

|  |  |
| --- | --- |
|  |  |

**Datu publicēšanas platformas ieviešanas tehniskās specifikācijas un ar to saistīto vadlīniju izstrāde**

*DPP mašīnlasāmo atvērto datu kopu datu*

*struktūras standarts*

versija 1.1

|  |  |
| --- | --- |
| **Pasūtītājs:**  | **Valsts reģionālās attīstības aģentūra** Alberta iela 10, Rīga, LV-1010Tālrunis: +371 670 79 000Fakss: +371 670 79 001e-pasts: pasts@vraa.gov.lv  |
|  |  |
| **Izpildītājs:** | **Agile & CO SIA,** Citadeles iela 12, Rīga, LatvijaTālrunis: +371 200 67 009Fakss: +371 67 321 400e-pasts: info@agile.lv  |

**Rīga 2016**

**Dokumenta aizsardzība**

Šis dokuments un tā saturs pēc pieņemšanas un nodošanas akta parakstīšanas ir uzskatāms par Valsts reģionālās attīstības aģentūras īpašumu.

Valsts reģionālās attīstības aģentūrai šo dokumentu atļauts lietot bez ierobežojumiem Latvijas valsts varas un pārvaldes institūcijās.

Augstāk minēto nosacījumu pārkāpšana ir uzskatāma par LR tiesību aktu pārkāpumu un vainīgā persona var tikt saukta pie atbildības atbilstoši LR tiesību aktos noteiktajai kārtībai.

**Kontaktpersona:**

Krišjānis Jurģelis

SIA „Agile & Co”

Citadeles iela 12, Rīga, LV – 1010

Tālr.: +371 26321422

E-pasts: krisjanis@agile.lv

**Tirdzniecības zīmes**

Visas tekstā izmantotās tirdzniecības zīmes pieder to īpašniekiem un ir izmantotas tikai kā atsauces.

**Dokumenta izmaiņu vēsture**

| **Nr.p.k.** | **Versijas numurs** | **Izmaiņu datums** | **Izmaiņu apraksts** |
| --- | --- | --- | --- |
| 1. | 0.1. | 18.08.2016. | Sagatavota un iesniegta izskatīšanai dokumenta struktūra  |
| 2. | 0.5. | 10.10.2016. | Pasūtītājam izskatīšanai iesniegta pirmā dokumenta versija atbilstīgi 03.10.2016. veiktajai intervijai ar DPP biznesa īpašnieku |
| 3. | 1.0. | 06.11.2016. | Veikti precizējumi vizuālajā noformējumā un pievienota saīsinājumu tabula |
| 4. | 1.1. | 14.11.2016. | Dokumenta gala versija ar precizējumiem atbilstoši komentāriem, kas no Pasūtītāja saņemti 07.11.2016. |

**Satura rādītājs**

[1 Vispārēja informācija par dokumentu 5](#_Toc466051803)

[1.1 Dokumenta mērķis 5](#_Toc466051804)

[1.2 Dokumenta lietotāji 5](#_Toc466051805)

[1.3 Dokumenta sfēra 5](#_Toc466051806)

[1.4 Dokumentā izmantotie saīsinājumi 5](#_Toc466051807)

[1.5 Saistītie dokumenti 7](#_Toc466051808)

[1.6 Pieņēmumi un ierobežojumi 7](#_Toc466051809)

[2 Ievads 8](#_Toc466051810)

[3 Datu struktūras aprakstīšanas prasības 10](#_Toc466051811)

[3.1 Metadati par datu kopu 10](#_Toc466051812)

[3.2 Metadati par datu struktūru 11](#_Toc466051813)

[4 Metadatu dokumenta uzbūve 12](#_Toc466051814)

[4.1 Standartā lietotā JSON notācija 12](#_Toc466051815)

[4.2 Dokumenta pamatstruktūra 12](#_Toc466051816)

[4.3 CSV dialekta definīcija 13](#_Toc466051817)

[4.4 Datu kopas metadati 13](#_Toc466051818)

[4.5 Datu shēma 15](#_Toc466051819)

[4.5.1 Datu shēmas apraksts 15](#_Toc466051820)

[4.5.2 Tabulas kolonnu definīcijas 16](#_Toc466051821)

[5 Metadatu dokumentu publicēšanas principi 20](#_Toc466051822)

[Pielikums Nr.1 21](#_Toc466051823)

[Pielikums Nr.2 23](#_Toc466051824)

[Kontaktinformācija un rekvizīti 24](#_Toc466051825)

# Vispārēja informācija par dokumentu

## Dokumenta mērķis

Dokumenta mērķis ir sniegt informāciju par esošās situācijas analīzes ietvaros iegūtajiem secinājumiem, Datu publicēšanas platformas ieviešanas mērķiem, biznesa prasībām un tās ieviešanas ieguvumiem.

Dokuments ir sagatavots atbilstoši pie atklāta konkursa *„Konsultantu piesaiste programmatūras kvalitātes kontrolei VRAA īstenoto projektu realizācijā” (id.Nr.VRAA/2013/14/ERAF/AK)* rezultātā 2014. gada 26. februārī noslēgtās Vispārīgās vienošanās (Pasūtītāja vienošanās reģ. Nr. 13-7/14/6) 2016. gada 4. augustā noslēgtā līguma par darba uzdevuma izpildi Nr.13-7/16/95 ietvaros, kas noslēgts starp Valsts reģionālās attīstības aģentūru un SIA „Agile & Co”.

## Dokumenta lietotāji

Dokumentam ir šādi lietotāji:

* VARAM atbildīgie darbinieki, kuri nodrošina Datu publicēšanas platformas biznesa prasību uzturēšanu un definēšanu;
* VRAA atbildīgie darbinieki, kuri nodrošina Datu publicēšanas platformas tehnisko ieviešanu;
* Atvērto datu kopu publicētāji;
* DPP izstrādātāji.

## Dokumenta sfēra

Dokumentā ir apskatīti šādi jautājumi:

* Datu kopu struktūras definēšanas un aprakstīšanas pamatprincipi;
* DPP iekļaujamo datu kopu tipu apskats;
* Datu kopu struktūras aprakstīšanas standarts.

## Dokumentā izmantotie saīsinājumi

Dokumentā izmantotie saīsinājumi, kā arī to skaidrojums ir sniegti tabulā 1.

*Tabula 1* **Dokumentā izmantotie saīsinājumi**

| **Saīsinājums/ Termins** | **Nozīme** |
| --- | --- |
| CSVW grupa | *W3C CSV on the Web* darba grupa, kura ir izstrādājusi standartu tabulveida datu aprakstīšanai tīmeklī. |
| CSVW modeļaspecifikācija | CSVW izstrādātā tabulveida datu standarta specifikācija *Model for Tabular Data and Metadata on the Web.* Skat. <http://www.w3.org/TR/tabular-data-model/> |
| CSVW vārdnīcasspecifikācija | CSVW izstrādātā tabulveida datu standarta specifikācija *Metadata Vocabulary for Tabular Data.* Skat. <http://www.w3.org/TR/tabular-metadata/> |
| DPP | Datu publicēšanas platforma  |
| EDP | Eiropas datu portāls  |
| IRI | *Internationalized Resource Identifier* - URI identifikatoru formāta vispārinājums, kurš pieļauj ne-ASCII simolu lietošanu. Šajā dokumentā visur, kur ir pieminēti URI, var lietot arī IRI.Skat. <http://www.ietf.org/rfc/rfc3987.txt> |
| Izpildītājs | Sabiedrība ar ierobežotu atbildību „Agile & Co” |
| JSON | *JavaScript Object Notation* - atvērts tekstveida datu formāts, kas ļauj izteikt informāciju "atribūts - vērtība" pāru formā. |
| JSON-LD | *JSON-based Serialization for Linked Data* - W3C standarts, kurš definē ērti lietojamu veidu saistīto datu (RDF) informācijas izteikšanai JSON formā. Skat. <https://www.w3.org/TR/json-ld/>  |
| LR | Latvijas Republika |
| Metadati | Dati, kas apraksta citus datus. Papildus informācija, kas apraksta datu struktūru vai īpašības. |
| Metadatu dokuments | TDS standartam atbilstošs dokuments, kas apraksta tabulveida datu kopas struktūru un metadatus |
| Pasūtītājs | Valsts reģionālās attīstības aģentūra |
| PIKTAPS | Publiskās pārvaldes informācijas un komunikācijas tehnoloģiju arhitektūras pārvaldības sistēma |
| RDF | *Resource Description Framework* (Resursu aprakstīšanas ietvars) - W3C standarts, kas ļauj tīmeklī aprakstīt informāciju par objektiem, to īpašībām un savstarpējo saistību. Skat. <https://www.w3.org/RDF/>  |
| TDS standarts | Tabulveida datu struktūras standarts, kurš ir definēts šajā dokumentā |
| URI | *Unified Resource Identifier* - vienots resursu identifikatoru formāts, kurš parasti tiek izmantots, lai norādītu uz tīmekļa un saistīto datu objektiem. Skat. arī IRI. |
| VARAM | Vides aizsardzības un reģionālās attīstības ministrija  |
| VRAA | Valsts reģionālās attīstības aģentūra |
| W3C | *World Wide Web Consortium* (Globālā tīmekļa konsorcijs) - tīmekļa standartu organizācija. Skat. <https://www.w3.org/>  |

## Saistītie dokumenti

Šis dokuments ir skatāms kontekstā ar 2.tabulā apkopotajiem nodevumu dokumentiem.

*Tabula 2* **Saistītie dokumenti**

| **Nr.p.k.** | **Nosaukums (identifikators, versija)** |
| --- | --- |
|  | DPP Biznesa prasību izpētes, analīzes un detalizēšanas apkopojums |
| 1.
 | EDP izvērtējuma ziņojums ar aprakstu par EDP izmantotajiem rīkiem un to funkcijām |
|  | DPP tehniskā specifikācija  |
|  | DPP metadatu standarts |
|  | Ceļvedis DPP metadatu aprakstīšanai |
|  | Ceļvedis DPP datu struktūras izveidei un aprakstīšanai |
|  | Tehniskās vadlīnijas datu publicētājiem |
|  | Prasību saraksts par DPP nepieciešamajiem tehniskajiem resursiem |
|  | Creative Commons CC0 (Public Domain) un Attribution 4.0 International atvērto licenču tulkojums latviešu valodā |
|  | DPP ieviešanas plāns |

## Pieņēmumi un ierobežojumi

Dokuments ir sagatavots, ņemot vērā šādus pieņēmumus un ierobežojumus:

1. Metodes datu kopu struktūras definēšanai un aprakstīšanai ir atkarīgas no datu kopas veida. Šis mašīnlasāmo atvērto datu kopu datu struktūras standarts definē līdzekļus, ar kuriem var aprakstīt tabulveida datu kopu struktūru.
2. Standarta metode tabulveida datu struktūras aprakstīšanai ir CSVW 2015. gadā publicētais W3C standarts tabulveida datu aprakstīšanai tīmeklī.

# Ievads

Liels apjoms tīmeklī publicēto atvērto datu kopu ir pieejamas tabulveida (CSV, TSV, Excel, u.c.) datu formā.

**Tabulveida dati** ir tādi dati, kuros informācija ir strukturēta vienas vai vairāku tabulu veidā. Katra tabula sastāv no rindām (ierakstiem) un kolonnām. Katra rinda satur vienu ierakstu, kurš sastāv no kolonnām atbilstošajām šūnām. Visas šūnas, kas atrodas vienā kolonnā, apraksta viena un tā paša veida informācija.

Šis dokuments definē TDS standartu – veidu kā atbilstoši DPP vajadzībām aprakstīt tabulveida datu kopu saturu un uzbūvi. TDS standarta metadatu formāts galvenokārt ir paredzēts CSV un citu teksta formātu, kur kolonnas tiek atdalītas ar iepriekš definētu atdalītājsimbolu, datu aprakstīšanai, tomēr to ir iespējams pielāgot arī citu tabulveida datu (piem., rēķintabulu, HTML tabulu vai SQL datņu) struktūras aprakstīšanai[[1]](#footnote-1).

TDS standarts balstās uz W3C izstrādāto standartu tabulveida datu aprakstīšanai, kura pamata specifikācijas ir:

* **Model for Tabular Data and Metadata on the Web[[2]](#footnote-2)** - datu modeļa specifikācija, kura definē vispārīgu modeli tabulveida datu un to metadatu aprakstīšanai tīmeklī.
	+ Šajā dokumentā tai tiks lietoti apzīmējumi "CSVW modelis" un "CSVW modeļa specifikācija".
* **Metadata Vocabulary for Tabular Data[[3]](#footnote-3)** - metadatu vārdnīca tabulveida datu aprakstīšanai tīmeklī, kura definē uz JSON bāzes veidotu formātu tabulveida datu kopu metadatu aprakstīšanai.
	+ Šajā dokumentā tai tiks lietoti apzīmējumi "CSVW vārdnīca" un "CSVW vārdnīcas specifikācija".

TDS standarts izmanto CSVW modeli un vārdnīcu, precizējot šajās specifikācijās aprakstīto elementu lietojumu (t.sk. struktūras aprakstīšanai nepieciešamo elementu apakškopu). Jebkuram TDS standartā aprakstītā formā datu struktūras aprakstam ir jāatbilst CSVW vārdnīcas specifikācijas prasībām.

* Lai veidotu standartam atbilstošus metadatu dokumentus, to publicētājiem ir pietiekami sekot šajā dokumentā iekļautajām prasībām.
* Sistēmām, kuras apstrādā šajā standartā aprakstītos metadatus dokumentus, ir jāspēj bez kļūdām apstrādāt CSVW vārdnīcas specifikācijā definētajā formātā noformētus metadatu dokumentus.

Sīkāka informācija par JSON formātu, kurā ir aprakstāmi tabulveida datu kopu metadati, ir atrodama turpmākajās nodaļās.

# Datu struktūras aprakstīšanas prasības

Tabulveida datu kopu publicētājiem katrai datu kopai ir jāizveido metadatu dokuments, kas satur datu kopas metadatus un struktūras aprakstu. Metadatu dokumentam ir jāatbilst šajā standarta iekļautajām prasībām.

Katrs metadatu dokuments apraksta vienu tabulu.

Prasības pēc to svarīguma ir iedalāmas trīs veidos:

1. **Obligāta** (tekstā apzīmēta ar "ir jāiekļauj") - prasība ir obligāta;
2. **Būtiska** (apzīmēta ar "ir ieteicams") - prasība ir svarīga, to ir ieteikts ievērot;
3. **Brīvprātīga** (apzīmēta ar "var iekļaut") - prasības izpilde ir brīvprātīga.

## Metadati par datu kopu

Metadatu dokumentā par datu kopu **ir jāiekļauj** šāda informācija:

* URL, kas norāda uz datni, kuru apraksta šis metadatu dokuments[[4]](#footnote-4);
* nosaukums latviešu un angļu valodā;
* apraksts latviešu un angļu valodā. Aprakstā var iekļaut papildus informāciju, kas ir būtiska datu kopas lietotājiem (piem., datu atjaunošanas biežums, izmantotie aprēķina algoritmi, metodes datu iegūšanai u.c.);
* **datu kopas struktūras (shēmas) definīcija.**

Metadatu dokumentā **ir ieteicams** papildus iekļaut šādu informāciju (šī informācija ir jāiekļauj tad, ja par doto datu kopu tā ir pieejama):

* atslēgas vārdu saraksts, kas raksturo šo datu kopu;
* datu kopas oficiālās publicēšanas datums;
* pēdējo izmaiņu datums;
* informācija par organizāciju, kas ir publicējusi datu kopu (nosaukums un tīmekļa adrese);
* licence, kā vērtību norādot standarta licences identifikatoru (URI);
* URL tīmekļa lapai vai datnei, kurā ir atrodama informācija par šo datu kopu. Šis lauks var norādīt uz tīmekļa lapu vai uz datni, kas satur datu kopas aprakstu.

Papildus par datu kopu **var norādīt** šādu informāciju:

* papildus metadati, kuru iekļaušana ir atļauta CSVW vārdnīcas standartā;
* tehniska informācija par datu kopas dialektu gadījumos, ja datu kopa lieto nestandarta CSV formātu.

Datu publicētāji drīkst izmantot arī citus CSVW vārdnīcas specifikācijā atļautus atribūtus, lai aprakstītu papildus informāciju par datu kopu (skat. 4.4. nodaļu). Piemēram, tajā var iekļaut DPP metadatu standarta informāciju par datu kopu.

## Metadati par datu struktūru

Datu kopas struktūras (shēmas) definīcijā **ir jāiekļauj**:

* **visu tabulas kolonnu definīcija** (skat. zemāk);
* norāde uz tabulas primāro atslēgu (vienu vai vairākām kolonnām) – šis lauks ir jāiekļauj tad, ja šāda informācija ir pieejama.

Struktūras definīcijā ir jāiekļauj informācija (kolonnu apraksti) par visām tabulas kolonnām tādā secībā, kādā tās atrodas datnē.

Datu kopas struktūras (shēmas) aprakstā papildus **var iekļaut** šādu informāciju:

* informācija par tabulas rindu URI identifikatoru veidošanu (*aboutUrl*);
* informācija par tabulas ārējām atslēgām (*foreign keys*).

Katras kolonnas definīcijā **ir jāiekļauj**:

* nosaukums, kurš atbilst datnē lietotajam kolonnas nosaukumam;
* virsraksts (latviešu valodā);
* apraksts (latviešu valodā).

Kolonnas definīcijā **ir ieteicams** papildus iekļaut šādu informāciju (šī informācija ir jāiekļauj tad, ja šāda informācija ir pieejama):

* virsraksts angļu valodā;
* apraksts angļu valodā;
* teksta kolonnām – valodas kods ("lv", "en", utt.);
* informācija par kolonnas nozīmi (semantiku), norādot kolonnai atbilstošā jēdziena vai saistīto datu īpašības URI (*propertyUrl* atribūts)
* kolonnas datu tips, kuru vajadzības gadījumā papildina ar informāciju par lauka formāta nosacījumiem;
* pazīme vai kolonna ir obligāta.

Datu publicētāji kolonnas definīcijā papildus **var iekļaut** arī citus CSVW vārdnīcas specifikācijā atļautus atribūtus (skat. 4.5.2. nodaļu)*.*

# Metadatu dokumenta uzbūve

Metadatu dokumentiem, kas apraksta tabulveida datu kopu struktūru un metadatus, ir obligāti jāatbilst metadatu dokumenta pamatstruktūras principiem un tam ir jāsatur vismaz datu kopas metadati un datu shēmas apraksts.

Vajadzības gadījumā metadatu dokuments var papildus saturēt CSV datnes dialekta definīciju (skat. 4.3. nodaļu).

## Standartā lietotā JSON notācija

Metadatu dokuments ir JSON dokuments, kurš atbilst CSVW metadatu specifikācijā definētajam JSON dialektam[[5]](#footnote-5). Lai metadatu dokuments būtu atbilstošs DPP datu struktūras standartam, tam ir jāizpilda visas 3. nodaļā iekļautās obligātās datu struktūras aprakstīšanas prasības.

JSON formātā objekti tiek izteikti kā figūriekavās ietverta kopa ar objekta atribūtu nosaukumiem un to vērtībām, kuras var būt citi objekti, vērtību saraksti vai vienkāršas vērtības.

Turpmākās apakšnodaļas satur tabulas, kurās ir aprakstīti metadatu dokumentā lietoto JSON objektu atribūti. Kolonnās ar nosaukumu "Kard." ir norādīta īpašību kardinalitāte (piem., 0..1 nozīmē neobligātu atribūtu, kuram var būt ne vairāk kā 1 vērtība, savukārt 1..n nozīmē obligātu atribūtu, kuram var būt vairākas vērtības).

Šeit lietotajiem JSON objektiem nedrīkst būt atribūti ar vienādiem nosaukumiem, tādēļ gadījumā, ja atribūtam ir vairākas vērtības, tās tiek norādītas JSON saraksta formā.

## Dokumenta pamatstruktūra

Metadatu dokumenta pamatā ir viens augšējā līmeņa JSON objekts, kurš apraksta vienu tabulveida datni (tabulu)[[6]](#footnote-6). Dokumenta pamatstruktūrai ir jāatbilst šādam šablonam:

{

 "@context": [ "http://www.w3.org/ns/csvw", {"@language": "lv"} ],

 "url": "*datnes URL*",

*datu kopas metadati (skat. 4.4. nodaļu),*

*CSV dialekta definīcija (neobligāta, skat. 4.3. nodaļu),*

 "tableSchema": *datu shēmas apraksts (skat. 4.5. nodaļu)*

}

Augšējā līmeņa JSON objektam ir jāsatur atribūts @context ar šeit redzamo vērtību, kura sevī ietver CSVW formāta deklarāciju un informāciju par dokumentā iekļauto dabīgās valodas vērtību noklusēto valodu (latviešu valoda).

Augšējā līmeņa JSON objekts satur arī citus atribūtus, kuri ir aprakstīti turpmākajās nodaļās.

## CSV dialekta definīcija

Publicējot datu kopas, **ir ieteicams** izmantot CSV datu formātu, kurš atbilst CSVW datu modeļa dokumentā aprakstītajām prasībām (skat. nodaļu "Best Practice CSV")[[7]](#footnote-7), no kurām galvenās ir:

* izmantot UTF-8 kodējumu;
* datnes 1. rindai ir jāsatur kolonnu nosaukumi;
* kā kolonnu atdalītājsimbols ir jālieto komats;
* katrai tabulas rindai ir jāsatur vienāds kolonnu skaits;
* publicējot datni tīmeklī, kā HTTP satura veids ir jānorāda "text/csv" MIME tips.

Ja CSV datne neatbilst norādītajām labās prakses prasībām, tad metadatu dokumentā papildus ir jāiekļauj CSV formāta dialekta definīcija atbilstoši CSVW vārdnīcas specifikācijas (skat. nodaļu "Dialect Descriptions")[[8]](#footnote-8).

Dialekta definīcija tiek noformēta kā JSON objekts, kurš ir iekļauts augšējā līmeņa JSON objektā kā tā atribūts "dialect".

Piemēram, datnei Windows-1257 kodējumā, kurā kā kolonnu atdalītāji tiek lietoti tabulācijas simboli, atbilst šāda dialekta definīcija:

{

 "@context": [ "http://www.w3.org/ns/csvw", {"@language": "lv"} ],

 "url": "datnes URL",

 **"dialect": {**

 **"delimiter": "\t",**

 **"encoding": "windows-1257"**

 **}**

}

## Datu kopas metadati

Datu kopas (tabulas) metadati tiek norādīti kā augšējā līmeņa JSON objekta atribūti:

*Tabula 3* **Datu kopas (tabulas) metadati**

| Atribūts | Vērtība | Lietojuma piezīmes | Kard. |
| --- | --- | --- | --- |
| url | obligāts teksta lauks ar tabulveida datnes URL | Ir atļauts lietot relatīvos URL. Piem., datnei, kas atrodas vienā mapē ar metadatu dokumentu, pietiek norādīt datnes nosaukumu. | 1..1 |
| dc:title | Datu kopas virsraksti latviešu un angļu valodā | Skat. piezīmi par teksta vērtībām vairākās valodās.  | 2..n |
| dc:description | Datu kopas apraksti latviešu un angļu valodā | Skat. piezīmi par teksta vērtībām vairākās valodās. | 2..n |
| tableSchema | JSON objekts, kurš satur datu shēmas aprakstu | Skat. 0. nodaļu. | 1..1 |
| dcat:keyword | Atslēgvārdi | Atslēgvārdi, kas apraksta datu kopu | 0..n |
| dc:publisher | JSON objekts ar atribūtiem "schema:name" un "schema:url" | Kā objekta atribūtu vērtības tiek norādīti datu publicētāja organizācijas nosaukums un tīmekļa lapas URI. | 0..1 |
| dc:license | Licences specifikācijas URI | URI, kurš norāda uz mašīnlasāmu licences specifikāciju. | 0..1 |
| dc:issued | Datu kopas oficiālās izdošanas (piem., publicēšanas) datums | Datums vai datums un laiks ISO 8601 standarta formātā | 0..1 |
| dc:modified | Datu kopas pēdējo izmaiņu datums | Datums vai datums un laiks ISO 8601 standarta formātā | 0..1 |
| foaf:page | Teksta lauks, kurš satur tīmekļa lapas vai dokumentācijas URL | URL tīmekļa lapai vai datnei, kurā ir atrodama informācija par šo datu kopu. Drīkst norādīt vairākus URL. | 0..n |

Ja atribūtam ir vairākas vērtības, tad tās tiek norādītas kā saraksts:

 "dcat:keyword": [ "Budžets", "Finanšu ministrija" ]

Ja teksta lauka vērtība ir valodā, kas atšķiras no metadatu dokumenta noklusētās valodas, to norāda kā JSON objektu ar "@lang" (valodas ISO 639-1 kodu) un @value (teksta saturs) atribūtiem:

"dc:title": {

 "@lang": "lv",

 "@value": "Valsts budžets"

}

Ja atribūta vērtības tiek norādītas vairākās valodās, tās jānorāda saraksta veidā:

"dc:title": [{

 "@lang": "lv",

 "@value": "Valsts budžets"

 },{

 "@lang": "en",

 "en": "State budget"

 }]

Atribūti, kuru vērtības ir URI, kas identificē norādīto objektu (licenci, organizāciju u.c.), tiek norādīti kā JSON objekti, kuros URI ir iekļauts izmantojot "@id" atslēgas atribūtu:

 "dc:license": { "@id": "https://creativecommons.org/licenses/by/4.0/lv/" }

Atribūtiem, kuru vērtībām ir noteikts datu tips (piem., datums vai datums un laiks), vērtība tiek norādīta kā JSON objekts, kurš ir ietverta informācija par vērtības datu tipu (piem., xsd:date vai xsd:dateTime):

 "dc:issued": {"@value": "2010-12-31", "@type": "xsd:date"}

Sīkāka informācija par standartā paredzētajiem datu tipiem ir atrodama CSVW vārdnīcas specifikācijā[[9]](#footnote-9) un DPP datu struktūras aprakstīšanas ceļvedī.

Šī dokumenta 1. pielikumā ir atrodams metadatu dokumenta piemērs, kurš atbilst šī standarta prasībām un demonstrē nodaļā aprakstīto atribūtu lietojumu.

Vajadzības gadījumā, papildus šeit aprakstītajiem atribūtiem, datu publicētāji var lietot arī citus CSVW vārdnīcas specifikācijā norādītos atribūtus tabulas metadatu aprakstīšanai. Parasti šim nolūkam tiks izmantoti CSVW *vispārējie* atribūti[[10]](#footnote-10).

## Datu shēma

### Datu shēmas apraksts

Datu shēmas apraksts tiek iekļauts augšējā līmeņa JSON objektā kā "tableSchema" atribūts. Šim aprakstam ir noteikti jāsatur "columns" atribūts, kurš satur visu tabulas kolonnu definīcijas. Standartā ir paredzēti šādi datu shēmas apraksta objekta atribūti:

*Tabula 4* **Datu shēmas apraksta objekta atribūti**

| Atribūts | Vērtība | Lietojuma piezīmes | Kard. |
| --- | --- | --- | --- |
| columns | JSON objektu saraksts, kurš satur tabulas kolonnu definīcijas |  | 1..1 |
| aboutUrl | URI šablons, ar kura palīdzību tabulas rindām var izveidot tām atbilstošu URI identifikatoru  | Šablons ir teksta rinda, kurā var tikt ievietotas rindas kolonnu vērtības. Šo atribūtu var izmantot, lai tabulas rindām norādītu to saistīto datu URL | 0..1 |
| foreignKeys | Saraksts ar JSON objektiem, kas apraksta tabulas ārējās atslēgas  | Ārējās atslēgas norāda, ka šīs tabulas kolonnu vērtības norāda uz citām kolonnām tajā pašā vai citā tabulā. | 0..n |
| primaryKey | Tabulas primārā atslēga (vienas vai vairāku kolonnu nosaukumi) | Primārā atslēga norāda uz kolonnu vai sarakstu ar kolonnām, kuras ir šīs tabulas primārā atslēga. Tabulai var būt tikai viena primārā atslēga, bet tā var saturēt vairāku kolonnu nosaukumus. | 0..1 |

URI šabloni (piem., **aboutUrl**) tiek norādīti kā teksta rinda, kurā ar figūriekavām tiek apzīmēta šablona mainīgā daļa, kas tiek aizstāta ar atbilstošā tabulas elementa vērtību[[11]](#footnote-11):

 "tableSchema": {

 **"**aboutUrl**": "http://example.org/country/{code}",**

 "columns": [{

 "titles": "country",

 **"**name**": "**code**"**

 }],

 }

Piezīme: Gadījumā, ja tabula ir publicēta arī saistīto datu (*linked data*) formā, tad ar šablonu izveidotajiem URI jāatbilst tabulas objektu URI identifikatoriem tīmeklī.

Primārā atslēga (**primaryKey**) norāda uz vienu kolonnu vai sarakstu ar kolonnām, kuras ir tabulas primārā atslēga un viennozīmīgi identificē tabulas rindas. Primārās atslēgas vērtībām datu kopā ir jābūt unikālām. Datu kopu validātori pārbauda šo nosacījumu un atgriezīs kļūdu, ja tās nebūs unikālas.

Ārējās atslēgas (**foreignKey**) tiek norādīta atbilstoši CSVW vārdnīcas specifikācijā noteiktajam formātam[[12]](#footnote-12). Ārējās atslēgas ir jālieto uzmanīgi, jo datu kopu validātori šīs saites pārbauda un var atgriezt kļūdas ziņojumu (piem., ja nav atrodama ārējā atslēgā norādītā ārējā tabula vai tās kolonnas).

### Tabulas kolonnu definīcijas

Datu shēmas apraksta atribūts "column" satur sarakstu ar JSON objektiem ar tabulas kolonnu definīcijām tādā secībā kādā tās atrodas tabulveida datnē. Katra kolonnas definīcija atbilst vienai tabulas kolonnai un var saturēt šādus atribūtus:

*Tabula 5* **Kolonnas definīcijas apraksta atribūti**

| Atribūts | Vērtība | Lietojuma piezīmes | Kard. |
| --- | --- | --- | --- |
| **name** | **kolonnas nosaukums** | Kolonnu nosaukumiem vienas tabulas ietvaros ir jābūt unikāliem. Ja datne satur nosaukumu rindu, lauka vērtībai ir jāatbilst datnē norādītajam kolonnas nosaukumam. | 1..1 |
| titles | kolonnas virsraksti (dabīgās valodas teksts) | Virsraksti var būt vairāki un tie var būt norādīti vairākās valodās. Virsrakstam ir jābūt norādītam vismaz latviešu valodā. | 1..n |
| dc:description | plašāks kolonnas apraksts (dabīgās valodas teksts) | Apraksti var būt vairāki un tie var būt norādīti vairākās valodās. Aprakstam ir jābūt norādītam vismaz latviešu valodā. | 1..n |
| Mantojamie atribūti (skat. definīcija zem tabulas), kas definē kolonnas šūnu īpašības: |
| datatype | kolonnas datu tips  | Kolonnas datu tipa atribūtā ir iespējams norādīt arī formātu, kuram ir jāatbilst kolonnas vērtībām. | 0..1 |
| lang | kolonnas teksta valoda | ISO 639-1 valodas kods, kurš norāda valodu, kurā ir izteiktas dotās teksta kolonnas vērtības. | 0..1 |
| propertyUrl | kolonnas īpašības URI:URI šablons, ar kura palīdzību tabulas kolonnām var piekārtot tām atbilstošu URI identifikatoru | Šo atribūtu var izmantot, lai norādītu kolonnas nozīmi un tai atbilstošās saistīto datu īpašības identifikatoru. Šablons ir teksta rinda, kurā var tikt ievietots kolonnas nosaukums.  | 0..1 |
| required | pazīme, kas norāda vai kolonnas vērtības ir obligātas | Kolonnas vērtībai ir jābūt *true* (jā) vai *false* (nē), kas netiek likta pēdiņās. | 0..1 |

Dabīgās valodas saturs tiek norādīts atbilstoši 4.4. nodaļā minētajiem noteikumiem par vairāku teksta vērtību un, aprakstu gadījumā, daudzvalodu teksta vērtību pierakstīšanu. Kolonnu nosaukumus vairākās valodās var pierakstīt vienkāršākā formātā – tie tiek norādīti kā JSON objekts, kura atribūti ir valodu ISO 639-1 kodi un to vērtības ir atbilstošā valodas teksta vērtība vai vērtības:

 {

 "name": "latitude",

 "titles": {

 "lv": "ģeogrāfiskais platums",

 "en": "geo latitude"

 }

 }

Kolonnas definīcija var saturēt tā sauktos *mantojamos* atribūtus, kuri sniedz papildus informāciju par kolonnas šūnu vērtībām (t.i. šie atribūti attiecas uz katru kolonnas šūnu). Pilns mantojamo atribūtu saraksts ir atrodams CSVW vārdnīcas specifikācijā[[13]](#footnote-13). Šajā nodaļā ir iekļauti mantojamie atribūti, kuri ir būtiski DPP datu kopu kontekstā (skat. kolonnas definīcijas atribūtu tabulu).

Datu tipa (**datatype**) mantojamais atribūts ļauj norādīt kolonnas vērtību tipu (piem., vesels skaitlis). Tā vērtībai ir jābūt kādam no CSVW vārdnīcas standartā definētajiem datu tipiem[[14]](#footnote-14) vai no tiem atvasinātam datu tipam[[15]](#footnote-15). Standarta datu tipi tiek norādīti kā teksta vērtība ar dotā tipa identifikatoru (piem., "date"), bet atvasinātie tipi – kā JSON objekts, kura atribūti ir pamattipa identifikators (**base**) un papildus atribūti, kas norāda atvasinātā datu tipa ierobežojumus, piemēram:

{

 "name": "latitude",

 "titles": "ģeogrāfiskais platums",

 "datatype": {

 "**base**": "**number**",

 **"minimum": "-90",**

 **"maximum": "90"**

 }

}

Kolonnas īpašības URI atribūts (**propertyUrl**) var tikt izmantots:

1. lai aprakstītu tabulas kolonnu nozīmi, to norādot ar viennozīmīga RDF īpašības identifikatora (URI) palīdzību;
2. kā papildus informācija, kas palīdz tabulveida datus automātiski pārveidot formātos, kuros kolonnu atribūti un to vērtības tiek aprakstīti ar URI palīdzību (piem., RDF vai JSON-LD formā).

Piemēram, šādi var norādīt, ka vērtība ir ģeogrāfiskais platums (schema.org izpratnē)[[16]](#footnote-16):

{

 "name": "latitude",

 **"propertyUrl": "schema:latitude"**

 "datatype": "number"

}

Plašāka informācija par *datatype* un *propertyUrl* atribūtu lietojumu ir atrodama DPP datu struktūras aprakstīšanas ceļvedī [6].

Vajadzības gadījumā, papildus šeit aprakstītajiem atribūtiem, datu publicētāji var lietot arī citus CSVW vārdnīcas specifikācijā norādītos atribūtus. Parasti šim nolūkam tiek izmantoti *vispārējie* atribūti vai šajā nodaļā pieminētie *mantojamie* atribūti.

Šī dokumenta 1. pielikumā ir atrodams metadatu dokumenta piemērs, kurš atbilst šī standarta prasībām un kurā ir izmantoti šajā nodaļā aprakstītie kolonnas definīcijas atribūti.

##

# Metadatu dokumentu publicēšanas principi

CSVW modeļa specifikācijas nodaļa "Locating Metadata" definē standarta procedūru, kura ir jāizmanto, lai tīmeklī atrastu tabulveida datu kopu aprakstošo metadatu dokumentu[[17]](#footnote-17). Procedūra definē vairākus veidus kā publicēt metadatu dokumentus.

TDS standarts rekomendē izmantot vienkāršāko no veidiem un metadatu dokumentu publicēt ar tādu pašu URL kā CSV datnei, tam galā pievienojot "-metadata.json".

Cits veids, kuru var izmantot kopā ar iepriekš minēto veidu, ir sniegt informāciju par metadatu dokumentu HTTP atbildes sākuma daļas "Link:" laukā tad, kad tīmekļa lietotājs pieprasa dokumentā aprakstīto datu kopu:

HTTP/1.1 200 OK

Content-Type: text/tab-separated-values

...

Link: <metadata.json>; rel="describedBy"; type="application/csvm+json"

Papildus informācija par šo un citiem metadatu dokumentu publicēšanas veidiem ir atrodama CSVW modeļa specifikācijā.

Tīmeklī publicējot tabulveida metadatu dokumentus, HTTP pieprasījuma atbildē ir jānorāda tiem atbilstošais MIME tips: "application/csvm+json".

Sistēmām, kurās ir jāveic tabulveida datņu metadatu dokumentu atrašana, ir jāievēro visi CSVW metadatņu atrašanas procedūrā definētie principi.

# Pielikums Nr.1

**Metadatu dokumenta piemērs**

*Piezīme: pēc standarta pamatsatura apstiprināšanas šeit iekļautais piemērs tiks aizvietots ar Latvijas datu kopas metadatu dokumenta piemēru.*

Šajā pielikumā var atrast metadatu dokumenta piemēru, kurš apraksta CSV datni *tree-ops.csv* ar šādu saturu (fragments):

GID,On Street,Species,Trim Cycle,Inventory Date

1,ADDISON AV,Celtis australis,Large Tree Routine Prune,10/18/2010

2,EMERSON ST,Liquidambar styraciflua,Large Tree Routine Prune,6/2/2010

Metadatu dokuments, kas apraksta šo datni, var izskatīties šādi:

{

 "@context": ["http://www.w3.org/ns/csvw", {"@language": "lv"}],

 "url": "tree-ops.csv",

 "dc:title": "Tree Operations",

 "dcat:keyword": ["tree", "street", "maintenance"],

 "dc:publisher": {

 "schema:name": "Example Municipality",

 "schema:url": {"@id": "http://example.org"}

 },

 "dc:license": {"@id": "https://creativecommons.org/licenses/by/4.0/lv/"},

 "dc:modified": {"@value": "2010-12-31", "@type": "xsd:date"},

 "tableSchema": {

 "columns": [{

 "name": "GID",

 "titles": ["GID", "Generic Identifier"],

 "dc:description": "An identifier for the operation on a tree.",

 "datatype": "string",

 "required": true

 }, {

 "name": "on\_street",

 "titles": "On Street",

 "dc:description": "The street that the tree is on.",

 "datatype": "string"

 }, {

 "name": "species",

 "titles": "Species",

 "dc:description": "The species of the tree.",

 "datatype": "string"

 }, {

 "name": "trim\_cycle",

 "titles": "Trim Cycle",

 "dc:description": "The operation performed on the tree.",

 "datatype": "string"

 }, {

 "name": "inventory\_date",

 "titles": "Inventory Date",

 "dc:description": "The date of the operation that was performed.",

 "datatype": {"base": "date", "format": "M/d/yyyy"}

 }],

 "primaryKey": "GID",

 "aboutUrl": "#gid-{GID}"

 }

}

# Pielikums Nr.2

**Standartā izmantotās nosaukumvietas**

Nosaukumvietas (*namespaces*) ir standarta līdzeklis kā saīsināti atsaukties uz standarta vārdnīcu un XML shēmas terminiem. Tā ir tehniska informācija, kas datu publicētājiem nav ikdienā nepieciešama, tomēr tā ir būtiska pareizai standartā lietoto terminu interpretācijai.

CSVW vārdnīcas specifikācijā un uz tā pamata veidotajā TDS standartā tiek izmantots kompaktais IRI un URI pieraksta veids, kurš ir definēts JSON-LD standarta sadaļā Compact IRIs[[18]](#footnote-18).

TDS standartā tiek izmantotas klases un īpašības no šādām nosaukumvietām:

csvw: http://www.w3.org/ns/csvw#

dc: http://purl.org/dc/terms/

dcat: http://www.w3.org/ns/dcat#

foaf: http://xmlns.com/foaf/0.1/

rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

schema: http://schema.org/

xsd: http://www.w3.org/2001/XMLSchema#

# Kontaktinformācija un rekvizīti

**SIA „Agile&Co” kontakti**

Citadeles iela 12, Rīga, LV – 1010, Latvija

Telefons: +371 200 67 009

Fakss: + 371 673 21 400

e-pasts: info@agile.lv

web: [www.agile.lv](http://www.agile.lv)

**SIA „Agile&Co” rekvizīti**

Citadeles iela 12, Rīga, LV – 1010, Latvija

Reģ. Nr. 40003939574

Maksājumu rekvizīti:

Banka: a/s Swedbank

Kods HABALV22

Konts LV11HABA0551016601272

1. Lai gan CSVW modelis ļauj aprakstīt jebkādas tabulveida datu kopas, CSVW vārdnīcas specifikācijā ir paredzēti līdzekļi tieši CSV un līdzīgu formātu datņu dialektu aprakstīšanai. Citu tabulveida formātu atbalstam var būt nepieciešams papildus standartizācijas darbs. [↑](#footnote-ref-1)
2. <http://www.w3.org/TR/tabular-data-model/> [↑](#footnote-ref-2)
3. <http://www.w3.org/TR/tabular-metadata/> [↑](#footnote-ref-3)
4. Vajadzības gadījumā norādei uz tabulveida datni var izmantot relatīvos URL. [↑](#footnote-ref-4)
5. <https://www.w3.org/TR/2015/REC-tabular-metadata-20151217/#json-ld-dialect> [↑](#footnote-ref-5)
6. CSVW metadatu standarts pieļauj arī gadījumu, kad metadatu dokuments vienlaikus apraksta vairākas tabulas, tomēr DPP datu struktūras standartā ir paredzēts, ka katrs metadatu dokuments apraksta vienu tabulu. [↑](#footnote-ref-6)
7. <https://www.w3.org/TR/tabular-data-model/#syntax> [↑](#footnote-ref-7)
8. <https://www.w3.org/TR/2015/REC-tabular-metadata-20151217/#dialect-descriptions> [↑](#footnote-ref-8)
9. <https://www.w3.org/TR/tabular-metadata/#datatypes> [↑](#footnote-ref-9)
10. <https://www.w3.org/TR/tabular-metadata/#common-properties> [↑](#footnote-ref-10)
11. <https://www.w3.org/TR/tabular-metadata/#uri-template-properties> [↑](#footnote-ref-11)
12. <https://www.w3.org/TR/tabular-metadata/#schema-foreignKeys> [↑](#footnote-ref-12)
13. <https://www.w3.org/TR/tabular-metadata/#dfn-inherited-property> [↑](#footnote-ref-13)
14. <https://www.w3.org/TR/tabular-metadata/#built-in-datatypes> [↑](#footnote-ref-14)
15. <https://www.w3.org/TR/tabular-metadata/#derived-datatypes> [↑](#footnote-ref-15)
16. <http://schema.org/latitude> [↑](#footnote-ref-16)
17. <https://www.w3.org/TR/tabular-data-model/#locating-metadata> [↑](#footnote-ref-17)
18. <http://www.w3.org/TR/json-ld/#compact-iris> [↑](#footnote-ref-18)