

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

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

*Ceļvedis atvērto datu kopu struktūras izveidei un aprakstīšanai atbilstoši DPP standartam*

versija 1.1.

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

**Kontaktpersona:**

Krišjānis Jurģelis

SIA „Agile & Co”

Citadeles iela 12, Rīga, LV – 1010

Tālr.: +371 26321422

E-pasts: [krisjanis@agile.lv](mailto: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. | 0.2. | 01.09.2016 | Precizēts dokumenta nosaukums atbilstīgi komentāriem |
| 3. | 0.5. | 10.10.2016. | Sagatavots dokumenta pamatsaturs |
| 4. | 1.0. | 06.11.2016. | Pasūtītājam akceptēšanai iesniegtā versija ar precizējumiem atbilstoši komentāriem, kas no Pasūtītāja saņemti 26.10.2016. |
| 5. | 1.1. | 14.11.2016. | Dokumenta gala versija ar precizējumu atbilstoši komentāram, kas no Pasūtītāja saņemts 07.11.2016. |

**Satura rādītājs**

[1 Vispārēja informācija par dokumentu 5](#_Toc463874695)

[1.1 Dokumenta mērķis 5](#_Toc463874696)

[1.2 Dokumenta lietotāji 5](#_Toc463874697)

[1.3 Dokumenta sfēra 5](#_Toc463874698)

[1.4 Dokumentā izmantotie saīsinājumi 5](#_Toc463874699)

[1.5 Saistītie dokumenti 6](#_Toc463874700)

[1.6 Pieņēmumi un ierobežojumi 7](#_Toc463874701)

[2 Informācija dokumenta izmantotājam 8](#_Toc463874702)

[3 Ievads 9](#_Toc463874703)

[3.1 Kas ir tabulveida dati? 9](#_Toc463874704)

[3.2 Kādi ir ieteicamie tabulveida datu formāti? 10](#_Toc463874705)

[3.3 Vai TDS standartu var lietot datu kopām, kas nav CSV formātā? 10](#_Toc463874706)

[3.4 Kas ir TDS standarta metadatu dokuments? 11](#_Toc463874707)

[3.5 Kāda informācija ir jānorāda obligāti? 11](#_Toc463874708)

[3.6 Vai DPP lietotājiem datu kopas ir jāapraksta "ar roku"? 12](#_Toc463874709)

[3.7 Kā datu kopai pievienot tās metadatu dokumentu? 12](#_Toc463874710)

[4 Datu kopu struktūras aprakstīšana 14](#_Toc463874711)

[4.1 Kolonnas datu tipa norādīšana 14](#_Toc463874712)

[4.2 Kā norādīt loģiskās pazīmes (jā / nē) vērtības? 15](#_Toc463874713)

[4.3 Kā norādīt datuma formātu? 15](#_Toc463874714)

[4.4 Kā norādīt pieļaujamo vērtību intervālu? 16](#_Toc463874715)

[4.5 Kā norādīt, ka kolonna satur vairāku vērtību sarakstu? 17](#_Toc463874716)

[4.6 Kā norādīt lauka vērtību nozīmi? 17](#_Toc463874717)

[4.7 Kā norādīt tabulas ierakstu nozīmi? 18](#_Toc463874718)

[Kontaktinformācija un rekvizīti 19](#_Toc463874719)

# Vispārēja informācija par dokumentu

## Dokumenta mērķis

Dokumenta mērķis ir sniegt informāciju par mašīnlasāmu atvērto datu kopu datu struktūras izveidošanu un aprakstīšanu atbilstoši struktūras datu aprakstīšanas standartam.

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:

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

## Dokumenta sfēra

Dokumentā ir apskatīti šādi jautājumi:

* Labā prakse datu kopu struktūras definēšanai un aprakstīšanai;
* Datu kopu struktūras standarta pielietošana datu kopu struktūras aprakstīšanai;
* Piemēri kā atsevišķa sadaļa, pilnai kopai (datnes saturs).

## Dokumentā izmantotie saīsinājumi

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

*Tabula 1* **Dokumentā izmantotie saīsinājumi**

| **Saīsinājums** | **Nozīme** |
| --- | --- |
| CSVW grupa | W3C CSV on the Web darba grupa, kura ir izstrādājusi standartu tabulveida datu aprakstīšanai tīmeklī. |
| CSVW modeļa specifikācija | 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 vārdnīcas specifikācija | 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> |
| 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 | DPP tabulveida datu struktūras standarts, kurš ir definēts [6] |
| 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 metadatu standarts |
|  | Ceļvedis DDP metadatu aprakstīšanai |
|  | DPP mašīnlasāmo atvērto datu kopu datu struktūras standarts |
|  | Tehniskās vadlīnijas datu publicētājiem |
|  | Creative Commons CC0 (Public Domain) un Attribution 4.0 International atvērto licenču tulkojums latviešu valodā |

## 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 dokuments paredz, ka datu struktūras aprakstīšanai tiek izmantots TDS standartā aprakstītais metadatu formāts, kura pamatā ir W3C standarts tabulveida datu aprakstīšanai.

# Informācija dokumenta izmantotājam

Ceļveža ievada daļā ir apskatīti būtiskākie datu kopu struktūras un metadatu aprakstīšanas jautājumi – kas ir tabulveida dati, kāda veida datu kopas var aprakstīt TDS standartā, kāda informācija ir jānorāda obligāti un tas kā informācijas sistēmas (piem., DPP) var palīdzēt datu publicētājiem datu struktūras aprakstu izveidē.

Tālāk dokumentā ir apskatīti konkrētāki datu struktūras aprakstīšanas uzdevumi un parādīts kā ar TDS standarta līdzekļiem var norādīt kolonnas datu tipu, pieļaujamās vērtības un citus būtiskus atribūtus.

Dokumenta noslēgumā ir apskatīts, kā mašīnlasāmā formā pierakstīt informāciju par datu kopas tabulas un tās kolonnu nozīmi.

# Ievads

Šis ceļvedis satur informāciju, kas var noderēt datu publicētājiem un programmatūras izstrādātājiem DPP datu kopu struktūras aprakstīšanā un TDS standarta pielietošanā.

TDS metadatu formāta pamatā ir CSVW metadatu vārdnīca un JSON metadatu formāts, kurš ir definēts W3C specifikācijā **Metadata Vocabulary for Tabular Data[[1]](#footnote-1)** (tālāk tekstā "CSVW vārdnīca" vai "CSVW vārdnīcas specifikācija").

Aprakstot tabulveida datu kopas, to struktūras aprakstam ir jāatbilst CSVW metadatu vārdnīcas specifikācijai. TDS metadatu standarts definē precīzāku, DPP vajadzībām atbilstošu CSVW metadatu formāta lietojumu (piem., nosakot obligāti aprakstāmo informāciju).

## Kas ir tabulveida dati?

**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 un sastāv no tabulas kolonnām atbilstošajām šūnām. Visas šūnas, kas atrodas vienā kolonnā, apraksta viena un tā paša veida informācija.

Atbilstoši TDS standartam tabulas (piem., CSV datnes) pirmajā rindā ir jānorāda saraksts ar kolonnu nosaukumiem.

Piemērs:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **country** | **country group** | **name (en)** | **name (fr)** | **name (lv)** | **latitude** | **longitude** |
| at | eu | Austria | Autriche | Austrija | 47.6965545 | 13.34598005 |
| be | eu | Belgium | Belgique | Beļģija | 50.501045 | 4.47667405 |
| bg | eu | Bulgaria | Bulgarie | Bulgārija | 42.72567375 | 25.4823218 |

Šī tabula sastāv no 7 kolonnām un 4 rindām, kur pirmā rinda satur tabulas kolonnu nosaukumus. Apskatot tabulu, var secināt, ka pirmās divas kolonnas satur īsu tekstveida kodu, kam seko cilvēkiem lasāma informācija vairākās valodās. Pēdējās divas kolonnas satur daļskaitļus.

Tabulu saturu un struktūru ir nepieciešams dokumentēt, lai to varētu saprast visi potenciālie lietotāji. Kaut arī daļu informācijas lietotāji var nojaust no tabulas kolonnu nosaukumiem un pašiem datiem, tomēr šie secinājumi var būt nepareizi vai nepilnīgi.

Labā prakse ir pienācīgi dokumentēt publicējamo datu kopu saturu un struktūru. Šī dokumentācija ir jānodrošina gan cilvēkiem lasāmā formā (datu kopas apraksts), gan mašīnlasāmā un automātiski apstrādājamā formā.

Cilvēkam lasāmas dokumentācijas piemērs:

LR Uzņēmumu reģistra atvērtie dati: <http://dati.ur.gov.lv/>

## Kādi ir ieteicamie tabulveida datu formāti?

Tabulveida dati (tabulas) var būt pierakstīti un publicēti tīmeklī dažādos formātos: CSV un tam līdzīgos formātos, rēķintabulu formā (MS Excel, LibreOffice u.c.), kā tīmekļa lapas (HTML) un citos.

Lai pilnvērtīgi izmantotu TDS standarta datu kopu struktūras aprakstīšanai, datu kopām ir jābūt publicētām CSV teksta formātā vai tā variantos (dialektos).

TDS standarts **rekomendē** izmantot CSV datu formātu, kurš atbilst CSVW datu modeļa dokumentā aprakstītajām prasībām (skat. TDS standarta 4.3. nodaļu un CSVW modeļa specifikācijas nodaļu "Best Practice CSV"[[2]](#footnote-2)), no kurām galvenās prasības ir:

* datnes 1. rindai ir jāsatur kolonnu nosaukumi;
* kā kolonnu atdalītājsimbols ir jālieto komats (,);
* izmantot UTF-8 kodējumu.

Datu lauku vērtības, kuras satur atdalītājsimbolus (komatus), pēdiņas (") vai jaunas rindas simbolus, datnē ir jānorāda tās pilnībā iekļaujot pēdiņās ("vērtība"). Lauka vērtībās esošos pēdiņu simbolus ir jāaizvieto ar diviem pēdiņu simboliem ("").

Piezīme: par CSV formāta dialektu ir uzskatāms jebkāds tabulveida datu teksta formāts, kurā kolonnu vērtības ir atdalītas ar iepriekš norunātu atdalītājsimbolu (piem., TSV jeb ar tabulāciju atdalīts datu formāts). TDS standarts ļauj aprakstīt informāciju par jebkurā CSV dialektā noformētām datu kopām ar nosacījumu, ka tiek norādīta informācija par izmantoto CSV dialektu (skat. TDS standarta 4.3. nodaļu).

## Vai TDS standartu var lietot datu kopām, kas nav CSV formātā?

Gadījumā, ja tabulveida datu kopa nav izteikta CSV (vai tā dialektā), ieteicamā pieeja ir to pārveidot CSV formā un CSV datni aprakstīt atbilstoši TDS standartam.

Kaut arī CSVW datu modeli var izmantot lai aprakstītu jebkādas tabulveida datu kopas, tomēr, šādi rīkojoties, rodas papildus grūtības (piem., Excel datne var saturēt vairākas tabulas un ir nepieciešams veids kā ar tabulas URL atsaukties uz katru tabulu atsevišķi).

## Kas ir TDS standarta metadatu dokuments?

TDS standarta izpratnē metadatu dokuments ir JSON dokuments, kas satur datu kopas metadatus un struktūras aprakstu, un atbilst standartā iekļautajām prasībām.

Metadatu dokumentā iekļautā informācija palīdz datu lietotājiem saprast datu kopas nozīmi un struktūru, kas savukārt veicina datu kopas tālāku izmantošanu.

Katrs metadatu dokuments (atbilstoši TDS standartam) apraksta vienu tabulu.

Metadatu dokumenta piemērs (bez datu struktūras apraksta):

{

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

"url": "countries.csv",

"dc:title": "Valstu saraksts",

}

No CSVW specifikācijas viedokļa šis dokuments jau ir pilnvērtīgs CSVW metadatu dokuments, tomēr tas neatbilst TDS standartam, jo tajā nav iekļauta visa obligāti prasītā informācija.

## Kāda informācija ir jānorāda obligāti?

Informācija par obligāti norādāmajiem metadatiem ir atrodama TDS standarta 3. nodaļā (obligātās prasības) un 4. nodaļā (skat. atribūtu tabulas).

Piemēram izvēlēsimies nelielu CSV datni:

**"country","country group","name (en)","name (lv)"**

"at","eu","Austria","Austrija"

"be","eu","Belgium","Beļģija"

"bg","eu","Bulgaria","Bulgārija"

Tai atbilstošs metadatu dokuments, kurš satur minimālo TDS standartā obligāti prasīto informāciju:

{

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

"url": "valstis.csv",

"dc:title": [{

"@lang": "lv", "@value": "Valstu saraksts"

}, {

"@lang": "en", "@value": "List of countries"

}],

"dc:description": [{

"@lang": "lv", "@value": "Saraksts ar valstu kodiem un nosaukumiem ..."

}, {

"@lang": "en", "@value": "A list of country codes and names ..."

}],

"tableSchema": {

"columns": [{

"name": "country",

"titles": "Valsts kods",

"dc:description": "Valsts kods atbilstoši ISO 3166-1 alpha-2 kodifikatoram"

}, {

"name": "country\_group",

"titles": "Valstu grupa",

"dc:description": "Valstu grupa, kam pieder šī valsts"

}, {

"name": "name (en)",

"titles": "Nosaukums (angļu val.)",

"dc:description": "Valsts nosaukums angļu valodā"

}, {

"name": "name (lv)",

"titles": "Nosaukums (latviešu val.)",

"dc:description": "Valsts nosaukums latviešu valodā"

}]

}

}

## Vai DPP lietotājiem datu kopas ir jāapraksta "ar roku"?

*Piezīme: precīzu atbildi uz šo jautājumu būs iespējams dot pēc DPP izveides.*

Nē – metadatu dokumentu izveide būtu pēc iespējas jāautomatizē tā, lai lietotāji, kuri to vēlas, datu kopu struktūras aprakstus varētu izveidot automātiski, ar DPP atbalstu.

Tādā gadījumā metadatu dokumentu veidošana ārpuse DPP sistēmas būtu nepieciešama vienīgi gadījumos, ja lietotāji grib izteikt par datu kopu un tās struktūru sīkāku papildus informāciju, kuru DPP sistēma nepievieno automātiski.

Viena iespēja kā varētu izpausties šī automatizācija ir, ka DPP pēc tabulveida datu kopas pievienošanas datu publicētājam piedāvā dialogu ar tabulas pamatdatiem (informāciju, ko ir iespējams automātiski iegūt no datnes vai DPP informācijas par to), kurā ir paredzēti ievades lauki informācijai, kas vēl ir nepieciešama no datu publicētāja (piem., kolonnu nosaukumi un apraksti).

## Kā datu kopai pievienot tās metadatu dokumentu?

Lai datu struktūras apraksts būtu praktiski lietojams, tam ir jābūt atrodamam, tādēļ ir nepieciešams veids kā lietotājiem "pateikt" kurš metadatu dokuments attiecas uz doto datu kopu. Šis jautājums ir apskatīts TDS standarta 5. nodaļā (*Metadatu dokumentu publicēšanas principi*).

Vienkāršākais risinājums ir turēt abus dokumentus vienkopus un publicēt tos kopā. Tādā gadījumā var izmantot CSVW modeļa specifikācijā aprakstīto likumu par to, ka pēc noklusējuma metadatu dokuments tiek meklēts URL adresē, kuru iegūst no datnes URL tam galā pievienojot "-metadata.json".

Piemērs metadatu dokumenta datnes URL veidošanā:

Datne: http://dati.ur.gov.lv/register/register.csv

Metadatu dokuments: http://dati.ur.gov.lv/register/register.csv**-metadata.json**

Vajadzības gadījumā šo noklusēto metadatu dokumentu URL veidošanas likumu var modificēt.

Cita alternatīva ir CSV datnei atbilstošo metadatu dokumentu norādīt atbildē uz datnes tīmekļa pieprasījumu – HTTP atbildes sākuma daļā ("Link" laukā). Tas ir tehniski sarežģītāks risinājums, tomēr tas var noderēt DPP un tam līdzīgu sistēmu izveidē.

Sīkāka informācija par šo un citiem veidiem kā publicēt metadatu dokumentus tā, lai saglabātu to saistību ar tajos aprakstīto datni, ir atrodama CSVW modeļa specifikācijas nodaļā "Locating Metadata".

# Datu kopu struktūras aprakstīšana

Dokumentējot datu kopu struktūru, ir vēlams norādīt pēc iespējas precīzāku informāciju par tās kolonnu saturu un nozīmi.

Papildus cilvēkam lasāmai informācijai par kolonnu (virsraksts, apraksts) ir vērtīgi norādīt *mašīnlasāmu* informāciju, kas apraksta kolonnas datu tipu, formātu un nozīmi. Šo informāciju ir iespējams izmantot, lai:

* validētu datnes saturu atbilstoši norādītajam kolonnu datu tipam un formātam;
* datu ievades formās – lai pareizi attēlotu datus un veiktu formāta pārbaudi;
* pareizi interpretētu un atkalizmantotu šīs datu kopas.

TDS standarta 4.5.1 nodaļā (*Datu shēmas apraksts*) ir atrodama informācija par galvenajiem laukiem, ko ir ieteicams norādīt datu kopu struktūras aprakstā.

## Kolonnas datu tipa norādīšana

Kolonnas datu tipu norāda ar atribūtu "datatype". Tā vērtības var būt CSVW metadatu specifikācijā definētie pamattipi vai no tiem atvasināti datu tipi.

Plaši izmantoti datu tipi ir:

* string (teksts);
* boolean (loģiskā vērtība - patiess vai aplams; jā vai nē);
* integer (vesels skaitlis);
* number (skaitlis - atbilst xsd:double);
* date (datums);
* datetime (datums un laiks);
* time (laiks);
* html (HTML satura fragments);
* xml (XML satura fragments);
* json (vērtība JSON formātā).

Piezīme: kolonnām, kuru tips ir teksts, datu tipu var nenorādīt, jo teksts ("string") ir noklusētais datu tips.

Daļai datu tipu (boolean, date, u.c.) ir savs noklusētais formāts, kādā ir jāpieraksta informācija. Norādot kolonnas datu tipu, ir jāpārliecinās, ka tās formāts atbilst dotā datu tipa formātam. Gadījumā, ja kolonna tiek lietots citāds formāts, tas ir īpaši jānorāda (ar atvasināta datu tipa un tā "format" atribūta palīdzību).

Pilns pamattipu saraksts ir atrodams CSVW modeļa specifikācijas 4.6. nodaļā[[3]](#footnote-3).

## Kā norādīt loģiskās pazīmes (jā / nē) vērtības?

Ar loģiskās pazīmes (*boolean*) datu tipu apzīmē kolonnas, kurām var būt tikai divas vērtības – patiess (kolonnas nosacījums izpildās) vai aplams (kolonnas nosacījums neizpildās).

Piemēram, datu kopā var būt pazīme, kas norāda vai uzņēmums ir aktīvs:

{

"name": "is\_active",

"titles": "Uzņēmums ir aktīvs",

"dc:description": "Pazīme, kas norāda vai uzņēmums ir aktīvs (1=jā, 0=nē).",

**"datatype": "boolean"**

}

Loģiskās pazīmes tipa noklusētās vērtības ir:

* patiess: 1, *true*
* aplams: 0, *false*

Gadījumā, ja kolonnā šīs vērtības tiek norādītas savādāk (piem., "J" un "N"), tad tas ir jānorāda kolonnas metadatos izmantojot lauku "format"[[4]](#footnote-4):

{

"name": "is\_active",

"titles": "Uzņēmums ir aktīvs",

"dc:description": "Pazīme, kas norāda vai uzņēmums ir aktīvs (J=jā, N=nē).",

"datatype": {

**"base": "boolean",**

**"format": "J|N"**

}

}

Šādi tika izveidots jauns, no "boolean" atvasināts datu tips, kuram atbilstošās loģiskās vērtības ir "J" (patiess) un "N" (aplams).

## Kā norādīt datuma formātu?

Standarta formāts, kuru ir ieteicams izmantot datuma un laika norādīšanai mašīnlasāmā formā, ir ISO 8601[[5]](#footnote-5):

* Datums: 2016-11-18;
* Laiks: 12:05:37;
* Datums un laiks: 2016-11-18T12:05:37;

Tas ir noklusētais formāts datuma un laika datu tipiem TDS standartā. Papildus ieguvums no šo formātu lietošanas ir tas, ka, kārtojot datuma laukus kā teksta rindas, tiek saglabāta vērtību hronoloģiskā secība.

Ja kolonnas vērtībām tiek izmantots cits formāts, tas ir jānorāda ar atribūta "format" palīdzību:

{

"name": "date",

"titles": "Datums",

"dc:description": "Publicēšanas datums (dd/mm/gggg).",

"datatype": {

**"base": "date",**

**"format": "dd/MM/yyyy"**

}

}

Savietojamības nolūkā ir jāizmanto tikai tādus datuma un laika formātus, kuri ir norādīti CSVW modeļa specifikācijā[[6]](#footnote-6). Gadījumā, ja datu kopā ir izmantots nesavietojams datu formāts, tad ir ieteicams datu kopu pārveidot tā, lai izmantotu standarta formātu.

## Kā norādīt pieļaujamo vērtību intervālu?

Datu pārbaudes vajadzībām var norādīt kolonnas pieļaujamās vērtības:

"datatype": {

"base": "integer",

**"minimum": "1",**

**"maximum": "5"**

}

Arī datu tipu definīcijas var dokumentēt norādot to nosaukumu un aprakstu (tie pieder pie *vispārējiem* atribūtiem, kurus var lietot ar jebkuru metadatu dokumentam objektu):

"datatype": {

**"dc:title": "Novērtējums",**

**"dc:description": "Novērtējums 5 baļļu skalā (no 1 līdz 5)."**

"base": "integer",

"minimum": "1",

"maximum": "5"

}

Pieļaujamo vērtību intervālu var norādīt arī datuma un laika kolonnām. Piemēram, šādi norāda, ka datumam ir jābūt ne mazākam par 2000. gada 1. janvāri:

"datatype": {

"base": "date",

**"minimum": "2000-01-01"**

}

Datuma un laika intervālu norādīšanai vienmēr tiek izmantots ISO 8601 formāts.

## Kā norādīt, ka kolonna satur vairāku vērtību sarakstu?

Dati var saturēt kolonnas, kuras satur sarakstu ar vērtībām (piem., ar ";" atdalīts atslēgvārdu saraksts). Tālākai datu apstrādei var būt noderīgi zināt, ka kolonna satur sarakstu un to kāds atdalītājsimbols tajā tiek lietots.

Šo informāciju var norādīt ar atribūta "separator" palīdzību:

{

"name": "keywords",

"title": "atslēgvārdi"

**"separator": ";",**

}

Gadījumā, ja ir būtiska secība kādā vērtības atrodas sarakstā, to var norādīt atsevišķi:

{

"name": "latlong",

"dc:description": "koordinātes - piem.: 50.501045 4.47667405",

**"separator": " ",**

**"ordered": true,**

}

## Kā norādīt lauka vērtību nozīmi?

Lai varētu pilnvērtīgi izmantot datu kopu, ir jāsaprot tās nozīme (semantika) – kas ir aprakstīts datu kopā un katrā tās laukā. Šī informācija ļauj datus izmantot citās sistēmas un lietojumos. Piemēram, ja ir zināms, ka tabula ir par grāmatām un satur lasītāju vērtējumus par tām, tad šo tabulu ir iespējams izmantot grāmatu portālā vai meklēšanas rīkā.

Lai to varētu darīt automātiski (piem., lai sistēma automātiski saprastu, ka kolonnā ir grāmatas ISBN kods), nepietiek ar cilvēkam saprotamu aprakstu. Ir nepieciešams veids kā *mašīnlasāmā* formā "pateikt", ka viena kolonna satur ISBN kodu, bet cita – grāmatas novērtējumu.

To var izdarīt izmantojot metadatu vārdnīcas – standarta kodifikatorus, kuros ir definēti identifikatori lietām un to īpašībām, par kuriem ir nepieciešams kaut ko "pateikt". Parasti tam izmanto URI identifikatorus.

Lai aprakstītu datu nozīmi ir jāatrod (vai jāizveido) metadatu vārdnīca ar vajadzīgajiem jēdzieniem. Piemēram, schema.org ļauj aprakstīt informāciju par grāmatām[[7]](#footnote-7):

{

"name": "isbn",

**"propertyUrl": "schema:isbn"**

}

## Kā norādīt tabulas ierakstu nozīmi?

Lai aprakstītu tabulas nozīmi, ir būtiski pateikt kāda veida objekti ir aprakstīti tabulā. Tas nozīmē, ka *mašīnlasāmā* formā ir "jāpasaka", ka tabulas rindas satur, piemēram, informāciju par grāmatām.

Tātad, ir nepieciešams norādīt rindās aprakstīto objektu tipu. Līdzīgi kā kolonnu nozīmes aprakstīšanā (iepriekšējā nodaļā), to dara, norādot uz atbilstošo metadatu vārdnīcas jēdzienu – šajā gadījumā grāmatu.

Tā, kā tabulā parasti nav kolonna, kurā jau būtu norādīts atbilstošā jēdziena URI, to var izdarīt ar CSVW datu pārveidojumu līdzekļiem, izveidojot jaunu "virtuālo" kolonnu un iekļaujot to tabulas shēmas aprakstā:

{

"virtual": true,

"propertyUrl": "rdf:type",

"valueUrl": "schema:Book"

}

Virtuālās kolonnas vērtība ir izvēlētā jēdziena URI identifikators (schema:Book). Šo informāciju varēs izmantot automātiski izmantot citi, piemēram, izmantojot CSVW standarta procedūras tabulveida datu pārveidošanai JSON vai RDF formā.

Plašāks datu nozīmes tēmas apskats ir ārpus šī dokumenta uzdevumiem un ir atrodams:

* [Linked Data: Evolving the Web into a Global Data Space](http://linkeddatabook.com/editions/1.0/)
* [RDF 1.1 Primer](https://www.w3.org/TR/rdf11-primer/)

# 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](mailto: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. <http://www.w3.org/TR/tabular-metadata/> [↑](#footnote-ref-1)
2. <https://www.w3.org/TR/tabular-data-model/#syntax> [↑](#footnote-ref-2)
3. <https://www.w3.org/TR/tabular-data-model/#datatypes> [↑](#footnote-ref-3)
4. <https://www.w3.org/TR/tabular-data-model/#formats-for-booleans> [↑](#footnote-ref-4)
5. <https://en.wikipedia.org/wiki/ISO_8601> [↑](#footnote-ref-5)
6. <https://www.w3.org/TR/tabular-data-model/#formats-for-dates-and-times> [↑](#footnote-ref-6)
7. <http://schema.org/Book> [↑](#footnote-ref-7)