

Europeana duomenų modelis (EDM)

Martin DOERR

Kompiuterijos mokslų institutas, Mokslinių tyrimų ir technologijos fondas – Hellas, Kreta, Graikija,
el. p. martin@ics.forth.gr

Stefan GRADMANN

Bibliotekininkystės ir informacijos mokslų institutas, Humboldt universitetas, Berlynas, Vokietija,
el. p. stefan.gradmann@ibi.hu-berlin.de

Steffen HENNICKE

Humboldt universitetas, Berlynas, Vokietija, el. p. steffen.hennicke@ibi.hu-berlin.de

Antoine ISAAC

Europeana, Haga, Nyderlandai, el. p. aisaac@few.vu.nl

Carlo MEGHINI

Informacijos ir technologijų institutas, Pisa, Italija, el. p. carlo.meghini@isti.cnr.it

Herbert van de SOMPEL

Mokslinė biblioteka, Los Alamos nacionalinė laboratorija, New Mexico, JAV, el. p. Herbert@hvdsomp

Europeana duomenų modelis (*Europeana Data Model – EDM*) – tai naujas požiūris į duomenų, kuriuos Europeana teikia įvairios kultūros paveldo institucijos, struktūravimą ir pateikimą. Šiuo modeliu siekiama didesnės raiškos ir lankstumo, palyginti su dabar taikomu Europeana Semantic Elements (ESE), kurį jis turėtų pakeisti. Esminiai EDM projektiniai principai grindžiami pagrindiniais semantinio saityno ir susietų duomenų iniciatyvų, kurias Europeana siekia praturtinti, principais ir gerąja praktika. Paties modelio pagrindą sudaro galiojantys standartai, tokie kaip RDF(S), OAI-ORE, SKOS ir Dublin Core. Jis veikia kaip bendra aukščiausio lygmens ontologija, išlaikanti originalius duomenų modelius ir informacijos perspektyvą, tačiau kartu suteikianti ir sąveikumo galimybę. Straipsnyje išsamiai nagrinėjami minėti aspektai ir projektiniai principai, skatinę EDM plėtrą.

Reikšminiai žodžiai: Europeana duomenų modelis; kultūros paveldas; semantinis saitynas.

1. Įvadas į Europeana

Europeana dažnai viešai pristatoma kaip portalas, teikiantis prieigą prie milijonų įvairiausioms kultūros paveldo sritims priklausančių objektų. Nors toks pristatymas iš dalies ir teisingas, jis neatspindi kai kurių esminių ypatybių to, kuo Europeana turėtų galiausiai tapti. Concordia et al. (2010) bando išaiškinti, kad išskirtinė Europeana portalo savybė yra ne tiek jo apimtis, kiek siekiamybė padaryti duomenų įvairovę ir funkcionalumą prieinamą per API sąsają. Tai leistų įvairioms portale nedalyvaujančioms bendruomenėms pritaikyti mūsų išsamias (ir gausias) Europos kultūros paveldo vertybių pateiktis savo reikmėms, o Europeana portalas (<http://www.europeana.eu>)

galiausiai taptų vienu iš dalyvių, besinaudojančių šiais duomenų ir funkcionalumo lobiais per API.

Toliau įgyvendinant sumanymą, Europeana objektų pateiktims bus suteikiamas išsamus semantinis kontekstas, leidžiantis su šiais ištekliais atlikti sudėtingas semantines operacijas, nesuderinamas su tradicine skaitmeninių bibliotekų sąsaja. Tokiam funkcionalumui pasiekti reikalinga nuolatinė Europeana objektų pateikčių sąsaja su susietais atvirais duomenimis saityne arba semantinį kontekstą teikiančiais ištekliais, esančiais Europeana duomenų erdvėje, tokiais kaip į SKOS konvertuoti tezaurai ir struktūruoti žodynai.

Europeana techninė strategija orientuota į besiplėtojančios susietų duomenų paradigmos praturtinimą ir naudojimąsi jos privalumais.

2. Nuo ESE prie EDM

Tačiau pagrindinė kliūtis, kurią teko įveikti einant link susietų atvirų duomenų, buvo *Europeana* semantinių elementų (ESE) formatas, grindžiamas paprastu ir patikimu duomenų modeliu. ESE sukūrimas ir įgyvendinimas buvo būtina pirmojo ankstyvojo *Europeana* prototipo realizavimo pradžios 2008 m. lapkričio mėn. sąlyga, tačiau dabar išaiškėję jo trūkumai paradoksaliai trukdo pereiti prie semantinių požiūriu išsamaus funkcionalaus modelio (visų pirma ESE visai netinka vadintis „semantiniu“...).

Esminė ESE problema yra „plokščias“ jo modeliavimo pobūdis, neleidžiantis įterpti saitų į išorinius saityno išteklius ir tai, kad jis negali būti praplečiamas pritaikant jį labiau specializuotiems, išplėtotos struktūros modeliams. Be to, ESE neįmanoma įdiegti į susietų duomenų vardų erdves taip, kaip reikėtų siekiant, kad *Europeana* taptų tokių būsimų paskirstytos informacijos struktūrų dalimi. Tokia būtų trumpa svarbiausių su ESE susijusių probleminių dalykų apžvalga.

Netrukus po darbo su *Europeana* semantinių duomenų lygmeniu pradžios įgyvendinant *EuropeanaConnect* projekto 1-ąją veiksmų programą tapo aišku, kad ESE turės būti skubiai pakeistas tuo metu dar tebesitęsiančio *Europeana* duomenų modelio (EDM) specifikacijos rengimo proceso rezultatais. Todėl buvo nutarta pagreitinti šį procesą, siekiant skubių vertingų ir praktiškai pritaikomų rezultatų. Šio veiklos etapo, prasidėjusio vėlyvą 2009 m. vasarą, rezultatas buvo 5-oji EDM versija, kuri ir aptariama šiame straipsnyje.

3. EDM konstrukciniai principai

3.1. W3 standartai ir semantinis saitynas

Pirmoji saityno realizacija iš esmės buvo didžiulė hipertekstinė taikomoji programa su daugybe vienas su kitu susietų „dokumentų“ (tinklalapių), kurių kiekvienam buvo priskirtas URI. Situacija ėmė keistis atsiradus semantiniam saitynui, kuris, be kitų dalykų, apima realiojo pasaulio daiktus (vadinamuosius neinformacinius išteklius, plg. <http://www.w3.org/TR/cooluris>), kuriems irgi priskirti URI. Pastaruoju metu toliau mėginant aiškinti semantinio saityno esminius principus, susiformavo „susietų duomenų“ (<http://linkeddata.org/>) arba „duomenų saityno“ sąvokos, teikiančios pagrindą atsirasti saityno dokumento (arba vadinamojo „informacijos išteklius“), išreiškiančio neinformacinį išteklių, ir kreipties iš neinformacinio išteklius į saityno dokumentą idėjoms.

Lemiamas pokytis, kurį patirs *Europeana* pradėjus taikyti EDM, bus suderinamumas su šia semantinio saity-

no paradigma. Pastaraisiais metais Saityno konsorciumas (W3C) parengė daug standartų, sudarančių galimybes kompiuterio skaitomų, struktūruotų duomenų pateikimui saityne ir jų mainams. Pirmasis iš jų yra *Resource Description Framework* (RDF), įgalinantis pateikti struktūruotą informaciją apie bet kokį išteklių nesudėtingų trijų dalių teiginių forma (veiksny, tarinys, papildinys). RDF remiasi idėja, kad ištekliai gali būti aprašomi pasitelkiant semantiškai prasmingus juos siejančius ryšius. Pavyzdžiui, tripletas „*ec:ulysses, ex:author, ex:james_joyce*“¹ aprašo knygą „Ulysses“ – jos identifikatorių susieja su kitu identifikatoriumi, žyminčiu James Joyce, naudojamas autoriaus tipo nuorodą, žymintį knygos ir jos autoriaus santykį. EDM visiškai atitinka šį tripletais pagrįstą aprašo metodą.

Nuorodų tipai yra pagrindinis veiksnys, teikiant galimybę RDF perteikti reikšmingas žinias. Šios savybės kartu su tipais, kurie gali būti priskiriami subjekto ir objekto ištekliams (klasėms), apibrėžiami ontologijose. Žodis „ontologija“ čia suprantamas kaip „konceptualizacijos“ (Gruber 1993) sinonimas, o ne taip, kaip priimta filosofijoje: „mokslas apie tai, kas yra, visų tikrovės sričių objektų, savybių, įvykių, procesų ir santykių rūšis bei struktūras“ (Smith 2003). Ontologijos, apibrėžiamos taikant RDF schemas (RDFS) ir Saityno ontologijos kalbos (OWL) standartus, apima neformalias apibrėžtis žmogaus skaitomų dokumentų forma ir formalias apibrėžtis ribojimų ir taisyklių, įgalinančių aptikti prieštaravimus arba iš pateiktų faktų išvesti naujus, forma. Tarkime, ontologija gali apibrėžti knygų, paveikslų ir asmenų klases bei vieną autoriaus savybę ir formaliai teigti, kad visi ištekliai, susieti su knyga per autoriaus savybę, priklauso asmens tipui. Ji taip pat gali ir formaliai apibrėžti kitai klasei priklausančią objektą kaip knygos ir paveikslų viršklasę. Taikant loginio išvedimo modulį knygų ir paveikslų rinkinio duomenims ir atliekant visų vieno asmens sukurtų objektų užklausa, galima gauti visus šiuos objektus iš anksto nežinant, kokiam konkrečiam tipui jie priklauso. Ši ypatybė turi lemiamą reikšmę, kai būtina informacijos integracija.

Semantinio saityno koncepcija leidžia tame pačiame apraše derinti įvairias ontologijas. Galima teikti skirtingus požiūrius į tuos pačius tvirtinimus arba, siekiant atitikti bendresnius taikymo reikalavimus, formuluoti tvirtinimus, kurie apima terminus iš skirtingų žodynų, pritaikytų konkrečioms reikmėms. Todėl, siekiant pateikti esminius RDF duomenis, knygų katalogui gali būti iš naujo panaudojama knygų aprašo ontologija, o tiksliam aprašymui knygų autorių, koks yra autoritetiniame faile, – asmenų ontologija.

EDM bus naudojamos kai kurios jau sukurtos pavyzdinės ontologijos, tokios kaip W3C remiama Paprasta

¹ Čia ex: reiškia sutartinę pavyzdžio vardų erdvę.

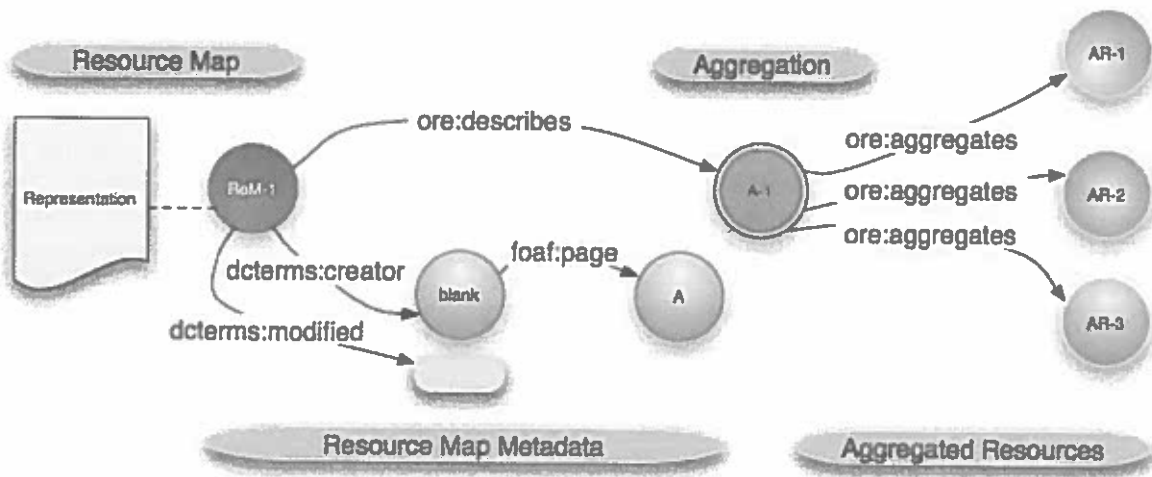
žinių organizavimo sistema (*Simple Knowledge Organization System* – SKOS). SKOS apibrėžia modelį, reprezentuojantį Paprastos žinių organizavimo sistemos elementus: tezaurus, klasifikacijos sistemas ir panašius dalykus. Viena iš SKOS pagrindinių klasių skirta koncepcijoms. Koncepcijomis grindžiamiems modeliavimo metodams taikant tokius standartus kaip ISO 2788, SKOS formuluojamos savybės šių koncepcijų žymėms (pvz., „skos:prefLabel“ pirminei koncepcijos žymei, „skos:altLabel“ – alternatyvioms žymėms), šių koncepcijų semantiniams ryšiams (skos:narrower, skos:broader, skos:related) ir bendrajai koncepcijos dokumentacijai (skos:scopeNote, skos:definition ir t. t.). Svarbu tai, kad SKOS leidžia derinti skirtingas koncepcijų schemas, pvz., susiejamos skirtingų semantiškai lygiaverčių tezaurų koncepcijos naudojant „skos:exactMatch“ savybę. Tai sudaro technines galimybes programoms orientuotis semantiniame skirtingų šaltinių koncepcijų lygmenyje, išlaviruoti tokiam koncepcijų tinkle ir pasiekti objektus, aprašytus vartojant skirtingas – tačiau semantiškai giminingas – koncepcijas.

Tarp kitų ontologijų, kurios specifiskai reikšmingos EDM, yra *Dublin Core* (DC) ir *Friend-of-a Friend* (FOAF). *Dublin Core* pateikia konceptualų žodyną, skirtą esminėms kultūros objektų savybėms (kūrėjams, santykiams su kitais šaltiniais, dalykinimui ir kt.) aprašyti taip, kaip leidžia semantinis saitynas, tenkinant visą reikmių spektrą. DC teikia ESE pagrindą: naudojant ją kaip sudedamąją modelio dalį įmanomas tiesioginis suderinamumas su paveldėtais *Europeana* duomenimis. Be to, tai leidžia teikėjams, nenorintiems teikti išsamesnių aprašų, naudoti paprastą duomenų teikimo būdą. Galiausiai tai leidžia EDM duomenis labiau pritaikyti mainams ir daugkartiniam naudojimui: DC naudojama daugelyje programų, kurios labai lengvai gali būti pritaikytos EDM duomenims įsisavinti.

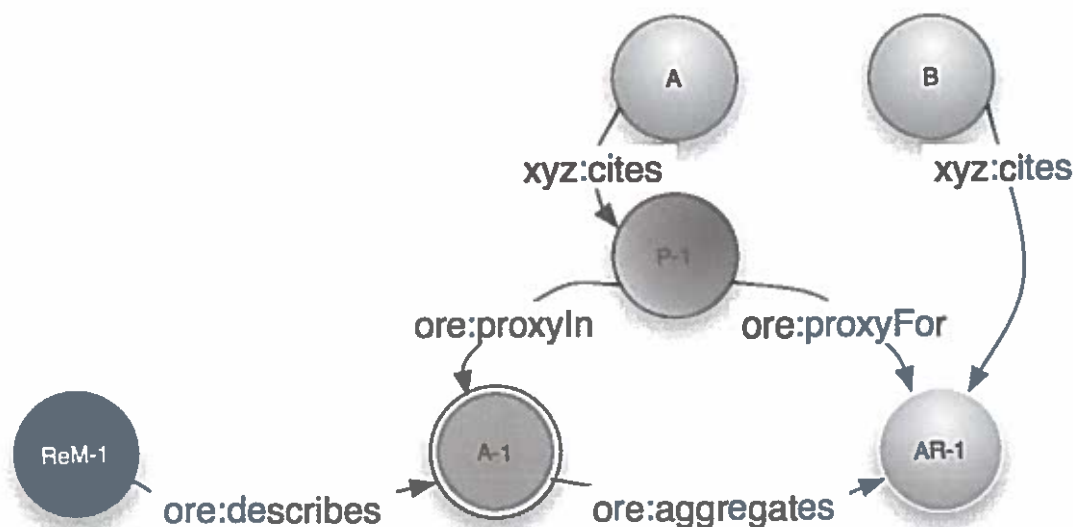
FOAF – tai ontologija, naudojama asmenims RDF aprašyti, jos ištakos buvo reikalavimai saityno profilio aprašui. Todėl ji gali būti pritaikyta (nors tam reikėtų ją šiek tiek adaptuoti arba praplėsti) daugybei asmenų, kuriems tenka ypatingas vaidmuo *Europeana* objektų kontekste.

3.2. Object Reuse and Exchange standartas

Tipiškos objektų pateiktys *Europeana* dažniausiai bus sudėtiniai entitetai, susidedantys iš keleto dalių, pavyzdžiui, metaduomenų atributų, paveikslo miniatiūros ir statinio HTML įėjimo tinklalapio (tai tik keli iš daugybės pavyzdžių). Dėl šios ir kitų priežasčių struktūriniu EDM ontologijos modeliavimo standartu buvo pasirinktos OAI *Object Reuse and Exchange* (OAI-ORE) specifikacijos (<http://www.openarchives.org/ore/1.0/to>). OAI-ORE apibrėžia saityno išteklių grupių identifikavimą ir aprašą. Siekiant sukurti vienareikšmišką kreiptį į saityno išteklių grupę, įvedamas naujas *išteklis*, atstovaujantis kitų *ištekliai grupėi*. Šiam naujam *ištekliai*, vadinamam *sankaupa*, kaip ir bet kokiam kitam saityno ištekliai, priskiriamas URI. Kadangi *sankaupa* yra konceptualus konstruktas, jis modeliuojamas kaip neinformacinis ištekliai, neturintis savo *išraiškos*, o aprašomas kito *ištekliai*. Pastarasis *ištekliai* vadinamas *ištekliai žemėlapiu*; jis turi URI ir kompiuterio skaitomą *išraišką*, teikiančią išsamią informaciją apie *sankaupą*. Iš esmės *ištekliai žemėlapis* nurodo, kurią *sankaupą* jis aprašo (1 schema „ore:describes“ santykis), ir išvardija *sukauptus išteklius*, esančius *sankaupos* dalimi (2 schema „ore:aggregates“ santykis, esantis „determs:hasPart“ daline savybe). Be to, *ištekliai žemėlapis* gali reikšti santykius ir savybes, būdingus visiems *sukauptiems ištekliai*, taip pat ir paties *ištekliai žemėlapiu* metaduomenis. Pavyzdžiui, 2 schema iliustruoja, kad nurodoma *ištekliai žemėlapis*



1 schema. Esminiai OAI-ORE duomenų modelio komponentai



2 schema. OAI-ORE įgaliojimas: kreiptis į sukauptą išteklių kontekste

lapio autorystė ir modifikacijos laikas (atitinkamai „dcterms:creator“ ir „dcterms:modified“ santykiai). Išteklių žemėlapis gali išreikšti ir sankaupos, sukauptų išteklių bei paties išteklių žemėlapio santykius su bet kuriuo kitu sutartiniu ištekliumi.

Be to, duomenų modelis leidžia išreikšti, kad sukauptas išteklius ir pats yra sancaupa (viena į kitą įdėtos sankaupos). Tuo tikslu išreiškiamas sukaupto išteklių ir išteklių žemėlapio santykis „ore:isDescribedBy“ (kuris yra inversinis santykiui „ore:describes“ ir „rdfs:seeAlso“ dalinė savybė), teigiantis, kad sukauptas išteklius pats yra sancaupa. Aptikimo tikslais duomenų modelis taipogi sudaro galimybę išteklių žemėlapij išreikšti, kad tam tikros sankaupos sukauptas išteklius taip pat yra ir kitos sankaupos dalis. Tai pasiekama naudojant sukaupto išteklių ir tos kitos sankaupos santykį „ore:isAggregatedBy“ (inversinį santykiui „ore:aggregates“). Kreipčiai į kultūros paveldo vertybes gana dažnai naudojami ir neprotoliniai identifikatoriai, kurie gali būti išreiškiami kaip URI (pvz., URN). Siekiant tai įgyvendinti praktiškai, išreiškiamas sankaupos ir iš dalies lygiavertio išteklių, identifikuojamo neprotolinio URI, santykis „ore:similarTo“. „ore:similarTo“ specifiškumą būtų galima nusakyti taip: tarp „rdfs:seeAlso“ ir „owl:sameAs“. Būtina pabrėžti, kad Susietų duomenų bendrija tebesvarsto, ar reikia kitojų negu „owl:sameAs“ lygiavertiškumo išraiškų (http://events.linkedata.org/ldow2010/papers/ldow2010_paper09.pdf).

Pažymėtina, kad konkrečios sankaupos sukauptą išteklių žymintis URI nesiskiria nuo URI, identifikuojančio šį išteklių neatsižvelgiant į sancaupą. Tačiau daugeliu taikymo atvejų – taip pat ir *Europeana* – būtina skirti kreiptį į patį išteklių ir kreiptį į tą patį išteklių, esantį kokios nors sankaupos sukauptu ištekliumi. To pavyzdžiai ga-

lėtų būti citavimas kontekste ir proveniencijos. Tokiam atskirymui pasiekti OAI-ORE įvedama įgaliojimo sąvoka. Įgaliojimas – tai išteklius, atstovaujantis sukauptam ištekliui tam tikros sankaupos kontekste. Įgaliojimo URI naudojamas kaip priemonė ištekliui kontekste reikšti. 2 schemeje vaizduojami „ore:ProxyFor“ ir „ore:ProxyIn“ santykiai, siejantys įgaliojimą atitinkamai su sukauptu ištekliumi ir sancaupa. Ji taip pat iliustruoja ir tai, kuo skiriasi sukaupto išteklių ir jo įgaliojimo citavimas: pirmuoju atveju cituojamas išteklius „toks, koks jis yra“, antruoju – išteklius, koks jis yra konkrečios sankaupos kontekste. Siekiant sklandžiai orientuotis saityne ir teikti kontekstinę informaciją OAI-ORE suprantantiems klientams, nustatant įgaliojiniams priskirtus HTTP URI, turėtų būti nukreipiama į sukauptą išteklių, o atsakymas turėtų apimti HTTP nuorodos antraštę (<http://datatracker.ietf.org/doc/draft-nottingham-http-link-header/>), kuri nukreiptų į sancaupą.

3.3. EDM ir susiję atviri duomenys

Susietų duomenų tikslas – sudaryti galimybę struktūruotų duomenų mainams saityne. Esminį pagrindą jo sėkmingam įgyvendinimui teikia saityno struktūra ir pateikimo kalba – RDF. Jie padeda sukurti prieigą prie RDF duomenų saityne, taip įgyvendinant visuotinę saityno, kaip duomenų sanglaudos, kuri paprastai vadinama *semantiniu saitynu*, viziją.

Semantinio saityno vizija turėjo didelę įtaką EDM projektavimui. Ji ypač svarbi pasirenkant RDF kaip EDM metamodelį ir nusprendžiant, kad bet koks *Europeana* informacinėje erdvėje reikšmingas objektas (kultūros paveldo objektas ar konceptualus entitetas, toks kaip vieta, sąvoka ir pan.) turi būti išteklius, identifikuojamas HTTP

URI. Toks pasirinkimas leidžia normalizuoti *Europeana* aprašų vertes taikant HTTP URI formatą – *de facto* standartą, suderinamą su saityno struktūra.

Susieti duomenys šiai vizijai suteikia labai svarbų mastą, nes jie įgalina *Europeana* savo informacinėje erdvėje naudoti HTTP URI ir kaip nuorodas, suteikiančias prieigą prie atitinkamų objektų struktūruotų aprašų. Todėl šios nuorodos atlieka jungčių, susiejančių *Europeana* informacinę erdvę su kitų autoritetinių duomenų telkiniu informacine erdve, funkciją, taip suteikdamos galimybę *Europeana* kaupti papildomus duomenis apie asmenis, vietas, sąvokas ir t. t. Neabejotina, kad tokiu būdu surinktos žinios turėtų atlikti svarbų vaidmenį tobulinant svarbius *Europeana* tinkamumo aspektus, tokius kaip galimybė realizuoti atradimo funkcijas.

3.4. Svarbių semantinių santykių nustatymas

EDM nesaisto įkeltų metaduomenų pateikimo su kokia nors viena bendra schema, o veikia kaip bendra aukščiausio lygmens ontologija, kurią naudojant gali būti išreiškiami metaduomenys, derantys su kitais, originaliais duomenų modeliais. Juo siekiama kuo išsamesnės, o ne bendriausios sričių aprėpties. Tai leidžia integruoti skirtingų *Europeana* duomenis teikiančių bendruomenių atskiras informacines perspektyvas ir poreikius. Taip gali būti išlaikoma originali skirtingų bendruomenių taikomų standartų, tokių kaip LIDO, CIDOC CRM, MARC ir EAD, įvairovė, o vidiniai aukščiausio lygmens santykiai garantuoja beprecedentę įvairiais formatais pateiktų duomenų atitiktį ir tikslumą.

Ypač nelengva buvo rasti semantinių santykių, kurie būtų išskirtinai svarbūs informacijai apie socialinių ir kultūrinių artefaktų struktūrą ir prieigą prie jų, būtų pakankamai abstraktūs ir galėtų integruoti tūkstančius labiau specializuotų santykių, tačiau ir pakankamai reikšmingi dabar esantiems tikslumo skirtumams tarp paieškos pagal reikšminius žodžius ir prieigos pagal dalykines sritis pašalinti. Nors pagrindinės schemos, tokios kaip DC, VRA ir CIDOC CRM ontologija, yra labai tipinės, nė viena jų iki galo neapėrija kitos.

Paliekant nuošalėje tapatumo valdymą, galiausiai būtų galima išskirti penkis esminius semantinius santykius, kuriuos išsamiau diferencijuoja tokie standartai, kaip *Dublin Core*:

- klasifikavimas naudojant kategorijas, kurios gali būti reikšiamos SKOS;
- skaidymas į dalis ir informacinių išteklių inkorporavimas vieno į kitą;
- panašumas, t. y. santykis tarp objektų arba informacinių išteklių, pasižyminčių kai kuriomis bendromis savybėmis – atsitiktinai, dėl kieno nors įtakos ar susijusios

išvedimo istorijos, kaip apibūdinama FRBR (Doerr and LeBoeuf, 2007);

- „apie“ santykis, t. y. entitetai arba idėjos, kurias įkūnija ar pateikia objektas arba informacinis išteklius arba kuris į jas nukreipia ar „yra apie“ jas;

- elemento istorija, t. y. objektai, asmenys, vietos, laikotarpiai, įvykiai, su kuriais kažkas turėjo ryšį, kurių metu buvo, su kuriais „susitiko“. Žvelgiant labiau analitiškai, visi istoriniai santykiai gali būti paaiškinami ir praplečiami kaip buvimas įvykiuose bei susijusiuose įvykių parametruose.

Visos paminėtos šių santykių vertės gali būti pateikiamos kaip URI, nukreipiantys į atitinkamus susietus atvirus duomenis, pavyzdžiui, VIAF (www.viaf.org) asmenims ir geografinių vardų žodyną – vietoms. Asmenys gali būti pateikiami ir naudojant FOAF ontologiją bei jų santykius su EDM įvykiais.

4. EDM ir bendruomenėms skirtos pateikimo schemos

„Saityno kalba“ RDF leidžia teigti savybių integracinius ryšius, kitais žodžiais tariant – tai, kad santykis, susiejantis kokį nors elementą su konkrečia verte, apima ir bendresnius santykius su šia verte (yra „<kieno> posavybis“), lygiai kaip tezauro „platesni terminai“. Todėl įmanoma tokiais ryšiais susieto elemento paieška bendresnio santykio, kuris nėra deklaruotas duomenyse, kontekste. Šia RDF savybe kol kas plačiau nesinaudoja duomenis kuriančių vartotojų bendruomenės, nors iš tikrųjų tai yra svarbiausia priemonė, integruojanti bendruomenėms skirtas ir tipines arba „esmines“ pateikimo schemas. Pavyzdžiui, „dcterms:created“ apima arba detalizuoja „dc:date“. EDM suteikia galimybę radikaliau negu kada nors iki šiol apibendrinti metaduomenų savybes.

Europeana numato, kad bet kuri bendruomenė, siekdama didesnio EDM esančių duomenų tikslumo, gali paskelbti taikomąjį profilį (Heery and Patel, 2000). Visi tokių profilių nustatomi iš esmės svarbūs ryšiai turėtų išsamiai detalizuoti vieną arba kelias EDM savybes ir taip užtikrinti, kad specifinę bendruomenės semantiką atspindėtų ir apie ją neinformuotų vartotojų atliekamų užklausų rezultatai. Savo ruožtu *Europeana* įdėta daug pastangų renkant informaciją apie praktinius metaduomenų reikalavimus iš bendruomenių, t. y. apie metaduomenų formatus, grindžiamus gausiais faktiniais duomenimis, kurie gali būti pateikiami *Europeana*, ir tikrindama jų suderinamumą su EDM. Pagaliau ir pats EDM yra šių pradinių duomenų galutinio apibendrinimo rezultatas.

Su *Europeana* susijusios bendruomenės gali būti skirstomos į bibliotekas, skaitmenines bibliotekas, garso ir vaizdo, archyvų ir muziejų sektorius. Muziejuose šiuo

metu yra didžiausia dalinių disciplinų ir metaduomenų formatų įvairovė, atspindinti jų administruojamų materialių objektų įvairovę. Muziejai teikia mažiausiai *Dublin Core* metaduomenų. Jungtinėse Valstijose labiau populiarūs yra VRA ir CDWA, nes taikant šiuos standartus fiksuojama fizinė objektų buvimo vieta. EDM integruoja visus tris formatus. Tačiau išsamūs muziejų metaduomenys labai sudėtingi, todėl buvo susitarta taikyti CIDOC CRM (ISO 21127:2006), kuris apibrėžia turtingą bendrą metaduomenų elementų semantiką RDF atitinkančia forma, tačiau savaime nenustato laukų naudojimo. Tik neseniai svarbiausios tarptautinių suinteresuotų muziejų sektoriaus grupės sutarė, kad muziejų duomenų surinkimui bus taikomas LIDO formatas. Jis suderinamas su CRM ir jį taikant Europos ATHENA projekto metu jau sukurti gausūs duomenų telkiniai. LIDO sukurtas pertvarkant *CDWA-Lite* į įvykį orientuotą schemą. LIDO duomenys gali būti transformuojami į EDM, taip pat ir į daug išsamesnį CRM. Į CRM gali būti transformuojami ir šimtai kitų muziejų formatų. CRM savo ruožtu integruoja EDM. Dėl šios priežasties muziejų bendruomenės atstovai pasisakė už CRM, kaip EDM muziejams skirtą taikomąjį bendruomenės profilį. Taikant tam tikras taisykles, CRM metaduomenų pagrindu gali būti kuriami *Dublin Core* metaduomenys (Kakali et al. 2007).

Archyvuose svarbus vaidmuo tenka rinkinio lygmens aprašams, tokiems kaip EAD. Jie puikiai dera su EDM; ypač daug galimybių archyviniams „fondams“ aprašyti teikia sąvoka „ore:aggregation“. Tarptautinė archyvų taryba neseniai inicijavo diskusiją apie bendrą conceptualų modelį, kuris būtų panašus į FRBR arba CRM. Kol kas taikomas CRM, įgalinantis išsamiau aprašyti su archyvinio turiniu susijusius istorinius faktus lyginant su vien tik EDM lygmeniu (Stasinopoulou et al. 2007). Be to, gana patogūs yra ir vis didesnį populiarumą įgyja rinkinio lygmens *Dublin Core* aprašai. Bibliotekose vis dar tebevyrauja MARC formatas, tačiau daugelį metaduomenų galima būtų teikti ir paprastesniu MODS formatu, kurį būtų galima naudoti EDM kaip taikomąjį profilį, tačiau MODS įrašams jau beveik prilygsta *Dublin Core*. Pradedant praėjusio amžiaus pabaiga, atsiradus FRBR, bibliotekų dėmesys pradėjo krypti išsamesnių metaduomenų link. Nors FRBR pasižymi plačiu taikymo spektru, tebėra ginčytina tiksli „kūrinio“ ir „išraiškos“ semantika, be to, praktinis FRBR diegimas bibliotekose atsilieka. Tačiau neginčijama tai, kad būtina kaupti turinį, atsižvelgiant į išvedimo hierarchinius lygmenis, ir tai galima daryti ne-taikant kontroversiškos kaupimo lygmenų klasifikacijos. FRBR_{oo}, kurie yra ontologinė FRBR interpretacija (Doerr and LeBoeuf, 2007), identifikuoja esminius santykius, esančius sudedamąja EDM „panašumo“ savybių dalimi. Todėl EDM iš tikrųjų gali atspindėti neginčytinas esmines nekontroversiškas FRBR sąvokas.

5. Į objektą orientuotas ir į įvykį orientuotas požiūris

Natūralu tai, kad, projektuojant metaduomenų schemas, dažniausiai susitelkiama ties požymių, kaip „radimo priemonių“, tiesioginiu priskyrimu rinkiniui priklausančiam objektui. Tokia nuostata remiasi ne vieną šimtmetį gyvuojančia gera ir labai sėkminga bibliotekininkystės tradicija. Ji grindžiama paradigma, kad vartotojui žinoma objekto tema, medžiaginės savybės arba kokie nors objekto ryšiai su giminingu elementu, tokiu kaip autorius arba redaktorius. Be to, pagal šią nuostatą, vartotojui gavus prieigą prie objekto, pats objektas patenkins jo informacinius poreikius. Šią paradigmą galima pritaikyti skaitmeniniams rinkiniams, ir tai buvo padaryta, tuo labiau kad, be prieigos naudojant metaduomenis, galima paieška ir pagal turinį.

Laikui bėgant mūsų informacinės sistemos tampa galingesnės; imama suvokti, kad metaduomenys patys savaime turi didžiulę su dokumentine informacija susijusią vertę. Be to, esant nepaprastai didelėms šiuolaikinių rinkinių apimtims, didėjantis sinonimiškumas labai riboja paprastų požymių identifikacines galimybes. Galiausiai pradedami aprašinėti didžiuliai „nekalbiniai“ objektų rinkiniai, tokie kaip vaizdai ar muziejų objektai, kurių suvokimas gali visiškai priklausyti nuo metaduomenų.

Tai skatino suinteresuotumą teikti turiningesnius ir nuoseklesnius įrašus, atspindinčius objektų kilmę ir istoriją. Pradedant dešimtojo dešimtmečio viduriu, kai kurios tarptautinės bendruomenės (IndeCs, CRM, ABC, OPM) pradėjo suvokti, kad sudėtingą istorinių sąryšių aprašymą galima normalizuoti ir ženkliai supaprastinti, jeigu, užuot daugiausia dėmesio apraše skyrus objektui, dokumentuojant tiek pat dėmesio būtų skiriama įvykiams, kurie savotiškai „sujungia“ visus dinamiškus santykius tarp asmenų, objektų, laiko ir vietos. Tai ne tik suteikia daugiau išsamumo, bet ir leidžia tiksliai aptikti objektus, kuriuos sieja bendra istorija, o tai yra tas pats, kas dalyvauti įvykiuose.

Detaliai aprašant įvykius, blogiausiu atveju gali padvigubėti kai kurių metaduomenų apimtis, nes jie pasi-pildytų įvairiais dažnai pasikartojančiais požymiais (pvz., „creation_date, creation_place, creator“), susijusiais su tuo pačiu įvykiu, iš kitos pusės – aprašai taptų išsamesni. Tarkime, tiksli senovės Egipto objektų, bronzos amžiuje atvežtų į Kretą, paieška neįmanoma, jeigu taikysime tipinius metodus neaprašydami įvykių. Reikalavimą dokumentuoti įvykius kelia kultūros paveldo bendruomenė. Galima abipusė į objektą orientuotų ir į įvykį orientuotų to paties elemento aprašų transformacija, reguliuojant tikslumo praradimą, transformuojant į objektą orientuotą kryptimi. Tokios transformacijos – tai *Europeana* modelio vidinė savybė. Todėl abiem aspektams būdingas akivaizdus kompromisas tarp sudėtingumo ir

turtingos raiškos gali būti lanksčiai pritaikomas vartotojų reikmėms tenkinti ir įkeltų šaltinių pateikimo kokybei gerinti.

Šia prasme „susitikimo“ santykis yra veiksminga EDM inovacija. Jis paprastai ir aiškiai susieja „į objektą orientuotą“ ir „į įvykį orientuotą“ modeliavimą bei praplečia „DC:data“ tipišumą iki kitų esminių neinformacinių išteklių kategorijų (įvykių, veiksmų, vietų, materialiu ir nematerialiu objektu), todėl jo semantinei aprėpčiai iki šiol negali prilygti joks duomenų standartas. Ši aprėptis garantuoja išplečiamumą, neprarandant prieigos per esminius semantinius santykius, kuriuos atspindi daugelis specializuotų duomenų laukų.

6. EDM patikra ir galimybės

Su 5-ąja versija EDM įgavo stabilumą. Ekspertų grupei pasirodė būtina šiuo etapu pradėti modelio įvertinimą ir patikrą, kad galiausiai būtų galima įtraukti jį į *Europeana* „Danube“ laidos specifikacijas; patikros pagrindą turėjo sudaryti realūs skirtingiems kontekstams atstovaujantys pavyzdžiai.

Todėl *Europeana* surengė keturis „bendruomenių susitikimus“, kuriuose dalyvavo archyvų, garso ir vaizdo archyvų, bibliotekų ir muziejų atstovai. Kiekviena bendruomenė pateikė tipinių savo rinkinių duomenų pavyzdžių. Buvo siekiama sužinoti, kaip sektųsi parengti skirtingų bendruomenių standartų ir EDM sankirtas. Buvo gauta vilčių teikiančių rezultatų. Savaimė suprantama, kad kiekviena bendruomenė įnešė savų elementų, kurie daugiausia buvo susiję su vaizdavimu ir paieška. EDM, kaip bendra aukščiausio lygmens ontologija, pasirodė esanti labai lanksti ir stabili bei galinti suderinti atskiroms bendruomenėms būdingas klases ir savybes kaip specializacijas. Paskutinis iš šių susitikimų serijos bus surengtas Pizoje 2010 m. birželio mėn., kuomet bendruomenių atsiliepimai bus integruoti į EDM.

6.1. Patikra

Archyvai pateikė archyvinės medžiagos, koduotos EAD, paieškos priemonių pavyzdžių. Tokių archyvinių aprašų skiriamasis bruožas yra giluminė hierarchinė sandara ir tai, kad daug dėmesio skiriama struktūrotam bei kontekstiniam aprašui. EDM savybės skaidyti į dalis ir inkorporuoti parodė modelio galimybes apdoroti rinkinių aprašus, apimančius keletą žemesnio lygmens aprašų, kurių kiekvienas tarpinis lygmuo apima kontekstinę informaciją.

Muziejai daugiausiai pateikė pavyzdžius, koduotus *museumdat* ir LIDO. Tvirtai į įvykį orientuota *museumdat* / LIDO koncepcija labai gerai derėjo su EDM. Pateiktos klasės ir į įvykį orientuotos savybės suteikė pa-

kankamai lanksčių modeliavimo galimybių, leidžiančių integruoti turiningus į įvykį orientuotus LIDO aprašus tipologizuojant įvykius ir sukuriant poklasius bei dalines savybes kaip EDM poklasių ir klasių specializacijas. Tačiau muziejai pasiūlė EDM klases ir savybes pakeisti, kur įmanoma, CRM klasėmis ir savybėmis bei naudoti CRM entitetus be jokių atitikmenų EDM muziejų taikomajame profilyje.

Garso ir vaizdo archyvų bendruomenė labai nevienalytė, teikianti labai skirtingus objektus ir taikanti daug skirtingų kodavimo standartų. Garso ir vaizdo sferoje aprašomą entitetą dažnai sunku identifikuoti, ir tai atitinka jos prigimtį. Garso ir vaizdo bendruomenės perspektyva nėra tokia aiški kaip muziejų (LIDO) ar archyvų (EAD), be to, jos medžiagą sudaro daug daugiau objektų sukurtų skaitmenine forma. Vis dėlto EDM įrodė galintis integruoti šios bendruomenės pateiktų pavyzdžių įvairovę ir turiningumą.

Bibliotekos pateikė daug, dažnai labai sudėtingų pavyzdžių, kurių visų sutaptis su EDM buvo sėkmingai nustatyta. Tačiau buvo aišku, kad bibliotekų duomenų modeliavimui EDM daug naudos suteiks modelio praplėtimas, kuris apims ir FRBR pirmosios grupės kategorijas „kūriny“, „išraiška“, „apraiška“ ir „vienetas“. Bibliotekų ekspertai sutarė, kad EDM įdiegus ir pradėjus taikyti RDA, ženkliai išaugs būtinybė įtraukti ir FRBR kategorijas, kurios galiausiai taptų bendruomenės taikomojo profilio dalimi.

Visų bendruomenių pasitarimų rezultatas buvo išvada, kad patikrintas ir įdiegtas EDM taps patikima integruojančia priemone, skirta skirtingų kultūrinių artefaktų, atstovaujančių įvairiems kontekstams, modeliavimui.

6.2. Galimybės

Naujasis *Europeana* duomenų modelis (EDM) pakeis *Europeana* semantinius elementus (ESE), kuriais dabar grindžiama *Europeana* duomenų erdvė. EDM teikia daugiau išraiškos ir lankstumo galimybių bei įgalina įvairesnes ir tikrai semantines milijonų objektų iš įvairiausių kultūros paveldo bendruomenių pateiktis *Europeana*.

Palyginti su ankstesniais duomenų modeliais, EDM būdingas labai aukštas abstrakcijos lygmuo. Jis radikaliau negu bet kada iki šiol apibendrina kultūros paveldo srities metaduomenų savybes ir nesaisto įkeltų metaduomenų pateiktis su kokia nors viena bendra schema. Siekiant pateikti turiningus ir funkciškai suderintus *Europeana* objektų aprašus, EDM nuosekliai integruojamos tokios plačiai pripažintos ontologijos, kaip SKOS, *Dublin Core* ir FOAF. Būdamas bendra aukščiausio lygmens ontologija, EDM leidžia integruoti aiškias įvairių į *Europeana* duomenis teikiančių bendruomenių informacines perspektyvas ir reikmes bei išlaikyti bendruomenių standartų,

tokių kaip LIDO, CIDOC CRM, MARC ar EAD, įvairovę.

Struktūruotai informacijai apie kultūros paveldo objektus identifikuoti EDM kaip metamodelį taiko RDF(S) ir URI. Struktūrinio modeliavimo pagrindu EDM ontologijai teikia OAI-ORE (*OAI Object Reuse and Exchange*) specifikacijos. Atvira EDM struktūra daro *Europeana* suderinamą su semantinio saityno paradigma ir leidžia jai tapti naujai atsirandančios susietų atvirų duomenų bendruomenės dalimi. Iš tikrųjų EDM kultūros paveldo institucijoms atveria kelią iš dabartinių, dažniausiai uždarų, informacinių struktūrų į atviras, susietas aplinkas, taip pasitarnaudamas šioms institucijoms ir saityno bendruomenei.

Bendruomenių pasitarimų metu atlikti patikros bandymai parodė, kad EDM gali sėkmingai funkcionuoti kaip bendra aukščiausio lygmens ontologija, skirta dau-

geliui labiau specializuotų duomenų modelių iš įvairių žinijos sričių. Jeigu bendruomenės, siekdamos didesnio savo EDM duomenų rinkinių tikslumo, susitartų dėl taisykomojo profilio, taptų įmanoma palyginti „be praradimų“ integruoti bet kokį duomenų modelį.

Ir galiausiai, EDM atveria naują naudojimo perspektyvą įvairiausiems į *Europeana* įkeliamiems duomenims. Tarkime, EDM grindžiama struktūra leidžia suteikti kontekstą objektų pateiktims, taip sukurdamą naujus euristinius veiksmo planus mokslininkams (tiksliau, „skaitmeninių humanitarinių mokslų“ atstovams), tačiau tai jau atskiro darbo tema...

Iš anglų kalbos vertė T. Auškalnis

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

Nuorodos

Cesare Concordia, Stefan Gradmann, Sjoerd Siebinga (2010): Not just another portal, not just another digital library: A portrait of Europeana as an application program interface. In: International Federation of Library Associations and Institutions 36(1), pp. 61-69. (<http://dx.doi.org/10.1177/0340035209360764>)

Martin Doerr, Patrick LeBocuf (2007): Modelling Intellectual Processes: the FRBR - CRM Harmonization. In: C. Thanos, F. Borri, and L. Candela (eds.): Digital Libraries: R&D, LNCS 4877, pp. 114-123, 2007. (First DELOS Conference on Digital Libraries, February 2007 Tirrenia, Pisa, Italy.)

Thomas R. Gruber (1993): Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: Nicola Guarino and Roberto Poli (eds.): Formal Ontology in Conceptual Analysis and Knowledge Representation.

Rachel Heery, Manjula Patel (2000): Application profiles: mixing and matching metadata schemas. Ariadne 25. (<http://www.ariadne.ac.uk/issue25/appprofiles/intro.html>)

Constantia Kakali, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies. In: Proceedings of the International Conference on Dublin Core and metadata Applications (DC-2007), pp. 128-139. 27-31 August, 2007. Singapore.

Barry Smith (2003): Ontology. In: L. Floridi (ed.): Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, pp. 155-166.

Thomais Stasinopoulou, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Ontology-based Metadata Integration in the Cultural Heritage Domain. In: Proceedings of the 10th International Conference on Asian Digital Libraries, pp. 165-175. Hanoi, Vietnam, December 10-13, 2007.