

Socialinių tinklų tyrimų duomenų archyvo funkciniai reikalavimai, projektavimas ir įgyvendinimas

Costis Dallas

Vilnius University, Faculty of Communication; University of Toronto, Faculty of Information

Ingrida Kelpšienė

Vilnius University, Faculty of Communication

Rimvydas Laužikas

Vilnius University, Faculty of Communication

Justas Gribovskis

Vilnius University, Faculty of Communication

Santrauka. Skaitmeninės komunikacijos eroje socialinių tinklų platformos tapo svarbiomis kolektyvinės atminties ir tapatybės kūrimo bei sklaidos priemonėmis. Tačiau dinamiška šių platformų prigimtis kartu su jų komerciniu pobūdžiu ir ribotomis archyvavimo galimybėmis kelia nemažai iššūkių mokslininkams, siekiantiems tyrinėti ir suprasti socialinių tinklų svetainėse vykstančias diskusijas paveldo, istorijos, atminties ir tapatybės tema. Šiame straipsnyje pristatomas skaitmeninių mokslinių tyrimų duomenų archyvas, kuris padeda spręsti šiuos iššūkius, taip pat dalinamasi šio archyvo kūrimo bei praktinio pritaikymo gerąja praktika, kuri įgalina semantiškai ir išsamiai rinkti, saugoti, atvaizduoti ir analizuoti socialinių medijų diskusijas.

Pagrindiniai žodžiai: tyrimo duomenų archyvas; socialinės medijos; skaitmeninė atmintis; skaitmeninis paveldas; skaitmeninis archyvavimas

Functional Requirements, Design and Implementation of a Social Network Research Data Archive

Abstract. In the era of digital communication, social networking platforms have become important tools for the creation and dissemination of collective memory and identity. However, the dynamic nature of these platforms, together with their commercial nature and limited archiving capabilities, pose a number of challenges for

Straipsnis parengtas įgyvendinant projektą „Jungiančioji skaitmeninė atmintis paribiuose: kultūrinės tapatybės, paveldo komunikacijos ir skaitmeninės kuratorystės praktikų socialiniuose tinkluose tyrimas“. Projektas bendrai finansuotas iš Europos socialinio fondo lėšų (projekto Nr. 09.3.3-LMT-K-712-17-0027) pagal dotacijos sutartį su Lietuvos mokslo taryba (LMT).

Received: 2023-10-11. **Accepted:** 2024-05-16.

Copyright © 2024 Costis Dallas, Ingrida Kelpšienė, Rimvydas Laužikas, Justas Gribovskis. Published by Vilnius University Press. This is an Open Access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

researchers seeking to research conversations on heritage, history, memory and identity on social networking sites. This article presents a digital research data archive that helps address these challenges, and also shares best practices in creating and practically applying this archive, enabling the semantic and comprehensive collection, storage, visualization, and analysis of social media discussions.

Keywords: research data archive; social media; digital memory; digital heritage; digital archiving

Įvadas

Šiuolaikinėje skaitmeninėje erdvėje socialinių tinklų platformos (toliau – STP) tapo svarbiomis visuomenėje vykstančių diskusijų vietomis, kurios atskleidžia skirtingų žmonių požiūrius ir supratimą apie paveldą, istoriją bei prisideda prie kolektyvinės atminties ir tapatybės formavimo. Vis dėlto STP turinio efemeriškumas ir platformų komercializacija kelia didelių iššūkių tyrėjams, siekiantiems tirti ir suprasti jose vykstančias diskusijas, nes tradiciniai mokslo duomenų archyvavimo ir tyrimų metodai susiduria su įvairiais apribojimais komercinių STP kontekste dėl prieigos prie duomenų ir jų kontrolės (Ben-David, 2020).

Straipsnio tikslas ir tyrimo klausimai

Šio straipsnio tikslas – pristatyti skaitmeninį mokslinių tyrimų duomenų archyvą, kuris padeda spręsti šiandieninius iššūkius bei leidžia semantiškai ir išsamiai rinkti, saugoti, atvaizduoti bei analizuoti STP diskusijas. Straipsnyje taip pat siekiama geriau suprasti ir paaiškinti tokioms sistemoms keliamus reikalavimus bei galimus sprendimus, bandant atsakyti į pagrindinius tyrimo klausimus:

1. Kokie funkciniai reikalavimai leistų tokiam archyvui reprezentuoti reikšmingas STP esančios informacijos savybes ir sudarytų sąlygas tyrėjams analizuoti tokią informaciją, atliekant mokslinius tyrimus?
2. Kaip būtų apibrėžta tokio mokslinio archyvo architektūra, jo pagrindinės charakteristikos ir funkcijos?

Literatūros apžvalga

Straipsnyje pristatoma STP turinio archyvavimo tema mokslinėje literatūroje yra nagrinėjama įvairiuose kontekstuose. Specializuotuose tyrimuose daugiausia yra analizuojami iššūkiai bei siūlomi sprendimai, skirti STP turinio rinkimui (Comer & Copeland, (n.d.); Littman ir kt., 2018), jo išsaugojimui (Thomson, 2017; Hemphill ir kt., 2021; Madhava, 2011; Raad ir kt., 2016; Tilbury, 2021); valdymui ir duomenų kuravimui (Mosweu, 2019; Sabharwal, 2021); prieigai (Thomson & Kilbride, 2015); panaudojimui moksliniuose tyrimuose (Hemphill ir kt., 2022; Colombo ir kt., 2023). Taip pat yra parengta ir archyvavimo tematikai skirtų sisteminių literatūros ar taikomųjų iniciatyvų analizių bei apžvalgų (Córdula ir kt., 2019; Michel ir kt., 2021). Kaip pažymi A. Acker ir A. Kriesbergas savo atliktoje JAV praktikos apžvalgoje, socialinių tinklų archyvavimo iniciatyvos „yra susietos

su API grindžiamais priegos režimais, taikant visiems vienodą požiūrį į vartotojų motyvaciją rinkti duomenis“ (Acker & Kriesberg, 2020: 118), tačiau, kaip žinome, poreikiai archyvuoti socialinių medijų duomenis gali būti įvairūs, pradedant archyvavimu pramoginiais ar asmeniniais tikslais (Zhao & Lindley, 2014; Cannelli & Musso, 2022) ir baigiant pilietiniais sumetimais arba viešąja atskaitomybe, kai siekiama užtikrinti vyriausybės ir valstybės pareigūnų veiksmų skaidrumą (Franks, 2016).

Socialinių medijų archyvų panaudojimą moksliniuose tyrimuose apibendrina atlikti literatūros apžvalgos tyrimai. Pavyzdžiui, K. Weller pastebi, kad tyrėjai dirbdami su „Twitter“ platformos duomenimis dažniausiai remiasi „pranešimų rinkiniais (gautais atsitiktinai arba pagal konkrečius paieškos kriterijus), – [t. y.] vartotojų profiliais / vartotojų tinklais, – ir duomenimis iš eksperimentų, apklausų ar interviu“ (Weller, 2014: 244). Kita vertus, [autorių] atlikta sisteminė sunkiojo paveldo socialinių tinklų svetainėse (įskaitant „Facebook“, „Instagram“, „Twitter“ ir „YouTube“) literatūros apžvalga rodo, jog naudojamų tyrimų metodų spektras yra labai platus ir apima ne tik kokybinę bei kiekybinę turinio analizes, bet ir audiovizualinę, tekstinę, diskurso, naratyvų, klasterių, kolokacijų, didžiųjų duomenų (angl. *big data*) ir kt. analizes (Kelpšienė ir kt., 2022). Tačiau galimybės sukurti skaitmeninius archyvus, kurie įgalintų atlikti visapusiškus STP tyrimus, panaudojant įvairius tyrimų metodus, tebėra antraeilė ir nepakankamai ištirta sritis.

Naujausiame pasauliniame archyvavimo iniciatyvų tyrime daugiausia dėmesio skiriama nacionalinių bibliotekų ir archyvų organizacijų projektams, ir daroma išvada, kad „viešai prieinamų nacionalinių archyvų ir didelio masto socialinių medijų archyvų naudojimas moksliniuose tyrimuose dar tik atsiranda“ (Vlassenroot ir kt., 2021). Negana to, neseniai atlikta dažniausiai naudojamų socialinių medijų archyvavimo sistemų apžvalga nepateikia jokių išvalgų, kaip tokios sistemos galėtų padėti tyrėjams (Borji ir kt., 2022). Naudoti STP kaip mokslinių tyrimų platformas nėra perspektyvi alternatyva, nes, kaip teigia tyrėja A. Ben-David, šiuo metu populiariausia socialinių tinklų svetainė „Facebook“ iš esmės yra nearchyvuojama ir netgi reikalauja „priešingo archyvavimo“ (angl. *counter-archiving*) metodo (Ben-David, 2020).

Mokslinės literatūros analizė šia tema parodė, kad literatūroje siūlomi STP duomenų archyvavimo ir valdymo sprendimai dažniausiai atitinka nacionalinių bibliotekų ir archyvų institucijų tikslus. Juose vyraujantis priegos modelis remiasi prielaida, jog tyrėjas turi labai aiškiai apibrėžtą ir labai konkretų poreikį ieškoti tam tikro objekto, o skaitmeninė biblioteka ir (arba) archyvas pateikia metodus, kaip gauti šį konkretų šaltinį. Deja, egzistuojantys modeliai nepateikia tinkamų sprendimų ar gerosios praktikos pavyzdžių, kuriais remiantis galėtų vykti STP turinio archyvavimas pasirinktoje teminėje srityje. Nors mokslinėje literatūroje yra paliečiami su STP turinio techniniu išsaugojimu bei prieiga susiję klausimai, vis dėlto yra matomas akivaizdus žinių trūkumas sprendžiant kitus klausimus. Pavyzdžiui, kaip suteikti prieigą prie tarpusavyje susijusių duomenų įvairovės, t. y. prie sąsajų ir jų sklaidos struktūrų ar prie komunikacinių reikšmių formavimo? Kaip atskleisti semantinius ryšius tarp duomenų struktūrų, esančių STP? Arba kaip sudaryti sąlygas duomenų interpretavimui, remiantis išankstinėmis žiniomis ir išoriniais, bet susijusiais informacijos šaltiniais? Ši žinių spraga egzistuoja todėl, kad esančios duomenų

archyvavimo sistemų paradigmos yra orientuotos į statišką turinio saugojimą, bet nėra modeliuojamos ir kuriamos kaip dinamiškos ir aktyvios mokslinių tyrimų infrastruktūros.

Tyrimo kontekstas

Šiame straipsnyje yra pristatomas mokslinio STP archyvo kūrimo procesas ir tyrimai, kurie buvo atlikti įgyvendinant projektą CONNECTIVE. Atsižvelgiant į multimodalinį socialinių medijų sąveikų pobūdį, projekto CONNECTIVE propaguojama tyrimo duomenų archyvavimo praktika siekė užfiksuoti STP pokalbių turtingumą ir sudėtingumą, kurie leistų tyrėjams atlikti išsamią turinio analizę, neapsiribojant vien tekstine analize. Be to, siūloma infrastruktūra atsižvelgė į dinamišką ir kintantį duomenų interpretacijų ir klasifikacijų pobūdį, pripažįstant įvairius kontekstus, kuriuose šios diskusijos vyksta.

Projekto metu buvo atliktas dviejų populiariausių STP („Facebook“ ir „Instagram“) turinio surinkimas ir jo lyginamoji analizė. Dėmesys buvo skiriamas lietuviškų STP turiniui, kuris yra aktualus paveldo, istorijos, atminties ir tapatybės tyrimams, apimantiems platų temų spektrą. Renkant duomenis buvo siekiama įtraukti tiek „autoritetų balsų“, tiek individualias „paprastų žmonių“ bei neformalių visuomenės grupių nuomones. Duomenų rinkimas taip pat buvo grindžiamas supratimu, kad paveldas, atmintis ir tapatybė nėra pastovios ar monolitinės sąvokos, o veikia per nuolatines dinamiškas sąveikas ir (re) interpretacijas socialiai konstruojamos ir formuojamos realijos.

Metodai ir darbo planas

Siekdami nustatyti pagrindinius funkcinis reikalavimus, kurie įgalintų sukurti tinkamą skaitmeninių duomenų archyvą, skirtą moksliniams tyrimams atlikti, vadovavomės užklausomis grindžiama reikalavimų nustatymo metodika (Potts ir kt., 1994). Pirmiausia buvo atlikta leksinė analizė, siekiant nustatyti pasikartojančius terminus, susijusius su socialinių medijų duomenimis (subjektus, ryšius ir savybes), taip pat ir identifikuoti teorinius konceptus bei tyrimų metodus, naudojamus STP tyrimuose. Analizei atlikti buvo pasitelktas 274 paskelbtų mokslinių tyrimų rinkinys, kurio publikacijose orientuojamasi į socialinių medijų veikimą įvairiose srityse, pavyzdžiui, archeologinės bendruomenės, atmintis, tapatybė, sunkusis paveldas, Holokaustas ir kt. Šią analizę taip pat papildė projekto tyrėjų praktinių poreikių analizė, kurios pagrindu buvo nustatytos duomenų, susijusių su atminties socialinių medijų praktikose tyrimais, charakteristikos, stebint dešimties projekto tyrėjų, dirbančių su Lietuvos socialinių tinklų duomenimis, praktiką. Formuluojant šiuos reikalavimus buvo atsižvelgta į: 1) identifikuotas STP tematinę pogrupių užklausas, kurias tyrėjai formulavo savo tyrimui atlikti; 2) jų atsakymais į klausimą, kuriame buvo prašoma išsamos informacijos apie vykdomo tyrimo apimtį, mokslinius klausimus, pirminius STP duomenis, analizės metodus ir preliminarias išvadas, kurių pagrindu bus rengiami moksliniai straipsniai bendraautorių monografijoje; 3) vertinant jau parengtas atvejų studijas (būsimus monografijos straipsnius), kuriose analizuojami STP duomenys. Antrajame etape, siekiami nustatyti funkcinis reikalavimus, taikėme

scenarijais grindžiamą metodą (Sutcliffe ir kt., 1998) ir nagrinėjome du lietuviškų STP duomenų mokslinio naudojimo tyrimuose scenarijus (aptariami kitame skyriuje), kurie sujungia STP duomenų aspektus ir identifikuotus tyrimų praktikos metodus bei CON-NECTIVE projekto tyrėjų poreikius.

Atsižvelgiant į etinius reikalavimus, STP duomenų masyvas buvo formuojamas tik iš viešai prieinamų pokalbių, įrašų ir komentarų, siekiant juos išsaugoti kaip svarbias Lietuvos kultūros paveldo, atminties ir tapatybės praktikų reprezentacijas. Šis pasirinkimas etikos požiūriu kontekste dera su EU erdvėje veikiančiomis socialinių ir humanitarinių mokslų tyrimų gairėmis (Europos Komisija, 2018) bei darbo su socialinių medijų duomenimis rekomendacijomis (Weller & Watteler, 2021), nes užtikrina teisę į mokslinių tyrimų laisvę bei tarnauja viešajam interesui. Tačiau, laikantis tų pačių etinių reikalavimų, duomenys nebuvo renkami iš asmeninių profilių ir privačios ar ribotos prieigos grupių – kas buvo vienas iš esminių STP duomenų rinkimo apribojimų, nes privačiose grupėse esanti informacija taip pat gali būti svarbi tiek moksliniu, tiek istoriniu, politiniu ar kt. požiūriu.

Schemos ir duomenų bazės projektavimo specifikacija buvo parengta susiejant ankstesniame etape (Kirtiklis ir kt., 2023) išplėtotą socialinių medijų komunikacijos conceptualųjį (ontologijos) modelį (tiksliau, jame identifikuotas esybes, ryšius ir savybes) su duomenų bazės schema, panaudojant ontologijos įrankį „Protégé“ ir „arrows.app“ modeliavimo įrankį (Jones & Karaca, 2023), ir siekiant, kad schema reprezentuotų grafų modelį, kuriame būtų galima naudoti „Cypher“ užklauso kalbą (Robinson ir kt., 2015). Pradinė modelio versija buvo išbandyta „Neo4J“ grafų duomenų bazėje, siekiant atvaizduoti: 1) duomenų rinkinį, kuris buvo išeksportuotas iš nedidelės mums prieinamų „Facebook“ pokalbių ir „Instagram“ įrašų imties; ir 2) tyrimu grindžiamas socialinių medijų duomenų savybes bei jų analitines kategorijas, kurios buvo naudojamos mokslinėms žinioms reprezentuoti. Peržiūrėta galutinė schemos versija buvo parengta iteratyviai, išbandant duomenų bazės gebėjimą atvaizduoti informaciją ir atlikti tinkamas informacijos paieškas per užklausas „Neo4J“ duomenų bazėje (Lal, 2015).

„Neo4J“ pasirinkimą lėmė kelios svarbios priežastys. Visų pirma, „Neo4J“ sistemos (įskaitant ir semantinį) išraiškingumas, kuris leidžia apibrėžti mazgus ir ryšius bei jų savybes, kurios atliepia įvairius STP turinio sudėtingumo laipsnius, numatytus STP ontologijoje, taip pat leidžia priskirti tam tikras „etiketes“ (angl. *labels*) duomenims. Antra, sistemos išplečiamumas, kuris įgalina pridėti reikalingas ypatybes ir ryšius, atsiradus naujai ar papildomai informacijai STP duomenų masyve. Ne mažiau svarbus buvo ir sistemos pritaikomumas, nes jos suderinamumas su „Cypher“ užklauso kalba leido tiesiogiai „Neo4J“ duomenų bazėje įdiegti išplėtotą schematinį modelį ir palaikyti lanksčias užklauso ir schemos keitimo operacijas. Be to, savybių grafų modelio ir „etikečių“ derinys leido atvaizduoti STP sąveikas kaip žinių grafus. Semantinės schemos ypatybės, kurių neapėmė sudarytas grafų modelis (pvz., unikalumas, kardinalumas, referencinis vientisumas), buvo įgyvendinamos algoritmiškai, taikant procedūrinę logiką.

Skaitmeninio archyvo diegimui buvo pasitelkta „Python“ programavimo kalba ir „kwas“ scenarijų ir duomenų apdorojimo užduotims atlikti. JSON ir CSV buvo naudojami kaip duomenų saugojimo ir keitimosi jais formatai. Transformuotiems duomenims saugoti

grafų duomenų bazėje „Neo4J“ buvo naudojama „Amazon Web Services“ (AWS) archyvo saugojimo ir valdymo saugykla, kuri leidžia saugoti ne tik tekstinę, bet ir kitą svarbią informaciją, pvz., nuotraukas, vaizdo, garso įrašus ir kt. AWS paslaugos paketas buvo pasirinktas todėl, kad jis turi vartotojo prieigos, tapatybės ir saugos strategijų valdymo funkcijas, kurios leidžia vartotojams pasirinkti savo saugos poziciją pagal konkrečius poreikius ir užtikrinti duomenų matomumą bei kontrolę.

Specifinis žinių atvaizdavimo aktyviame STP archyve metodas apima duomenų transformavimo procesą, kurio tikslas – atskleisti konceptualią socialinių medijų duomenų struktūrą. Duomenų archyvo naudojimo eiga yra daugiapakopis procesas, t. y. nuo pradinio šaltinio duomenų surinkimo ir įtraukimo iki konkrečios analizės rezultatų generavimo, todėl infrastruktūra turi palaikyti atskirus informacijos paketus, skirtus duomenims surinkti, archyvuoti ir pateikti (CCSDS, 2007). Informacijos surinkimo paketuose duomenys buvo saugojami JSON formatu ir apėmė kilmės bei išsaugojimo metaduomenis, taip pat originalių pokalbių STP faksimilines kopijas. Archyvinės informacijos paketai buvo kuriami atvaizduojant duomenis žinių grafais, leidžiančias atlikti išsamią ir labiau struktūruotą analizę. Galiausiai, informacijos pateikimo paketai buvo generuojami įvairiais formatais, pvz., .graphml (<http://graphml.graphdrawing.org/>), skirtais socialinių tinklų analizei (SNA), ar .csv, kuris įgalina eksportuoti ir toliau analizuoti duomenis su kita programine įranga kaip MAXQDA (<https://www.maxqda.com>), Excel ar SPSS. Duomenų archyve taip pat buvo taikomas žodynu grindžiamas temų priskyrimo metodas, kuris leido kategorizuoti ir priskirti temas skirtingiems duomenų subjektams.

Vartotojo reikalavimai

Norint tinkamai atvaizduoti STP sąveikas mokslinių duomenų archyve, jos turėjo apimti tiek pačius duomenų objektus, tiek jų istorinį ir kultūrinį kontekstą, taip pat ir išsaugojimo metaduomenis bei vykdomos mokslinės tyrimų veiklos apraiškas. Pirmiausia juose turėjo būti dokumentuojamas duomenų surinkimo, vertinimo ir įtraukimo procesas, pateikiant informaciją apie tai, kaip duomenys buvo surinkti ir atrinkti į archyvą. Jame taip pat turėjo būti fiksuojami ir atvaizduojami duomenų objektai, tiesiogiai susiję su šiomis STP diskusijomis, įskaitant pačius pranešimus ir juose dalyvaujančius dalyvius. Tai reiškia, kad reikėjo išsaugoti ne tik žinučių turinį ir retorinę formą, bet taip pat ir dalyvių reakcijas, tokias kaip pasidalijimai įrašu, komentarai, bei įvairias emocines išraiškas (pvz., emotikonai „patinka“, „nuliūdęs“ ir kt.). Be pirminių objektų duomenų, į archyvą taip pat turėjo būti įtraukti ir susiję duomenys, suteikiantys šioms diskusijoms istorinį-kultūrinį kontekstą. Tokia informacija apėmė vartotojų profilių išsaugojimą ir datos bei laiko fiksavimą, matomą kitose diskusijose ir kituose susijusiuose internetiniuose ištekliuose. Užfiksavus šią kontekstinę informaciją, archyvas galėjo padėti visapusiškai susieti STP sąveikas ir jų reikšmę platesniame kultūriniame ir istoriniame kontekste.

Siekdami suprasti STP archyvo naudotojų poreikius, taip pat atlikome scenarijais grindžiamą vartotojų (tyrėjų) poreikių tyrimą (Sutcliffe ir kt., 1998), kurio metu buvo analizuojami du scenarijai: 1) kritinė diskurso analizė ir vartotojų įsitraukimas į „Facebook“

diskusiją Lietuvos partizanų istorijos tema; 2) semiotinė analizė ir agentiško abdukcija (angl. *abduction of agency*) prieštaringoje STP diskusijoje dėl sovietinio rašytojo Petro Cvirkos paminklo nukėlimo.

Pirmojo scenarijaus atveju nagrinėjome socialinių medijų įrašą „Facebook“ grupėje „Lietuvos partizanų istorija“, kuriame minimas Tauro apygardos Perkūno rinktinės vadas Vaclovas Navickas-Rytas. Šis įrašas sulaukė pakankamai didelio vartotojų aktyvumo, t. y. 117 vartotojų spaudė „patinka“, 23 išreiškė „liūdžiu“, 13 „myliu“, o 4 „susirūpinęs“ reakcijos. Įrašą taip pat lydėjo 7 pasidalijimai ir 6 komentarai. Analizuodami esamas vartotojų įsitraukimo formas, galime suprasti, kokias emocijas ir socialines reakcijas sukėlė šis įrašas ir pastebėti įvairias Vaclovo Navicko-Ryto pagerbimo, pagarbos ir susižavėjimo juo išraiškas. Pavyzdžiui, vienas vartotojas išreiškia pagarbą partizanui teigdamas: „Garbė Lietuvos didvyriui!“, o ši frazė palydima Lietuvos vėliavos emotikono. Jo komentarui pritaria 7 kiti vartotojai, spausdami „patinka“. Taikydami kritinės diskurso analizės metodą, galime nustatyti tam tikras lingvistines ir konceptualiąsias strategijas, naudojamas vartotojų komentaruose, kuriuose atribucijos sąvoka tampa akivaizdi, nes Vaclovui Navickui-Rytui priskiriamas „Lietuvos didvyrio“, „Lietuvos patrioto“ ir „Lietuvos sūnaus“ statusas. Diskusijoje taip pat išryškėja ir predikacijos samprata, kur nuorodose į Lietuvos partizanus vyrauja teiginiai, jog jie „paaukojo savo gyvybes už laisvę“, pabrėžiant partizanų auką. Nėgana to, nagrinėdami vartotojų komentarus, galime išskirti ir konkrečias konceptualiąsias metaforas. Pavyzdžiui, frazė „Lietuvos sūnus“ siejama su TĖVYNĖS KAIP MOTINOS metafora, kur Tėvynė vaizduojama kaip globojanti ir sauganti figūra, pabrėžiant gilų emocinį individų ryšį su ja. Šios metaforos vartojimas parodo visuomenėje egzistuojantį suvokimą, kad Vaclovas Navickas-Rytas yra herojinė figūra, glaudžiai susijusi su Lietuvos tapatybe ir gerove, ir taip įkūnija tautos vertybes ir siekius.

Antrojo scenarijaus atveju nagrinėjome STP diskusiją dėl Vilniaus miesto tarybos vykdomo prieštarigai vertinamo sovietinio rašytojo Petro Cvirkos paminklo nukėlimo, kurioje analizuojamas semiosferinis turinio vertimas ir veikimo abdukcija, o diskursas nagrinėjamas per komunikacinio ryšio teorijos objektyvą (Laužikas, Dallas, forthcoming). Kokybinė STP diskusijos turinio analizė atskleidžia skirtingas paminklo pašalinimo šalininkų ir priešininkų perspektyvas. Paminklo nukėlimo šalininkai P. Cvirką apibūdina kaip „kruviną kolaborantą“ ir „išdaviką“, pabrėždami jo pripažinimą sovietų valdžios institucijose bei dalyvavimą atvežant „Stalino saulę“ į Vilnių. Jų oponentai teigia, kad P. Cvirkos paminklo pašalinimas naikina Lietuvos istoriją, ir pabrėžia, kad paminklas priklauso jų kolektyvinei istorijai, nepriklausomai nuo jos teigiamų ar neigiamų aspektų. Jie kritikuoja įamžinimo politizavimą ir teigia, kad P. Cvirka turėtų būti vertinamas tik kaip poetas.

Taikant semiosferos ir Art Nexus teorines prieigas ir traktuojant diskusijas STP kaip turinio reikšmės vertimą, Petro Cvirkos, kaip sovietų kolaboranto, suvokimas skatina kasdiniuos agentiško abdukcijos procesus. Tai apima perėjimą nuo pagrindinės autoriaus žinutės (indekso) ir joje komunikuojamos reikšmės (prototipo) prie žinutės „perskaitymo“ ir „vertimo“, kurį atlieka vartotojai (gavėjai, suvokimo agentai). Pastarieji prisideda prie naujų žinučių (indeksų) konstravimo su jau kitokiomis reikšmėmis (prototipais) nei tos,

kurios buvo pagrindinėje autoriaus žinutėje. Tokiu būdu vartotojai įvykdo agentiško abdukciją ir perima autoriaus iniciatyvą komunikuoti savo reikšmes (prototipus). Mokslinio pažinimo prasme yra svarbu, kad toks tyrimo požiūris atskleidžia dinamišką agentiško ir suvokimo pobūdį STP diskusijoje, parodant, kaip dalyvių veiksmas ir pasisakymas daro įtaką kitų dalyvių interpretacijoms ir reakcijoms.

Abu pateikti scenarijų pavyzdžiai leido geriau suprasti STP diskusijų specifika ir jų sudėtingumą mokslo duomenų infrastruktūros modeliavimo požiūriu. Taip pat pavyzdžiai leido identifikuoti funkcinis duomenų infrastruktūros reikalavimus, į kuriuos būtina atsižvelgti renkant STP duomenis ir kuriant skaitmeninę mokslo duomenų infrastruktūrą. Pavyzdžiui, labai svarbu yra turėti ne tik tikslų pagrindinio įrašo, paskelbto autoriaus, tekstą, bet taip pat ir tiksliai perkopijuotus vartotojų komentarus, jų eiliškumą išsidėstymą laike, identifikuotas emocines reakcijas (emotikonas), kiekybinius reakcijų rodiklius ir kitus svarbius duomenų atributus.

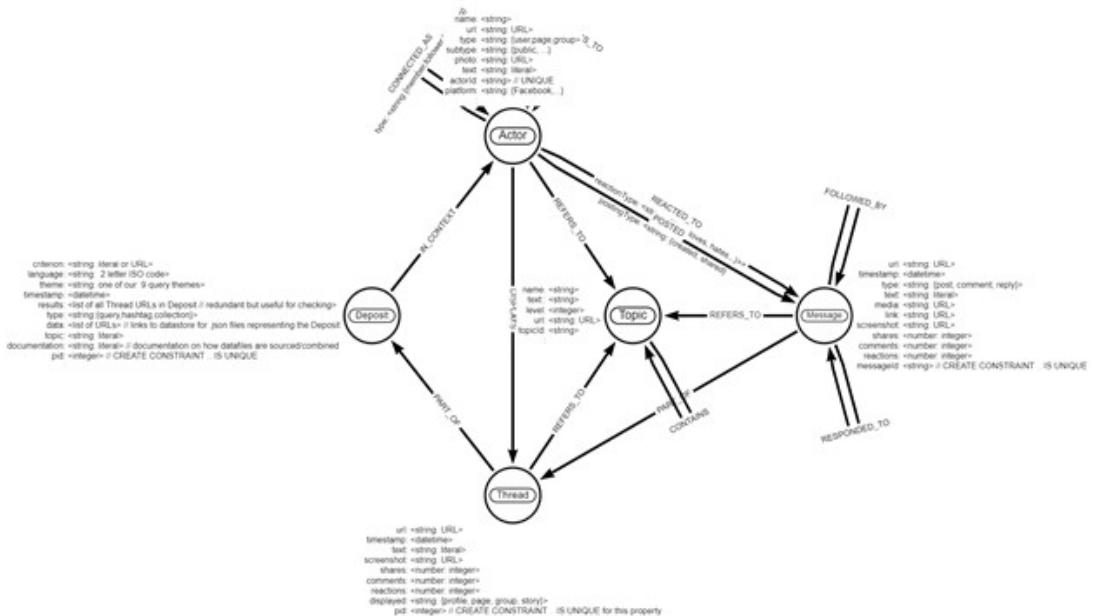
Visi išsaugoti metaduomenys atliko svarbų vaidmenį, užtikrinant archyvuojamų STP duomenų vientisumą ir autentiškumą. Be to, norint tinkamai atvaizduoti STP sąveikas kaip įrašus, taip pat reikėjo įtraukti nuolatines kategorizacijas, įrašų praturtinimus ir anotacijas, kurias atliko tyrėjai, naudodamiesi archyvu ir taip prisidėdami prie duomenų organizavimo ir interpretavimo proceso. Tai apėmė pokalbių temų ar potemių priskyrimą, reikšmingų elementų anotavimą ir įrašų praturtinimą papildoma kontekstine informacija. Pavyzdžiui, diskurso konstrukcijų ir kultūrinių metaforų, įsivirtinusių STP diskusijose, nustatymas svarbus tuo, jog jis atskleidžia naratyvus ir struktūras, kurios formuoja paveldo, atminties ir tapatybės konceptualizavimo ir aptarimo būdus. Analizuodami šiuos konstruktus tyrėjai gali identifikuoti dominuojančius naratyvus, nustatyti galios dinamiką ir ideologines pozicijas, kurios daro įtaką socialinių medijų pokalbiams. Kita vertus, kultūrinės metaforos apima ir simbolinės kalbos, vaizdinių ir nuorodų, parodančių ir įtvirtinančių tam tikras kultūrinės vertybes, įsitikinimus ir supratimą, vartojimą, todėl šios metaforos dažnai atlieka svarbų vaidmenį formuojant paveldo, atminties ir tapatybės reikšmes ir interpretacijas socialinių medijų pokalbiuose. Iš esmės šios diskurso konstrukcijos yra priemonės, jungiančios individą ir bendruomenę bei veikiančios kaip asmens ir grupės (bendruomenės) tapatumo pagrindas. Dėl šios priežasties labai svarbus reikalavimas duomenų archyvui buvo jo gebėjimas atstovauti tokioms diskursyvinėms, semiotinėms ir simbolinėms struktūroms.

Duomenų modelis

CONNECTIVE projekto duomenų archyvo modelis (1 pav.) ir jo realizavimas remiasi ankstesniame straipsnio skyriuje išdėstytais funkciniais duomenų infrastruktūros reikalavimais bei grindžiamas projekte išplėtotą paveldo ir atminties praktikų semiotinės veiklos socialiniuose tinkluose ontologija (Kirtiklis ir kt., 2023).

Pagrindinės duomenų archyvo charakteristikos apima depozito esybių ir IN_CONTEXT ryšių atvaizdavimą, kuris fiksuoja archyvavimo ir metodologiniais tikslais svarbius pro-veniencijos ir išsaugojimo metaduomenis. Semiotinė socialinių medijų diskusijų forma

ir turinys taip pat reprezentuojami per tokias esybes, kaip „gija“ (angl. *thread*) ir „pranešimas“ (angl. *message*), bei jų ryšius, tokius kaip PART_OF, RESPONDED_TO, FOLLOWED_BY, DISPLAYS ir SHARED_FROM. Šie elementai fiksuoja socialinių medijų sąveikų sintagminę, sklaidos ir diskusinę struktūrą, leidžiančią atlikti įvairias diskurso analizės operacijas. Archyve egzistuoja ir veikėjo (angl. *Actor*) esybė ir jo ryšiai, pvz., POSTED, REACTED_TO ir CONNECTED_AS, kurie fiksuoja veikėjo narystę grupėje, semiotines veiklas ir vartotojų, dalyvaujančių socialinių medijų sąveikoje, emocinę poziciją. Šios esybės įgalina įvairius STP analitinius metodus, leidžiančius geriau suprasti socialinių tinklų dinamiką ir santykius.

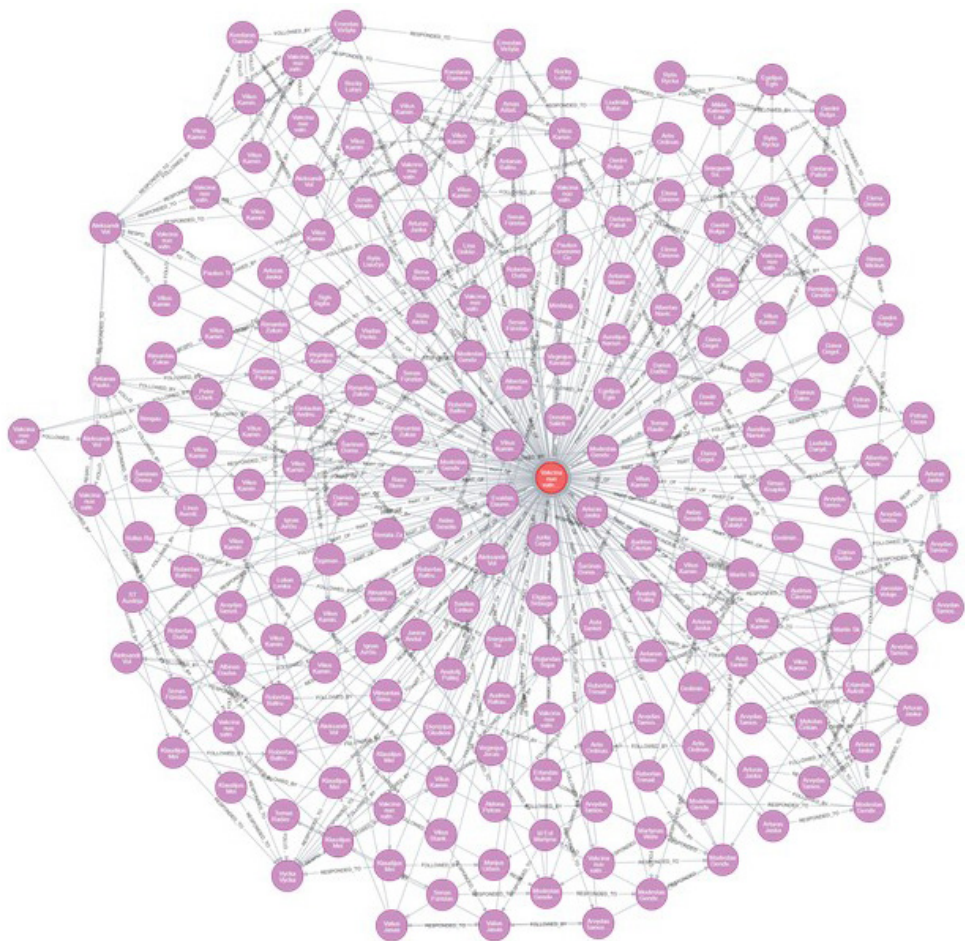


1 pav. Duomenų modelis

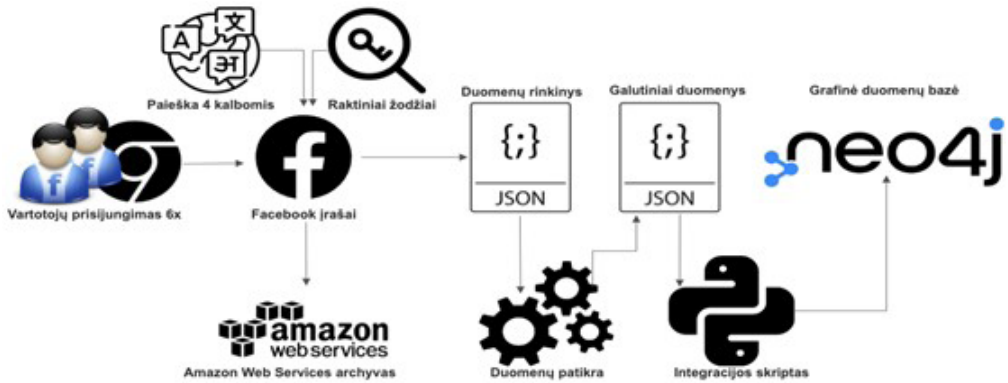
Kad būtų galima atlikti teminę analizę ir suskirstyti duomenis į kategorijas, taip pat priskirti įvairių rūšių analitines kategorijas, tokias kaip predikacija, metafora, prielaida, aliuzija ir kt., konkretiems duomenų elementams, schemeje numatytas papildomas objekto tipas „Tema“. Temos gali būti susijusios taksonominėje hierarchijoje per CONTAINS ryšį, nustatant izoliavimo rūšį (bendroji, mereologinė, geografinė ar teminė įtrauktis). Duomenų archyvas turi žodynu pagrįstą temų priskyrimo metodą, leidžiantį suskirstyti duomenis į kategorijas arba priskirti juos skirtingoms temoms. Šis priskyrimas remiasi duomenų bazių užklausomis, kuriose atsižvelgiama į įvairius veiksmus (pvz., paskyros, puslapio ar grupės ID), bei leksiniais modeliais, kurie vyrauja įrašuose, pranešimuose ir komentaruose. Veikėjams, pranešimams ir diskusijoms priskiriami „temų mazgai“ (angl. *topic nodes*), palengvinantys tematinį duomenų organizavimą ir analizę. Šis atvaizdavimas reprezentuoja referencinę reikšmę STP diskusijose ir jų istorinius, semantinius bei interpretacinius aspektus. Kiti duomenų bazės objektai (pranešimai, gijos, agentai) gali

būti susiję su viena ar keliomis temomis pagal IS_ABOUT ryšį. Kai duomenų masyve esantys įrašai yra suskirstomi pagal temas, tyrėjai gali lengviau atsirinkti duomenis, kurie yra reikalingi tyrimui, ir kartu kelti kur kas platesnio pobūdžio klausimus tam tikrame tematiniame bloke. Pavyzdžiui, kokios asociacijos, siejamos su didvyriškumu, yra priskiriamos asmenims, kurie socialinių medijų žinutėse buvo įvardijami kaip partizanai? Arba kokios konceptualios metaforos yra vartojamos esamuose pranešimuose, kuriuose yra minimas Holokaustas? „Temų mazgai“ padeda socialinių medijų įrašus tyrinėti ir per laiko prizmę, bandant atsekti tam tikrų idėjų ar konceptų sklaidą. Pavyzdžiui, kada pirmą kartą P. Cvirka buvo įvardijamas kaip sovietų kolaborantas ir kaip ši idėja toliau buvo eskaluojama įrašuose ir vartotojų komentaruose?

Duomenų archyvo projektavimas ir įgyvendinimas

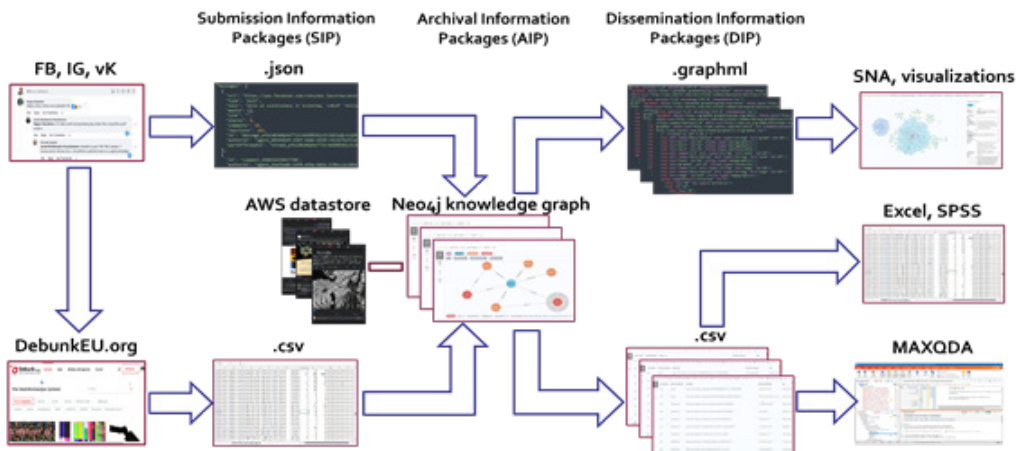


2 pav. Ryšių tinklas grafų schemeje „Neo4J“ (atvaizduota 10 000 ryšių iš 812 692)



4 pav. Duomenų rinkimo iš STP ir integracijos schema

Darbas su duomenų archyvu vyko keliais etapais, t. y. nuo pradinio šaltinio duomenų fiksavimo ir gavimo iki konkrečios analizės duomenų generavimo (5 pav.). Archyvo architektūra buvo grindžiama Atviros archyvų informacinės sistemos (OAIS) principais, paremtais gerąja archyvavimo praktika. Šiuo metu OAIS yra etaloninis atvirosios prieigos archyvų modelis, atitinkantis ISO standartą, kuris apibrėžia svarbiausius skaitmeninės informacijos nuolatinio ir ilgalaikio išsaugojimo principus (CCSDS, 2012). Sukurta architektūra sujungė procesus, informacijos struktūras ir sistemas, kurios buvo naudojamos projekto CONNECTIVE STP duomenims įvertinti, užfiksuoti, gauti, praturtinti ir suteikti prieigą prie jų pagal OAIS standartą.



5 pav. Mokslo duomenų infrastruktūros architektūra

Archyvas palaiko atskirus informacijos surinkimo paketus (angl. *Submission Information Packages*, SIP), į kuriuos įeina JSON srautai, apimantys duomenų kilmės ir išsaugojimo metaduomenis, taip pat originalių STP diskusijų faksimilines kopijas. Archyvinės informacijos paketai (angl. *Archival Information Packages*, AIP) sukuriama, suskirstant

duomenis į žinių grafus, leidžiančius atlikti išsamesnę ir struktūriškesnę analizę. Galiausiai, informacijos pateikimo paketai (angl. *Dissemination Information Packages*, DIP) generuojami įvairiais formatais, pvz., .graphml socialinių tinklų analizei (SNA) atlikti ir .csv eksportai turinio analizei atlikti, naudojant tokias priemones kaip MAXQDA, Excel ar SPSS. Minėtų funkcinių duomenų infrastruktūros reikalavimų realizavimas sukūrė skaitmeninių priemonių, metodų ir duomenų valdymo procesų derinį, kuris leido socialinių medijų archyvui veiksmingai fiksuoti, transformuoti ir analizuoti STP duomenis, taip pat ir tyrėjams dirbti su saugomais duomenimis bei tyrinėti įvairius socialinės sąveikos aspektus bei reiškinius.

Įgyvendinant CONNECTIVE projektą, archyvas buvo sukurtas kaip įrankis, kurį naudojant buvo saugojami ir analizuojami projekto metu surinkti duomenys. Jo pagrindu taip pat buvo sukurta prototipinė internetinė skaitmeninio tyrimų archyvo sąsaja¹, pritaikyta „Neo4J“ duomenų bazei ir skirta dviejų tipų vartotojams, t. y. tyrėjams (registruotiems vartotojams) ir plačiajai visuomenei (neregistruotiems vartotojams). Pastaroji prieiga leidžia neregistruotiems vartotojams atlikti: a) laisvo teksto užklausų paiešką; b) paiešką pagal temą; c) užklausą pagal pavyzdį, įvedant kelis kriterijus. Neregistruoti vartotojai taip pat gali pasiekti duomenų bazę naudodami temų katalogą, kuriame grupuojami ir pateikiami rezultatai remiantis plečiama temų, susijusių su tyrimo objektu, hierarchija (pvz., partizanai, Holokaustas ir kt.). Gijų ir pranešimų, atitinkančių užklausą, rezultatai pateikiami kaip URL nuorodos, nukreipiančios į jų pradinę vietą atitinkamoje socialinių tinklų platformoje, o informacija apie STP vartotojus yra redaguojama, užtikrinant, kad būtų laikomasi teisinių ir etinių duomenų apsaugos parametrų, atitinkančių BDAR reglamentavimą.

Registruota (slaptažodžiu apsaugota) prieiga yra skirta tyrėjams, o šiuo metu prieinama tik projekto CONNECTIVE tyrimų grupės nariams, kurie, be minėtų trijų tipų paieškų, taip pat gali pasiekti duomenis duomenų bazėje naudodami vartotojo apibrėžtą šifravimo užklausą. Užklausos rezultatai taip pat gali būti eksportuojami .csv, .json arba .graphml formatais. Registruoti vartotojai taip pat gali matyti STP vartotojus ir originalias STP pokalbių, saugomų AWS duomenų saugykloje, ekrano kopijas. Ateityje archyvas galėtų būti pildomas atliekamo tyrimo arba įgyvendinamo mokslinio projekto pagrindu, surenkant papildomus STP duomenų paketus, importuojant juos į archyvo infrastruktūrą ir analizuojant duomenis, pasitelkiant archyve įdiegtus funkcinius įrankius.

Diskusija ir išvados

Standartiniai skaitmeninio archyvavimo iniciatyvų metodai paprastai yra riboti, jei archyve norima turėti standartinio formato, pvz., WARC (ISO 2017), ir į tikslų skaitmeninio ištekliaus atkuriamumą orientuoto skaitmeninio pakaitalo bei metaduomenų įrašo, apimančio tiek katalogavimo, tiek išsaugojimo informaciją, derinį, kuris užtikrintų archyvuojamų duomenų integralumą ir autentiškumą (Vlassenroot ir kt., 2021). CONNECTIVE pro-

¹ Šiuo metu archyvas yra pasiekiamas adresu: <http://46.101.102.232>.

jekte propaguojamas požiūris į STP duomenų archyvavimą yra grindžiamas kitokiu nei „tradicinis“ archyvavimas teoriniu ir praktiniu požiūriu, kuris remiasi kritiniais moksliniais tyrimais, atliktais skaitmeninio kuravimo ir žiniatinklio archyvavimo srityje. Šiam modeliui būdingos kelios savybės, kurios jį išskiria kaip specifinę STP tyrimų duomenų infrastruktūrą, peržengiant tradicinio skaitmeninio archyvavimo ribas. Visų pirma, mūsų propaguojamas požiūris atitinka „priešingo archyvavimo“ (Ben-David, 2020) idėjas, ragindamas rinkti, pertvarkyti ir publikuoti STP duomenis, nesilaikant „tradicinio“ archyvavimo principų, bet siūlant alternatyvų archyvavimo modelį. Antra, šis modelis pripažįsta, jog svarbu fiksuoti ne tik pirminius socialinių medijų duomenis, bet ir jų referencinį bei episteminį turinį ir kontekstą, kuris leistų atlikti išsamią analizę, vykdant mokslinius tyrimus. Be to, sukurtas skaitmeninis archyvas laikosi Atviros archyvų informacinės sistemos (OAI) principų, užtikrinant duomenų autentiškumą ir vientisumą bei aiškiai atribojant pateikimo, archyvavimo ir sklaidos informacijos paketus. Galiausiai, sistema įgalina atlikti įvairių tipų analizes, pertvarkant ir praturtinant esamas schemas.

Savo prigimtimi duomenų archyvas gali būti apibūdintas kaip silpnai susietas, besikeičiantis ir netvarkingas, tačiau tik tokiu būdu jis gali integruoti įvairiapusių ir dinamiškus socialinių medijų duomenis, reprezentuojančius nuolat kintantį skaitmeninių sąveikų pobūdį. STP duomenys yra sudėtingi, juose gali egzistuoti įvairių tipų turinio ir sąveikų, todėl juos gali reikėti nuolat atnaujinti ir pritaikyti, kad būtų užfiksuoti atsirandantys nauji modeliai ir reiškiniai. Aprašytas duomenų valdymo procesų ir metodų derinys leidžia STP archyvui efektyviai fiksuoti, transformuoti ir analizuoti socialinių medijų duomenis, taip mokslininkams sukuriama visapusiška ir prisitaikanti infrastruktūra, skirta įvairiems socialinių sąveikų ir reiškinų aspektams tyrinėti. Mūsų siūlomas modelis taip pat remiasi naudojimo atvejais ir įrodymais pagrįstais reikalavimais, pabrėžiant praktinį aktyvaus STP archyvo taikymą ir naudingumą. Tačiau žinoma, jog, tyrėjams naudojantis archyvu, dirbant su duomenimis ir tyrinėjant įvairias jo dimensijas, gali atsirasti naujų išvalgų ir reikalavimų.

Tai nereiškia, kad mūsų siūlomas alternatyvus archyvo modelis neturi apribojimų. Šiuo atžvilgiu techniniai klausimai, ypač susiję su patikimų duomenų rinkimu, susiduria su nuolatiniais iššūkiais. Be to, siekiant užtikrinti atitiktį teisės aktams ir atsakingą duomenų valdymą, reikia atidžiai spręsti teisinius ir etinius klausimus, apimančius asmens duomenų apsaugą, intelektinę nuosavybę ir dalijimąsi tarpinių etapų tyrimo duomenimis.

Apibendrinant galima teigti, kad mūsų požiūris į STP archyvavimą ir jo kaip mokslo duomenų infrastruktūros realizacija siūlo patikimą sprendimą, kuris atsižvelgia į teorinius, metodologinius ir praktinius aspektus. Spręsdami STP duomenims būdingus iššūkius ir sudėtingus šios srities klausimus, mes praktiškai prisidedame prie skaitmeninės kuratorystės srities plėtos, gilindami supratimą apie socialinę ir kultūrinę skaitmeninių sąveikų dinamiką. Integruojant įrašų tęstinumo idėjas, pragmatinį požiūrį į skaitmeninę kuratorystę ir atminties, kaip gyvybiškai svarbios archyvinės paradigmos, pripažinimą, mūsų sistema suteikia tvirtą pagrindą CONNECTIVE projekto paveldo, atminties ir tapatybės tyrimams lietuviškose socialinių tinklų svetainėse atlikti. Sistema leidžia archyvuoti skirtingus vartotojų požiūrius ir perspektyvas bei spręsti etinius ir politinius iššūkius,

būdingus šiuolaikinei archyvavimo praktikai, taip pat dalinai išsprendžia ir socialinių medijų archyvavimo erdvėje egzistuojančius iššūkius.

Literatūra

Acker, A., & Kreisberg, A. (2020). Social media data archives in an API-driven world. *Archival Science*, 20(2), 105–123. <https://doi.org/10.1007/s10502-019-09325-9>

Ben-David, A. (2020). Counter-archiving Facebook. *European Journal of Communication*, 35(3), 249–264. <https://doi.org/10.1177/0267323120922069>

Borji, S., Asnafi, A. R., & Naeini, M. P. (2022). A Comparative Study of Social Media Data Archiving Software. *Preservation, Digital Technology & Culture*, 51(3), 111–119. <https://www.degruyter.com/document/doi/10.1515/pdct-2022-0013/html>

Cannelli, B., & Musso, M. (2022). Social media as part of personal digital archives: Exploring users' practices and service providers' policies regarding the preservation of digital memories. *Archival Science*, 22(2), 259–283. <https://doi.org/10.1007/s10502-021-09379-8>

Colombo, G., Bounegru, L., & Gray, J. (2023). Visual Models for Social Media Image Analysis: Groupings, Engagement, Trends, and Rankings. *International Journal of Communication*, 17(0), 1–28.

Comer, R. S., & Copeland, A. J. (n.d.). *Methods for Capture of Social Media Content for Preservation in Memory Organizations*. <https://bulletin.jcdl.org/Bulletin/v11n2/papers/comer.pdf>.

Consultative Committee for Space Data Systems (CCSDS). (2012). *Reference Model for an Open Archival Information System (OAIS): Recommended practice*. <http://public.ccsds.org/publications/archive/650x0m2.pdf>.

Córdula, F. R., Siebra, S. de A., & Araújo, W. J. de. (2019). Preservação Digital em Mídias Sociais: Uma revisão sistemática de literatura. *Revista Ibero-Americana de Ciência da Informação*, 13(1), 391–411. <https://doi.org/10.26512/rici.v13.n1.2020.24727>

European Commission. (2018). *Ethics in social science and humanities*. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf.

Franks, P. C. (2016). Applying records management principles to managing public government social media records. In S. M. Zavattaro & T. A. Bryer (eds.), *Social media for government: Theory and practice* (pp. 45–59). Routledge.

Hemphill, L., Hedstrom, M. L., & Leonard, S. H. (2021). Saving social media data: Understanding data management practices among social media researchers and their implications for archives. *Journal of the Association for Information Science and Technology*, 72(1), 97–109. <https://doi.org/10.1002/asi.24368>

Hemphill, L., Schöpke-Gonzalez, A., & Panda, A. (2022). Comparative sensitivity of social media data and their acceptable use in research. *Scientific Data*, 9(1), Article 1. <https://doi.org/10.1038/s41597-022-01773-w>

Jones, A., & Karaca, I. (2023). *Arrows.app* [JavaScript]. Neo4j-labs. <https://github.com/neo4j-labs/arrows.app> (Original work published 2020).

Kelpšienė, I., Armakauskaitė, D., Denisenko, V., Kirtiklis, K., Laužikas, R., Stonytė, R., Murinienė, L., & Dallas, C. (2022). Difficult heritage on social network sites: An integrative review. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448221122186>

Kirtiklis, K., Laužikas, R., Kelpšienė, I., & Dallas, C. (2023). An Ontology of Semiotic Activity and Epistemic Figuration of Heritage, Memory and Identity Practices on Social Network Sites. *SAGE Open*, 13(3). <https://doi.org/10.1177/21582440231187367>

Lal, M. (2015). *Neo4j Graph Data Modeling*. Packt Publishing Ltd.

Laužikas, R., & Dallas, C. (forthcoming). The message is the agent: Nexus and semiosphere in social media communication. In E. Zantides & S. Andreou (eds.), *Semiotics and Visual Communication IV: Myths of Today*. Cambridge Scholars Publishing.

Littman, J., Chudnov, D., Kerchner, D., Peterson, C., Tan, Y., Trent, R., Vij, R., & Wrubel, L. (2018). API-based social media collecting as a form of web archiving. *International Journal on Digital Libraries*, 19(1), 21–38. <https://doi.org/10.1007/s00799-016-0201-7>

- Madhava, R. (2011). 10 Things to Know About Preserving Social Media. *Information Management*, 45(5), 33–37, 54.
- Michel, A., Pranger, J., Geeraert, F., Lieber, S., Mechant, P., Vlassenroot, E., Chambers, S., Birkholz, J., & Messens, F. (2021). *WP1 report: An international review of Social Media Archiving initiatives* [Report]. <https://orfeo.belnet.be/handle/internal/7741>.
- Mosweu, T. (2019). The good, the bad and the ugly: Social media prospects and perils for records management. *ESARBICA Journal: Journal of the Eastern and Southern Africa Regional Branch of the International Council on Archives*, 38, 45–62. <https://doi.org/10.4314/esarj.v38i1.3>
- Potts, C., Takahashi, K., & Anton, A. I. (1994). Inquiry-based requirements analysis. *IEEE Software*, 11(2), 21–32. <https://doi.org/10.1109/52.268952>
- Raad, E., Al Bouna, B., & Chbeir, R. (2016). Preventing sensitive relationships disclosure for better social media preservation. *International Journal of Information Security*, 15(2), 173–194. <https://doi.org/10.1007/s10207-015-0278-9>
- Rieder, B., Abdulla, R., Poell, T., Woltering, R., & Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring “We are all Khaled Said.” *Big Data & Society*, 2(2). <https://doi.org/10.1177/2053951715614980>
- Robinson, I., Webber, J., & Eifrem, E. (2015). *Graph Databases: New Opportunities for Connected Data*. O'Reilly Media.
- Sabharwal, A. (2021). Functional Frameworks for Socialized Digital Curation: Curatorial Interventions and Curation Spaces in Archives and Libraries. *Library Trends*, 69(3), 672–695. <https://doi.org/10.1353/lib.2021.0009>
- Sutcliffe, A. G., Maiden, N. A. M., Minocha, S., & Manuel, D. (1998). Supporting scenario-based requirements engineering. *IEEE Transactions on Software Engineering*, 24(12), 1072–1088. <https://doi.org/10.1109/32.738340>.
- Thomson, S. D. (2017). Preserving social media: Applying principles of digital preservation to social media archiving. *Researchers, Practitioners and Their Use of the Archived Web*, 1–13. <https://doi.org/10.14296/resaw.0007>
- Thomson, S. D., & Kilbride, W. (2015). Preserving Social Media: The Problem of Access. *New Review of Information Networking*, 20(1–2), 261–275. <https://doi.org/10.1080/13614576.2015.1114842>
- Tilbury, J. (2021). Part 2: Preserving content from closed systems: Digital preservation of social media. *IQ: The RIMPA Quarterly Magazine*, 37(3), 27–28. <https://search.informit.org/doi/10.3316/informit.20211111056663>.
- Vlassenroot, E., Chambers, S., Lieber, S., Michel, A., Geeraert, F., Pranger, J., Birkholz, J., & Mechant, P. (2021). Web-archiving and social media: An exploratory analysis. *International Journal of Digital Humanities*, 2(1), 107–128. <https://doi.org/10.1007/s42803-021-00036-1>
- Weller, K. (2014). What do we get from Twitter—and What Not? A Close Look at Twitter Research in the Social Sciences. *Knowledge Organization*, 41(3), 238–248. <https://doi.org/10.5771/0943-7444-2014-3-238>
- Weller, K., & Watterler, O. (2021). *GDPR & ethical issues with social media data*. <https://doi.org/10.5281/zenodo.4558694>
- Zhao, X., & Lindley, S. E. (2014). Curation through use: understanding the personal value of social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI, 14)* (pp. 2431–2440). Association for Computing Machinery. <https://doi.org/10.1145/2556288.2557291>