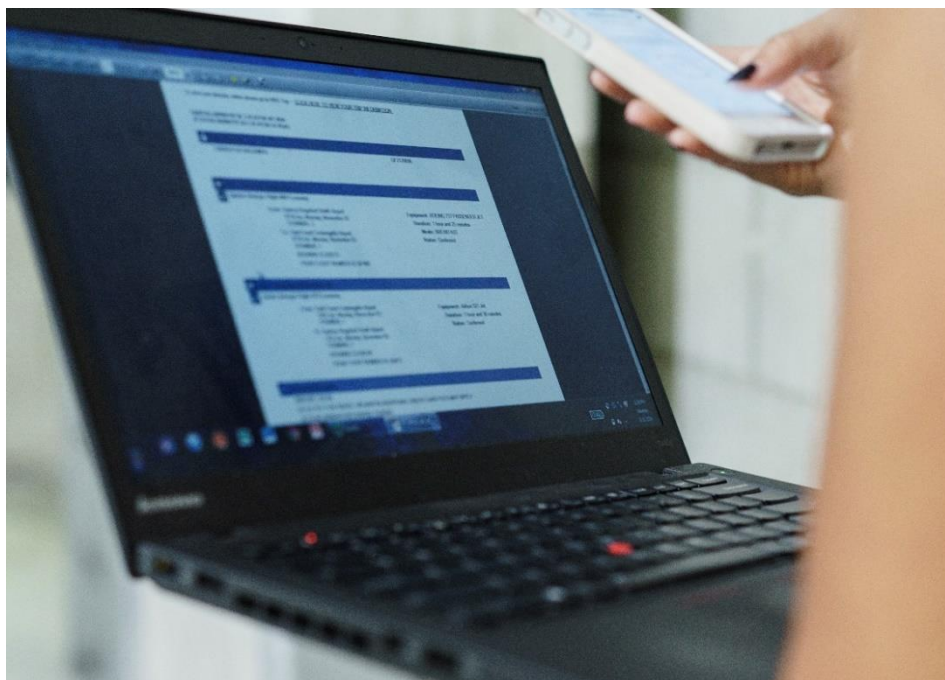


Informacinės visuomenės plėtros komitetas prie Susisiekimo ministerijos

Lietuvos atvirų duomenų formavimo
metodologinių ir teisinio reglamentavimo
priemonių įgyvendinimui bei tam reikalingų
valstybės institucijų darbuotojų kompetencijų
ugdymo paslaugos

2018 m. vasario 9 d.
Vilnius

Valstybės informacinių išteklių duomenų
metaduomenų standartinės aprašo struktūros ir jos
taikymo aprašas (R3) ir organizacinių bei techninių
reikalavimų dėl metaduomenų
perdavimo/apsikeitimo (integracijos) į Europos
atvirų duomenų portalą aprašas (R4)



Dokumento pakeitimai

Lentelė Nr. 1. Dokumento pakeitimų valdymas

Data	Aprašymas	Redaguotos dalys
2018-01-17	Pateiktas dokumento projektas derinimui	-
2018-01-24	Pateiktas papildytas dokumentas	Papildytos visos dalys
2018-02-01	Pateiktas papildytas dokumentas	Papildytos viso dalys

Informacinės visuomenės plėtros komitetui
prie Susisiekimo ministerijos
Gedimino pr. 7,
LT-01103 Vilnius

2018 m. sausio 17 d.

Projekto „Lietuvos atvirų duomenų formavimo metodologinių ir teisinio reglamentavimo priemonių įgyvendinimui bei tam reikalingų valstybės institucijų darbuotojų kompetencijų ugdymo paslaugos“ Valstybės informacinių išteklių duomenų metaduomenų standartinės aprašo struktūros ir jos taikymo aprašas ir organizacinių bei techninių reikalavimų dėl metaduomenų perdavimo/apsikeitimo (integracijos) į Europos atvirų duomenų portalą aprašas

Gerbiamieji,

Šią ataskaitą parengė UAB „PricewaterhouseCoopers“ (toliau – Paslaugų teikėjas arba PwC), remiantis 2017 m. gruodžio 11 d. sudaryta pirkimo sutartimi Nr. 6F-52 (toliau – Sutartis), sudarytą su Informacinės visuomenės plėtros komitetu prie Susisiekimo ministerijos (toliau – Perkančioji organizacija, IVPK).

Darbo apimtis

Valstybės informacinių išteklių duomenų metaduomenų standartinės aprašo struktūros ir jos taikymo aprašo (R3) ir organizacinių bei techninių reikalavimų dėl metaduomenų perdavimo/apsikeitimo (integracijos) į Europos atvirų duomenų portalą aprašo (R4) (toliau – Ataskaita) apimtis buvo apibrėžta pirkimo „Lietuvos atvirų duomenų formavimo metodologinių ir teisinio reglamentavimo priemonių įgyvendinimui bei tam reikalingų valstybės institucijų darbuotojų kompetencijų ugdymo paslaugos“ techninėje specifikacijoje (toliau – Techninė specifikacija). Ataskaitoje pateikiama informacija yra susijusi su Valstybės informacinių išteklių duomenų metaduomenų standartinė struktūra, jos taikymu, organizaciniais bei techniniais reikalavimais dėl metaduomenų perdavimo/apsikeitimo (integracijos) į bei kitomis sritimis, susijusiomis su Lietuvos atvirų duomenų formavimo metodologinių ir teisinio reglamentavimo priemonių įgyvendinimui bei tam reikalingų valstybės institucijų darbuotojų kompetencijų ugdymo paslaugomis (toliau – Projektas, Paslaugos) ir Techninėje specifikacijoje nurodytais reikalavimais.

Darbo metodika

Ataskaita parengta remiantis Techninėje specifikacijoje nurodytais dokumentais bei reikalavimais, PwC pateiktu techniniu pasiūlymu ir susitikimų metu gauta medžiaga. PwC nesiėmė priemonių šaltinių patikimumui nustatyti ir papildomai netikrino pateiktos informacijos. Todėl PwC nesuteikia tiesioginių ar numanomų garantijų kitiems asmenims (išskyrus IVPK pagal Sutartį) dėl Ataskaitos tikslumo ar išsamumo.

Darbus pagal Sutartį atlikome iki 2018 m. sausio 17 d. Ataskaitoje neaptariami vėlesnių įvykių ar aplinkybių padariniai, taip pat vėliau paaiškėjusi informacija. Negalime nustatyti, kokią įtaką šie darbai arba tyrimai būtų turėję Ataskaitos parengimui.

Atkreipiame Jūsų dėmesį į Ataskaitoje išdėstytas svarbias mūsų pateikiamas prielaidas. Paslaugų teikėjas neatsako kitiems asmenims (išskyrus IVPK pagal Sutartį) už Ataskaitos rengimą. Be to, Paslaugų teikėjas neprisiima sutartinės, deliktinės ir kitokios atsakomybės (neatsižvelgiant į ieškinio formą) taikytinos teisės leistina apimtimi ir neprisiima atsakomybės už kitų asmenų patirtus padarinius (išskyrus IVPK Sutarties pagrindu) ar kitus pagal šią Ataskaitą priimtus sprendimus ar atsisakymą juos priimti.

Dokumento platinimas ir atsakomybė

Asmuo, susipažinęs ir perskaitęs šią Ataskaitą, sutinka ir įsipareigoja laikytis šių sąlygų:

- Ataskaitos skaitytojas supranta, kad Ataskaita pagrįsta ekspertiniu vertinimu;
- Ataskaitos skaitytojas supranta, kad Paslaugų teikėjo darbai yra atlikti pagal mūsų kliento – IVPK – nurodymus ir skirti tik mūsų klientui;
- Ataskaitos skaitytojas pripažįsta, kad Ataskaita parengta mūsų kliento – IVPK – nurodymu, todėl į ją gali būti įtraukta ne viskas, kas svarbu skaitytojui;
- Ataskaitos skaitytojas sutinka, kad PwC, jos partneriai, įgaliotojai, darbuotojai ir atstovai neprisiima sutartinės ar deliktinės atsakomybės (įskaitant aplaidumą ir teisės aktuose nustatytų įsipareigojimų pažeidimą ir kt.) ir neatsako už žalą, išlaidas ar nuostolius, patirtus skaitytojui vadovaujantis šiuo dokumentu ar su juo susipažinus.

Bendroji informacija

Jeigu kiltų klausimų dėl šios Ataskaitos turinio, prašome kreiptis į Justą Urboną telefonu +37065511056 arba el. paštu justas.urbonas@pwc.com.

Pagarbiai

Audrius Leipus
UAB „PricewaterhouseCoopers“
Konsultacijų skyriaus vyr. projektų vadovas

Turinys

Sąvokos ir sutrumpinimai	8
1 Metaduomenų aprašymui taikomų standartų apžvalga	9
1.1 Metaduomenų pristatymas	9
1.1.1 Atvirų duomenų metaduomenys	9
1.1.2 Metaduomenų standartų problematika	9
1.1.3 Svarbiausios institucijos ir iniciatyvos Europoje, susijusios su atvirų duomenų metaduomenų standartais	10
1.2 Atvirų duomenų metaduomenų standartai	10
1.3 Rekomenduojami standartiniai formatai publikavimui	12
1.4 Europos rekomenduojamas standartas	12
2 ES atvirų duomenų portalo ir kitose ES šalyse veikiančių atvirų duomenų portalų metaduomenų aprašymo praktika	14
2.1 Atvirų duomenų portalo metaduomenų praktika	14
2.1.1 ES mastu naudojami atvirų duomenų portalai	14
2.1.2 Tarptautinė atvirų duomenų metaduomenų praktika	14
2.1.3 Susietų atvirų duomenų skelbimas	16
2.1.4 Metaduomenų paruošimo ir nustatymo techniniai standartai	17
2.2 Nacionalinių duomenų portalų gerieji pavyzdžiai naudojant metaduomenis	18
2.2.1 Nacionaliniai duomenų portalai Europoje	18
2.2.2 Metaduomenų aprašymas nacionaliniuose atvirų duomenų portaluose Europoje	18
3 Valstybės informacinių išteklių duomenų metaduomenų standartinio aprašo struktūra	22
3.1 Lietuvai rekomenduojami naudoti metaduomenų standartai	22
3.2 DCAT-AP žodynas	23
3.2.1 DCAT-AP žodyno pristatymas	23
3.2.2 DCAT-AP plėtiniai	24
3.2.3 DCAT standarto duomenų modelis	26
3.3 Siūloma metaduomenų struktūra Lietuvoje	27
3.3.1 Suderinimas su DCAT ir pritaikymas	27
3.3.2 Bendri laukelių užpildymo reikalavimai	27
3.3.3 DCAT-AP v1.1 standarto pakeitimai Lietuvoje	28
3.3.4 Atvirų duomenų metaduomenų hierarchinė struktūra	30
3.3.5 Pagrindinės techninės klasės	30
3.3.6 Kontroliuojami žodynai	32
3.4 Metaduomenų struktūros aprašymo pavyzdžiai	35
3.4.1 DCAT pagrindinių metaduomenų aprašymo pavyzdys	35
3.4.2 RDF formatas	36
3.4.3 JSON formatas	36

4	Standartinės metaduomenų struktūros taikymo atvirų duomenų metaduomenų formavimui procedūros	37
4.1	Siūlomos atvirų duomenų metaduomenų valdymo procedūros Lietuvoje	37
4.1.1	Duomenų rinkimas.....	37
4.1.2	Duomenų ir metaduomenų paruošimas	39
4.1.3	Metaduomenų skelbimas	40
4.1.4	Metaduomenų surinkimas (angl. Harvesting metadata)	40
4.1.5	Metaduomenų priežiūra.....	41
4.2	DCAT standarto pritaikymas.....	41
5	Reikalavimai metaduomenų tvarkymui ir perdavimui į atvirų duomenų portalą	43
5.1	Atvirų duomenų portalų pristatymas	43
5.1.1	Funkciniai reikalavimai.....	43
5.1.2	Duomenų pateikimo funkcijos.....	45
5.1.3	Techniniai reikalavimai.....	46
5.2	Atvirų duomenų platformos	46
6	Reikalavimai metaduomenų perdavimui / apsikeitimui su Europos atvirų duomenų portalu.....	48
6.1	Organizaciniai reikalavimai.....	48
6.1.1	Atvirų duomenų portalų ruošimas	48
6.1.2	Duomenų surinkimo prašymo registravimas	49
6.1.3	Duomenų surinkimo metu	49
6.2	Techniniai atvirų duomenų skelbimo Europos atvirų duomenų portale reikalavimai.....	50
6.2.1	Bendrieji techniniai reikalavimai	50
6.2.2	Techniniai CKAN reikalavimai	52
	Priedas Nr. 1. Klasių savybių detalizavimas	54
	Catalogue klasė	54
	Catalogue Record klasė	55
	Dataset klasė	57
	Distribution klasė	60
	Agent klasė.....	62
	Category Scheme klasė	62
	Category klasė.....	63
	Checksum klasė	63
	Identifier klasė.....	63
	Licence Document klasė.....	64
	Period of Time klasė	64
	Priedas Nr. 2. RDF failo metaduomenų struktūros aprašymas.....	65
	Priedas Nr. 3. JSON failo metaduomenų struktūros aprašymas.....	67
	Priedas Nr. 4. Europos atvirų duomenų portalui tinkamo RDF failo pavyzdys	68
	Priedas Nr. 5. Kontrolinis sąrašas dėl Europos atvirų duomenų portalų duomenų surinkimo.....	70

Sąvokos ir sutrumpinimai

Sąvoka	Paaškinimas
Ataskaita	Valstybės informacinių išteklių duomenų metaduomenų standartinės aprašo struktūros ir jos taikymo aprašas ir organizacinių bei techninių reikalavimų dėl metaduomenų perdavimo/apsikeitimo (integracijos) į Europos atvirų duomenų portalą aprašas
IT	Informacinės technologijos
IVPK, Perkančioji organizacija	Informacinės visuomenės plėtros komitetas prie Susisiekimo ministerijos
Projektas, Paslaugos	Lietuvos atvirų duomenų formavimo metodologinių ir teisinio reglamentavimo priemonių įgyvendinimui bei tam reikalingų valstybės institucijų darbuotojų kompetencijų ugdymo paslaugos
PwC, Paslaugų teikėjas	UAB „PricewaterhouseCoopers“
Sutartis	Informacinės visuomenės plėtros komiteto prie Susisiekimo ministerijos ir UAB „PricewaterhouseCoopers“ 2017 m. gruodžio 11 d. sudaryta pirkimo sutartis Nr. 6F-52
Techninė specifikacija	Lietuvos atvirų duomenų formavimo metodologinių ir teisinio reglamentavimo priemonių įgyvendinimui bei tam reikalingų valstybės institucijų darbuotojų kompetencijų ugdymo paslaugų techninė specifikacija

1 Metaduomenų aprašymui taikomų standartų apžvalga

1.1 Metaduomenų pristatymas

1.1.1 Atvirų duomenų metaduomenys

Metaduomenys¹ – struktūriškai apibrėžta informacija, kuri apibūdina, paaiškina dokumentą ar informacinį išteklių. Jais pateikiama informacija, kuri padeda vartotojams geriau suprasti duomenų reikšmę, struktūrą ir nurodo tokią informaciją, kaip naudojimo teisės ir licencijos sąlygos, duomenis sukūrusi organizacija, duomenų kokybė, duomenų prieigos būdai, duomenų rinkinių atnaujinimo tvarkaraščiai ir kt.

Metaduomenys palengvina tokias užduotis, kaip duomenų rinkinio atradimas ar pakartotinis naudojimas, ir gali būti skirstomi pagal skirtingus išsamumo lygius.

Metaduomenys gali būti kelių tipų, kurie klasifikuojami į skirtingas taksonomijas. Pavyzdžiui, specifinė taksonomija apibrėžia tris metaduomenų tipus pagal jų aprašomasias, struktūrines ir administracines savybes. Aprašomieji metaduomenys skirti atpažinti duomenų rinkinius, struktūriniai metaduomenys skirti suprasti duomenų rinkinio formatą, kuriuo yra publikuojamas, ir administraciniai metaduomenys pateikia informaciją apie versiją, atnaujinimo tvarkaraštį, ir kt. Kita taksonomija gali apibrėžti metaduomenų tipus pagal uždavinius, kuriems jie yra naudojami, pavyzdžiui, atradimui ir pakartotiniam naudojimui.

Duomenų naudotojai turi galėti:

- rasti duomenis;
- suprasti duomenų pobūdį ir struktūrą, pavyzdžiui, ką duomenys apibūdina ir kaip tai daro;
- sužinoti apie duomenų šaltinį ir kokios jų naudojimo sąlygos.

Metaduomenys turi būti prieinami ne tik duomenų naudotojams, bet ir kompiuterinėms programoms, kad jos galėtų automatiškai apdoroti metaduomenis susietus su tam tikru duomenų rinkiniu. Taip pat svarbu, kad būtų naudojamos standartinės sąlygos, apibūdinančios metaduomenis, tokiu būdu palengvinant automatinį metaduomenų apdorojimą ir pagerinant duomenų sąveiką.

1.1.2 Metaduomenų standartų problematika

Metaduomenų standartai yra naudojami atvirų duomenų metaduomenų apibūdinimui ir taip palengvina duomenų sąveiką, leidžia suprasti kas ir kaip yra apibūdinama, duomenys labiau struktūrizuoti, lengviau randami, pasiekiami ir automatiškai apdorojami.

Atvirų duomenų standartų skaičius didėja, dalis jų susiję su bendromis sąvokomis, kaip žmonės, adresai ir organizacijos, kiti standartai skirti specifiniams domenams, kaip transportas, sveikata ar mokslas, dar kiti standartai naudojami tam tikrose šalyse ar regionuose.

Atvirų duomenų standartai² yra kuriami skirtingais būdais. Juos gali sukurti paprastos bendruomenės, kurioms svarbi duomenų integracija ir bendradarbiavimas kuriant ir tobulinant standartus, pavyzdžiui, *OpenSpending* ir *Open311* standartai. Kitas standarto nustatymo būdas yra formalus standartizacijos institucijos procesas, pavyzdžiui, W3C, ISO arba IETF. Standartizacijos įstaigos turi pranašumą, nes gali pateikti tvarią standartų nustatymo ir palaikymo sistemą. Tačiau piliečiams ar visuomeninėms organizacijoms gali kilti tam tikrų kliūčių prisijungti. Standartizacijos įstaigos neturi lankstumo ar specializacijos, skirtos spręsti naujus ar besikeičiančius skirtingų bendruomenių poreikius. Dažnai kuriant standartą, naudojamas abiejų metodų derinys.

Nors duomenų standartų yra daug, nėra vieno paprasto paieškos mechanizmo jiems surasti. Standartizacijos institucijos skelbia informaciją apie jų sukurtus standartus jų tinklapiuose ar

¹ Šaltinis: <https://www.w3.org/TR/2015/WD-dwbp-20150224/#ProvideMetadataMachine>

² Šaltinis: <https://www.linkedin.com/pulse/open-data-standards-steven-adler>

kataloguose. Taip pat yra kuriami atskiri tinklapiai skirti vienam standartui, pavyzdžiui, *OpenRefferal* ar *OpenContracting Data Standard*³.

Kitas sunkumas yra tai, kaip išsirinkti, kuris atvirų duomenų standartas yra tinkamas. Tai gali priklausyti nuo domeno, standarto naudotojų, išsamumo lygio, kuriam standartas yra skirtas, ar standartas yra aktyviai palaikomas ir kt. Todėl svarbu, kad šie metaduomenys būtų prieinami kartu su atvirų duomenų standartu, ir jie būtų bendrai apibūdinami visuose standartuose, kad duomenys būtų lengvai randami ir palyginami.

Dėl šių daugybės priežasčių, Europos Sąjunga iniciavo ISA programą⁴, kuri teikia bendrą atvirų duomenų standartą ir gerąsias praktikas.

1.1.3 Svarbiausios institucijos ir iniciatyvos Europoje, susijusios su atvirų duomenų metaduomenų standartais

Šioje dalyje apžvelgiamos svarbiausios Europos institucijos ir iniciatyvos, darančios įtaką atvirų duomenų metaduomenų standartų formavimui. Keli pavyzdžiai:

- **Europos Komisijos Informatikos generalinis direktoratas (DG DIGIT), Ryšių tinklų, turinio ir technologijų generalinis direktoratas (DG CONNECT), Biudžeto generalinis direktoratas (DG BUDGET)** – tai yra Europos Komisijos padaliniai, kurie atsako už skaitmeninių paslaugų teikimą, ES bendrosios skaitmeninės rinkos, interneto saugos, skaitmeninio mokslo ir inovacijų politiką ir biudžetą.
- Tarptautinės asociacijos:
 - **Atvirų duomenų institutas** (*Open data institute*) – nepelno siekianti organizacija, įkurta Jungtinėje Karalystėje. Jos tikslas sujungti, parengti ir įkvėpti žmones iš viso pasaulio kurti inovacijas, susijusias su duomenų panaudojimu.
 - **Atvirų duomenų barometras** (*Open data barometer*) – Pasaulinio žiniatinklio fondo projektas, kurio tikslas parodyti visuotinių atvirųjų duomenų iniciatyvų paplitimą ir poveikį visame pasaulyje. Jis vertina, kaip valstybės skelbia ir naudoja atvirus duomenis atsakomybei, inovacijoms ir socialiniam poveikiui didinti.
 - **Pasaulinio Žiniatinklio Konsorciumas (W3C)** – tarptautinė bendruomenė, kurios narės organizacijos, darbuotojai ir visuomenė, glaudžiai dirba kartu kuriant Interneto standartus;
- Pagrindinės iniciatyvos:
 - **Viešojo sektoriaus informacijos dalinimosi iniciatyva „Share-PSI“** – tinklas, kuriame kartu dirbo valstybės skyriai atsakingi už PSI direktyvos įgyvendinimą visoje Europoje su standartų institucijomis, akademinėmis institucijomis, komercinėmis organizacijomis, prekybos organizacijomis ir interesų grupėmis;
 - **ISA² Programa** – remia skaitmeninių sprendimų kūrimą, kurie leidžia Europos viešoms administracijoms, verslui ir piliečiams pasinaudoti sąveikiomis tarpvalstybinėmis ir tarp-sektorinėmis viešosiomis paslaugomis;
 - **INSPIRE direktyva** – siekia sukurti ES erdvinį duomenų infrastruktūrą, naudojamą aplinkos politikoje ar veiklose, kurios gali turėti įtaką aplinkai. Ši infrastruktūra leidžia viešojo sektoriaus organizacijoms dalintis aplinkos erdvine informacija, palengvinti viešą prieigą prie erdvinės informacijos visoje Europoje ir prisidėti formuojant politiką tarpvalstybiniu lygiu.

1.2 Atvirų duomenų metaduomenų standartai

Pasaulinio Žiniatinklio Konsorciumas (W3C) rekomenduoja duomenų ir metaduomenų kodavimui naudoti terminus iš standartinių žodynų. Tai padeda kurti ir naudoti atvirų duomenų metaduomenų standartus, taip mažinant kliūtis duomenų leidėjams ir naudotojams.

³ Šaltinis: https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_29

⁴ Šaltinis: https://ec.europa.eu/isa2/home_en

Dažnai kaip duomenų standartai yra įvardijami: principų ir politikos standartai; Atvirasis apibrėžimas (ang. *The Open Definition*); duomenų failų formatai; specifiniai domeno politikos standartai; bendrieji techniniai standartai; specifiniai domeno techniniai standartai.

Pasaulinio Žiniatinklio Konsorciumas apibrėžia DCAT-AP kaip metaduomenų standartą atvirų duomenų aprašymui. Jis yra naudojamas Europos ir nacionaliniuose⁵ lygmenyse ir įgalina atvirų duomenų aprašymą ir prieigą visoje Europoje.

Taip pat galimi kiti standartai:

- *Asset Description Metadata Schema* (ADMS)⁶ - metaduomenų schema, kurią sukūrė Europos viešojo administravimo institucijų sąveikumo sprendimų (ISA) programa. ADMS, yra DCAT profilis, naudojamas aprašant semantinius išteklius, t. y. dažnai pakartotinai naudojamus metaduomenis (pvz., XML schema, bendriniai duomenų modeliai) ir atskaitinius duomenis (pvz., kodų sąrašai, taksonomijos, žodynai). ADMS yra naudojamas semantinių išteklių aprašymui, pvz., pateikiamų Joinup platformoje.
- *Dublin Core* schema - terminų žodynas, naudojamas įvairiems tikslams nuo interneto išteklių aprašymo iki metaduomenų žodynų iš skirtingų metaduomenų standartų sujungimo, užtikrinant metaduomenų žodynų sąveiką susietuose duomenyse.
- CKAN - plačiausiai paplitusi atvirų duomenų portalų programinė įranga ir yra naudojama daugybės nacionalinių portalų, tokių kaip *gov.co.uk*.
- *Data Catalog Vocabulary* (DCAT) - RDF žodynas, skirtas palengvinti sąveiką tarp duomenų katalogų, kurie yra skelbiami internete.
- *INSPIRE Metadata Schema* - Europos Parlamento ir Tarybos direktyva, kuria siekiama sukurti ES erdvinių duomenų infrastruktūrą, suteikiančią prieigą prie informacijos, kuri gali būti naudojama ES aplinkos politikai skirtingose šalyse ir viešajame sektoriuje remti. Šios informacija apima 34 aplinkosaugos temas, turinčias įvairias sritis, pvz., adresai, pastatai, gyventojų pasiskirstymas ir demografija.
- *Common Core Metadata Schema* (CCMS) atvirų duomenų projekte - paremtas DCAT standartu ir teikia bendrą žodyną, į kurį galima kopijuoti skirtingas atvirų duomenų metaduomenų schemas.
- *Data Catalog Interoperability Protocol* (DCIP) - Vienas iš pagrindinių DCIP tikslų yra sukurti CKAN įskiepi (ang. *plugin*), kad CKAN duomenys būtų rodomi kaip DCAT, bet šis projektas dar nebaigtas.
- *Vocabulary of Interlinked Datasets* (VoID) - RDF schema žodynas, skirtas RDF duomenų rinkinių metaduomenų aprašymui. Jo pirminis tikslas yra panaikinti atotrūkį tarp duomenų leidėjų ir duomenų naudotojų, naudojančių išskirtinį žodyną skirtingų duomenų rinkinių požymių aprašymui.
- *Schema.org* - schemų rinkinys (RDF/*Microdata* formatams), kurį žiniatinklio valdytojai gali naudoti HTML puslapių žymėjimui paieškos variklių pripažintais būdais.
- *Google Dataset Publishing Language* - duomenų ir metaduomenų pateikimo kalba. Duomenų rinkiniai, aprašyti šiuo formatu, gali būti tiesiogiai rodomi Google viešųjų duomenų naršyklėje.
- *Data Quality Vocabulary* (DQV) - žodynas, naudojamas duomenų ir jų metaduomenų kokybės aprašymui, pvz., kokybės sertifikatų, standartų ir kokybės politikos aprašymui.
- *Provenance Ontology* - naudojama provenencijos (kilmės) informacijos aprašymui ir gali būti specialiai konkrečios paskirties provenencijos modeliavimui įvairiose srityse.

Šioje dalyje apžvelgiami techniniai atvirų duomenų standartai, sutelkiant dėmesį į Europos Komisijos rekomenduojamus standartus.

⁵ Išskyrus kelias šalis, kurios vis dar įgyvendina arba vertina standartus, pvz., Lietuva.

⁶ <https://www.w3.org/TR/vocab-adms/>

1.3 Rekomenduojami standartiniai formatai publikavimui

Atvirų duomenų brandos lygį nusako 5-ių žvaigždučių schema, kurią 2010 m. pristatė Pasaulinio žiniatinklio konsorciumo įkūrėjas ir direktorius seras Tim Berners-Lee⁷. Penkių žvaigždučių schema iš esmės remiasi Maslow poreikių piramide ir nusako 5 brandos lygmenis nuo žemiausio iki aukščiausio. Kuo daugiau žvaigždučių – tuo didesnis pasiektas atvirų duomenų brandos lygis.

Metaduomenų standartai ir automatiškai apdorojami standartų formatai⁸ yra naudojami atvirų duomenų brandos lygio didinimui. Pavyzdžiui, 5 brandos lygis yra pasiekiamas sujungus visus duomenų rinkinius. Kad šis lygis būtų pasiektas, reikia skelbti aiškius metaduomenis pagal konkretų formatą, kad SPARQL varikliai galėtų surasti reikalingus duomenis.

Toliau esanti lentelė pateikia skirtingų atvirų duomenų brandos lygių ir skirtingų standartinių formatų apibendrinimą. Kiekviename lygyje, metaduomenų standartas gali būti naudojamas duomenų rinkinių aprašymui, bet aukštesniame lygyje reikalaujamas griežtesnis ir platesnis šių standartų naudojimas.

Lentelė Nr. 2. 5 žvaigždučių brandos lygio schema

Brandos lygis	Paaiškinimas	Naudojami publikavimo formatai
★ 1 žvaigždutė	Turinys publikuojamas žiniatinklyje (bet kokiu formatu) pagal atvirą licenciją.	PDF
★★ 2 žvaigždutės	Turinys teikiamas struktūrizuotų duomenų forma (pvz., Excel formatu, o ne kaip skenuoti duomenys).	Excel
★★★ 3 žvaigždutės	Naudojami atviri formatai (pvz., CSV, o ne Excel), konvertuojama į automatiškai nustatomus formatus (pvz., JSON, XML)	CSV
★★★★ 4 žvaigždutės	Naudojami URI elementai žymėjimui, jog žmonės galėtų sukurti nuorodas į šiuos duomenis.	RDF, JSON
★★★★★ 5 žvaigždutės	Susiejami turimi duomenys su kitais duomenimis, taip pateikiama daugiau informacijos.	Linked Open Data (LOD)

Bendrai, duomenis rekomenduojama skelbti bent 3 brandos lygio, kad būtų juos galima pakartotinai naudoti. Iš tiesų, kaip privatus ar valstybinės institucijos galėtų pakartotinai naudoti atvirose duomenis, jei jie nėra automatiškai apdorojami⁹?

1.4 Europos rekomenduojamas standartas

DCAT-AP (*Data CAtalogue – Application profile*) – tai specifikacija, paremta DCAT standartu ir naudojama Europos viešojo sektoriaus sukauptų duomenų rinkmenų apibūdinimui. DCAT-AP pagalba galima harmonizuoti duomenų rinkmenų katalogus ir apsikeisti duomenų rinkmenomis tarp duomenų portalų. Tai palengvina duomenų paiešką tarpvalstybinių mastu ir skirtinguose sektoriuose.

DCAT-AP esminiai principai:

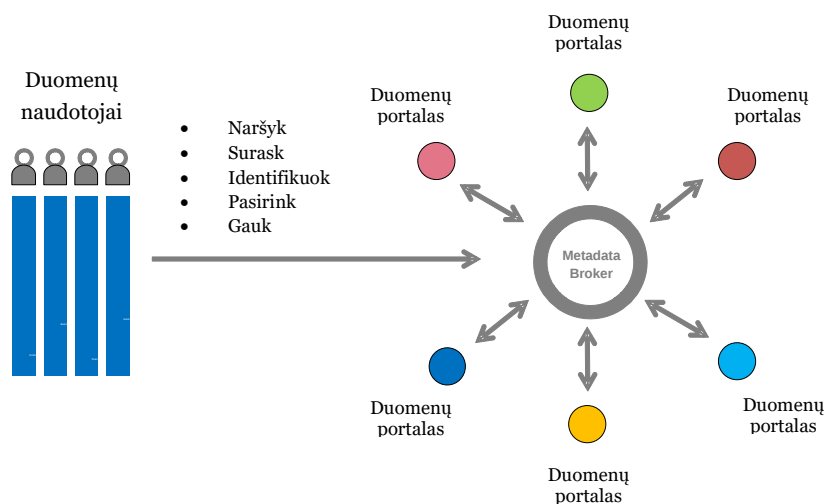
- Nustato esminius DCAT elementus ir požymius Europos kontekste;
- Nurodo žodyną, naudojamą Europos kontekste;

⁷ Šaltinis: <http://5stardata.info/>

⁸ Šaltinis: <https://www.europeandataportal.eu/en/providing-data/goldbook/preparing-data>

⁹ Kaip laikiną techninį sprendimą, galima naudoti procesų automatizavimo (robotizavimo) sprendimus, skirtus PDF, Excel ar CSV failų, ir duomenų suderinimui. Tačiau tai nėra geriausias sprendimas.

- Nurodo griežtus reikalavimus minimaliai aprašytus metaduomenis, kuriais bus keičiamasi tarp Europos atvirų duomenų portalų (parodyta toliau esančiame paveiksle)



Paveikslas Nr. 1. Prieigos suteikimo prie tipiška apibūdintų duomenų rinkinių metaduomenų procesas

2 Europos atvirų duomenų portalų ir kitose ES šalyse veikiančių atvirų duomenų portalų metaduomenų aprašo praktika

Portalai yra nustatomi pasauliniu ir nacionaliniu lygiu. Komunikacija tarp portalų yra paremta specifine aplikacijų programavimo sąsaja (API), kuri automatiškai surenka duomenis iš nacionalinių portalų ir sinchronizuoja juos su pasauliniu portalu.

Šioje dalyje pateikiamos gerosios metaduomenų naudojimo praktikos Europos ir nacionaliniame lygyje.

2.1 Atvirų duomenų portalų metaduomenų praktika

2.1.1 ES mastu naudojami atvirų duomenų portalai

ES naudojami keli atvirų duomenų portalai. **ES atvirų duomenų portalas**¹⁰ (angl. *European Union Open Data Portal*) yra skirtas ES lygmens institucijų ir įstaigų kaupiamos informacijos atvėrimui. Kadangi šio portalų pirminė paskirtis nėra nacionalinių atvirų duomenų portalų informacijos sklaida, šis portalas toliau ataskaitoje nebus nagrinėjamas.

ES šalių nacionalinių atvirų duomenų portalų duomenys yra skelbiami **Europos atvirų duomenų portale**¹¹, kurį 2015 m. Europos Komisija įkūrė iš Integruotos Europos Infrastruktūros (angl. *Connecting Europe Facility*) programos lėšų. Šis portalas yra pagrindinė Europos viešojo sektoriaus informacijos saugykla, kurios duomenys yra atviri pakartotiniam naudojimui Europoje. **Toliau ataskaitoje, minint Europos atvirų duomenų portalą, bus nagrinėjamas būtent šis Europos atvirų duomenų portalas.**

Pagrindinė Europos atvirų duomenų portalų funkcija yra suteikti vieną bendrą prieigą prie informacijos, kurią skelbia visų lygių Europos (ES šalių, Europos Ekonominės zonos šalių ir kitos Europos šalių) viešojo administravimo institucijos, visomis 24 oficialiomis ES kalbomis. Skatinant tarpvalstybinį lygiu skelbiamų duomenų suderinamumą, portale suteikiamos bendro metaduomenų formato, naudojant RDF technologiją, nuorodos (DCAT programos profilis).

Metaduomenų aprašymai išverčiami į visas 24 kalbas¹², naudojant automatinio vertimo technologiją.

Europos atvirų duomenų portalas siekia tapti vieta, kurioje visi ES šalių narių atvirų duomenų portalai dalinasi savo atvirų duomenų metaduomenimis. Tokiu būdu, žmonės ir įmonės gali vienoje vietoje pasiekti duomenų metaduomenis, arba kitais žodžiais, pasiekti duomenis apie duomenis prieinamus visoje Europoje. Lietuva taip pat turėtų prisijungti prie Europos atvirų duomenų portalų, kad galėtų keistis duomenimis.

Europos Komisija rekomenduoja laikytis Pasaulinio Žiniatinklio Konsorciūmo (W3C) skelbiamų metaduomenų naudojimo praktikų. Jie tai taip pat skelbia papildomus patarimus, kaip tvarkyti metaduomenis, kurie yra aprašomi šioje dalyje.

2.1.2 Tarptautinė atvirų duomenų metaduomenų praktika

Šias praktikas paskelbė Europos Sąjunga, jos šalys narės ir Pasaulinio Žiniatinklio konsorciūmas (W3C). Šioje dalyje siekiama pateikti IVPK svarbiausias¹³ tinkamų metaduomenų ruošimo praktikas.

Toliau lentelėje pateikiami metaduomenų naudojimo praktikos pavyzdžiai¹⁴.

Lentelė Nr. 3. Metaduomenų naudojimo praktikos pavyzdžiai

Geroji praktika	Aprašymas
-----------------	-----------

¹⁰ Šaltinis: <http://data.europa.eu/euodp/en/home>

¹¹ Šaltinis: <https://www.europeandataportal.eu/en/>

¹² Vertimas yra įdėtas į RDF failą, naudojant įprastą XML žymą; tai nėra automatinis vertimas.

¹³ Daugiau informacijos W3C gerųjų praktikų tinklapyje ir Europos atvirų duomenų portale.

¹⁴ W3C taip pat skelbia ir kitas gerąsias praktikas, kuris galima rasti: <https://www.w3.org/TR/dwbp/#metadata>

1	Pateikti metaduomenis	Metaduomenų pateikimas yra pagrindinis reikalavimas skelbiant duomenis internete, nes tai padeda duomenų naudotojams ir kompiuterinėms programoms suprasti duomenis ir kitus svarbius aspektus, kurie apibūdina duomenų rinkinius ar jų platinimą.
2	Pateikti aprašomuosius metaduomenis	Aprašomųjų metaduomenų pateikimas leidžia programoms automatiškai surasti duomenų rinkinius, prieinamus internete, ir naudotojams geriau suprasti duomenų rinkinio pobūdį ir platinimą.
3	Pateikti struktūrizuotus metaduomenis	Struktūrizuoti metaduomenys pateikia informaciją apie vidinę duomenų sklaidos struktūrą, kuri yra reikalinga duomenų rinkinio naudojimui ir užklausų kūrimui.
4	Duomenų licencija	Licencija yra naudinga informacija, kuri gali būti pridėta prie duomenų internete. Priklausomai nuo leidėjo paskirtos licencijos tipo, gali skirtis apribojimai duomenų dalinimuisi ir pakartotiniam naudojimui. Duomenų internete kontekste, duomenų rinkinio licencija gali būti nurodyta metaduomenyse arba atskirame susietame dokumente. Licencijavimas yra apibūdintas dalyje 4 <i>Standartinės metaduomenų struktūros taikymo atvirų duomenų metaduomenų formavimui procedūros</i> .
5	Duomenų kilmė (angl. provenance)	Duomenų kūrėjas nebūtinai yra duomenų leidėjas, todėl atitinkamų metaduomenų surinkimas ir pateikimas yra svarbus. Be kilmės duomenų, naudotojai neturi kito būdo patikrinti duomenų vientisumo ir patikimumo. Kilmė yra reikšmė, pagal kurią naudotojai sprendžia apie duomenų rinkinio kokybę. Savybės dct:creator, dct:publisher ir dct:issued yra naudojamos informacijos apie duomenų rinkinio kilmę pateikimui. Kilmei apibūdinti galima naudoti PROV-O žodyną ¹⁵ .
6	Duomenų kokybė	Duomenų kokybė gali paveikti duomenų tinkamumą konkrečioms programoms. Duomenų kokybės dokumentacija žymiai palengvina duomenų pasirinkimo procesą, taip pat padidina pakartotinio naudojimo galimybes. Automatiškai apdorojami duomenų rinkinio kokybės metaduomenys gali būti pateikti naudojant <i>Data Quality Vocabulary</i> ¹⁶ .
7	Duomenų versijos	Versijos informacija padaro duomenų rinkinio peržiūra atpažįstama. Duomenų naudotojai gali naudoti unikalumą, sprendžiant, ar ir kaip duomenys keitėsi laikui bėgant ir kuri duomenų rinkinio versija yra tinkamiausia. Geras duomenų versijavimas leidžia naudotojams suprasti ar nauja duomenų rinkinio versija yra prieinama. Išsamus versijavimas leidžia pakartoti tyrimus, palyginti duomenis ir išvengti susipainiojimo. Unikalių versijų numerių pagal standartinį metodą naudojimas leidžia suprasti, kuo versijos skiriasi. Tinkamiausias metodas versijų informacijos pateikimui: • Įtraukti unikalų versijos numerį ar datą į duomenų rinkinio metaduomenis. • Naudoti nuoseklią numeravimo schemą, pvz., [SchemaVer]. • Įgalinti skirtingus URI¹⁷, kad surinktų paskutinę ir senesnes versijas.

¹⁵ Šaltinis: <https://www.w3.org/TR/prov-overview/>

¹⁶ Daugiau informacijos W3C tinklapiuose: <https://www.w3.org/TR/vocab-dqv/> ir <https://www.w3.org/TR/vocab-dqv/#dqv:QualityMetadata>

¹⁷ Tam Lietuva turėtų sukurti atvirų duomenų URI strategiją.

8 Nuoseklių URI, kaip duomenų rinkinio identifikatorių ir kaip identifikatorių tarp duomenų rinkinių, naudojimas	Bendros identifikacijos sistemos priėmimas leidžia patikimus pagrindinių duomenų identifikacijos ir palyginimo procesus. Tai yra esminė išankstinė sąlyga tinkamui duomenų valdymui ir pakartotiniam naudojimui. Nuoseklumui užtikrinti, URI taip pat turi būti sukurti pagal tai. Šia tema yra išsamiai aprašyta, pvz., Europos Komisijos tyrime dėl nuoseklių URI¹⁸, kuris taip pat suteikia daug nuorodų į kitus šaltinius. URI duomenų rinkinio atpažinimo pavyzdys: http://data.mycity.example.com/transport/dataset/bus/stops
9 Automatiškai apdorojami standartiniai duomenų formatai	Nestandardinių duomenų formatų naudojimas yra brangus ir neefektyvus, ir duomenys gali prarasti prasmę, kai jie yra perkeltami. Priešingai, standartiniai duomenų formatai leidžia sąveiką ir yra pritaikyti ateities naudojimui, kaip vizualizacija, dažnai yra negalima, kai duomenys yra paskelbiami pirmą kartą. Taip pat svarbu paminėti, kad dauguma automatiškai apdorojami standartiniai formatai yra ir neutralūs vietos atžvilgiu. Duomenų galima pateikti automatiškai apdorojamais standartiniais duomenų formatais: CSV, XML, HDF5, JSON ir RDF sintakse, pvz., RDF/XML, JSON-LD, ar Turtle. Dalyje 2.1.4 <i>Metaduomenų paruošimo ir nustatymo techniniai standartai</i> pateikiama daugiau informacijos apie naudojamus standartinius formatus.
10 Duomenų pateikimas keliais formatais	Duomenų pateikimas daugiau nei vienu formatu sumažina išlaidas, kylančias dėl duomenų pakeitimų. Tai taip pat sumažina klaidų atsiradimo galimybę pakeitimo metu. Reiktų įvertinti labiausiai reikalingus duomenų formatus ir kitas alternatyvas, kurie formatai gali būti naudingi ateityje. Duomenis rekomenduojama skelbti RDF ir JSON formatais.

2.1.3 Susietų atvirų duomenų skelbimas¹⁹

Susieti duomenys (angl. *Linked Data*) yra viena pagrindinių semantinio žiniatinklio kertinių idėjų. Semantinis žiniatinklio tikslas yra sukurti sąsajas tarp duomenų rinkinių, kurios yra suprantamos ne tik žmonėms, bet ir mašinoms, ir susieti duomenys pateikia gerąsias praktikas, kaip sukurti šias sąsajas. Susieti duomenys yra kūrimo principų rinkinys, skirtas automatiškai apdorojamų tarpusavyje susietų duomenų rinkinių pasidalijimui.

Duomenų išreiškimas, susietu pagal visuotinius principus būdu, gali padidinti duomenų kokybę. Taip pat semantinė grafinė duomenų bazė, atvaizduojanti susietus duomenis, sukuria semantines sąsajas tarp skirtingų šaltinių ir formatų ir leidžia naujoms žinioms susikurti iš jau esamų faktų.

Susieti atviri duomenys yra efektyvus mišinys tarp susietų duomenų ir atvirų duomenų – duomenys yra susieti ir naudoja atvirų duomenų šaltinius. Grafinė duomenų bazė gali apdoroti didelius pirminių duomenų rinkinių kiekius iš skirtingų šaltinių ir sujungti juos su atviraisiais duomenimis, ir tai leidžia pateikti išsamesnes užklausas ir praplečia duomenų paieškos galimybes. Vienas iš susietų atvirų duomenų šaltinių pavydžių yra DBpedia, visuomeninis projektas, kuriuo siekiama išgauti struktūrinę informaciją iš Vikipedijos ir pateikti šią informaciją internete.

Pagrindiniai atvirų duomenų skelbimo kaip susietųjų duomenų principai yra:

- HTTP URI naudojimas, kad visi objektai būtų pasiekiami per HTTP
- URI peržiūrai, pateikti naudingą informaciją, naudojant RDF ir/ar SPARQL: semantinė grafinė duomenų bazė yra technologija sukurta susietų duomenų saugojimu ir tarpusavyje susietų duomenų paaiškinimui, taip semantiškai praturtinant duomenų bazes. Priešingai nei sąryšinės

¹⁸ Šaltinis: <https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>

¹⁹ Šaltinis: <https://www.w3.org/2013/Talks/0902-Lisbon-IH/>

duomenų bazės, trilypė saugykla (angl. *triplestore*)²⁰ atvaizduoja skirtingus ryšius tarp subjektų grafinėse duomenų bazėse. Iš kitos pusės, SPARQL yra W3C standartizuota užklausų kalba RDF trilypė saugyklai.

- Nuorodų į kitas URI sąsajas įtraukimas, kad būtų galima atrasti daugiau dalykų.

Europos Komisija taip pat išanalizavo skirtingus verslo modelius, kaip galima panaudoti susietus atvirojo valstybės duomenis²¹.

2.1.4 Metaduomenų paruošimo ir nustatymo techniniai standartai

Metaduomenys turi būti paskelbti taikant techninius standartus. Metaduomenų aprašymo standartas Europoje yra DCAT, o techninis formatas yra RDF. Toliau pateiktoje lentelėje aprašomi rekomenduojami atvirų duomenų metaduomenų publikavimo formatai.

Lentelė Nr. 4. Rekomenduojami atvirų duomenų metaduomenų publikavimo formatai

Formatas	Aprašymas
CSV	Kableliu atskirtų reikšmių (CSV) failas pateikia lentelės duomenis (skaičius ir tekstą) kaip paprastą tekstą. Kiekviena failo eilutė yra duomenų įrašas. Kiekvienas įrašas yra sudarytas iš vieno ar kelių laukelių, atskirtų kableliu.
RDF	RDF (angl. <i>Resource Description Framework</i>) yra pagrindinis susietųjų duomenų formatas, naudojamas bendra sintaksė pateikti duomenis tinkle. Ši sintaksė yra ryšys (URI, angl. <i>Unique Resource Identifier</i>), kuris sudarytas iš trilypių deskriptorių: subjekto, predikato ir objekto. Aprašant objektą šiais trimis deskriptoriais jis tampa susietuoju.
RDFa	Duomenų portaluose dažnai naudojama RDF technika RDFa - RDF įterpimas į HTML.
JSON²²	JSON yra lengvas duomenų keitimosi formatas, kurį lengva skaityti, analizuoti ir sugeneruoti. JSON yra teksto formatas, paremtas JavaScript programavimo kalba, kuris yra optimizuotas duomenų mainams. JSON yra kuriamas iš dviejų struktūrų: 1. pavadinimų/reikšmių porų ir 2. reikšmių sąrašo.
XML	XML yra plačiai naudojamas duomenų mainų formatas, nes tai suteikia galimybių saugoti duomenų struktūrą ir failų kūrimo būdus bei leidžia kūrėjams kurti duomenų dalis, neapsunkinant jų skaitymo.
HDF5²³	HDF5 yra duomenų modelis, biblioteka ir failų formatas, skirtas saugoti ir valdyti duomenims. Jis palaiko neribotą skaičių duomenų tipų ir yra sukurtas lanksčiai ir efektyviai įvesčiai/išvesčiai ir dideliems ir sudėtingiems duomenų kiekams.
URI	URI reiškia unikalų išteklių identifikatorius (angl. <i>Unique Resource Identifier</i>) ir gali nukreipti į tekstą, URN (angl. <i>Uniform Resource Name</i>), arba URL (angl. <i>Uniform Resource Locator</i>).

ES atvirų duomenų portalas skatina naudoti ES duomenų rinkinius trečiųjų šalių programoms kurti. Siekiant šio tikslo, siūlomos dvi programavimo sąsajos: REST API ir SPARQL *endpoint*.

²⁰ Trilypė saugykla (angl. *triplestore*) arba RDF saugykla yra specialiai sukurta duomenų bazė, skirta trilypiams objektams (angl. *triple*) saugoti ir atkurti naudojant semantines užklausas. Trilypis objektas yra duomenų objektas, kuris susideda iš subjekto-predikato-objekto, pvz., "Jonui yra 35" ar "Jonas pažįsta Tomą".

²¹ Šaltinis: <https://joinup.ec.europa.eu/sites/default/files/document/2013-04/EDF%20-%20How%20the%20ISA%20Programme%20promotes%20LOGD.pdf>

²² Daugiau apie JSON formatą: <https://www.w3.org/TR/json-ld/>

²³ Šaltinis: <https://support.hdfgroup.org/HDF5/>

REST API

Vienas iš būdų pasiekti ES OAP yra naudojant REST API. Visos portalo pagrindinės funkcijos yra prieinamos naudojant API, apimant tai, ką galima padaryti su žiniatinklio sąsaja ir kt. Gautai informacijai galima naudoti išorinį kodą, kuris susisieks su portalo API. Pranešimų formatas tiek užklauskoms, tiek atsakymams yra JSON²⁴.

SPARQL endpoint

Kitas būdas pasiekti ES OAP yra SPARQL *endpoint*, naudojantis automatinį apdorojimą ir leidžiantis pateikti užklauskas RDF duomenų rinkinių aprašymuose. Detalesnę SPARQL *endpoint* specifikaciją galima rasti W3C tinklalapyje²⁵.

2.2 Nacionalinių duomenų portalų gerieji pavyzdžiai naudojant metaduomenis

2.2.1 Nacionaliniai duomenų portalai Europoje

ES šalys skelbia viešus duomenis apie nacionalinius duomenų portalus ir geografinius (angl. *geospatial*) portalus²⁶. Atvirų duomenų portalai yra žiniatinklio sąsajos, skirtos lengviau rasti pakartotinai naudojamą informaciją. Kaip ir bibliotekų katalogai, juose yra kaupiami metaduomenų įrašai, skirti pakartotiniam duomenų rinkinių panaudojimui. Kartu su konkrečiomis paieškos funkcijomis jie palengvina dominančių duomenų rinkinių paiešką.

API programinė įranga suteikia tiesioginę ir automatizuotą prieigą prie duomenų. Šalys gali laisvai pasirinkti atvirų duomenų skelbimo standartą, tačiau Europoje rekomenduojama naudoti DCAT standartą. Skirtingi standartai yra plačiau aprašyti skyriuje *Metaduomenų aprašymui taikomų standartų apžvalga*.

DCAT-AP taikymo aprašas pateikia bendrą specifikaciją viešojo sektoriaus duomenų rinkinių Europoje aprašymui, kad būtų galima keistis duomenų rinkinių metaduomenų aprašais tarp duomenų portalų. Ši specifikacija skirta vienodam viešojo sektoriaus duomenų rinkinių aprašymui Europoje, kad būtų įmanoma iškart ieškoti tų pačių duomenų rinkinių skirtinguose portaluose, tokiu būdu palengvinant viešojo sektoriaus duomenų prieinamumą tarpvalstybiniu mastu ir tarp skirtingų sektorių²⁷.

DCAT-AP suteikia šias galimybes:

- Duomenų katalogai gali apibūdinti duomenų rinkinius, taikydami standartizuotą aprašymą, tuo pat metu išlaikydami savo sistemą dokumentavimui ir saugojimui.
- Turinio surinkėjai, pvz., Europos atvirų duomenų portalas, gali lengvai sujungti tokius aprašus į vieną prieigos tašką.
- Duomenų vartotojai dėl suderinto žodyno gali lengviau ieškoti ir rasti duomenų rinkinius, taikydami vienodus parametrus skirtinguose portaluose.

Dabartinė DCAT-AP versija buvo įgyvendinta daugelyje Europos šalių, pavyzdžiui, Airijoje, Nyderlanduose, Ispanijoje, Prancūzijoje, Vokietijoje ir Italijoje. Taip pat vis daugiau duomenų portalų Europos, regionų ir vietos lygiu atitinka DCAT-AP standartą.

2.2.2 Metaduomenų aprašymas nacionaliniuose atvirų duomenų portaluose Europoje

Šioje dalyje pristatomi trys atvirų duomenų portalų Europoje pavyzdžiai:

- Ispanijos nacionalinis duomenų portalas: <http://datos.gob.es/>
- Airijos nacionalinis duomenų portalas: <https://data.gov.ie/>

²⁴ Šaltinis: <http://www.ietf.org/rfc/rfc4627.txt>

²⁵ Šaltinis: <http://www.w3.org/TR/rdf-sparql-query/>

²⁶ Šaltinis: <https://www.geoportal.lt/>

²⁷ Šaltinis: <https://joinup.ec.europa.eu/news/final-version-dcat-app>

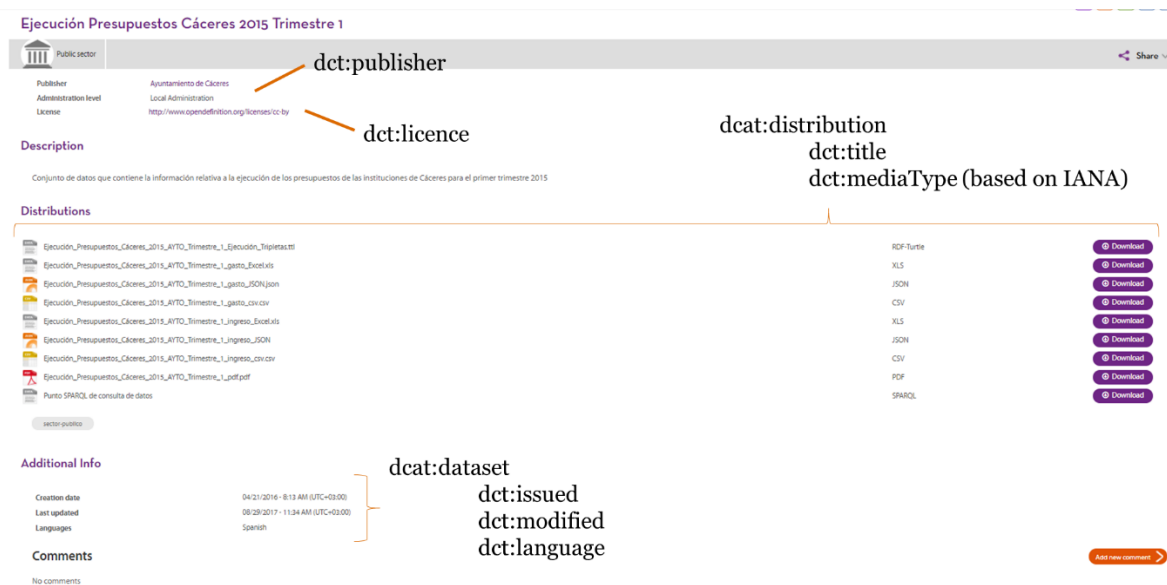
- Nyderlandų nacionalinis duomenų portalas: <https://data.overheid.nl/>

Pateikiamas kiekvieno portalų bendras metaduomenų vaizdas ir pagrindinės funkcijos, kurios gali būti naudojamos kaip pavyzdys Lietuvai.

2.2.2.1 Ispanija

Ispanijos nacionalinis duomenų portalas (<http://datos.gob.es/en>) turi šias pagrindines parinktis, kurios nurodytos toliau esančioje lentelėje.

Toliau pateikiamas Ispanijos duomenų portalų ekranvaizdis (*angl. screenshot*).



Paveikslas Nr. 2. Ispanijos duomenų portalų ekranvaizdis

Lentelė Nr. 5. Ispanijos duomenų portalų pagrindinės parinktys

Parinktis	Aprašymas
Duomenų katalogas	Duomenų bazė talpina virš 16000 duomenų rinkinių, suskirstytų pagal kategoriją, duomenų formatą, leidėją ir kitus kriterijus.
API	API yra mechanizmas leidžiantis daryti užklausas <i>datos.gob.es</i> semantinėje duomenų bazėje.
SPARQL endpoint	Naudojant šią paslaugą, galima konfigūruoti savo užklausas naudojant SPARQL užklausų kalbą <i>datos.gob.es</i> RDF diagramoms.
Puslapio statistika	Vizualiai pateikiama puslapio statistika apie iniciatyvų skaičių, duomenų pakartotinio naudojimo atvejus, publikuojamų duomenų rinkinių skaičių, populiariausių duomenų rinkinių ir kita statistika, kurią galima interaktyviai peržiūrėti puslapyje arba atsisiųsti PNG, JPG, SVG ir PDF formatais arba komponentų duomenys gali būti išsaugoti CSV, XLSX arba JSON formatais.
Iniciatyvų žemėlapis	Iniciatyvų pasiskirstymas šalyje rūšiuojant pagal administracijos lygį, sektorių, pakartotinio naudojimo dažnumą ir t.t.
Aplikacijų duomenų bazė	Aplikacijos duomenų bazė, kurioje yra virš 200 aplikacijų, naudojančių atvirus duomenis

Airijos nacionalinis duomenų portalas (<https://data.gov.ie/data>) turi šias pagrindines parinktis, kurios nurodomos toliau esančioje lentelėje.

General Election 2007- Data on Female Candidates

Published by Department of Housing, Planning and Local Government. Licensed under Creative Commons Attribution 4.0

Openness rating: ★★★★☆

48 views

Government and Public Sector

General Election 2007- Data on Female Candidates This includes data on the number and percentage of female candidates in each constituency and that of those elected.

Data Resources (1)



General Election 2007- Data on Female Candidates

Preview

Download

Data Resource Preview - General Election 2007- Data on Female Candidates

Constituency	No. of Seats	Total No. of Candidates	No. of Women Candidates	Women as % of Candidates	No. of Women Elected	Women as % of Members Elected
Additional Information						

Additional Information

Title	General Election 2007- Data on Female Candidates
Description	General Election 2007- Data on Female Candidates This includes data on the number and percentage of female candidates in each constituency and that of those elected.
Contact Point	name: Helena Campbell email: helena.campbell@envirom.ie phone: -
Keywords	2007, candidates, constituency, elected, female, general election
Theme	Government and Public Sector
Publisher	Department of Housing, Planning and Local Government
Issue Date	21 Nov 2017
Last Modification Date	
Publication Frequency	never
Language	en
Landing Page	http://www.housing.gov.ie/sites/default/files/publications/files/2007_general_election_female_candidates.csv
Geographic Coverage	national
Temporal Extent	From: 2007, To: n/a

dc:dataset

dct:title

dct:description

foaf:agent

dc:keyword

dc:theme

dct:publisher

dct:issued

dct:modified

dct:temporal (frequency)

dct:language

dc:accessURL

dct:spatial

dct:accrualPeriodicity

Paveikslas Nr. 3. Airijos duomenų portalo ekranvaizdis

Lentelė Nr. 6. Airijos duomenų portalo pagrindinės parinktys

Parinktis	Aprašymas
Duomenų rinkiniai	Duomenų bazė talpina virš 5000 duomenų rinkinių, suskirstytų pagal temą, duomenų formatą, leidėją ir kitus kriterijus.
Leidėjų sąrašas	Duomenų bazė talpina virš 100 leidėjų, atveriančių savo duomenis <i>data.gov.ie</i> svetainėje
Užklauso pateikimas	Funkcija, leidžianti pateikti užklausą viešai atverti dar nepublikuojamą duomenų rinkinį.
Puslapio statistika	Vizualiai pateikiama puslapio statistika apie duomenų rinkinių skaičiaus kitimą, populiariausių duomenų rinkinius, pagrindinius naudojamus raktazodžius, atitikimą techninei sistemai ir kita statistika.
Gerieji pavyzdžiai	Gerųjų pavyzdžių skiltyje pateikiami geriausi duomenų rinkinių panaudojimo pavyzdžiai

Nyderlandų nacionalinio duomenų portalo (<https://data.overheid.nl/>) pagrindinės parinktys yra aprašomos toliau esančioje lentelėje.

Toliau esančiame paveiksle pateikiamas Nyderlandų duomenų portalų ekranvaizdis (*angl. screenshot*).

Kustlijnkaartboek 2016: Af te beelden trendwaarden schaal (50000)

Bereken de trendwaarden horizontale positie kustlijn ten opzichte van de Basiskustlijn voor het jaar 2017. Cijfer en kaarten worden jaarlijks vervoerd in het kustlijnkaartboek dat wordt uitgegeven door de Watersdienst. Op basis van dit boek wordt het suppletieschema voor het jaar 2 jaar na meting vastgesteld. Afgebeeld is de trend in uitwijking van de positie van de te toetsen kustlijn (TKL) ten opzichte van de Basiskustlijn. In de onderliggende tabel alle toetsparameters die door de toetsprogramma's WINKUST worden uitgerekend. De kleur geeft de richting van de trend (zeewaarts/landwaarts) en geeft de ligging weer van de SI (zeewaarts/landwaarts).

Tags

Vind datasets met dezelfde tags

Kust	Veiligheid	Water	Zee
------	------------	-------	-----

Downloads	Webdienst	Links	Overige	Documentatie	Metadata
-----------	-----------	-------	---------	--------------	----------

A. Organisatie

Data-eigenaar	Ministerie van Infrastructuur en Milieu Rijkswaterstaat
Contact data-eigenaar	servicedesk-data@rws.nl
Catalogus	Nationaal Georegister
Catalogusreferentie	http://nationaalgeoregister.nl/geonetwork?uuid=c499bdc0-23fa-40ea-99eb-24370d12f56
Catalogus-identificatie	c499bdc0-23fa-40ea-99eb-24370d12f56

B. Verplichte velden

Status	Beschikbaar
Licentie	CC-0
Taal	Nederlands
Wijzigingsdatum	31-12-2016

C. Vindbaarheid	
D. Kenmerken data	
Locatie	Nederland
Dekking in tijd	01-01-2016 tot 31-12-2016
Verversingsgraad	Maandelijks
Linked Data Score	0 sterren: onbekend
High value dataset	Nee

dcat:dataset

dct:spatial

dct:temporal (frequency)

dct:accrualPeriodicity

dct:publisher

dcat:catalogue

dct:status

dct:licence

dcat:dataset

dct:issued

dct:language

Paveikslas Nr. 4. Nyderlandų duomenų portalų ekranvaizdis

Lentelė Nr. 7. Nyderlandų duomenų portalų pagrindinės parinktys

Parinktis	Aprašymas
Duomenų rinkiniai	Duomenų bazė talpina virš 12000 duomenų rinkinių, suskirstytų pagal statusą, prieinamumą, temą, raktažodžius, leidėją ir kitus kriterijus.
Duomenų kalendorius	Kalendoriuje pateikiama informacija apie planuojamus publikuoti duomenų rinkinius. Duomenis galima filtruoti pagal datą, temą arba leidėją.
Leidėjų peržiūra	Leidžia filtruoti duomenų rinkinius pagal leidėją, analizuoti, kiek ir kokios kokybės duomenų rinkinius paskelbė konkretus leidėjas.
Užklausos pateikimas	Funkcija, leidžianti pateikti užklausą viešai atverti dar nepublikuojamą duomenų rinkinį. Pateikus užklausą ji iškart patalpinama <i>data.overheid.nl</i> puslapyje, kad kiti puslapio lankytojai matytų, jog užklausa apie tokių duomenų atvėrimą jau yra pateikta.
Puslapio statistika	Vizualiai pateikiama puslapio statistika apie duomenų kokybę, rinkinių skaičius kitimą, populiariausių duomenų rinkinius, daugiausiai duomenų rinkinių paskelbusius leidėjus, populiarumą pagal temą ir kita statistika.

3 Valstybės informacinių išteklių duomenų metaduomenų standartinio aprašo struktūra

3.1 Lietuvai rekomenduojami naudoti metaduomenų standartai

Toliau esančioje lentelėje aprašomi metaduomenų standartai, kurie rekomenduojami naudoti Lietuvos atvirų duomenų aprašymui.

Lentelė Nr. 8. Metaduomenų standartai, rekomenduojami Lietuvos atvirų duomenų aprašymui

Standartas	Naudojimo paskirtis
DCAT	DCAT programos profilis duomenų portalams (DCAT-AP) yra specifikacija paremta <i>Data Catalogue vocabulary</i> (DCAT) žodynu. DCAT-AP suteikia bendrą specifikaciją Europos viešojo sektoriaus duomenų rinkinių aprašymui, taip įgalinant duomenų rinkinių aprašymų mainus tarp duomenų portalų. Šis standartas yra rekomenduojamas visų Lietuvos atvirų duomenų metaduomenų aprašymui.
StatDCAT-AP²⁸	StatDCAT-AP ²⁹ yra DCAT plėtinys, kurio tikslas yra suteikti specifikaciją ir įrankius, kurie padidintų sąveiką tarp statistinių duomenų aprašymų tarp statistinių duomenų ir atvirų duomenų portalų. StatDCAT-AP apibrėžia tam tikrą skaičių DCAT-AP modelio papildymų, kurie gali būti naudojami aprašant duomenų rinkinius bet kuriuo formatu, pvz., duomenis skelbiamus kaip Statistiniai duomenis ir metaduomenis eXchange (SDMX ³⁰) standartu, skirtu statistinių duomenų mainams. StatDCAT-AP turėtų būti naudojamas aprašant statistinius atvirus duomenis. Tačiau, šiuo atveju, kadangi StatDCAT-AP yra DCAT-AP variacija, iš pradžių rekomenduojama įdiegti DCAT-AP-LT ir tik tada įdiegti StatDCAT-AP, remiantis DCAT įdiegimu.
GeoDCAT-AP³¹	GeoDCAT-AP yra DCAT-AP plėtinys, skirtas geo-erdvinių duomenų rinkinių, duomenų serijų ir paslaugų aprašymui. Jis pateikia RDF sintaksę, privalomą metaduomenų elementų sąjungai, kaip apibrėžta ISO 19115:2003 pagrindiniame profilyje ir INSPIRE direktyvoje. Jo pagrindinis naudojimo atvejis yra erdvinių duomenų rinkinių, serijų ir paslaugų atradimo duomenų portaluose palengvinimas, taip padarant geo-erdvinę informaciją geriau atrandamą tarp valstybių ir sektorių. Tai gali būti pasiekta duomenų aprašymui mainais tarp duomenų portalų. GeoDCAT-AP specifikacija nepakeičia INSPIRE metaduomenų reglamento ar INSPIRE metaduomenų techninių gairių, paremtų ISO 19115 ir ISO19119 standartais. Jo tikslas yra leisti geo-erdvinių metaduomenų savininkams pasiekti daugiau, pateikiant papildomą privalomą RFD sintaksę. GeoDCAT-AP naudojimui taikoma ta pati rekomendacija, kaip ir dėl StatDCAT-AP (pateikiama stulpelyje aukščiau).

²⁸ Šaltinis: https://joinup.ec.europa.eu/sites/default/files/distribution/2016-12/statdcat-ap_version_1.0.0_1.pdf

²⁹ Šaltinis: https://ec.europa.eu/isa2/sites/isa/files/docs/publications/statdcat-ap_poster.pdf

³⁰ Šaltinis: <https://sdmx.org/>

³¹ Šaltinis: <https://joinup.ec.europa.eu/release/geodcat-ap/v101>

EU Budget Vocabulary³²

ES biudžeto žodynas (angl. *EU Budget Vocabulary*) yra skirtas sukurti, palengvinti mainus, pagerinti suprantamumą ir skatinti pakartotinį biudžetinės ES informacijos panaudojimą. Žodynas siekia padidinti vyriausybės skaidrumą, pagerinant ES biudžeto prieinamumą, naudojimą ir supratimą.

Svarbu paminėti, kad šis žodynas yra skirtas tik ES biudžetui, o ne šalių narių biudžetui. Tai reiškia, kad šis standartas neturi būti naudojamas atvirų duomenų portale.

Šis standartas buvo sukurtas duomenų, susijusių su ES biudžetu, valdymui Europos lygmenyje, todėl standartas neturėtų būti naudojamas aprašant finansinius duomenis Lietuvoje.

INSPIRE³³³⁴

INSPIRE direktyva siekia sukurti ES erdvinių duomenų infrastruktūrą, naudojamą ES aplinkosaugos politikoje ir kitose veiklose, kurios gali turėti įtaką aplinkai. Ši Europos Erdvinių Duomenų Infrastruktūra leidžia dalintis aplinkosaugos erdvine informacija tarp viešojo sektoriaus organizacijų, palengvina viešą prieigą prie erdvinės informacijos visoje Europoje ir padeda politikos formavimui tarp šalių³⁵.

Šio dokumento pagrindas yra DCAT-AP Lietuvai pritaikymas.

3.2 DCAT-AP žodynas

3.2.1 DCAT-AP žodyno pristatymas

DCAT-AP apima privalomų, rekomenduojamų³⁶ ir neprivalomų klasių ir savybių rinkinį, leidžiantį apibūdinti vertybes:

- *Privaloma klasė*: duomenų gavėjas turi sugebėti tvarkyti klasės egzemplioriaus informaciją (duomenis); duomenų siuntėjas turi pateikti informaciją apie klasės egzempliorius.
- *Rekomenduojama klasė*: duomenų gavėjas turi sugebėti tvarkyti klasės egzemplioriaus informaciją (duomenis); duomenų siuntėjas turi pateikti informaciją apie klasės egzempliorius, jei ji prieinama.
- *Neprivaloma klasė*: duomenų gavėjas turi sugebėti tvarkyti klasės egzemplioriaus informaciją (duomenis); duomenų siuntėjas gali pateikti informaciją, bet neprivalo tai daryti.
- Kiekvienai klasei nustatomi privalomi, rekomenduojami ir neprivalomi savybių rinkiniai.

³² Šaltinis: https://joinup.ec.europa.eu/sites/default/files/distribution/2016-06/specification_of_the_eu_budget_vocabulary_v1.02.pdf

³³ Šaltinis: <https://joinup.ec.europa.eu/collection/inspire>

³⁴ Šaltinis: <https://inspire.ec.europa.eu/>

³⁵ Šaltinis: https://inspire.ec.europa.eu/sites/default/files/presentations/20170908_LHQ_INSPIRE_in_RDF.pdf

³⁶ Šaltinis: <https://www.w3.org/TR/vocab-dcat/#vocabulary-overview>

3.2.2 DCAT-AP plėtiniai

3.2.2.1 DCAT-AP pritaikymo sąlygos

DCAT-AP gali būti pritaikytas, jei remiamasi toliau nurodytomis taisyklėmis³⁷:

- Plėtinių naudojimo sąlygų, nurodytas DCAT-AP v1.1, negali išplėsti, tik susiaurinti taip, kad visa plėtiniu teikiama informacija galėtų DCAT-AP v1.1.
- Plėtinyje galima pridėti klases, kurios nėra nurodytos DCAT-AP, tačiau plėtinyje neturėtų būti pridedamos klasės, panašios į DCAT-AP klases.
- Plėtinyje galima pridėti savybes kurios nėra nurodytos DCAT-AP, tačiau plėtinyje neturėtų būti pridedamos savybės, panašios į DCAT-AP savybes.
- Plėtinyje galima keisti DCAT-AP v1.1 nurodytas savybes, atsižvelgiant į šias taisykles:
 - DCAT-AP v1.1 privalomos savybės turi išlikti privalomomis plėtinyje;
 - DCAT-AP v1.1 rekomenduojamos savybės gali būti nustatytos privalomomis arba neprivalomomis plėtinyje;
 - DCAT-AP v1.1 neprivalomos savybės gali būti nustatytos rekomenduojamomis arba privalomomis plėtinyje;
 - DCAT-AP v1.1 rekomenduojamos ir neprivalomos savybės gali būti pašalinamos iš plėtinio.
- Plėtinyje turi būti naudojami kontroliuojami žodynai, kurie yra nurodomo DCAT-AP v1.1 specifikacijos 5 dalyje.
- Plėtinyje galima pridėti kontroliuojamus žodynus.

3.2.2.2 Plėtinių pavyzdžiai

Belgija³⁸

Belgijos nacionalinis atvirų duomenų portalas laikosi DCAT-AP be papildomų savybių. Tačiau, kadangi Belgijoje be nacionalinio atvirų duomenų portalo veikia ir regioninės (lokalios) duomenų atvėrimo iniciatyvos, duomenų teikėjai gali nurodyti papildomas savybes.

Vokietija³⁹

Vokietija pritaikė DCAT, kad būtų palaikoma daugiau licencijų, autorių teisių ir kitų įstatymų apribojimų.

Airija⁴⁰

Airija pakeitė geo-erdvinių metaduomenų savybes *Dataset* klasei. Kai kurios iš jų buvo pridėtos prie DCAT-AP versijoje 1.1: *SpatialResolution* (erdvinė išraiška), *SpatialReferenceSystem* (erdvinio sprendimo sistema) and *GeographicBoundingBox* (geografinių apribojimų langelis).

Nyderlandai⁴¹

Šis plėtinys apima *Dataset*, *Agent*, *Catalogue*, *Record* ir *Distribution* klasių atnaujinimus. Į šią specifikaciją, Nyderlandų plėtinyje taip pat įtraukiami visi privalomi ISO 19115 geo-duomenys.

3.2.2.3 Nacionalinių plėtinių kūrimo tendencijos

Toliau esančiame sąrašė nurodomi įprasti DCAT-AP nacionaliniai plėtiniai.

³⁷ Šaltinis: https://www.mkm.ee/sites/default/files/digital_agenda_2020_estonia_engf.pdf

³⁸ Informacija paremta komunikacija su Fedict.

³⁹ Šaltinis: <http://www.dcat-ap.de/def/>

⁴⁰ Šaltinis: Atvirų duomenų techninė struktūra (angl. *Open Data Technical Framework*) DATA.GOV.IE tinklapyje

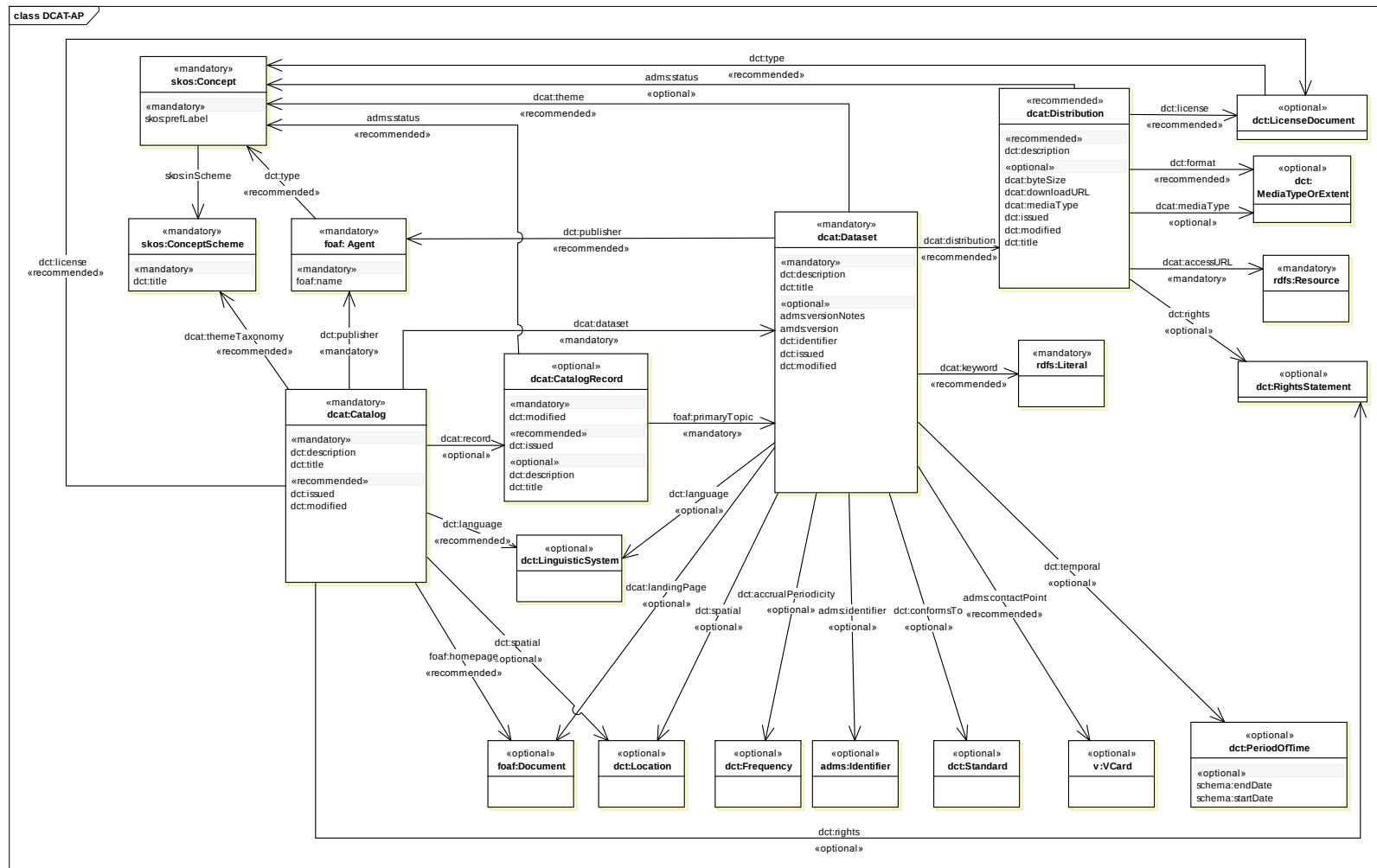
⁴¹ Šaltinis: <http://dcat-nl.info/nl/latest/>

- dct:identifier *Dataset* klasėje, *Agent* klasėje, ir *Standard* klasėje;
- dct:publisher *Dataset* klasėje padaromas privalomas;
- dct:license *Distribution* klasėje ir *Catalogue* klasėje;
- dcat:theme *Dataset* klasėje padaromas privalomas;
- vcard:fn *Organization* klasėje ir *Contact* klasėje yra atnaujinimas;
- dct:modified *Dataset* klasėje ir *Catalogue* klasėje;
- dct:issued, dcat:mediaType, rdf:type, dct:format ir dcat:accessURL yra dažnai atnaujinami;
- dct:spatial *Dataset* klasėje ir *Catalogue* klasėje;
- dct:license *Dataset* klasėje ir *Distribution* klasėje;
- dct:identifier *Agent* klasėje ir *Catalogue* klasėje;
- dct:subject *Dataset* klasėje;
- dct:creator *Dataset* klasėje;
- dct:description *Distribution* klasėje;
- *Agent* klasė: kartais⁴² rekomenduojama savybė dct:type yra pašalinama;
- *Distribution* klasė: toliau nurodytos rekomenduojamos ir neprivalomos savybės kartais⁴³ yra pašalinamos: spdx:checksum, foaf:page, dct:language, dct:conformsTo, dct:rights, adms:status, dct:issued, and dcat:mediaType, license, conformsTo, language, page, checksum, type, ir rights;
- *Dataset* klasė: toliau nurodytos rekomenduojamos ir neprivalomos savybės kartais¹ yra pašalinamos: dct:relation, dct:source, dct:accessRights, dct:provenance, foaf:page, dct:hasVersion, adms:sample, dct:type, ir adms:versionNotes;
- *Catalogue* klasė: toliau nurodytos rekomenduojamos ir neprivalomos savybės yra pašalinamos¹: hasPart, isPartOf. dcat:record, dct:spatial, dct:license, ir dct:rights;
- Neprivaloma savybė mediaType *Distribution* klasėje yra pašalinama.

⁴² DCAT-AP naudojimas bent jau Italijoje.

⁴³ DCAT-AP naudojimas bent jau Italijoje ir Nyderlanduose

3.2.3 DCAT standarto duomenų modelis



Paveikslas Nr. 5 DCAT-AP duomenų modelis

3.3 Siūloma metaduomenų struktūra Lietuvoje

3.3.1 Suderinimas su DCAT ir pritaikymas

Toliau apibūdinami DCAT suderinimo reikalavimai. Kad metaduomenys atitiktų DCAT-AP standartą, reikia pateikti šiuos metaduomenis:

- Katalogo aprašymą, apimant bent privalomas savybes;
- Informaciją apie privalomas savybes;
- Duomenų rinkinio, esančio Kataloge, aprašymą, apimant bent privalomas savybes;
- Duomenų rinkinių, esančių Kataloge, Platinimo aprašymą (jei toks yra), apimant bent privalomas savybes;
- Visų organizacijų aprašymus, įtrauktų į Katalogų ir Duomenų rinkinių aprašus, apimant bent privalomas savybes;
- Visų kategorijų schemas, kuriose yra Duomenų rinkinių aprašymuose patvirtintos kategorijos, apimant bent privalomas savybes;
- Visų kategorijų aprašymus, kurios naudojamos Duomenų rinkinio Kataloge aprašymui, apimant bent privalomas savybes.

3.3.2 Bendri laukelių užpildymo reikalavimai⁴⁴

Toliau esančioje lentelėje pateikiami siūlomi atvirų duomenų metaduomenų aprašymai.

Lentelė Nr. 9. Metaduomenų aprašymo sritys

Sritis	Aprašymas
Bendri laukeliai	
Pavadinimas (arba Vardas)	Duomenų pavadinimas. Jis turėtų būti parengiamas lietuvių kalba (duomenis skelbiant Europos atvirų duomenų portale – ir anglų kalba), nurodant pakankamai detalių, palengvinančių duomenų rinkinio paiešką ir suradimą. Reiktų vengti sutrumpinimų.
Aprašymas	Detalesnis aprašymas (pvz., santrauka), kuris padeda naudotojams greitai suprasti, ar rinkinys atitinka paiešką.
Kategorija (arba Tema)	Pagrindinė tematinė duomenų rinkinio kategorija, dažnai pasirenkama iš anksto nustatyto sąrašo ⁴⁵ . Kai kurie atvirų duomenų portalai leidžia pasirinkti vieną kategoriją, kiti kelias.
Raktiniai žodžiai	Raktiniai žodžiai dažnai yra keli atskiri esmę arba pagrindinius duomenų elementus nusakantys žodžiai, kurie padeda naudotojams atrasti duomenis. Reikia įtraukti žodžius suprantamus ir naudojamus techniniams ir netechniniams naudotojams. Raktiniai žodžiai taip pat gali būti kuriami rekomendacijų pagrindu, kurios padėtų naudotojams atrasti panašius duomenų rinkinius.
Pakeitimo data	Naujausia data, kai duomenų rinkinys buvo keistas ar atnaujintas.
Leidėjas	Organizacija, kuri skelbia metaduomenis

⁴⁴ „Getting meta with metadata“, What Works Cities, Bloomber Philantropics.

⁴⁵ Tai nėra hierarchinis sąrašas (pvz., šiame lygmenyje nėra subkategorijų), bet gali būti pridedamas prie žodyno, pavyzdžiui, nustatant „Kategorija“ kaip privalomą laukelį ir „Subkategorija“ kaip rekomenduojamą

Licencija	Dažnai atvirų duomenų portalai yra prieinami be jokių apribojimų pakartotiniam naudojimui (tai pažymima Paslaugų taisyklėse arba Duomenų politikoje), tačiau gali būti, kad tam tikras duomenų rinkinys yra prieinamas naudojant skirtingą licenciją.
Sudėtingesni laukeliai	
Dažnis	Duomenų rinkinio atnaujinimo dažnis, pvz., Niekada, Kas valandą, Kasdieną, Savaitės dienomis, Kas savaitę, Kas pusę mėnesio, Kas mėnesį, Kas ketvirtį, Kas pusę metų, Kas metus ir pan. Tai padeda lankytojams suprasti, kaip dažnai jie turėtų ieškoti naujos informacijos, taip pat programuotojams, kurie gali nustatyti automatinius atsisiuntimus.
Laiko intervalas	Laiko intervalas, kurį apima duomenų rinkinys. Gali būti nurodomas bendras intervalas visiems įrašams arba apimti anksčiausio ir vėliausio duomenų rinkinio laiko intervalus.
Erdvinė aprėptis	Vieta, geografinė sritis, kuriai duomenų rinkinys yra aktualus. Dažniausiai naudojamas vietos pavadinimas, kuris turi aiškias ribas (pvz., miesto, rajono).

3.3.3 DCAT-AP v1.1 standarto pakeitimai Lietuvoje

Toliau esančioje lentelėje pateikiami standartiniai DCAT-AP v1.1 klasės tipai ir argumentai dėl šių tipų naudojimo Lietuvoje.

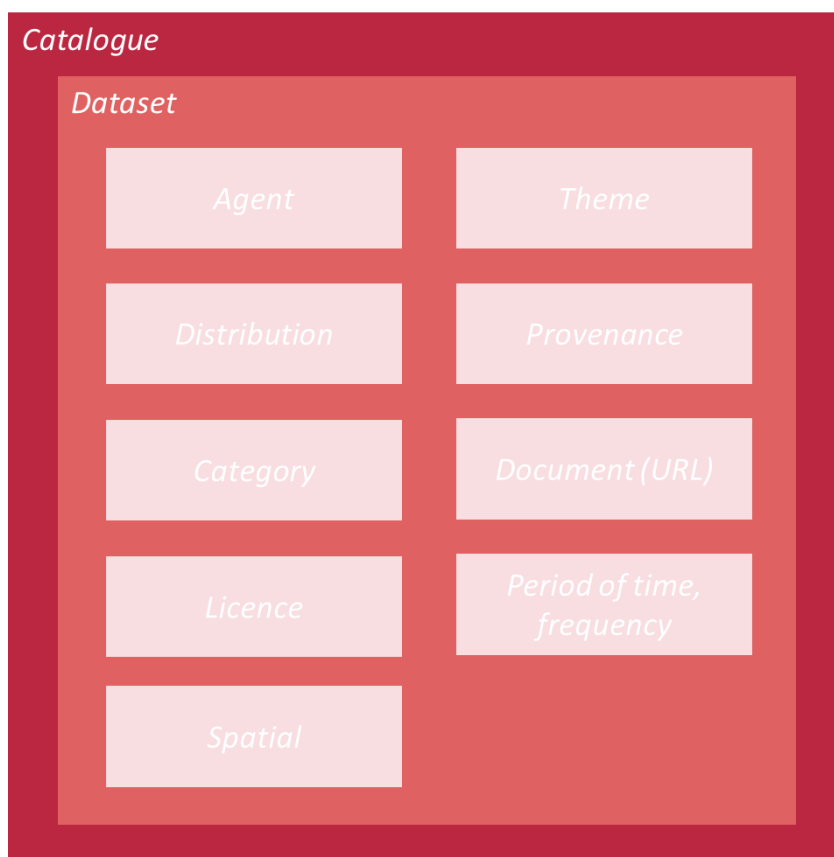
Lentelė Nr. 10. Siūlomi DCAT-AP v1.1 standarto pakeitimai Lietuvoje

Klasė	DCAT standarto nutaitytas klasės tipas	Lietuvos standartui siūlomas klasės tipas	Argumentas
Distribution (Platinimas)	Rekomenduojama	Privaloma	Platinimas nurodo būdą, kuriuo yra duomenys yra publikuojami, todėl svarbu nurodyti publikavimo formatą ir pateikti URL nuorodą numatytose savybėse.
Period of time (laiko periodas)	Neprivaloma	Rekomenduojama	Paskelbtų duomenų galiojimo laikas.
Catalogue Record: release date (paskelbimo data)	Neprivaloma	Rekomenduojama	Šis laukelis yra neprivalomas, nes <i>Catalogue</i> klasėje jis jau yra privalomas.
Dataset:release date (paskelbimo data)	Neprivaloma	Rekomenduojama	Šis laukelis leidžia identifikuoti atvirų duomenų paskelbimo datą ir taip įvertinti jų aktualumą. Tai gali daryti įtaką atvirų duomenų pakartotiniam naudojimui.

Dataset:update/ modification date (atnaujinimo/ pakeitimo data)	Neprivaloma	Rekomenduojama	Duomenų rinkinio atnaujinimo data, kai paskelbiama nauja duomenų rinkinio versija
Dataset:version (versija)	Neprivaloma	Rekomenduojama	Šis laukelis leidžia turėti skirtingą duomenų rinkinio versiją. Tai gali padidinti duomenų pakartotinį naudojimą.
Dataset:version notes (versijos pastabos)	Neprivaloma	Rekomenduojama	Jei duomenis skelbianči organizacija paskelbtų naują versiją, jie taip pat turėtų tai pažymėti.
Distribution: media type (medijos tipas)	Neprivaloma	Rekomenduojama	Medijos tipas gali būti lemiamas kriterijus potencialiam duomenų naudotojui renkantis duomenis. Todėl rekomenduojama užpildyti šį laukelį. Medijos tipas nurodo failo medijos tipą (pvz. paveikslėlis, Excel, kt.)

3.3.4 Atvirų duomenų metaduomenų hierarchinė struktūra

Toliau nurodyta schema vaizduoja pagrindinių metaduomenų klasių ir savybių hierarchinę struktūrą.



Paveikslas Nr. 6. Pagrindinių metaduomenų klasių ir savybių hierarchinė struktūra (techniniai pavadinimai pateikiami anglų kalba)

Kaip matoma, aukščiausia klasė yra Katalogas. Duomenų rinkinys apima klases ir savybes, kurios yra reikalingos duomenų rinkinio aprašymui. Pastaba: šiame paveiksle nurodomi tik pagrindiniai elementai⁴⁶.

3.3.5 Pagrindinės techninės klasės^{47 48}

Toliau aprašomos pagrindinės techninės klasės paremtos DCAT, kurios yra naudojamos metaduomenų struktūroje. Ši dalis yra paremta DCAT specifikacija ir pritaikyta Lietuvos kontekstui. Kitose lentelėse pateikiamos Lietuvai rekomenduojamos techninės klasės.

Lentelė Nr. 11. Pagrindinės DCAT techninės klasės ir siūloma jų svarba Lietuvoje

Techninis pavadinimas	Svarba	Aprašymas
Catalogue (Katalogas)	Privaloma	Katalogas ar saugykla, kurioje yra aprašomų duomenų rinkinių. Katalogas apibūdina jame esančius duomenų rinkinius.

⁴⁶ Daugiau informacijos pateikiama: https://www.w3.org/2011/gld/wiki/Dcat_examples

⁴⁷ Pilna specifikacija yra nurodoma W3C tinklapyje, taip pat *Joinup* platformoje: <https://dvc.w3.org/hg/gld/raw-file/default/dcat/index.html>

⁴⁸ Toliau esančiose lentelėse, visi techniniai pavadinimai nurodomi *kursyvu*

Techninis pavadinimas	Svarba	Aprašymas
Dataset (Duomenų rinkinys)	Privaloma	Konceptualus subjektas, kuris apibūdina skelbiamą informaciją.
Agent (Agentas)	Privaloma	Subjektas, kuris yra susijęs su Katalogais ir/su Duomenų rinkiniais. Jei Agentas yra organizacija, rekomenduojama naudoti Organizacijos ontologiją.
Literal (Reikšmė)	Privaloma	Reikšmė, pvz., ženklų eilutė arba sveikasis skaičius
Resource (Šaltinis)	Privaloma	Šaltiniai, gali būti skirstomi į grupes, apibūdinami RDF.
Distribution (Platinimas)	Privaloma ⁴⁹	Fizinis duomenų rinkinio paskelbimas tam tikrame formate.
Category scheme (Kategorijos schema)	Rekomenduojama	Rinkinys, kuriame apibrėžiama kategorija (pvz., kontroliuojamas žodynas).
Category (Kategorija)	Rekomenduojama	Duomenų rinkinio tema.
License document (Licencijos dokumentas)	Privaloma ⁵⁰	Teisinis dokumentas, kuris leidžia naudoti išteklių tam tikru būdu.
Period of time (laiko periodas)	Rekomenduojama ⁵¹	Laiko intervalas, apibrėžiamas pradžios ir pabaigos datomis.
Catalogue record (Katalogo įrašas)	Neprivaloma	Ši klasė yra neprivaloma ir naudojama ne visuose kataloguose. Ji yra naudojama, kai duomenų rinkinio metaduomenys ir duomenų rinkinio įrašo kataloge metaduomenys yra skirtingi. Pavyzdžiui, duomenų rinkinio skelbimo datos savybė nurodo datą, kada organizacija pirmą kartą paskelbė informaciją, o katalogo įrašo skelbimo datos savybė nurodo kada duomenų rinkinys buvo pridėtas į katalogą. Tokiu atveju, kai abi datos skiriasi, arba kai yra žinoma tik antroji, skelbimo data turėtų būti pažymėta tik prie katalogo įrašo.
Checksum (Kontrolinė suma)	Neprivaloma	Reikšmė, pagal kurią autentifikuojamas failo turinys. Ši klasė nurodo skirtingų kontrolinių ir kriptografinių pranešimų rinkimo algoritmų rezultatus.
Document (Dokumentas)	Neprivaloma	Teksto šaltinis, skirtas žmonėms, kuriame pateikiama informacija, pvz., apie duomenų rinkinio tinklą.

⁴⁹ Rekomenduojama DCAT-AP v 1.1

⁵⁰ Rekomenduojama DCAT-AP v 1.1

⁵¹ Neprivaloma DCAT-AP v 1.1

Techninis pavadinimas	Svarba	Aprašymas
Frequency (Dažnis)	Neprivaloma	Dažnis, kuriuo kas nors įvyksta, pvz., paskelbiamas duomenų rinkinys.
Identifier (Identifikatorius)	Neprivaloma	Identifikatorius, tam tikrame kontekste, sudarytas iš ženklų eilutės, kuri yra identifikuojama
Kind (Tipas)	Neprivaloma	vCard (el. vizitinės kortelės) specifikacijos aprašymas, pvz., nurodomas telefono numeris ir el. pašto adresas. Ši klasė yra pagrindinė keturiems vCard tipams (asmeninė, organizacija, vieta, grupė)
Linguistic system (Lingvistinė sistema)	Neprivaloma	Ženklų, simbolių, garsų, gestų ar taisyklių sistema, kuri yra naudojama komunikacijai, pvz., kalba
Location (Vieta)	Neprivaloma	Erdvinis regionas ar vietos pavadinimas. Gali būti pateiktas naudojant kontroliuojamą žodyną arba geografinėmis koordinatėmis.
Media type or extent (Medijos tipas ar apimtis)	Neprivaloma	Medijos tipas ar apimtis, pvz., kompiuterinio failo formatas
Publisher type (Leidėjo tipas)	Neprivaloma	Organizacijos, kuri skelbia duomenis, tipas. Šis tipas turėtų būti nurodytas naudojant kontroliuojamą žodyną, koduojamą kaip <i>SKOS Concept</i> ⁵² .
Rights statement (Teisių pareiškimas)	Neprivaloma	Intelektinės nuosavybės teisių pareiškimas, pateikiamas duomenų rinkinyje ar per jį, teisiniame dokumente, duodantis leidimą naudoti duomenis, arba prieigos teisių pareiškime
Standard (Standartas)	Neprivaloma	Standartas ar kita specifikacija, kurią atitinka duomenų rinkinio ar platinimo aprašymas
Status (Būsena)	Neprivaloma	Platinimo termino arba katalogo įrašo keitimo nurodymas.
Provenance Statement (Kilmės dokumentai)	Neprivaloma	Nurodo bet kokius duomenų nuosavybės ar saugojimo pasikeitimus nuo jų sukūrimo, kurie yra svarbūs jų autentiškumui, vientisumui ir interpretacijai.

Klasių savybių detalizavimas pateikiamas priede *Priedas Nr. 1. Klasių savybių detalizavimas*.

3.3.6 Kontroliuojami žodynai

Kontroliuojamas žodynas yra ribotas žodžių ar terminų sąrašas, naudojamas ženklinimui, indeksavimui ar suskirstymui pagal kategorijas. Jis yra kontroliuojamas, nes tik terminai iš sąrašo gali būti naudojami

⁵² Daugiau informacijos pateikiama: <http://www.w3.org/2009/08/skos-reference/skos.html#Concept>

subjekto srities aprašymui, kuriai yra taikomas kontroliuojamas žodynas. Kontroliuojami žodynai turėtų:

- Būti skelbiami su atvirąja licencija;
- Būti naudojami ir prižiūrimi ES institucijų, pripažinto standarto organizacijos ar kitų patikimos organizacijos;
- Būti tinkamai dokumentuojami;
- Būti paženklinami keliomis kalbomis, geriausiai, visomis oficialiomis ES kalbomis;
- Turėti palyginti mažai terminų (pvz., 10-25), kurie būtų pakankamai bendriniai ir jais būtų galima klasifikuoti išteklius;
- Turėti terminus, kuriuos URI gali identifikuoti, su kiekvienu URI nustatant termino dokumentaciją;
- Turėti susijusią išlaikymo ir versijų politiką.

Toliau lentelėje pateikiami URI savybių sąrašas ir standartinis žodynas.

Lentelė Nr. 12. URI savybės ir atitinkantis standartinis žodynas

URI savybė	Žodynas, aprašymas ir URI
dcat:mediaType	IANA <i>Media Types</i> (IANA medijų tipai): https://www.iana.org/assignments/media-types/media-types.xhtml Pvz.: cvs, pdf, vnd.ms-excel
dcat:theme	<i>Dataset Theme Vocabulary</i> : http://publications.europa.eu/mdr/resource/authority/data-theme/html/data-theme-eng.html Pvz.: AGRI - Žemės ūkis, žuvininkystė, miškininkystė ir maistas, EDUC - Švietimas, kultūra ir sportas.
dcat:themeTaxonomy	Pačio temų žodyno URI, pvz., sąvokų schema, bet ne sąvokų, esančių žodyne, URI.
dct:accrualPeriodicity	<i>MDR Frequency Named Authority List</i> http://publications.europa.eu/mdr/resource/authority/frequency/html/frequencies-eng.html Pvz.: ANNUAL – kasmet, DAILY – kas dieną.
dct:format	<i>MDR File Type Named Authority List</i> http://publications.europa.eu/mdr/authority/file-type/ Pvz.: PDF, XML, DOC ir kiti.
dct:language	http://publications.europa.eu/mdr/authority/language/index.html
dct:publisher	<i>Corporate bodies NAL</i> gali būti naudojami Europos institucijoms ir mažiems tarptautinių organizacijų rinkiniams. Kitų organizacijų atveju, reiktų naudoti nacionalinius, regioninius arba vietinius žodynus. Pvz.: AI – Tarptautinė amnestija, CURIA – Europos Sąjungos Teisingumo Teismas, UNCRD – Jungtinių Tautų regioninės plėtros centras.

dct:spatial	<i>MDR Continents Named Authority List, MDR Countries Named Authority List, MDR Places Named Authority List, Geonames.</i> Šie sąrašai turi būti naudojami aprašant žemynus, šalis ir vietas, kurie yra sąrašuose. Jei tam tikros vietos nėra sąrašuose, turi būti naudojami <i>Geonames</i> URI. Naudojant geografines koordinates, specifikacijoje rekomenduojama vadovautis <i>Core Location Vocabulary</i> aprašytu GeoDCAT-AP specifikacijoje. Pvz.: LTU_VNO – Vilnius, Lietuva
adms:status	<i>Catalogue Record</i> klasėje turi būti naudojamas ADMS pokyčių tipų žodynas: ADMS būsenų žodyno terminai yra nurodyti ADMS specifikacijoje. :created (sukurtas), :updated (atnaujintas), :deleted (ištrintas) https://www.w3.org/TR/vocab-adms/
adms:status	<i>Distribution</i> klasėje turi būti naudojamas ADMS būsenų žodynas. ADMS būsenų žodyno terminai yra nurodyti ADMS specifikacijoje. :created (sukurtas), :updated (atnaujintas), :deleted (ištrintas) https://www.w3.org/TR/vocab-adms/
dct:type	<i>Agent</i> klasėje turi būti naudojamas ADMS leidėjo tipų žodynas ADMS leidėjo žodyno terminai yra nurodyti ADMS specifikacijoje. https://www.w3.org/TR/vocab-adms/
dct:type	ADMS licencijos tipų žodynas ADMS licencijos žodyno terminai yra nurodyti ADMS specifikacijoje. https://www.w3.org/TR/vocab-adms/

Kategorizacijai Europos atvirų duomenų portalas rekomenduoja naudoti toliau lentelėje nurodomas kategorijas.

Lentelė Nr. 13. Europos atvirų duomenų portalų kategorijos

Kategorija	Vidinis kategorijos pavadinimas
Agrikultūra, žvejyba, miškininkystė, maisto pramonė	agriculture-fisheries-forestry-food
Švietimas, kultūra ir sportas	education-culture-and-sport
Aplinka	environment

Energija	energy
Transportas	transport
Mokslas ir technologijos	science-and-technology
Ekonomika ir finansai	economy-and-finance
Populiacija ir socialinės sąlygos	population-and-social-conditions
Sveikata	health
Vyriausybė, viešasis sektorius	government-public-sector
Regionai, miestai	regions-cities
Teisinė sistema, viešasis saugumas	justice-legal-system-public-safety
Tarptautiniai klausimai	international-issues

3.4 Metaduomenų struktūros aprašo pavyzdžiai

3.4.1 DCAT pagrindinių metaduomenų aprašo pavyzdys

Toliau pateikiamas DCAT-AP naudojimo pavyzdys ir pagrindinių terminų paaiškinimas.

Scottish Road Accident Statistics *dot:title (Dataset)*

Data about injury road accidents, accident costs, vehicles involved, drivers and riders, drink-drive accidents, drivers breath tested, casualties and international comparisons. *dot:description*

Source agency: Scottish Government

Designation: National Statistics

Language: English *dot:language*

Alternative title: Scottish Road Accident Statistics

Licence *dot:license*
UK Open Government Licence (OGL) [OPEN DATA](#)

Data Resources **2**

- Key statistics for 2007 *dot:accessURL*
- 2007 Volume *dot:title (Distribution)*

Additional Information

Openness score	★★★★★
Geographic coverage	Scotland <i>dot:spatial</i>
National statistic	yes
ONS Category	Travel and Transport <i>dot:theme</i>
Temporal coverage	No value <i>dot:temporal</i>
Date added computed	No value
Date updated computed	No value

Publisher
Scottish Government

Enquiries:
No details supplied

FOI Contact: *adms:contactPoint*
Web: <http://www.whatdotheyknow.com...>

Tags
accident, health-well-being-and-care, road, road-accidents, road-safety, roads, safety, transport, transport-accidents-and-casualties, travel-and-transport *dot:keyword*

About this dataset

- Added to data.gov.uk: 10/12/2011 *dot:issued*
- Modified on data.gov.uk: 10/06/2013 *dot:modified*
- History of changes
- JSON, API and URI for developers

Do more with this data

- Share your app
- Share an idea
- Request new data

Paveikslas Nr. 7. Junginės Karalystės duomenų portalio metaduomenų aprašo pavyzdys

Lentelė Nr. 14. Pagrindinių DCAT terminų paaiškinimai

DCAT terminai	Paiškinimas
dot:title (Dataset)	Duomenų rinkinio pavadinimas

dct:description	Duomenų aprašymas
dct:language	Kalba
dct:license	Licencija
dcat:accessURL	Prieigos nuoroda
dct:title (Distribution)	Duomenų rinkinio pavadinimas (čia „2007 Volume“)
dcat:downloadURL	Parsisiuntimo prieiga prie duomenų failo
dct:format	Formatas
dct:spatial	Geografinė aprėptis
dct:theme	Tema
dct:temporal	Laiko aprėptis
dct:publisher	Duomenų leidėjas
adms:contactPoint	Kontaktai
dcat:keyword	Raktiniai žodžiai
dct:issued	Paskelbimo data
dct:modified	Paskutinio pakeitimo data

3.4.2 RDF formatas

RDF formato metaduomenų elementai pateikiami priede *Priedas Nr. 2. RDF failo metaduomenų struktūros aprašymas*.

3.4.3 JSON formatas

JSON formato pavyzdys yra pateikiamas priede *Priedas Nr. 3. JSON failo metaduomenų struktūros aprašymas*.

4 Standartinės metaduomenų struktūros taikymo atvirų duomenų metaduomenų formavimui procedūros

4.1 Siūlomos atvirų duomenų metaduomenų valdymo procedūros Lietuvoje

Europos atvirų duomenų portalas⁵³ pateikia gaires dėl atvirų duomenų skelbimo. Procesas organizuojamas įvairiuose duomenų skelbimo sub-procesuose: duomenų surinkime, parengime, paskelbime ir palaikyme. Taikant šį procesą organizacijoje yra sukurama struktūrizuota atvirų duomenų sistema.

Bendrai, rekomenduojama vadovautis Europos rekomendacijomis dėl atvirų duomenų valdymo ir skelbimo. Kadangi metaduomenų skelbimas yra atliekamas skelbiant pačius atvires duomenis⁵⁴, metaduomenų aprašymas pateiktas atvirų duomenų skelbimo kontekste.

Toliau šioje dalyje aprašomi metaduomenų valdymo veiksmai.

4.1.1 Duomenų rinkimas

Duomenų rinkimas priklauso nuo organizacijos pobūdžio, bet dažniausiai apima prieinamų duomenų rinkinių atradimą, prioretizavimą, praktikavimą, temų ir kategorijų paskelbimą.

Visų pirma, svarbu atsirinkti esamus duomenis, kad būtų žinoma, kokie duomenys jau yra prieinami organizacijos viduje. Tai turėtų būti daroma kartu su darbuotojais, atsakingais už duomenų valdymą, sudarant visų duomenų sąrašą ir kur jie yra saugomi (duomenų inventORIZACIJA). Kai sąrašas yra paruoštas, duomenis gali būti prioretizuojami, nutariant kuriuos duomenų rinkinius reikia paskelbti anksčiausiai.

Atsakymai į šiuos klausimus gali padėti prioretizuoti duomenų rinkinius:

- Ar tai gali būti paskelbta (teisiškai, politiškai ir organizaciniu požiūriu)?
- Ar duomenys yra tinkamos kokybės (ir ar prieš paskelbiant nereikalingi pokyčiai)?
- Ar atliktas valymas, duomenys yra anoniminiai, geros kokybės ir tinkamo formato?
- Ar jie priklauso vienai iš vertingų temų?

Airijos Gerųjų praktikų vadovas (angl. *Irish Best Practice Handbook*) aprašė geriausias praktikas, susijusias su esamų duomenų auditu, ir pasiūlė būdus, kaip sužinoti apie organizacijoje prieinamus duomenų rinkinius (pateikiama toliau esančiuose poskyriuose).

4.1.1.1 Duomenų auditas

Pirmiausia rekomenduojama atlikti duomenų, kurie jau tinkami skelbimui, auditą. Atvirų duomenų kontekste, „duomenų rinkiniai“ apima duomenų bazines, registrus, vykstančius duomenų rinkimą ir tyrimus, geo-erdvinius duomenų rinkinius ir kita. Gryna tekstinė informacija, tokia kaip el. laišakai, užrašai, ataskaitos ir spaudos pranešimai, nėra laikomi „duomenimis“ šiame kontekste.

Galima vadovautis šiais duomenų audito būdais:

Vidinis duomenų auditas: vidinio duomenų audito tikslas yra nustatyti duomenų rinkinius, kurie nėra paskelbti tinklapiuose. Tai gali atlikti Atvirų duomenų pareigūnas, paskirtas kiekvienoje viešojoje organizacijoje. Vidinis duomenų auditas gali apimti darbuotojų interviu ir techninių išteklių apžvalgą, pvz., sąryšines duomenų bazines, bendras failų saugyklas ar asmeninius kompiuterių diskus.

⁵³ Šaltinis: <https://www.europeandataportal.eu/en/providing-data/goldbook/putting-place-open-data-lifecycle>

⁵⁴ CKAN pavyzdys: <http://docs.ckan.org/en/latest/user-guide.html#features-for-publishers>

Internetinis duomenų auditas: gali būti peržiūrėti viešųjų organizacijų tinklapiai, kad būtų nustatyti duomenų rinkiniai, kurie jau paskelbti internete. Atitinkami raktiniai žodžiai, kurių galima ieškoti, gali būti tokie terminai kaip „duomenys“, „duomenų bazė“, „atsisiųsti“, „API“, „eksportas“, „statistika“, „ataskaitos“. Paieškai tinkami failų formatai apima:

- *Microsoft Excel* (XLS, XLSX)
- *Comma-Separated Values* (CSV, TSV)
- *Extensible Markup Language* (XML)
- *Keyhole Markup Language and Geography Markup Language* (KML, GML)
- *Resource Description Framework* (RDF)
- *ESRI Shapefiles* (SHP)
- Aukščiau išvardintų formatų archyvuotos formos (ZIP, GZ, BZ2, kt.)

4.1.1.2 Duomenų apsaugos įstatymų laikymasis

Atrenkant duomenis paskelbimui, svarbu įsitikinti, ar atviri duomenys atitinka duomenų apsaugos įstatymus galiojančius Lietuvoje⁵⁵ ir Europoje. Atvirieji duomenys neturėtų įtraukti asmeninių ar slapčių duomenų, kurie gali būti susieti su asmeniu. Visais atvirų duomenų skelbimo atvejais, piliečio teisė į privatumą turi būti apsaugota ir duomenų leidėjas turi laikytis visų duomenų apsaugos principų.

4.1.1.3 Tinkamos licencijos pasirinkimas⁵⁶

Atvirų duomenų išskirtinumas yra tai, kad jie gali būti laisvai naudojami, įskaitant naudojimą komerciniais tikslais. Atvirosios licencijos ir atvirų duomenų susiejimas yra reikalingas, siekiant užtikrinti teisinį pagrindą pakartotiniam panaudojimui.

Kūrybinis turinys, kaip testas, fotografijos, skaidres ir kita, turėtų būti licencijuojamos naudojant *Creative Commons* licenciją. Toliau lentelėje nurodomos trys šios licencijos, naudojamos atvirajam turiniui.

Lentelė Nr. 15. *Creative Commons* licencijos

Licencijos lygis	<i>Creative Commons</i> licencija
Viešojo nuosavybė	CCo
Priskyrimas (angl. <i>attribution</i>)	CC-by
Priskyrimas ir panašaus turinio dalinimasis	CC-by-sa

Atvirų duomenų institutas rekomenduoja naudoti *Creative Commons 4.0* licenciją ne tik turiniui, bet ir duomenims. Taip pat galima naudoti panašias licencijas, kurios buvo sukurtos duomenų bazėms naudojant *Open Data Commons* licenciją. Toliau nurodomi trys licencijos lygiai, kuriuos galima pasirinkti.

Lentelė Nr. 16 *Open Data Commons* licencijos

Licencijos lygis	<i>Open Data Commons</i> licencija
Viešojo nuosavybė	PDDL
Priskyrimas (angl. <i>attribution</i>)	ODC-by

⁵⁵ Šaltinis: <https://www.e-tar.lt/portal/lt/legalAct/TAR.5368B592234C/XSpzxvEjIg>

⁵⁶ Kaip nurodoma Europos atvirų duomenų portale: <https://www.europeandataportal.eu/en/resources/training-companion/open-data-licensing> ir ODI tinklapyje: <https://theodi.org/guides/publishers-guide-open-data-licensing>

Priskyrimas ir panašaus turinio dalinimasis ODbL

Norint nurodyti turinio ar duomenų licenciją, reikia naudoti ir žmonėms suprantamą ir automatiškai apdorojamą metaduomenų aprašymą. Kuo aiškiau licenciją, kuri taikoma turiniui ar duomenims, yra pateikiama, tuo lengviau bus naudotojams sužinoti, kad jie gali naudoti licencijuojamą turinį ar duomenis. Žmonėms suprantami aprašymai ir žymos yra pateikiami *Creative Commons* ir *Open Data Commons* licencijų tinklalapiuose:

- *Creative Commons licence chooser*
- *Open Data Commons licences*

Pasirinkus licenciją, reikia nurodyti oficialią URI nuorodą, kaip parodyta toliau esančiame pavyzdyje iš Belgijos atvirų duomenų portalų.

Publisher

Organization:	IWEPS
Author:	IWEPS
E-mail:	walstat@iweps.be
License	CC Zero (CC 0)
Conditions:	https://creativecommons.org/publicdomain/zero/1.0/

Paveikslas Nr. 8. Licencijos pateikimo pavyzdys iš Belgijos atvirų duomenų portalų.

Daugiau informacijos apie europines licencijas galima rasti Europos atvirų duomenų portale: <https://www.europeandataportal.eu/en/content/show-license>.

4.1.2 Duomenų ir metaduomenų paruošimas

Pirminius duomenis, siekiant padidinti jų aktualumą naudotojui, reikia apdoroti. Metaduomenys yra svarbūs pakartotiniam atvirų duomenų panaudojimui. Jie padidina duomenų atradimą ir pakartotinį duomenų naudojimą. Todėl svarbu skirti laiko informuojant pakartotinį naudotoją apie duomenų rinkinio kokybę ir pateikti metaduomenis.

Aprašytus duomenų rinkinius yra lengviau interpretuoti. Susietų duomenų kontekste, pridėdam metaduomenis su URI, duomenys gali būti susieti. Tai pagerina duomenų atradimą ir sąsajas. Svarbu žinoti, jog siekiant, kad kiti atvirų duomenų portalai, tokie kaip Europos atvirų duomenų portalas, surinktų duomenis, metaduomenys yra privalomi.

Kokybiškų metaduomenų pateikimas gali būti sudėtingas, bet būtinas. Pasaulinio Žiniatinklio konsorciūmo (W3C) fondas sukūrė gaires ir geriausias praktikas, skirtas duomenų savininkams. Taip pat, sąveika su Europos atvirų duomenų portalu yra svarbi, todėl reikia vadovautis DCAP-AP specifikacija Lietuvai, kaip nurodyta dalyje 3. *Valstybės informacinių išteklių duomenų metaduomenų standartinio aprašo struktūra.*

Bendrai, metaduomenys turi būti skelbiami kartu su duomenimis automatiškai apdorojamu formatu ir standartiniais terminais, skirtais metaduomenų aprašymui. Taip pat, reikia aprašyti bendras duomenų rinkinio savybes apie vietinius parametrus, licenciją, kilmę ir kokybę.

Metaduomenų paruošimui reikia:

1. Pasirinkti kategoriją, kuriai priskiriami duomenys. Tam reikia naudoti standartinės kategorijas, pateiktas atvirų duomenų portale⁵⁷.

⁵⁷ Rekomenduojame naudoti G8 atvirų duomenų kategorijas (angl. *G8 Open Data Charter Category*): <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex#technical-annex>

2. Užpildyti visus privalomus laukelius ir standartinių reikšmių pateikimui naudoti kontroliuojamus žodynus, kur to reikalaujama. Aprašant metaduomenis, svarbu :
 - a. Pateikti aiškų duomenų rinkinio aprašymą;
 - b. Pateikti aišką ir prasmingą duomenų rinkinio pavadinimą;
 - c. Pateikti duomenų versijavimą ar datą;
 - d. Pateikti URI, kuriuo duomenis galima parsisiųsti;
 - e. Pateikti duomenų rinkinio licenciją.
3. Kur įmanoma, užpildyti rekomenduojamus laukelius pagal Lietuvos DCAT-AP. Tai padidins atvirų duomenų pakartotinį naudojimą ir suteiks daugiau vertės naudotojams.
4. Paruošti skelbiamus duomenis automatiškai apdorojamu formatu. Rekomenduojama viską skelbti RDF formatu. Jei tai negalima, aiškus CVS failas taip pat gali būti pakartotinai naudojamas.

4.1.3 Metaduomenų skelbimas

Metaduomenis galima paskelbti dviem būdais. Organizacijos gali pasirinkti tinkamiausią skelbimo būdą:

- Per atvirų duomenų portalo grafinę naudotoją sąsają (GUI) – leidėjas pateikia reikalingą informaciją pagal atvirų duomenų portale nurodytą procedūrą;
- Per atvirų duomenų portalo aplikacijų programavimo sąsają (API) – leidėjas turėtų vadovautis API ir Lietuvos DCAT-AP reikalavimais. Leidėjai turėtų naudoti REST API, kad susisiektų su atvirų duomenų portalu ir pateiktų užklausą dėl „duomenų rinkinio sukūrimo: paslaugos.

Taip pat, CKAN sistema pateikia standartinį metaduomenų įkėlimo mechanizmą. Tam yra išsamus ir funkcinis CKAN duomenų rinkinio schemas susiejimas su susietųjų duomenų formatais, įskaitant W3C rekomenduojamą DCAT žodyną. Norint pasiekti susietųjų duomenų formatus, CKAN API galima naudojama įprastu būdu, tačiau nurodant norimą gauti formatą.

Idealiu atveju, atviri duomenys turėtų būti skelbiami naudojant URI, kad duomenis būtų galima pasiekti internete. Jie leidžia atskirti konkretų išteklių keliuose duomenų rinkiniuose. Jie ypač svarbūs referencinių duomenų rinkiniams, kuriuos naudoja kelios agentūros, pvz., kelių numeriai, įmonių registracijos numeriai, teisės aktų numeriai. Taip pat svarbu sukurti bendrą URI struktūrą, kad jie būtų patvarūs, t. y. atsparūs pokyčiams. Be to, skelbiant duomenis su URI, jie gali būti sujungti su kitais duomenimis, taip sukuriant susietuosius duomenis. Šiuo metu, nacionalinė Lietuvos strategija dėl URI kūrimo nėra nustatyta.

4.1.4 Metaduomenų surinkimas (angl. Harvesting metadata)

Nors metaduomenų rinkiniai yra susiję su duomenų rinkiniais, kuriuos jie apibūdina, bet kartais gali būti naudinga pateikti metaduomenis keliose skirtingose vietose. Tai leidžia skirtingiems atvirų duomenų portalams pasiekti skirtingas auditorijas. Pavyzdžiui, Europos atvirų duomenų portalas nurodo ES narių atvirų duomenų portalų metaduomenis.

Taigi, dažnai metaduomenų rinkinius, apibūdinančius tuos pačius duomenų rinkinius, galima rasti skirtinguose atvirų duomenų portaluose.

Metaduomenų pateikimas skirtingiems portalams gali būti atliktas automatiškai arba rankiniu būdu.

Kai metaduomenys yra paskelbti, didesni portalai gali surinkti duomenis. Numatoma, kad Lietuvoje bus vienas centrinis atvirų duomenų portalas, kuris ateityje galėtų sutrinkti atvirose duomenis iš kitų šaltinių (pvz., privačių įmonių, savivaldybių). Tam yra svarbu, kad duomenys būtų skelbiami pagal DCAT-AP specifikaciją, kaip nurodyta šiame dokumente.

4.1.5 Metaduomenų priežiūra

Laikui bėgant duomenys keičiasi. Istoriniai duomenys išlieka tie patys, tačiau kiti duomenys gali būti atnaujinami, svarbu duomenis atnaujinti reguliariai.

Pavyzdžiui, kai metaduomenys keičiasi, reikia atnaujinti duomenų rinkinio metaduomenis, siekiant užtikrinti, kad duomenys būtų tinkamai randami ir pasiekiami.

Todėl naudinga sukurti atnaujinimo procesą, kuris užtikrintų, kad duomenys yra atnaujinami. Taip pat patariama turėti duomenų priežiūros procesą, kuris apima reguliarią duomenų ir metaduomenų priežiūrą, URI ir URL patikrinimą, vartotojų atsiliepimų patikrinimą ir nuolatinį tobulėjimą.

URI ir URL tikrinimas

Internetas yra greitai kintantis, todėl svarbu reguliariai tikrinti, ar visi pateikti URI ir URL (iš ir į duomenų rinkinius) yra veikiantys. Jei duomenų rinkinio URI ar URL pasikeis, naudotojai bus nukreipti į neegzistuojančius puslapius.

Naudotojų atsiliepimų tikrinimas ir nuolatinis gerinimas

Naudotojų atsiliepimai gali padidinti duomenų leidinių kokybę. Naudotojai gali pateikti atsiliepimus apie bet kokių duomenų aspektą, todėl įtraukiant atsiliepimus į savo procesus, galima pagerinti duomenų naudojimą ir radimą.

Metaduomenų atnaujinimo procedūra

Yra trys pagrindiniai veiksniai, kada reikia atnaujinti duomenų rinkinio metaduomenis:

1. Duomenų rinkinio atnaujinimas: kai duomenų rinkinys yra keičiamas, metaduomenys yra atnaujinami (automatiškai ir rankiniu būdu, pvz., aprašymas);
2. DCAT standarto atnaujinimas: kiekvieną mėnesį, asmuo turėtų patikrinti Europos atvirų duomenų portalą ir Joinup, ar nėra suplanuotų DCAT pakeitimų. Jei pakeitimai yra numatyti, turėtų būti įvertintas poveikis metaduomenims ir, jei reikia, atlikti pakeitimai. Jei keičiamas DCAT-AP-LT nacionalinis profilis, tada poveikis ir pakeitimai turėtų būti daromi remiantis siūlomos pakeitimo poveikio analize.
3. Remiantis vartotojų atsiliepimais: jei gaunami prasti atsiliepimai dėl metaduomenų, tada turėtų būti atliekami reikalingi pakeitimai.

4.2 DCAT standarto pritaikymas

DCAT pritaikymas turėtų būti atliekamas atidžiai ir tik po atvirų duomenų portalų paleidimo. Šio dokumento 3 dalyje, yra siūlomos minimalūs DCAT-AP standarto, išleisto ES, pakeitimai. Jei ateityje bus reikalingi kokie nors pakeitimai, jie turėtų būti pagrįsti konkrečiais reikalavimais, kylančiais iš paskelbtų duomenų ir nepatenkinamų vartotojų ir/arba viešųjų administracijų poreikių. Rekomenduojama kiek įmanoma laikytis DCAT-AP standarto, kad būtų užtikrintas nuolatinis suderinamumas su Europos atvirų duomenų portalu.

Jei pritaikymas yra reikalingas, DCAT pritaikymas turi būti atliktas vadovaujantis reikalavimais, nurodytais 3 skyriuje. Organizaciniu požiūriu, prieš keičiant DCAT-AP, rekomenduojama atlikti šiuos žingsnius:

1. Įsitikinti, ar pokyčiai reikalingi dėl konkrečių poreikių ir tai bus naudinga naudotojams ir/ar viešosioms įstaigoms. Nėra tikslinga apsunkinti reikalavimus, jeigu to reikalaujama.
2. Įsitikinti, ar pokyčiai atitinka minimalius DCAT-AP reikalavimus. Jei ne, Europos atvirų duomenų portalas negalės surinkti metaduomenų ir tai pakenks jų pakartotiniam panaudojimui.
3. Įsitikinti, ar pokyčiai yra suderinami su metaduomenų savininku. Pokyčiai turėtų būti aptarti su metaduomenų savininku, kad būtų galima įsitikinti, jog pakeitimai yra aktualūs.
4. Pritaikyti pokyčius šioje specifikacijoje ir juos pagrįsti. Nauja specifikacija turėtų būti paskelbta atvirų duomenų portale ir palikus laiko prisitaikyti prie įvestų pokyčių.

5. Paskelbti pokyčius viešai prieinamoje Lietuvos atvirų duomenų portalų dokumentacijoje ir oficialiai apie tai pranešti visoms atvirus duomenis teikiančioms institucijoms. Reiktų įsitikinti, kad visos viešosios institucijos supranta pakeitimus, nes kitaip pokyčiai neatneš naudos.
6. Pritaikyti visus pakeitimus visiems⁵⁸ metaduomenims, kurie yra pateikiami portale. Atsakomybė už tai gali būti padalinta tarp už atvirų duomenų portalą atsakingos institucijos kitų atvirus duomenis teikiančių institucijų, priklausomai nuo pokyčių apimtys ir esamų įsipareigojimų.
7. Po vieno metų, įvertinti pokyčių rezultatus, remiantis duomenų atvėrimo pažangos ir poveikio vertinimo metodika.

⁵⁸ Arba tuos, kurie yra svarbiausi

5 Reikalavimai metaduomenų tvarkymui ir perdavimui į atvirų duomenų portalą

5.1 Atvirų duomenų portalų pristatymas

Atvirų duomenų portalai palengvina prieigą prie viešojo sektoriaus informacijos ir pakartotinį šios informacijos panaudojimą.

Atvirų duomenų portalai yra žiniatinklio sąsajos, skirtos lengviau rasti informaciją pakartotiniam naudojimui. Juose yra duomenų rinkinių, kuriuos galima pakartotinai naudoti, metaduomenų įrašai, daugiausiai susiję su pirmine ar skaitine informacija, o ne tekstine informacija. Kartu su konkrečiomis paieškos funkcijomis, tai palengvina dominančių duomenų rinkinių paiešką. API taip pat dažnai naudojami, taip programinei įrangai suteikiant tiesioginę ir automatizuotą prieigą prie duomenų.

Atvirų duomenų portalai yra svarbūs elementai daugumoje atvirų duomenų iniciatyvose ir dažnai yra naudojami Europos viešųjų įstaigų ir vietiniu lygiu valstybėse narėse.

Apskritai, atvirų duomenų portalai turi šias tris funkcijas⁵⁹:

- **Randami duomenys**

Atvirų duomenų portalai skatina vartotojus naudoti atvirose duomenis. Platformos leidžia vartotojams greitai rasti ir naudoti duomenis. Jose yra optimizuotos paieškos, kad būtų lengviau randami reikiami ištekliai. Platformose taip pat teikiami duomenų kanalai, kurie yra atrandami ir indeksuojami internete. Šie duomenų kanalai leidžia paieškos varikliams, pvz., Engma.io, indeksuoti atvirose duomenis ir lengviau juos rasti.

- **Nuosekli vartotojo patirtis**

Yra ribotas atvirų duomenų platformų skaičius ir jos yra sukurtos atsižvelgiant į naudotoją – jų išdėstymas ir dizainas gali palengvinti duomenų atradimą.

- **Nuosekli prieiga prie duomenų**

Atvirų duomenų standartų sukūrimas paskatino pastovių API naudojimą atrandant, pasiekiant ir naudojant duomenis.

5.1.1 Funkciniai reikalavimai

5.1.1.1 Dalyviai

Toliau nurodomi dalyviai, kurie yra aprašomi šiame dokumente ir kuriems yra taikomas kiekvienas reikalavimas. Jais nesiekama atspindėti būsimų leidimų sistemos, bet tik nurodyti bendrus reikalavimus kiekvienam dalyviui:

- **Naudotojas** - visi, kurie lankosi portale su tikslu rasti, peržiūrėti ir naudoti duomenis.
- **Leidėjas** - valstybės tarnautojas, atsakingas už duomenų ir kito portalų turinio skelbimą ar atnaujinimą.
- **Administratorius** - valstybės tarnautojas, kuris valdo portalų technines ir konfigūracijos galimybes, pvz., vizualines temas, metaduomenų reikalavimus ir kt.

Leidėjas gali daryti viską, ką gali padaryti naudotojas, o administratorius gali daryti viską, ką gali padaryti leidėjas.

5.1.1.2 Duomenų katalogo funkcijos

Duomenų katalogas yra prieinamų duomenų sąrašas, kuriame kiekvienas katalogo įrašas atitinka vieną ar daugiau duomenų išteklių. Paprastai tai apima aukšto lygio informaciją apie kiekvieną duomenų rinkinį, pvz., pavadinimas, aprašymas, paskelbimo data, kategorija(-os) ir raktiniai žodžiai. Katalogo

⁵⁹ Pagal ES reikalavimus: <https://www.europeandataportal.eu/elearning/en/module8/#/id/co-01>

funkcijos priklauso nuo naudojamo metaduomenų standarto – čia informacija pateikiama pagal DCAT-AP standartą.

Toliau esančioje lentelėje nurodomos numatomos duomenų katalogo funkcijos.

Lentelė Nr. 17. Duomenų katalogo funkcijos

Dalyvis	Funkcija
Naudotojai	gali rasti duomenų rinkinius pagal vieną ar kelis terminus, nurodytus duomenų rinkinio metaduomenimis.
Naudotojai	gali naršyti ir tyrinėti duomenis pagal kategorijas.
Naudotojai	gali parsisiųsti katalogą automatiškai apdorojamu formatu, suderinimu su dažnai naudojama metaduomenų schema (dažniausiai JSON arba RDF).
Naudotojai	gali peržiūrėti duomenų rinkinio santrauką, kurioje pateikiama informacija apie duomenų išteklius, metaduomenis ir kitą susijusią dokumentaciją.
Naudotojai	gali ieškoti duomenų rinkinių pagal terminus, esančius duomenyse.
Leidėjai	gali pridėti, redaguoti ar pašalinti katalogo įrašus.
Leidėjai	gali pridėti ar redaguoti katalogo įrašo metaduomenis.
Leidėjai	gali pažymėti katalogo įrašus kaip privačius, kad jie nebūtų matomi naudotojams.
Administratoriai	gali kurti, redaguoti ir pašalinti metaduomenų kategorijas.
Administratoriai	gali konfigūruoti numatytąjį duomenų katalogų rūšiavimą.
Administratoriai	gali konfigūruoti bendrus duomenų katalogo nustatymus, įskaitant spalvų temą, logotipą ir pavadinimą.
Administratoriai	gali konfigūruoti papildomus bendrus duomenų katalogo nustatymus, įskaitant CSS stilius, pagrindiniame puslapyje rodomos elementus ir jų išdėstymą, rodomus katalogo elementus ir kt.

5.1.1.3 Duomenų saugojimo ir valdymo funkcijos

Duomenų saugojimo ir valdymo reikalavimai yra susiję su duomenų rinkinių prieiga ir archyvavimu. Toliau esančioje lentelėje nurodomos numatytos portalų funkcijos.

Lentelė Nr. 18. Duomenų saugojimo ir valdymo funkcijos

Dalyvis	Funkcija
Naudotojai	gali parsisiųsti saugomus duomenis.
Naudotojai	gali pasiekti saugomus duomenis naudojant REST API.
Leidėjai	gali sukurti naujus duomenų išteklius.
Leidėjai	gali pakeisti ar atnaujinti esamus duomenų išteklius.
Administratoriai	gali valdyti kitų administratorių ir leidėjų leidimus.

Administratoriai gali valdyti portalą turinį, redaguoti kategorijas ir atnaujinti metaduomenų standartus.

5.1.2 Duomenų pateikimo funkcijos

Atvirų duomenų nauda yra sukuriamą, kai jie yra pakartotinai naudojami ar bent jais remiamasi. Jei ne, tuomet jie neturi jokios ekonominės naudos⁶⁰.

Lentelė Nr. 19. Duomenų pateikimo funkcijos

Dalyvis	Funkcija
Naudotojai	gali peržiūrėti duomenų išteklius savo interneto naršyklėje (pvz., lentelės ar žemėlapiu forma).
Naudotojai	gali peržiūrėti duomenis savo mobiliuosiuose įrenginiuose (planšetiniu kompiuteriu, telefonu), automatiškai pertvarkant, kad portalas tilptų į įrenginį.
Naudotojai	gali vizualizuoti duomenų išteklius grafiniu būdu, įskaitant žemėlapius, grafikus, diagramas, ir kt.
Naudotojai	gali išsaugoti savo vizualizacijas su galimais konfigūracijos nustatymais.
Leidėjai	gali kurti duomenų vizualizacijas ir padaryti juos pasiekiamus duomenų kataloge arba duomenų išteklių puslapyje.
Administratoriai	gali kontroliuoti ar naudotojų sukurtos duomenų vizualizacijos yra matomos kitiems naudotojams.

5.1.2.1 Bendruomenės funkcijos

Bendruomenės funkcijos turi leisti dalyviams lengvai dalintis turiniu ir taip pagerinti atvirų duomenų pakartotinio naudojimo galimybes.

Lentelė Nr. 20. Bendruomenės funkcijos

Dalyvis	Funkcija
Naudotojai	gali pateikti komentarus ir atsiliepimus apie atskirus duomenų išteklius.
Naudotojai	gali pateikti užklausą dėl duomenų rinkinių viešo paskelbimo.
Naudotojai	gali pateikti komentarus ir atsiliepimus apie atskirus duomenų katalogus.
Leidėjai	gali peržiūrėti siūlomus duomenų išteklius
Leidėjai	gali atnaujinti siūlomų duomenų išteklių būsenas.
Leidėjai	gali atsakyti dėl siūlomų duomenų išteklių.
Leidėjai	gali peržiūrėti ir atsakyti į naudotojų komentarus ir atsiliepimus.
Leidėjai	gali peržiūrėti ir patvirtinti naudotojų komentarų ir atsiliepimų matomumą.

⁶⁰ Pastaba: Be tikrosios ekonominės vertės, vyriausybė turi užtikrinti savo skaidrumą piliečiams, nes tai yra demokratijos pagrindas ir pagal apibrėžimą, viešieji duomenys priklauso visuomenei.

Administratoriai gali sekti ir analizuoti naudotojų atsiliepimus.

5.1.2.2 Paskyrų ir profilių funkcijos

Toliau esančioje lentelėje nurodomi reikalavimai dėl duomenų portalų naudotojų ir profilių valdymo.

Lentelė Nr. 21. Paskyrų ir profilių funkcijos

Dalyvis	Funkcija
Naudotojai	gali registruoti paskyrą su el. paštu.
Naudotojai	gali sukurti, redaguoti ir pašalinti profilio informaciją, pasiekiamą kitiems naudotojams.
Leidėjai	gali pateikti komentarus ir atsiliepimus apie duomenų katalogą.
Administratoriai	gali valdyti kitų administratorių ir leidėjų leidimus.
Administratoriai	privalo galėti apriboti naudotojų paskyras (jei paskyros yra galimos duomenų portale).
Administratoriai	privalo galėti pasiekti vartotojo prašymu pateiktais asmens duomenis.

5.1.3 Techniniai reikalavimai

Toliau pateikiami bendrieji techniniai reikalavimai, kurie nepriskiriami konkrečiam dalyviui:

- Duomenų ištekliai privalo būti prieinami atsisiuntimui pirminiu formatu, kuriuo buvo paskelbtas portale.
- Portalas turėtų būti suderinamas su Europos atvirų duomenų portalu ir pasiekiamas per REST API.
- API užklauskos, kurios yra negalimos ar nesėkmingos, privalo grąžinti tinkamą HTTP atsakymo būseną.
- Portalas turi būti pasiekiamas interneto domeno pavadinimu (pvz., opendata.lt).
- Visiems veiksams turi būti naudojamas šifravimas, pvz., TLS (*Transport Layer Security*).
- Tinklapiai turi būti prieinami didelių ir mažų ekranų įrenginiuose.

5.2 Atvirų duomenų platformos

Rinkoje yra daugybė platformų, kuriose galima skelbti atvirose duomenis. Europos atvirų duomenų portalas rekomenduoja šias platformas:

- CKAN platforma – atviros kodo duomenų katalogas su minimaliomis funkcijomis ir paprasta sąsaja;
- *Git Data Publisher* – paprastas duomenų skelbimo įrankis;
- *Socrata* – aukštos kokybės komercinė duomenų platforma su pažangiais vizualizavimo įrankiais;
- *OpenDataSoft* – vidutinio lygio komercinė platforma su stipriais vizualizacijos įrankiais ir katalogų funkcijomis.

Lietuvai rekomenduojame išlaikyti kaip įmanoma paprastesnę. Todėl siūlome naudoti CKAN, kaip JK, Italijoje, Nyderlanduose, Australijoje ir kitur. CKAN suteikia visas funkcijas, išvardintas aukščiau, taip pat REST API, kuri leidžia šias paslaugas:

- Portalų duomenų rinkinių, grupių ar kitų CKAN objektų gavimą JSON formatu;
- Pilną duomenų rinkinio, išteklių ar kito objekto JSON pavaizdavimą;

- Paketų ar išteklių paieška pagal atitinkamas užklausas;
- Duomenų rinkinių, išteklių ir kitų objektų kūrimas, atnaujinimas ir pašalinimas;
- Neseniai portale pakeistų duomenų rinkinių srauto pateikimas.

Norėdami gauti išsamią CKAN naudotojo instrukciją, naudokite jos dokumentaciją, kuri detaliai aprašo, kaip skelbti atvirose duomenis, kaip kurti organizaciją, kaip naudoti API, ir kt.: <http://docs.ckan.org/en/latest/user-guide.html>.

6 Reikalavimai metaduomenų perdavimui / apsikeitimui su Europos atvirų duomenų portalu

6.1 Organizaciniai reikalavimai

Europos atvirų duomenų portalas pateikia procedūras, reikalingas atvirų duomenų paskelbimui. Procedūra vadinama „Duomenų teikėjo gairės“ (angl. *Data Supplier Guidelines*)⁶¹. Toliau pateikiami pagrindiniai aspektai Lietuvos kontekste.

6.1.1 Atvirų duomenų portalo ruošimas

Prieš pateikiant užklausą dėl duomenų surinkimo (angl. *harvesting*), reikalinga turėti paruoštą atvirų duomenų portalą ir atitikti Europos atvirų duomenų portalo reikalavimus. Visi duomenų rinkiniai turėtų būti aprašyti DCAT-AP standartu ir įdiegtos technologijos, reikalingos duomenų surinkimui. Dalyje 6.2 *Techniniai atvirų duomenų skelbimo Europos atvirų duomenų portale reikalavimai* nurodoma, kas turėtų būti atlikta iš techninės pusės.

Ruošiant portalą duomenų surinkimui, taip pat reiktų atsižvelgti į toliau išvardintus aspektus:

Surinkimo dažnis

Surinkimo procesai gali būti suskirstyti į grupes ir suplanuoti pagal fiksuotą tvarkaraštį (pvz., nakties metu), kad būtų išvengta apkrovos poveikis portalui reguliariomis naudojimo valandomis.

Surenkamų duomenų rinkinių kokybė

Dalinis/ diferencialinis / pasirinktinis duomenų surinkimas - reikia išsiaiškinti, ar portalas palaiko dalinį/ diferencialinį ir (arba) pasirinktinį duomenų surinkimą.

Duomenų rinkinių aprašymas ir atvirų duomenų skelbimas

Prieš Europos atvirų duomenų portalui surenkant duomenis, logiška paskelbti dalį atvirų duomenų Lietuvos atvirų duomenų portale. Tam Lietuvos institucijos turėtų vadovautis gerosiomis praktikomis, gairėmis ir standartais aprašytais šiame dokumente. Jei reikalinga papildoma informacija, jos turėtų įsitikinti, ar privaloma skelbti tą informaciją Europos atvirų duomenų portale, ir tik tada svarstyti galimybę papildyti šias rekomendacijas.

Toliau trumpai apibendrinami pagrindiniai metaduomenų reikalavimai:

- Duomenų rinkinių metaduomenys turėtų būti užpildyti pagal DCAT-AP standartą, užpildant bent jau privalomus laukelius, kaip nurodyta pavyzdyje priede.
- Metaduomenų kategorijos turėtų atitikti Europos atvirų duomenų portalo kategorijas. Dalyje 3 *Valstybės informacinių išteklių duomenų metaduomenų standartinio aprašo struktūra* aprašomi rekomenduojamos kategorijos suderinamumo užtikrinimui. Jei šios kategorijos turėtų būti pritaikytos, Europos atvirų duomenų portalui būtina pateikti kategorijų iš Lietuvos portalo į Europos portalą žemėlapi.
- Duomenų teikėjas turi įsitikinti, kad laukeliai „Pavadinimas“ ir „Aprašymas“ yra pateikti ir anglų kalba. Laukelius galima pateikti kita kalba, nurodant atitinkamą duomenų rinkinio objekto narį. Kalbos nurodymui reikia naudoti ISO 639-1 kalbų kodus. Pavyzdys:

```
"translation": {  
  "fr": {  
    "title": "Title in French",  
    "description": "Description in French"  
  },  
}
```

⁶¹ Šaltinis: https://www.europeandataportal.eu/sites/default/files/edp_s1_gdl_data-supplier-guidelines_v1.pdf

```
"es": {  
  "title": "Title in Spanish",  
  "description": "Description in Spanish"  
}
```

- Metaduomenų laiko žymės (angl. *Timestamps*) – duomenų rinkiniai turėtų turėti specialius duomenų laukelius (pvz., laiko žymė, leidžianti duomenų surinkimo procesui identifikuoti datą ir laiką, kada duomenų rinkinys buvo paskutinį kartą keistas), pagal kuriuos duomenų surinkimo procesai gali surinkti tik metaduomenų rinkinių sub-rinkinius.
- Duomenų rinkinių/ metaduomenų dubliavimas turėtų būti vengiamas.

Atitikimas pagal kontrolinį sąrašą

Europos atvirų duomenų portalas pateikia kontrolinį sąrašą, apibendrinantį visus pagrindinius reikalavimus. Rekomenduojame atlikti sąrašą, nurodytą šio dokumento prieduose, prieš pateikiant prašymą dėl duomenų surinkimo.

6.1.2 Duomenų surinkimo prašymo registravimas

Lietuvos atvirų duomenų pareigūnas turėtų susiekti su Europos atvirų duomenų portalu, užpildant šią formą: <https://www.europeandataportal.eu/en/feedback/form?type=2>

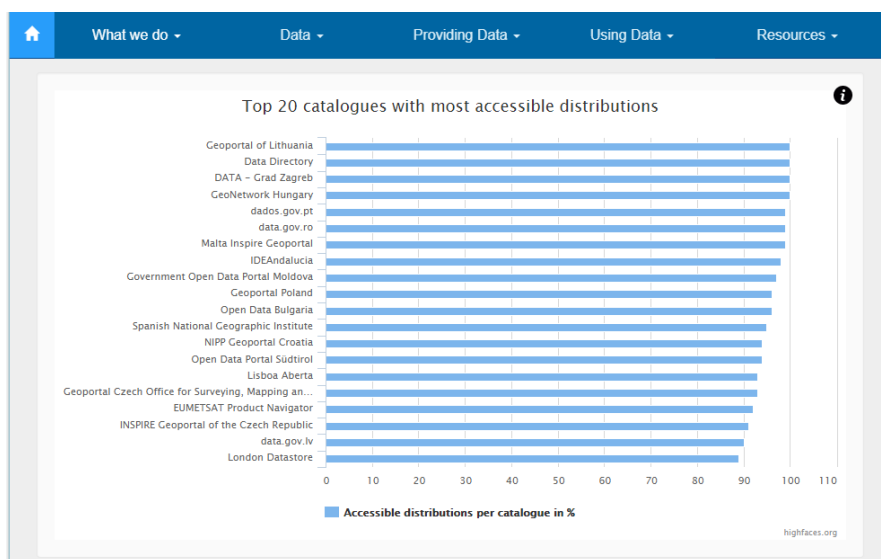
Europos atvirų duomenų portalo pareigūnas susieks su Lietuvos valstybės tarnautoju dėl tolimesnio duomenų surinkimo iš atvirų duomenų portalo proceso. Jei reikės Lietuvos atvirų duomenų portalo autentifikavimo, ši informacija bus perduota.

6.1.3 Duomenų surinkimo metu

Metaduomenų ir prieigos prie atvirų duomenų portalo tikrinimas

Kuomet Europos atvirų duomenų portalas surinkinės duomenis iš Lietuvos atvirų duomenų portalo, yra būtina užtikrinti, kad metaduomenų kokybę ir duomenų portalo suderinamumą. Tam atlikti, reikia kartą per savaitę patikrinti esamą sąsąją su Europos atvirų duomenų portalu būseną. MQA modelis teikia grafines ataskaitas apie surenkamų duomenų rinkinių metaduomenų kokybę ir pateikia pagrindinius kokybės indikatorius, pvz., platinimo prieinamumas ir pasiekiamumas (pavyzdys žemiau pateiktame paveiksle), duomenų rinkinių atitikimas pagal metaduomenų standartus, pažeidimų šaltiniai, ir kt.⁶²

⁶² Šaltinis: <https://www.europeandataportal.eu/mqa-service/en>



Paveikslas Nr. 9. Pavyzdys, kaip pateikiami katalogai su daugiausiai pasiekiamų platinimų

Klaidų tvarkymas ir pranešimas

Europos atvirų duomenų portalo naudojami duomenų surinkimo procesai automatiškai praneša apie įvairių tipų klaidas, kurios aptinkamos duomenų surinkimo metu, pateikdami *Jira* žinutę *HelpDesk* sistemoje⁶³.

Atsiradus klaidų pranešimui, Europos atvirų duomenų portalo darbuotojai yra pasiekiami anglų ir prancūzų kalbomis žemiau nurodytais kontaktais:

09:30 - 17:30 (GTM+1)

Anglų k.: +352 31 44 01-448

Prancūzų k.: +352 31 44 01-449

El.paštas: help@europeandataportal.eu

6.2 Techniniai atvirų duomenų skelbimo Europos atvirų duomenų portale reikalavimai

6.2.1 Bendrieji techniniai reikalavimai

Europos atvirų duomenų portalas pateikia aiškius techninius atvirų duomenų skelbimo reikalavimus, kad duomenys būtų įtraukti į šį portalą:

- **Duomenys turi turėti metaduomenis ir jie turi būti pateikti pagal CKAN, INSPIRE arba DCAT-AP standartus.**

Metaduomenys turi būti prieinami, kad būtų galima automatiškai apdoroti metaduomenis susietus su tam tikru duomenų rinkiniu. Taip pat svarbu, kad būtų naudojamos standartinės sąlygos, apibūdinančios metaduomenis, tokiu būdu palengvinant automatinį metaduomenų apdorojimą ir pagerinant duomenų sąveiką.

Pagrindiniai naudojami standartai yra:

- DCAT-AP ir CKAN – skirti įprastoms duomenų bazėms. Oficialus ES rekomenduojamas standartas pateikti metaduomenis yra DCAT-AP. CKAN metodas taip pat plačiai naudojamas atvirų duomenų portaluose ir yra tinkamas duomenų perdavimui Europos atvirų duomenų portalui.

⁶³ Šaltinis:

https://www.europeandataportal.eu/sites/default/files/edp_factsheet_portal_architecture_v1.4.pdf

- INSPIRE *Catalogue Services* ir *OpensSearch* (GEO/EOP) yra skirti geo-erdvinių duomenų bazėms.
- **Metaduomenys turi būti atvaizduoti naudojant XML arba JSON, jei norima, kad Europos atvirų duomenų portalas surinktų duomenis.**

Struktūrinių metaduomenys pateikia informaciją apie vidinę platinimo struktūrą, kuri yra reikalinga duomenų rinkinio naudojimui ir užklausų kūrimui. Automatiškai nuskaitomi metaduomenys priklauso nuo platinimo formato ir gali būti pateikiami XML schema arba JSON formatais.

- **Duomenų surinkimui į Europos atvirų duomenų portalą naudojamos CKAN, OAI-PMH arba *dump file* aplikacijų programavimo sąsajos.**

Duomenų surinkimui į ES duomenų portalą vyksta per aplikacijų programavimo sąsajas (API). Tam naudojama *Action API* sąsaja. Yra keli šios sąsajos naudojimo metodai:

- Duomenų teikėjas tiekia duomenų metaduomenis į Europos atvirų duomenų portalą, naudodamas CKAN, interneto pagrindu veikiantis katalogas duomenų laikymui ir platinimui. Svarbu, kad CKAN versija palaikytų *Action API*.
- Duomenų teikėjas siūlo su CKAN versiją, suderinamą su API, kur parametrai tiksliai atkartoja API funkcijas.

Duomenų surinkimui taip pat palaikomas su OAI-PMH (angl. *Open Archives Initiative Protocol for Metadata Harvesting*) suderinamas šaltinis arba metaduomenų išgavimas iš *dump file*, kuriame nurodytas duomenų bazės XML atvaizdavimas.

- **Naudojami standartiniai datos ir laiko formatai (ISO8601).**

Pagrindiniai formatai datai ir laikui (atitinkantys ISO *ISO8601* standartą) yra:

- MMMM-mm-DD
- hh:mm:ss:sss

- **Naudojamas pilnas skirstymo į kategorijas žodynas⁶⁴, arba kiti laukai, kuriose naudojami apibrėžti žodynai (pvz., atnaujinimo dažnumas)**

Europos Duomenų portalas naudoja kontroliuojamą kategorijų žodyną. Toliau pateikiami keli kategorijų pavyzdžiai, kurie yra naudojami portale.

Teikiant duomenis, tos pačios kategorijos turėtų naudojamos arba naudojamos kategorijos turėtų susietos su ES naudojamomis kategorijomis.

Kategorija	Vidinis kategorijos pavadinimas
AGRICULTURE, FISHERIES, FORESTRY, FOODS	agriculture-fisheries-forestry-food
EDUCATION, CULTURE AND SPORT	education-culture-and-sport
ENVIRONMENT	environment
ENERGY	energy
TRANSPORT	transport
SCIENCE AND TECHNOLOGY	science-and-technology
ECONOMY AND FINANCE	economy-and-finance
POPULATION AND SOCIAL CONDITIONS	population-and-social-conditions
HEALTH	health
GOVERNMENT, PUBLIC SECTOR	government-public-sector
REGIONS; CITIES	regions-cities
JUSTICE, LEGAL SYSTEM, PUBLIC SAFETY	justice-legal-system-public-safety

⁶⁴ Kontroliuojami žodynai yra aprašyti dalyje 3.3 *Siūloma metaduomenų struktūra Lietuvoje*

INTERNATIONAL ISSUES

international-issues

Paveikslas Nr. 10. ES kontroliuojamo kategorijų žodyno pavyzdys⁶⁵

- **Nurodyta, kaip dažnai Europos atvirų duomenų portalas turi įtraukti duomenis (kasdieną/ kas savaitę/ kas mėnesį).**

Dėl didelio metaduomenų kiekio, kuris turi būti surenkamas iš augančio duomenų teikėjų sąrašo, ir tam reikalingo laiko, ne visi duomenys yra surenkami kasdien. Duomenų surinkimas yra suskirstytas į grupes ir suplanuotas pagal fiksuotą grafiką (pvz., naktį), kad būtų išvengta galimo portalo apkrovos poveikio. Įprastai naudojamos registracijos bylos, kurias sukuria duomenų išgavimo proceso metu, nurodo optimalų laiką duomenų išgavimui.

Taip pat, yra reikia atsižvelgti ir į techninius apribojimus:

- Užklauskos - Šiuo metu duomenų surinkimas yra palaikomas pagal Atvirosios archyvų iniciatyvos protokolo (angl. *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*)⁶⁶, suderinamą šaltinį, arba iškelties failą (angl. *dump file*), kuriame yra duomenų rinkinių pateikimas RDF / XML formatu.
- Atsakymai - visi atsakymai iš atvirų duomenų portalo turi atitikti DCAT-AP.
- Klaidų valdymas – turėtų būti naudojami OAI-PMH protokolo pateikiami klaidų valdymo metodai.
- Duomenų šaltinio tinklapio API / galutiniai taškai – Duomenų šaltinio tinklapio REST API turėtų priimti užklauskas su, pavyzdžiui, startPos/maxRecs parametrais, skirtais surenkamo duomenų rinkinio atnaujinimui ar skaidymui.

6.2.2 Techniniai CKAN reikalavimai

Atviro kodo duomenų portalo platforma CKAN⁶⁷ yra plačiai naudojama kuriant atvirų duomenų platformas. Jo RPC⁶⁸-style API (Action API) yra palaikoma kaip Europos atvirų duomenų portalo duomenų teikėjų sąjaja. Iš esmės, yra šie sąsajos naudojimo pasirinkimai:

- Duomenų teikėjas naudoja CKAN, teikiant atvirų duomenų metaduomenis. Svarbu, kad naudojama CKAN versija palaikytų Action API⁶⁹ sąsają. Kitos CKAN API nėra palaikomos.
- Duomenų teikėjas pasiūlo API suderinamą su CKAN, kuriame reikiami galutiniai taškai pakartoja tikslų API elgesį.

Užklauskos ir atsakymai

Surenkant metaduomenis, reikia tik API galutinio taško paieškos paketo ("package_search"). Jo specifikacija yra detalai pateikta oficialioje dokumentacijoje⁷⁰.

Užklauskos pavyzdys: http://www.example.com/api/3/action/package_search?rows=50

Metaduomenų modelis

Pagrindinė CKAN duomenų schema buvo praplėsta ir pakeista, kad atitiktų dabartinius Europos atvirų duomenų portalo duomenų struktūros (DCAT-AP) reikalavimus. Atsakymas į paiešką

⁶⁵ Kiti standartai taip pat galimi, kaip skelbiama Joinup tyrime:

<https://joinup.ec.europa.eu/document/review-dcat-ap-draft-proposal-list-categorization-data>

⁶⁶ Šaltinis: <https://www.openarchives.org/pmh/>

⁶⁷ Šaltinis: <https://ckan.org/>

⁶⁸ Šaltinis: *Remote Procedure Call*

⁶⁹ Šaltinis: <http://docs.ckan.org/en/ckan-2.4.0/api/index.html#action-api-reference>

⁷⁰ Šaltinis: http://docs.ckan.org/en/ckan-2.4.0/api/index.html#ckan.logic.action.get.package_search

("package_search"), atskleidžia lauką "rezultatai", kuriame yra ieškomų duomenų rinkinių sąrašas. Metaduomenys turėtų atitikti DCAT-AP⁷¹ reikalvimus.

⁷¹ Pilnos gairės nurodytos: https://www.europeandataportal.eu/sites/default/files/edp_s1_gdl_data-supplier-guidelines_v1.pdf

Priedas Nr. 1. Klasių savybių detalizavimas

Catalogue klasė

Toliau lentelėje pateikiami detalūs *Catalogue* klasės savybių aprašymus.

Lentelė Nr. 22. *Catalogue* klasės savybės

Savybė	Svarba	Aprašymas, reikšmė/pavyzdys
dataset (duomenų rinkinys)	Privaloma	Sujungia Katalogą su Duomenų rinkiniu, kuris yra Katalogo dalis. Ryšys sudaromas RDF faile, kuris yra struktūrizuotas hierarchinis failas. Pvz.: a dcat:Dataset :dataset001 ;
publisher (leidėjas)	Privaloma	Organizacija ar asmuo, kuris paskelbė Katalogą. Pvz.: dct:publisher :transparency-office ;
description (aprašymas)	Privaloma	Laisvasis tekstas, apibūdinantis Katalogą. Gali būti nurodytas keliomis kalbomis. Pvz.: dct:description ;
title (pavadinimas)	Privaloma	Katalogo pavadinimas. Gali būti pakartotinai nurodoma kitų kalbų versija. Pvz.: dct:title "Imaginary Catalog" ;
homepage (tinklapis)	Rekomenduojama	Tinklapis, kuris yra pagrindinis katalogo puslapis. Pvz.: foaf:homepage <http://example.org/catalog> ;
language (kalba)	Rekomenduojama	Kalba, kuri yra naudojama aprašant duomenų rinkinio pavadinimus, aprašymus ir kt. Gali būti kartojama, metaduomenys yra pateikiami keliomis kalbomis. Pvz.: dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
licence (licencija)	Rekomenduojama	Licencija, pagal kuria Katalogas gali būti naudojamas ir pakartotinai naudojamas. Pvz.: dct:license rdf:resource= http://creativecommons.org/publicdomain/zero/1.0/ ;
release date (paskelbimo data)	Rekomenduojama	Oficiali katalogo išdavimo (pvz. paskelbimo) data. Pvz.: dct:issued "2011-12-05"^^xsd:date ;
themes (temos)	Rekomenduojama	Žinių organizavimo sistema, kuri yra naudojama duomenų rinkinio katalogui klasifikacijai.

		Pvz.: :catalog dcat:themeTaxonomy :themes . :themes a skos:ConceptScheme ; skos:prefLabel "A set of domains to classify documents" ;
update/ modification date (atnaujinimo/ pakeitimo data)	Rekomenduojama	Paskutinė data, kada Katalogas buvo keistas Pvz.: dct:modified "2011-12-05"^^xsd:date ;
has part (turi dalį)	Neprivaloma	Nurodo susijusį katalogą, kuris yra aprašomo katalogo dalis. Pvz.: dct:hasPart
is part of (yra dalis)	Neprivaloma	Nurodo susijusį katalogą, kuris yra fiziškai ar logiškai įtrauktas į aprašomą katalogą. Pvz.: dct:isPartOf
record (įrašas)	Neprivaloma	:catalog dcat:record :record-001 . :record-001 a dcat:CatalogRecord ; foaf:primaryTopic :dataset-001 ; dct:issued "2011-12-11"^^xsd:date ;
rights (teisės)	Neprivaloma	Teisių, susijusių su katalogu, nurodymas. Pvz.: dct:rights
spatial geographic (erdvinis/ geografinis)	/ Rekomenduojama ⁷²	Geografinė erdvė, kurią apima katalogo duomenys. Pvz.: dct:spatial <http://www.geonames.org/6695072> ;

Catalogue Record klasė

Ši klasė yra neprivaloma ir naudojama ne visiems katalogams. Ji naudojama, kai skiriasi duomenų rinkinio metaduomenys ir duomenų rinkinio įrašas kataloge. Tačiau yra mažai atvejų, kada turėtų būti daromas šis skirtumas (pvz., organizacijos paskelbimas lyginant su paskelbimu portale).

Lentelė Nr. 23. Catalogue Record klasės savybės

Savybė	Svarba	Aprašymas, reikšmė/pavyzdys
primary topic	Privaloma	Susieja katalogo įrašą su duomenų rinkiniu, aprašytu įrašė.

⁷² Pakeista iš *neprivaloma* remiantis IVPK pastabomis

(pagrindinė tema)		Pvz.: foaf:primaryTopic :dataset-001 ;
update/ modification date (atnaujinimo/ pakeitimo data)	Privaloma	Paskutinė data, kada katalogo įrašas buvo keistas. Pvz.: dct:modified: "2011-12-05"^^xsd:date ;
application profile (programos profilis)	Rekomenduojama	Programos profilis, kurį atitinka duomenų rinkinio metaduomenys Pvz.: dct:conformsTo
change type (pakeitimo tipas)	Rekomenduojama	Paskutinio duomenų rinkinio įrašo kataloge pakeitimo tipas Turi nurodyti vieną iš šių reikšmių :created, :updated or :deleted Pvz.: adms:status <http://purl.org/adms/status/Completed> ;
listing date (įrašo data)	Rekomenduojama	Data, kada duomenų rinkinio aprašymas buvo pridėtas į katalogą. Pvz.: dct:issued "2011-12-05"^^xsd:date ;
description (aprašymas)	Neprivaloma ⁷³	Įrašo aprašymas tekstu. Gali būti kartojamas kitomis kalbomis. Pvz.: dct:description
Language (kalba)	Neprivaloma	Kalba, naudojama nurodant duomenų rinkinio pavadinimus, aprašymus ir kt. Gali būti kartojama, jei metaduomenys yra nurodomi keliomis kalbomis. Pvz.: dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
source metadata (šaltinio metaduomenys)	Neprivaloma	Nurodo pirminių metaduomenų šaltinį, kuris buvo naudojamas kuriant duomenų rinkinio metaduomenis. Pvz.: dct:source
title (pavadinimas)	Neprivaloma	Katalogo įrašo pavadinimas Pvz.: dct:title "Imaginary Catalog Record" ;

⁷³ Šis laukelis yra neprivalomas, nes *Catalogue* klasėje jis jau yra privalomas.

Dataset klasė

Toliau esančioje lentelėje aprašomos *Dataset* klasės savybės.

Lentelė Nr. 24. *Dataset* klasės savybės

Savybė	Svarba	Aprašymas
description (aprašymas)	Privaloma	Duomenų rinkinio aprašymas tekstu.
title (pavadinimas)	Privaloma	Duomenų rinkinio pavadinimas Pvz.: dct:title "Imaginary dataset" ;
contact point (kontaktiniai duomenys)	Rekomenduojama	Kontaktinė informacija, kuria galima siųsti komentarus apie duomenų rinkinį. Pvz.: dcat:contactPoint <http://example.org/transparency-office/contact> ;
dataset distribution (duomenų rinkinio platinimas)	Rekomenduojama	Susieja duomenų rinkinį su prieinamu platinimo būdu. Pvz.: dcat:distribution :dataset-001-csv ;
keyword/tag (raktinis žodis/ žyma)	Rekomenduojama	Raktiniai žodžiai ar žymos, apibūdinantys duomenų rinkinį. Pvz.: dcat:keyword "accountability","transparency" ,"payments" ;
publisher (leidėjas)	Rekomenduojama	Organizacija ar įstaiga, kuri paskelbia duomenų rinkinį. Pvz.: dct:publisher :finance-ministry ;
theme/category (tema/ kategorija)	Rekomenduojama	Duomenų rinkinio kategorija. Duomenų rinkinys gali būti susietas su keliomis temomis. Pvz.: dcat:theme :accountability .
access rights (prieigos teisės)	Rekomenduojama ⁷⁴	Nurodo, ar duomenų rinkinio duomenys yra atviri, turi prieigos apribojimus ar nėra vieši. Tam ES leidinių biuras sukurs kontroliuojamą žodyną su trimis terminais (:public, :restricted, :non-public)

⁷⁴ Pakeista iš *neprivaloma* remiantis IVPK pastabomis

		Pvz.: dct:accessRights
conforms to (atitinka)	Neprivaloma	Įgyvendinimo normos ar kita specifikacija. Pvz.: dct:conformsTo <> .
documentation (dokumentacija)	Neprivaloma	Puslapis ar dokumentas apie aprašomą duomenų rinkinį. Pvz.: foaf:page <http://www.ebi.ac.uk/chembl/> .
frequency (dažnis)	Neprivaloma	Dažnis, kuriuo duomenų rinkinys yra atnaujinimas. Pvz.: dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-W> ;
has version (turi versiją)	Neprivaloma	Nurodo, ar susijęs duomenų rinkinys yra aprašomo duomenų rinkinio versija, pataisymas ar adaptacija. Pvz.: dct:hasVersion
identifier (identifikatorius)	Neprivaloma	Pagrindinis duomenų rinkinio identifikatorius, pvz., URI ar kitas unikalus identifikatorius katalogo kontekste. Pvz.: dct:identifier "chembl"@en .
is version of (yra versija)	Neprivaloma	Nurodo susijusį duomenų rinkinį, kurio aprašytas duomenų rinkinys yra versija, pataisymas ar adaptacija. Pvz.: dct:isVersionOf :chembl .
landing page (tinklapis)	Neprivaloma	Tinklapis, kuris suteikia prieigą prie duomenų rinkinio, jo platinimo ir/arba papildomos informacijos. Tinklapis turi nukreipti į pirminio duomenų teikėjo puslapį, o ne trečiosios šalies puslapį, pvz., kaupėjo. Pvz.: dcat:landingPage <http://example.org/dataset-002.html> ;
language (kalba)	Neprivaloma	Duomenų rinkinio kalba. Gali būti kartojama, jei duomenų rinkinyje naudojamos kelios kalbos. Pvz.: dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;

other identifier (kitas identifikatorius)	Neprivaloma	Antrinis duomenų rinkinio identifikatorius, toks kaip MAST/ADS18 , DataCite , DOI , EZID arba W3ID ⁷⁵ . Pvz.: adms:identifier
provenance (kilmė)	Neprivaloma	Duomenų rinkinio kilmės pareiškimas. Kilmė gali būti pateikiama kaip grynasis tekstas (kaip ir toliau pateiktame pavyzdyje). Europos Komisija Joinup platformoje nurodo: „Kadangi kilmės informacija nėra plačiai naudojama ir dažniausiai pateikiama kaip grynasis tekstas, šios informacijos surinkimo naudingumas yra abejotinas ir gali būti ignoruojamas“ ⁷⁶ . Pvz.: dct:provenance xml:lang="en">The first description of this dataset was created by the European Commission in 2016</dct:provenance
related resource (susijęs šaltinis)	Neprivaloma	Nurodo susijusį šaltinį. Pvz.: dct:relation
release date (paskelbimo data)	Rekomenduojama ⁷⁷	Oficiali duomenų rinkinio išdavimo (pvz. paskelbimo) data. Pvz.: dct:issued "2013-08-29T00:00:00"^^xsd:dateTime .
sample (imtis)	Neprivaloma	Duomenų rinkinio imties pasiskirstymas. Pvz.: adms:sample
source (šaltinis)	Neprivaloma	Nurodo susijusį duomenų rinkinį, iš kurio gautas aprašomas duomenų rinkinys. Pvz.: dct:source :pubchem-bioassay-09-01-2014 .
spatial/ geographical coverage (erdvinis/ geografinis paplitimas)	Neprivaloma	Geografinis regionas, kurį apima duomenų rinkinys. Pvz.: dct:spatial <http://www.geonames.org/6695072> ;

⁷⁵ Mikulski Archive for Space Telescopes (MAST). Referencing Data Sets in Astronomical Literature. http://archive.stsci.edu/pub_dsn.html, DataCite. <http://www.datacite.org/>, DOI. Digital Object Identifier. <http://www.doi.org/>, EZID. <http://n2t.net/ezipd>, W3C Permanent Identifier Community Group. Permanent Identifiers for the Web, <https://w3id.org/>

⁷⁶ <https://joinup.ec.europa.eu/release/dcat-ap-how-model-and-express-provenance>

⁷⁷ Pakeista iš *neprivaloma*

temporal coverage (laikotarpis)	Neprivaloma	Laikotarpis, kurį apima duomenų rinkinys. Pvz.: dct:temporal <http://reference.data.gov.uk/id/quarter/2006-Q1> ;
type (tipas)	Neprivaloma	Duomenų rinkinio tipas. Kontroliuojamo žodyno reikšmės dar nėra nustatytos.
update/ modification date (atnaujinimo/ pakeitimo data)	Rekomenduojama ⁷⁸	Paskutinė data, kada duomenų rinkinys buvo keistas. Pvz.: dct:modified "2011-12-05"^^xsd:date ;
version (versija)	Rekomenduojama ⁷⁹	Duomenų rinkinio versijos numeris ar kita versija. Pvz.: <owl:versionInfo>v 1.17 2003/02/26 12:56:51 mdean</owl:versionInfo>
version notes (versijos pastabos)	Rekomenduojama ⁸⁰	Nurodomi skirtumai tarp esamos ir ankstesnės versijos. Pvz.: adms:versionNotes

Distribution klasė

Toliau esančioje lentelėje aprašomos *Distribution* klasės savybės.

Lentelė Nr. 25. *Distribution* klasės savybės

Savybė	Svarba	Aprašymas
access URL (prieigos URL)	Privaloma	URL nuoroda, kuri suteikia prieigą prie duomenų rinkinio platinimo. Gali būti nurodoma tokia informacija, kaip gauti duomenį rinkinį. Pvz.: dcat:accessURL <http://example.org/dataset-002.html> ;
description (aprašymas)	Rekomenduojama	Platinimo aprašymas tekstu. Gali būti kartojamas keliomis kalbomis.
format (formatas)	Rekomenduojama	Platinimo failo formatas. Pvz.: dct:format "text/html";

⁷⁸ Pakeista iš *neprivaloma*

⁷⁹ Pakeista iš *neprivaloma*

⁸⁰ Pakeista iš *neprivaloma*

licence (licencija)	Rekomenduojama	Licencija, pagal kurią prieinamas platinimas. Pvz.: dct:license rdf:resource= http://creativecommons.org/publicdomain/zero/1.0/ ;
byte size (dydis baitais)	Neprivaloma	Platinimo dydis, nurodytas baitais. Pvz.: dcat:byteSize "5120"^^xsd:decimal ;
checksum (kontrolinė suma)	Neprivaloma	Mechanizmas, kuru galima patikrinti, ar platinimo turinys nėra pakeistas. Pvz.: <spdx:checksumValue>d6a770ba38583ed4bb4525bd96e50461655d2759</spdx:checksumValue>
documentation (dokumentacija)	Neprivaloma	Puslapis ar dokumentas apie platinimą. Pvz.: foaf:page < http://www.ebi.ac.uk/chembl/ >
download URL (parsisiuntimo nuoroda)	Neprivaloma	URL nuorodą į parsisiunčiamą failą nurodytu formatu. Pvz.: dcat:downloadURL < http://www.example.org/files/001.csv > ;
language (kalba)	Neprivaloma	Platinimo klasėje naudojama kalba. Gali būti kartojama, jei metaduomenys pateikiami keliomis kalbomis. Pvz.: dct:language < http://id.loc.gov/vocabulary/iso639-1/en > ;
linked schemas (susijusi schema)	Neprivaloma	Nustatyta schema, kurią atitinka aprašomas platinimas. Pvz.: dct:conformsTo <> .
media type (medijos tipas)	Rekomenduojama ⁸¹	Platinimo medijos tipas, kaip nurodyta oficialiame IANA medijų tipų registre. Pvz.: dcat:mediaType "text/csv" ;
release date (paskelbimo data)	Neprivaloma	Oficiali platinimo išdavimo (pvz. paskelbimo) data. Pvz.: dct:issued "2011-12-11"^^xsd:date ;
rights	Neprivaloma	Teisių, susijusių su platinimu, pareiškimas.

⁸¹ Pakeista iš *neprivaloma*

(teisės)		Pvz.: dct:rights
status (būsena)	Neprivaloma	Platinimo terminas. Pvz.: adms:status <http://purl.org/adms/status/Completed> ;
title (pavadinimas)	Neprivaloma	Platinimo pavadinimas. Gali būti kartojamas keliomis kalbomis. Pvz.: dct:title "ChEMBL" ;
update/ modification date (atnaujinimo/ pakeitimo data)	Neprivaloma	Paskutinė data, kada platinimas buvo keistas. Pvz.: dct:modified "2011-12-05"^^xsd:date ;

Agent klasė

Toliau esančioje lentelėje aprašomos *Agent* klasės savybės.

Lentelė Nr. 26. *Agent* klasės savybės

Savybė	Svarba	Aprašymas
name (vardas)	Privaloma	Agento pavadinimas. Gali būti kartojamas keliomis kalbomis. Pvz.: <foaf:name>Peter Parker</foaf:name>
type (tipas)	Rekomenduojama	Agento tipas, kuris leidžia katalogą ar duomenų rinkinį. Pvz.: dct:type

Category Scheme klasė

Toliau esančioje lentelėje aprašomos *Category Scheme* klasės savybės.

Lentelė Nr. 27. *Category Scheme* klasės savybės

Savybė	Svarba	Aprašymas
title (pavadinimas)	Privaloma	Kategorijos schemos pavadinimas. Gali būti nurodomos kelios pavadinimo versijos.

Pvz.:
dct:title

Category klasė⁸²

Toliau esančioje lentelėje aprašomos *Category* klasės savybė.

Lentelė Nr. 28. *Category* klasės savybė

Savybė	Svarba	Aprašymas
preferred label (pageidaujama žyma)	Privaloma	Pageidaujama kategorijos etiketė. Gali būti kartojama keliomis kalbomis. Pvz.: <skos:prefLabel xml:lang="en">Access to countryside</skos:prefLabel>

Checksum klasė

Toliau esančioje lentelėje aprašomos *Checksum* klasės savybė.

Lentelė Nr. 29. *Checksum* klasės savybė

Savybė	Svarba	Aprašymas
algorithm (algoritmas)	Privaloma	Algoritmas, naudojamas kontrolinės sumos (angl. <i>Checksum</i>) pateikimui. Šiuo metu, SHA-1 yra vienintelis palaikomas algoritmas. Pvz.: <spdx:algorithm rdf:resource="http://spdx.org/rdf/terms#checksumAlgorithm_sha1"/>

Identifier klasė

Toliau esančioje lentelėje aprašomos *Identifier* klasės savybė.

Lentelė Nr. 30. *Identifier* klasės savybė

Savybė	Svarba	Aprašymas
notation	Privaloma	Ženklų eilutė, kuri yra identifikuojama identifikatoriaus scheme, nurodant jo duomenų tipą.

⁸² Klasė „Kategorija“ apibrėžta Europos Komisijos standarte, kaip aprašyta čia. Vis dėlto W3C specifikacijoje pateikiamas kitas kategorijų valdymas, naudojant klases ""concept skos:Concept" ir "skos:ConceptScheme", pakeičiant šias klases. Kadangi oficiali šių klasių integracija nėra prieinama Joinup platformoje, rekomenduojama laikytis standarto, iki tolesnio pranešimo ar atnaujinimo. Norėdami nurodyti temą duomenų rinkinyje, siūloma naudoti klasės "duomenų rinkinio" nuosavybę "temą" ir taksanomijas, kurios nurodytos dalyje 3.3 *Siūloma metaduomenų struktūra* Lietuvoje.

(žymėjimas)

Pvz.:
skos:notation "K"^^<ChemicalSymbolNotation> .

Licence Document klasė

Toliau esančioje lentelėje aprašomos *Licence Document* klasės savybė.

Lentelė Nr. 31. *Licence Document* klasės savybė

Savybė	Svarba	Aprašymas
licence type (licencijos tipas)	Rekomenduojama	Licencijos tipas, pvz., nurodantis „viešas naudojimas“ arba „reikalaujamas autorinis honoraras“. Pvz.: dct:type

Period of Time klasė

Toliau esančioje lentelėje aprašomos *Period of Time* klasės savybės.

Lentelė Nr. 32 *Period of Time* klasės savybės

Savybė	Svarba	Aprašymas
start date/time (pradžios data/ laikas)	Neprivaloma	Periodo pradžia Pvz.: schema:startDate
end date/time (pabaigos data/ laikas)	Neprivaloma	Periodo pabaiga Pvz.: schema:endDate

Priedas Nr. 2. RDF failo metaduomenų struktūros aprašymas

Failo nuskaitymui, RDF faile turėtų būti nurodyti šie elementai:

- `xmlns:time="http://www.w3.org/2006/time#"`
- `xmlns:dct="http://purl.org/dc/terms/"`
- `xmlns:dcap="http://www.w3.org/ns/dcat#"`
- `xmlns:foaf="http://xmlns.com/foaf/0.1/"`
- `xmlns:xsd="http://www.w3.org/2001/XMLSchema#"`
- `xmlns:tema="http://datos.gob.es/kos/sector-publico/sector/"`
- `xmlns:auto="http://datos.gob.es/recurso/sector-publico/territorio/Autonomia/"`
- `xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"`
- `xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"`

<https://github.com/ckan/ckanext-dcat/blob/master/examples/catalog.rdf>

RDF failo pateikimo pavyzdys:

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
  xmlns:time="http://www.w3.org/2006/time#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:dcap="http://www.w3.org/ns/dcat#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:tema="http://datos.gob.es/kos/sector-publico/sector/"
  xmlns:auto="http://datos.gob.es/recurso/sector-publico/territorio/Autonomia/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <dcat:Catalog rdf:about="https://data.some.org/catalog">
    <dct:identifier>https://data.some.org/catalog</dct:identifier>
    <dct:title>An example catalog</dct:title>
    <dct:description>A longer description of what the catalog is, who maintains it, etc. Special
    characters should be properly handed (if the file is utf8 encoded), for instance for catalogs from Sant
    Adrià.</dct:description>
    <dct:issued      rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-04-
    26T00:00:00Z</dct:issued>
    <dct:modified    rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-05-
    22T09:38:45Z</dct:modified>
    <dct:language>en</dct:language>
    <foaf:homepage rdf:resource="https://data.some.org/the/actual/catalog"/>
    <dct:license rdf:resource="https://data.some.org/link/to/license"/>
    <dcat:dataset>
      <dcat:Dataset rdf:about="https://data.some.org/catalog/datasets/1">
        <dct:identifier>https://data.some.org/catalog/datasets/1</dct:identifier>
        <dct:title>Example dataset 1</dct:title>
        <dct:description>A longer description of dataset 1: includes three distributions, a CSV file, a PDF
        file and an API. Some special characters: Θεσσαλονίκη.</dct:description>
        <dcat:keyword>stats</dcat:keyword>
        <dcat:keyword>pollution</dcat:keyword>
        <dct:issued      rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-04-
        26T00:00:00Z</dct:issued>
        <dct:modified    rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-04-
        26T00:00:00Z</dct:modified>
```

```
< dct:language>en</dct:language>
< dct:license rdf:resource="https://data.some.org/link/to/license"/>
< dct:publisher>
  < foaf:Organization>
    < foaf:name>Publishing Organization for dataset 1</foaf:name>
    < foaf:mbox>contact@some.org</foaf:mbox>
  </foaf:Organization>
</dct:publisher>
< dcat:distribution>
  < dcat:Distribution>
    < dct:identifier>https://data.some.org/catalog/datasets/1/distribution/1</dct:identifier>
    < dct:title>A distribution for this dataset in CSV format</dct:title>
    < dct:description>A longer description for the distribution</dct:description>
    < dcat:downloadURL
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">https://data.some.org/the/actual
/catalog/datasets/1/resources/1.csv</dcat:downloadURL>
    < dcat:mediaType>text/csv</dcat:mediaType>
    < dct:license rdf:resource="https://data.some.org/link/to/license"/>
  </dcat:Distribution>
</dcat:distribution>
< dcat:distribution>
  < dcat:Distribution>
    < dct:identifier>https://data.some.org/catalog/datasets/1/distribution/2</dct:identifier>
    < dct:title>A distribution for this dataset in PDF format</dct:title>
    < dct:description>A longer description for the distribution</dct:description>
    < dcat:downloadURL
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">https://data.some.org/the/actual
/catalog/datasets/1/resources/2.pdf</dcat:downloadURL>
    < dct:format>
      < dct:IMT>
        < rdf:value>application/pdf</rdf:value>
      </dct:IMT>
    </dct:format>
    < dct:license rdf:resource="https://data.some.org/link/to/license"/>
  </dcat:Distribution>
</dcat:distribution>
< dcat:distribution>
  < dcat:Distribution>
    < dct:identifier>https://data.some.org/catalog/datasets/1/distribution/3</dct:identifier>
    < dct:title>A distribution for this dataset as an API</dct:title>
    < dct:description>A longer description for the distribution</dct:description>
    < dcat:accessURL
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">https://data.some.org/the/actual
/catalog/link/to/an/api</dcat:accessURL>
  </dcat:Distribution>
</dcat:distribution>
</dcat:Dataset>
</dcat:dataset>
</dcat:Catalog>
</rdf:RDF>
```

Priedas Nr. 3. JSON failo metaduomenų struktūros aprašymas

Toliau nurodomas metaduomenų turinys pateiktas JSON formatu⁸³:

```
[
  {
    "title": "Live traffic information from the Highways Agency",
    "license": "No license provided",
    "publisher": {
      "name": "Highways Agency",
      "mbox": "test@test.com"
    },
    "distribution": [
      {
        "downloadURL": "https://s3-eu-west-1.amazonaws.com/lmtesting2810/HATRIS_15MinuteMIDAS_YYYY-MM-DD_o.csv.zip",
        "title": "Hattris 15 Min Midas o",
        "format": "application/zip"
      }
    ]
  },
  {
    "title": "Roadworks locations",
    ...
  }
]
```

Pastaba: W3C pateikia daugiau informacijos apie JSON naudojimą čia: <https://www.w3.org/TR/json-ld/>.

⁸³ Šaltinis: http://guidance.data.gov.uk/dcat_fields.html

Priedas Nr. 4. Europos atvirų duomenų portalui tinkamo RDF failo pavyzdys

Toliau nurodomas duomenų rinkinys, aprašytas visomis privalomomis DCAT savybėmis, ir suderinimas su Europos atvirų duomenų portalu:

```
<dcat:Dataset>
  < dct:title xml:lang="en">v394d2705_emilia</dct:title>
  < dct:description xml:lang="en">Place: Emilia area; Orbit type:
Descending; Map type: Mean Ground Velocity; Number of Images: 1; Track:
394; Frame: 2705; SLC: 37; Master: 2008-05-19; Number of interferograms:
139; Processor Name: SBAS algorithm (SARScape); Parameter Name: n.a.;
Satellite: ASAR ENVISAT; Input File Name: n.a.;
  </dct:description>
  < dct:spatial>
    < dct:Location>
      < locn:geometry
        rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">
        POLYGON((10. 87649 45.23888,12.38418 45.23888,12.38418
44.144384,10.87649
44.144384,10.87649 45.23888))
      </locn:geometry>
    </dct:Location>
  </dct:spatial>
  < dct:temporal>
    < dct:PeriodOfTime>
      < schema:startDate
        rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2003-0210
      </schema:startDate>
      < schema:endDate
        rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-06-09
      </schema:endDate>
    </dct:PeriodOfTime>
  </dct:temporal>
  < dct:created
    rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-12-28
  </dct:created>
  < dct:conformsTo rdf:parseType="Resource">
    < dct:title xml:lang="en">Corrigendum to INSPIRE Metadata Regulation published in
the Official Journal of the European Union, L 328, page 83</dct:title>
    < dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2009-
12- 15
    </dct:modified>
  </dct:conformsTo>
  < dcat:theme rdf:resource="http://inspire.ec.europa.eu/theme/ge"/>
  < dcat:theme rdf:resource="http://eurovoc.europa.eu/100151"/>
  < dcat:theme
    rdf:resource="http://publicdata.eu/def/category/science-and-technology"/>
  < dcat:theme rdf:resource="http://eurovoc.europa.eu/1151"/>
  < dcat:keyword xml:lang="en">GEOMORPHOLOGY</dcat:keyword>
  < dct:language
    rdf:resource="http://publications.europa.eu/resource/authority/language/ENG"/>
  < dct:provenance>
    < dct:ProvenanceStatement>
```

```

                <rdfs:label xml:lang="en">Validation Flag: validated; Validation
                Time: 2013-01-12; Validation Responsible: Dr. Cristiano Tolomei
                (cristiano.tolomei@ingv.it)</rdfs:label>
            </dct:ProvenanceStatement>
        </dct:provenance>
        <dcat:landingPage>
            <foaf:Document
                rdf:about="http://185.12.7.250:8080/petascope/...">
                <dct:title
                    xml:lang="en">ASAR_EMILIADISPDESMEAN_32632_90</dct:title>
                <dct:description xml:lang="en">WCS request to
                download the coverage (raster, geotiff).</dct:description>
            </foaf:Document>
        </dcat:landingPage>
        <dcat:landingPage>
            <foaf:Document
                rdf:about="http://185.12.7.250:8080/petascope/...">
                <dct:title
                    xml:lang="en">ASAR_EMILIADISPDESMEAN_32632_90</dct:title>
                <dct:description xml:lang="en">WCS request to
                download the coverage (csv).</dct:description>
            </foaf:Document>
        </dcat:landingPage>
        <dct:identifier>VELISAR2.o_emilia_v394d2705</dct:identifier>
    </dcat:Dataset>

```

Priedas Nr. 5. Kontrolinis sąrašas dėl Europos atvirų duomenų portalų duomenų surinkimo

Reikalavimas	Reikšmė	Komentaras
Įsitikinkite, kad jūsų portalas nurodo metaduomenis!		Tik metaduomenys gali būti surinkti, o ne patys duomenys!
Kuris metaduomenų standartas yra naudojamas?	CKAN INSPIRE DCAT-AP	Jei naudojamas kitas standartas, prašome nurodyti išsamų schemos aprašymą.
Kuris metaduomenų atvaizdavimo formatas yra naudojamas?	XML/JSON	
Kokia API sąsaja yra naudojama duomenų perdavimui?		
Ar norint gauti prieigą prie API, reikalingas autentifikavimas?	Taip/Ne	
Pridėkite baigtinį kategorijų žodyną ar kitus laukelius, kuriuose naudojamas apibrėžtas žodynas (pvz., atnaujinimo dažnis)	ISO8601	Su vertinimu, kur taikoma
Naudokite standartinius datos ir laiko formatus	ISO8601	
Ar galimas dalinis ar diferencialinis duomenų surinkimas?	Taip/Ne	Dalinis ar diferencialinis duomenų surinkimas leidžia surinkti tik tuos duomenų rinkinius, kurie buvo pakeisti po tam tikros datos
Kaip dažnai portale turėtų būti surenkami duomenys (angl. <i>harvest data</i>)?		