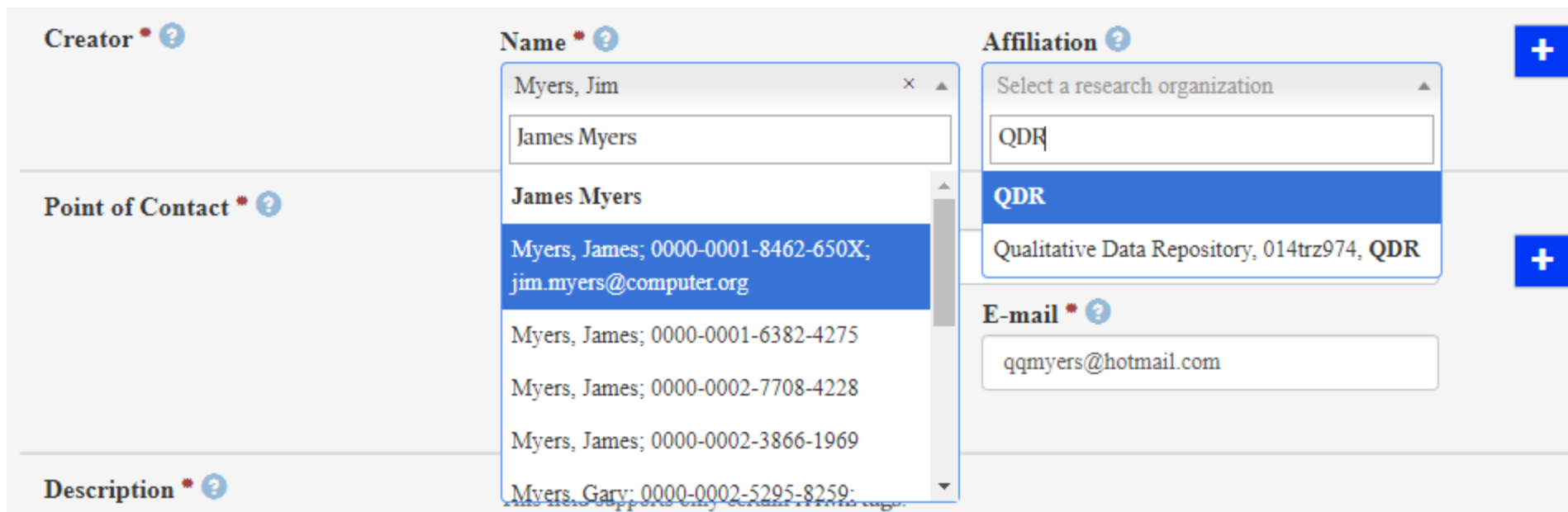
## gdcc kopienas skriptu katalogs (I2 + I3)

Skripts	Lauks	Vārdnīca
<code>orcid.js</code>	<code>authorIdentifier</code>	ORCID API
<code>ror.js</code>	<code>authorAffiliation</code>	ROR API
<code>ontoportal.js</code>	<code>keyword</code>	OntoPortal (NCBO BioPortal u.c.)
<code>geonames.js</code>	<code>producerAddress</code>	Geonames
<code>localcontexts.js</code>	<code>culturalLabels</code>	Local Contexts

Avots: [github.com/gdcc/dataverse-external-vocab-support](https://github.com/gdcc/dataverse-external-vocab-support) · MIT licence · DataverseNO, Borealis Toronto, [demo.dataverse.org](https://demo.dataverse.org) un Harvard Dataverse jau lieto.

Ja jums vajag MeSH vai SNOMED CT integrāciju — konfigurācijas jautājums uz [dv.dataverse.lv](https://dv.dataverse.lv) puses, ne pilnas funkcionalitātes izstrāde.

## CVOC darbībā — gdcc ORCID + ROR autoru piesaiste (I2 + I3)



**Creator** ?

**Name** ?

Myers, Jim

James Myers

**James Myers**

Myers, James; 0000-0001-8462-650X;  
jim.myers@computer.org

Myers, James; 0000-0001-6382-4275

Myers, James; 0000-0002-7708-4228

Myers, James; 0000-0002-3866-1969

Myers, Gary; 0000-0002-5295-8259

**Affiliation** ?

Select a research organization

QDR

**QDR**

Qualitative Data Repository, 014tr974, QDR

**E-mail** ?

qqmyers@hotmail.com

Saglabāts no [gdcc/dataverse-external-vocab-support README](#) · MIT licencēts · Input1 piemērs.

Lietotājs raksta autora vārdu → nolaižamais saraksts ar ORCID/ROR piedāvājumiem → izvēlas → **saglabātā vērtība ir URI, ne tikai teksts**. To F-UJI pārbaude [FsF-I2-01M](#) redz un atzīmē pass.

## Tālākai izpētei — saites un materiāli (I2 + I3)

### CVOC pilna dokumentācija

gdcc kopienas repo

[github.com/gdcc/dataverse-external-vocab-support](https://github.com/gdcc/dataverse-external-vocab-support)

### :CVocConf admin docs

Dataverse Guides

[guides.dataverse.org/en/latest/installation/config.html#cvocconf](https://guides.dataverse.org/en/latest/installation/config.html#cvocconf)

### Local Contexts

CARE principi · pamatiedzīvotāju datu  
marķēšana

[localcontexts.org](https://localcontexts.org)

### OntoPortal tehniskie piemēri

BioPortal API + integrācija

[ontoportal.github.io/documentation](https://ontoportal.github.io/documentation)

### gdcc kopiena

Slack + e-pasta saraksti

[gdcc.io](https://gdcc.io)

### VPC datu kuratori

Konsultācija par disciplīnas vārdnīcu

[vpc.lv](https://vpc.lv)



Fināns  
Eiropas Savienībā  
NextGenerationEU



2027  
Nacionālais  
atlabības plāns

# Praktiskais darbs

Vārdnīcas izvēle

## Datu kopas — izvēlieties vienu (*I2 + R1.3 ieskats*)

Variants	Datu kopa	DOI
(a)	Balodis · Biogrāfisko interviju transkripcijas (humanitāri)	<a href="https://doi.org/10.71782/DATA/QVUERT">10.71782/DATA/QVUERT</a>
(b)	Surikova et al. · Pieaugušo profesionālā attīstība (sociāli zin.)	<a href="https://doi.org/10.71782/DATA/SMGJHG">10.71782/DATA/SMGJHG</a>
(c)	Spinazzola et al. · <i>Athar</i> biosignal datu kopa (dzīvības zin.)	<a href="https://doi.org/10.5281/zenodo.15365515">10.5281/zenodo.15365515</a>
(d)	<b>Sava datu kopa</b> — jau publicēta vai plānošanas stadijā	jūsu DOI

**Strādāriet ar to, kas jūsu disciplīnai ir vistuvāk.** Pēc darba demonstrēšu Baloža (QVUERT) paraugu kā strādājošu piemēru — pārējiem trim ir paralēli raksti, kurus var pārlasīt README failā pēc sesijas.

## Protokols — kā strādāt (I2 + R1.3 ieskats)

1. **Atveriet datu kopas lapu.** `dv.dataverse.lv`: scroll uz *Citation Metadata* → **Keywords**. Zenodo: *Subjects* zem nosaukuma. Sava kopa: lapa vai API eksports (sk. README §5).
2. **Pierakstiet atslēgvārdus un to formātu** — cik ir? Brīvs teksts, URI, kombinācija? Atsevišķi lauki vai komatu virkne?
3. **Atrodiet kandidāta vārdnīcu(-as)** savai disciplīnai. Disciplīnas ceļš:
  - medicīna → **MeSH**; vide → **GEMET**; ES politika → **EuroVoc**; lauksaimniecība → **AGROVOC**; sociālo zin. → **ELSST**
  - humanitāri / vispārīgi / krustdisciplināri → **Wikidata**, **LCSH**, **GND**
  - reģistri meklēšanai: **LOV** ( `lov.linkeddata.es` ), **BARTOC** ( `bartoc.org` )
4. **Atrodiet konkrētus jēdzienus, kas atbilst jūsu atslēgvārdiem.** Atveriet vārdnīcas pārliuku (piem., `meshb.nlm.nih.gov` , `wikidata.org` , `eionet.europa.eu/gemet` ), meklējiet **katru** savu atslēgvārdu un pierakstiet jēdziena URI. Ja precīza URI nav — pierakstiet **plašāko jēdzienu**, ar kuru var savienot, vai atzīmējiet "nav atbilstošas URI". *Tas pats par sevi ir konstatējums.*
5. **Aptauja 4** fiksēs jūsu vārdnīcu izvēli. Tad demonstrēšu Baloža paraugu kā strādājošu piemēru.

Q&A panelis pieejams jautājumiem. Ja cilne neielādējas — atveriet `materials/practice/3-interoperable/data/topic3-block3.6-{paraugs}/` HTML failu.

## Aptauja 4 — vārdnīcas izvēle (I2 + R1.3 ieskats)

*Kuru kontrolēto vārdnīcu jūs apsvērāt savai izvēlētajai datu kopai?  
Anonīma · vairākas atbildes · ~90 sek balsošanai*

- (a) **MeSH** — Medical Subject Headings (medicīna)
- (b) **GEMET** — vides terminoloģija
- (c) **EuroVoc** — ES politikas terminoloģija
- (d) **AGROVOC** — lauksaimniecības terminoloģija
- (e) **Wikidata** vai **LCSH** (*humanitāri / vispārīgi*)
- (f) **ELSST** vai cita disciplīnas vārdnīca
- (g) Cita vārdnīca (*atstājiēt komentāru pēc sesijas*)
- (h) Nezinu / vajag konsultāciju

## Baloža paraugs — atslēgvārdu lauka stāvoklis (QVUERT)

Subject: "Arts and Humanities"  
Keyword: "filozofija Latvijā, padomju filozofija, konversijas, pārrāvumi, epistēmiskās stratēģijas"

- **Viens** klasifikators (Subject) un **viens** keyword lauks ar **piecu LV terminu komatu virkni**
- Nav `subjectScheme` · nav `valueUri` · vārdnīca **nav** izvēlēta
- F-UJI 2026-04-27: **FsF-I2-01M fail (0/2)** — empīriskais motivācijas avots šim blokam

**Divkārša nepilnība:** (1) vārdnīca nav izvēlēta — brīvs teksts; (2) pieci koncepti vienā virknē — metadatu modelis tos neredz kā piecus, bet kā vienu lauku.

## Baloža paraugs — kandidāta vārdnīcas un jēdzieni (I2)

Termins	Vārdnīca	Kandidāta jēdziens
filozofija Latvijā	<b>Wikidata</b>	<a href="#">wd:Q5891</a> <i>philosophy</i> + <a href="#">wd:Q211</a> <i>Latvia</i> kā vieta
padomju filozofija	<b>LCSH</b>	<i>Philosophy, Russian — Soviet period</i>
epistēmiskās stratēģijas	<b>Wikidata</b>	<a href="#">wd:Q9471</a> <i>epistemology</i> ( <i>plašāks; precīza jēdziena nav</i> )
konversijas ( <i>filozof.</i> )	<b>brīvs teksts</b> ar atrunu	konteksta-specifiska nozīme — Wikidata <a href="#">Q193292</a> ir reliģiska konversija, <b>ne šī</b>
pārrāvumi ( <i>rupture</i> )	<b>brīvs teksts</b>	nav viennozīmīga URI — Foucault-tradīcijas jēdziens
<i>Arts and Humanities</i>	<b>DataCite Subject</b>	augstā līmeņa klasifikators — jau ir laukā

**Realitāte:** vārdnīcas pārklāj ~3 no 5 terminiem ar URI; pārējie paliek brīvā tekstā ar disciplīnas-specifisku nozīmi. *Tas ir pareizi* — vārdnīcas nav universāls aizvietotājs, bet stiprā URI piesaiste tām, kas ir.

## I2 trūkuma divi slāņi (I2 + R1.3 ieskats)

Slānis	Risināmais
Depozitora izvēle ( <i>brīvs teksts</i> → <i>vārdnīcas termins ar URI</i> )	Jūs — rītdien
Repozitorija piesaiste (CVOC) ( <i>termins ar URI</i> ↔ <i>saglabāts URI</i> )	Repozitorija administrators

Baloža QVUERT: **depozitora izvēle nav notikusi** — brīvs teksts ar piecām nozīmēm vienā virknē. Pirmais slānis. (SMGJHG paralēls; Athar — depozitora izvēle **ir**, MeSH *ir* nosaukts, bet platformas CVOC nepieslēdzas — otrs slānis. Sk. README §7.)

Pirmo slāni jūs varat risināt rītdien — izvēlieties vārdnīcu savai disciplīnai.

Otro slāni risina repozitorija administrators — **runājiet ar VPC datu kuratoriem par CVOC iespējošanu [dv.dataverse.lv](https://dv.dataverse.lv).**

# Kopsavilkums un tilts

I-rinda pilna · R1.3 pa pusei · 4. sesija — R praksē

## FAIR karte — sesijas beigas

<b>F1</b> pastāvīgs ID	<b>F2</b> bagātīgi metadati	<b>F3</b> ID metadatos	<b>F4</b> reģistrēts
<b>A1</b> protokols	<b>A1.1</b> atvērts protokols	<b>A1.2</b> autorizācija protokolā	<b>A2</b> nezūdoši metadati
<b>I1</b> formālā valoda	<b>I2</b> FAIR vārdnīcas	<b>I3</b> jēgpilnas attiecības	
<b>R1</b> bagātīgs apraksts	<b>R1.1</b> licence	<b>R1.2</b> izcelsme	<b>R1.3</b> nozaru standarti

**I-rinda pilna.** R1.3 pa pusei — vārdnīcas izvēle, ko jūs tikko praktizējāt, ir kopienas standarta izvēle mikrolīmenī. R, R1.1, R1.2 — pēcpusdienas pēdējā sesija.

## Galvenās atziņas

1. **Saistītie dati = trijnieku grafs** — JSON-LD `@context` jau ir RDF; trijnieki savienojas ar kopējiem subjektiem/objektiem; jūsu metadatu eksports jau ir grafa fragments.
2. **Vārdnīcas un ontoloģijas ir atšķirīgi resursi** — vārdnīcas ir trijnieku subjekti/objekti (SKOS publicē); ontoloģijas ir predikāti + klases (OWL publicē).
3. **I2 ≠ R1.3** — I2 ir vērtību vārdnīca konkrētiem laukiem; R1.3 ir kopienas shēma visam ierakstam. Atšķirīgi F-UJI testi.
4. **I2 + I3 var būt dubulta loma** — viens ORCID URL `creator.sameAs` ir abi vienlaikus.
5. **CVOC ir iespējams uz [dv.dataverse.lv](https://dv.dataverse.lv)** — mehānisms gatavs ieslēgšanai; **runājiet ar VPC datu kuratoriem.**

## Tilts uz 4. sesiju

I-rinda pilna. R1.3 pa pusei. **R, R1.1 un R1.2 — atvērtas.** Tā ir 4. sesija pēc pārtraukuma.

- **R1.1** · Licences — kā SPDX licence padara automatizāciju iespējamu
- **R1.2** · Izcelsme — kā provenance ietekmē datu citēšanu un atkārtotu izmantošanu
- **R1.3** (*pilnā veidā*) · Nozaru standarti — DDI sociālajām zinātnēm · EML ekoloģijai · MIxS ģenētikai
- **Croissant** — kopienas standarts ML datu kopām · uz Dataverse upstream roadmap · vēl nav LV eksportos
- **Wikidata SPARQL praktiskais darbs** · ne tikai mehānisma demonstrācija — jūs paši rakstīsiet vaicājumus

*Pārtraukums — 15 minūtes. Atgriežamies 15:00 ar 4. sesiju.*

# Semināra 3. sesija

**I + R — Saistītie dati, vārdnīcas, RDF · pabeigts**

Pārtraukums līdz 15:00.