

SVEIKATOS (IR NE TIK) DUOMENŲ EŽERAS LIETUVOJE

Valstybės informacinėse sistemose kasdien sukuriamas didelis kiekis duomenų, bet galimybės juos naudoti ilgą laiką buvo ribotos. Lietuva per pastaruosius kelerius metus sukūrė Europoje pirmaujančią teisinę ir technologinę infrastruktūrą – valstybės duomenų ežerą, kuriame sujungiami įvairūs sveikatos ir kitų sričių duomenys ir sudaromos sąlygos juos analizuoti. Ką šiandien iš šių duomenų galime padaryti ir ko dar trūksta, kad duomenų analizė taptų kasdiene sveikatos sistemos praktika?

KELIAS IKI EŽERO

2016 m. Europos Sąjungoje įsigaliojęs Bendrasis duomenų apsaugos reglamentas nustatė bendrus asmens duomenų tvarkymo principus, tačiau paliko nemažai erdvės praktiniams sprendimams, ypač dėl sveikatos duomenų. Įvairios organizacijos skirtingai interpretavo, kokius duomenis, kokiais tikslais ir kokiomis sąlygomis galima teikti, todėl duomenų panaudojimas liko fragmentiškas. Dar 2018 m. Europos Komisija įvardijo poreikį kurti bendrą Europos sveikatos duomenų erdvę, kuri suvienodintų šias taisykles tarp valstybių narių ir sudarytų sąlygas saugiai bei efektyviai naudoti sveikatos duomenis mokslui, inovacijoms ir viešajam valdymui.

Šiuos iššūkius ypač išryškino COVID-19 pandemija. Nors Lietuvoje jau veikė Elektroninė sveikatos paslaugų ir bendradarbiavimo informacinė sistema (ESPBI IS), galimybės operatyviai susieti jos duomenis su kitais valstybės registrais ir naudoti juos analizei nebuvo aiškiai apibrėžtos. Pandemija tapo praktiniu įrodymu, kad vertę kuria ne vien sukaupti duomenys, bet ir gebėjimas juos panaudoti.

Lietuva atsakė į šį iššūkį visa grandine teisinių, institucinių ir technologinių pokyčių. 2020 m. sukurta Valstybės duomenų valdysenos informacinė sistema, paremta *Palantir Foundry* technologija. 2021 m. priimtas Lietuvos Respublikos pakartotinio sveikatos duomenų naudojimo įstatymas, o 2022 m. atnaujintas Lietuvos Respublikos oficialiosios statistikos ir valstybės duomenų valdysenos įstatymas. Pertvarkytas Lietuvos statistikos departamentas virto Valstybės duomenų agentūra (VDA), buvo įteisinta valstybės duomenų sąvoka ir nustatytos jų naudojimo taisyklės.

Šių pokyčių rezultatas yra vientisa Lietuvos duomenų pakartotinio naudojimo ekosistema. Joje dalyvauja duomenų valdytojai – Registrų centras, Valstybinė ligonių kasa (VLK), sveikatos priežiūros įstaigos ir kitos institucijos – bei duomenų naudotojai: ministerijos, Higienos institutas, mokslininkai ir verslo atstovai. Sistemos centre veikia Valstybės duomenų agentūra, kuri ne tik organizuoja duomenų gavimą ir teikimą, bet ir suteikia saugią bandomąją analizės aplinką (vadinamąją „smėliadėžę“), vysto analitinius sprendimus ir kaupia kompetenciją įvairiais duomenų naudojimo klausimais.

Tokios integruotos duomenų kolekcijos kartu su analizės įrankiais vadinamos duomenų ežerais. Panašių pavyzdžių galima rasti ir privačiame sektoriuje – gerai žinoma bendrovė IQVIA, kaupianti duomenis iš ligoninių, vaistinių ir kitų šaltinių bei teikianti jų analizės paslaugas. Vis dėlto Lietuvos nacionalinis ežeras išskirtinis ir pločiu – vienoje infrastruktūroje sujungiami ne tik sveikatos, bet ir kitų sričių duomenys lygiomis prieigos teisėmis visiems, t. y. tiek viešajam, tiek privačiam sektoriui. Tai sukuria unikalias galimybes moksliniams tyrimams, duomenimis grįstam viešajam valdymui ir inovacijoms.

EŽERO DUOMENYS

Lietuvoje veikia keli nacionalinio lygmens sveikatos duomenų šaltiniai: vadinamoji *E. sveikata* (Elektroninė sveikatos paslaugų ir bendradarbiavimo informacinė sistema, ESPBI IS), Valstybinės ligonių kasos informacinė sistema *Sveidra* bei teminiai registrai, tokie kaip Higienos instituto Mirties atvejų ir jų priežasčių valstybės registras ar Nacionalinio vėžio instituto Vėžio registras. Didelė dalis duomenų dar glūdi ligoninių, poliklinikų ir vaistinių informacinėse sistemose.

Unikalu tai, kad šiuos šaltinius galime ne tik analizuoti atskirai, bet ir susieti tarpusavyje. Pavyzdžiui, analizuojant pacientų eiles, galima sujungti siuntimų duomenis iš ESPBI IS, registracijos informaciją iš Išankstinės pacientų registracijos sistemos (IPR IS) ir galiausiai faktinį apsilankymą pagal *Sveidros* duomenis. Tokia analizė leidžia vertinti laukimo laiką pagal regioną, gydytojo specialybę, gydymo įstaigos tipą ir kitus pjūvius. Sveikatos duomenys gali būti jungiami ir su kitų valstybės informacinių

sistemų duomenimis, todėl atsiranda galimybė atsakyti į gerokai platesnius klausimus: kaip darbo netekimas veikia savižudybės riziką, ar modernūs gydymo metodai vienodai pasiekiami skirtingų socialinių ir ekonominių grupių gyventojams, kokie veiksniai lemia sveikatos netolygumus tarp regionų ir pan.

Didelė dalis šių duomenų jau integruota į VDA duomenų ežerą, o atsirandant naujiems poreikiams jis nuolat plečiamas. Šiandien ežere veikia 868 jungtys su duomenų teikėjais, per kurias saugiai ir operatyviai perduodama beveik 15 tūkst. duomenų lentelių. Nors dauguma šių duomenų yra administracinio pobūdžio – registrai ir informacinės sistemos, ežere kaupiami ir molekuliniai duomenys, pavyzdžiui, COVID-19 virusų sekoskaita, ne-struktūruoti tekstiniai dokumentai, o ateityje planuojama įtraukti ir medicininius vaizdus.

Prieiga prie šių duomenų organizuojama paraiškų principu. Tyrėjai, viešojo sektoriaus institucijos ar kiti naudotojai gali kreiptis per Oficialiosios statistikos portalą. Valstybės duomenų agentūra įvertina paraišką, prireikus padeda pasirinkti tinkamus duomenų šaltinius ir įvertinti planuojamų analizų įgyvendinamumą. Suderinus duomenų apimtį ir, jei taikoma, kainą, pareiškėjui suteikiama prieiga prie bandomosios aplinkos su duomenimis. Ši sistema naudojama aktyviai – vien per 2026 m. pirmąjį

pusmetį VDA gavo 133 paraiškas (tiesa, didžioji jų dalis buvo susijusi ne su sveikatos duomenimis). Svarbu tai, kad beveik visais atvejais duomenys analizuojami pačioje VDA platformoje ir nelieka tyrėjų kompiuteriuose ar USB atmintinėse.

Viešojo sektoriaus institucijoms ir įstaigoms duomenys teikiami nemokamai (tiksliau, teikimo sąnaudas padengia valstybės biudžetas). Kitais atvejais – moksliniams tyrimams, privataus sektoriaus inovacijoms, farmacijos ar kitoms veikloms – gali būti taikomi mokesčiai. Mokestis grindžiamas ne duomenų komercine verte, o jų parengimo sąnaudomis. Taigi planuojant mokslinius projektus reikia nepamiršti, kad duomenų gavimas yra visavertė tyrimo dalis, kuriai reikia laiko ir išteklių. Kita vertus, vertėtų pasvarstyti, ar šių išlaidų padengimas centralizuotai nepadėtų sukurti geresnių sąlygų duomenimis pasinaudoti platesniam mokslininkų ratui.

NUO PRIEIGOS PRIE POVEIKIO

Duomenų ežeras – tai ne tik sukaupti duomenys, bet ir technologinė platforma, leidžianti juos efektyviai analizuoti. VDA naudoja *Palantir Foundry* platformą, kuri leidžia kurti ir prižiūrėti jungtis su duomenų bazėmis, parinkti ir teikti duomenis, administruoti vartotojus ir jų projektus. Tyrėjui ar analitikui tai – naršyklėje veikianti

Metai

Q 2025

Jar Kodas E027

Q Enter values...

Kvalifikacija E027

Q Šeimos gydytojas

Diagnozė Pirmi Trys Simboliai E027

Q Enter values...

Savivaldybė E027

Q Enter value...

Kvalifikacija Target E027

Q Gydytojas kardiologas

Gydytojas oftalmologas

Gydytojas neurologas

Gydytojas oftalmologas

Gydytojas reumatologas

Gydytojas pulmonologas

Gydytojas dermatovenerologas

Gydytojas nefrologas

Gydytojas alergologas ir klinikinis imuno...

Gydytojas hematologas

Gydytojas dietologas

Gydytojas endokrinologas

Cancel

Apply

NAGRINĖJAMOS DIAGNOZĖS E027 ir E027-ats FORMOSE.

VERTA ATKREIPTI DĖMESĮ:

- Jeigu siunčiama įstaigos viduje, specialistai gali nerašyti atsakymo (specialistų užpildyta E025 gali būti atsakymas į E027 siuntimą)
- Į šias lenteles nepatenka E025 duomenys

E027 siuntimus išrašiusių specialistų kvalifikacijos

3 parameters

Displaying 1 rows

source_practitioner_qualification	siunt_sk	pac_sk
Šeimos gydytojas	1946591	1068109

Show data

Affects 10 boards

E027 kam siunčia (target - pasirinktos kvalifikacijos)?

3 parameters

Displaying 12 rows

target_practitioner_qualification	siunt_sk	pac_sk
Gydytojas oftalmologas	531172	427151
Gydytojas kardiologas	401442	343452
Gydytojas neurologas	387005	316918
Gydytojas endokrinologas	257091	214869
Gydytojas pulmonologas	92363	76153
Gydytojas reumatologas	77411	58536
Gydytojas dermatovenerologas	66029	61095
Gydytojas nefrologas	42687	35360
Gydytojas alergologas ir klinikinis imuno...	29800	27087
Gydytojas hematologas	25068	20929
Infekcinių ligų gydytojas	23921	21431
Gydytojas dietologas	11702	10850

E027 kam siunčia (target, % nuo pasirinktų kvalifikacijų)? centai

3 parameters

Show less + Show more

Showing records: 1 - 10

1 pav. Interaktyvi ataskaita duomenų ežero aplinkoje apie gydytojų išrašomus siuntimus; kairėje esantys filtrai leidžia analizuoti duomenis įvairiais pjūviais

bandomoji darbo aplinka, kurioje galima peržiūrėti duomenų lenteles, jas jungti, vizualizuoti, kurti interaktyvias ataskaitas ar atlikti statistinę analizę.

Šioje platformoje rengiama visa VDA oficialioji statistika, o ja naudojasi ir kitos institucijos savo darbams palengvinti. Higienos instituto specialistai taip pat jau yra automatizavę dalį statistinių procesų naudodami ežero įrankius. Automatizavimas nereiškia, kad procesai vyksta be žmogaus priežiūros, tačiau jis leidžia atsisakyti rutininių užduočių. Galima kurti automatinis duomenų srautus, kokybės patikras, programinius saugiklius ir reguliariai atsinaujinančias ataskaitas, kad specialistai daugiau laiko skirtų rezultatams interpretuoti ir sprendimams priimti, o ne mechaniniam duomenų apdorojimui. Svarbu, kad šiuos produktus gali kurti patys sveikatos sistemos specialistai. Tai iliustruoja 1 pav. pateiktas pavyzdys: Valstybinė ligonių kasa ežero aplinkoje savarankiškai sukūrė interaktyvią ataskaitą, leidžiančią analizuoti, kokias diagnozes šeimos gydytojai nurodo siųsdami pacientus į gydytojų specialistų konsultacijas. Ataskaita automatiškai atsinaujina, turi kelis interaktyvius filtrus ir buvo sukurta pačių VLK specialistų. Tam neprireikė nei sudėtingo programavimo, nei viešojo pirkimo, nei būtinybės įsipareigoti išoriniam IT paslaugų teikėjui.

Kai reikalingi techniškai sudėtingesni sprendimai, Valstybės duomenų agentūra gali prisidėti ir savo analitiniais įgūdžiais. Viena iš tokių pavyzdžių – kartu su Registrų centru ir gydymo įstaigomis įgyvendinamas medicininių klasterių duomenų mainų ir stebėsenos platformos projektas. Jo metu sukurta sistema, automatiškai renkanti, apskaičiuojanti ir vizualizuojanti šimtus insulto, miokardo infarkto ir kitų klasterių veiklą apibūdinančių rodiklių. Gydymo įstaigų specialistai šiais rodikliais naudojami kasdienėje veikloje, stebi savo rezultatus ir gali operatyviai identifikuoti sritis, kuriose būtini pokyčiai.

Higienos institutas naudoja kelias VDA sukurtas bandomąsias aplinkas.

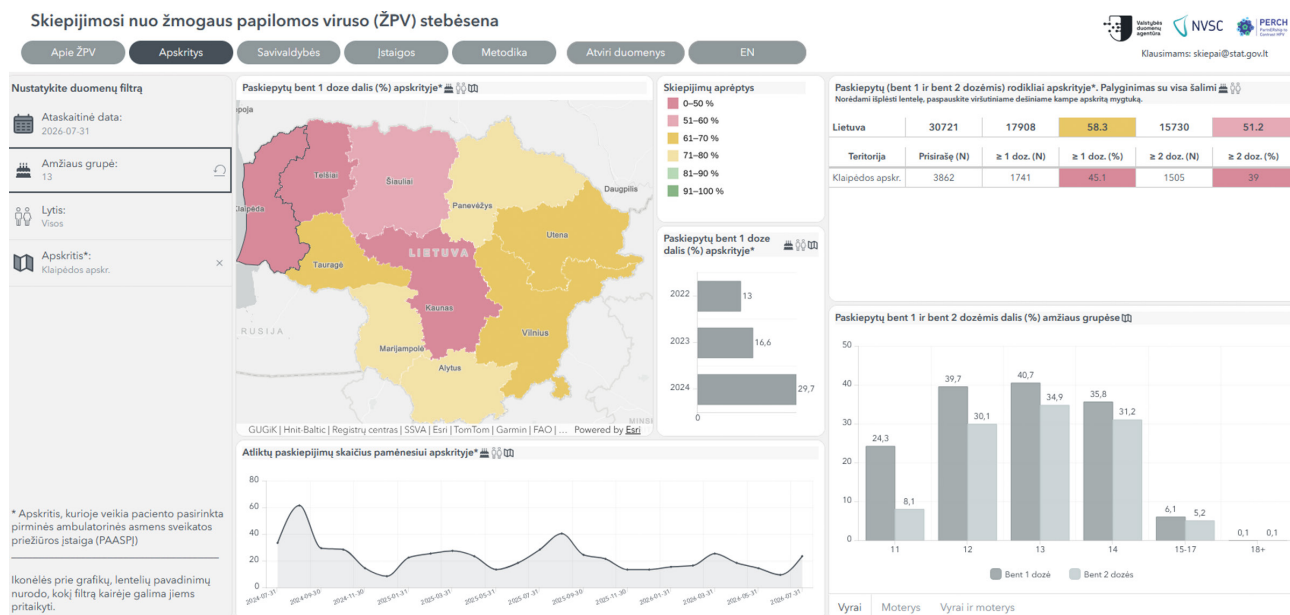
- Bandomojoje aplinkoje **Sveikatos priežiūros įstaigų personalas** sujungti *Sodros* įdarbinimų / socialinio draudimo įmokų, Valstybinės akreditavimo Asmens sveikatos priežiūrai tarnybos prie SAM licencijuotų specialistų bei įstaigų, Kompetencijų platformos duomenys naudojami detalesnei ir tiksliai sveikatos specialistų statistikai rengti.
- Bandomojoje aplinkoje **Sveikatos statistikos duomenys** sujungti *Sveidros*, Mirties priežasčių registro,

Profesinių ligų registro, ateityje: *E. sveikatos*, *Sodros* laikinojo nedarbingumo, Asmens su negalia teisių apsaugos agentūros prie SADM neįgalumo duomenys, taip pat gyventojų socioekonominiai duomenys (pajamos, išsilavinimas ir pan.) naudojami įvairiai statistinei analizei atlikti.

- Bandomojoje aplinkoje **Sveikatos specialistų poreikio prognozavimas** sujungti Švietimo sistemos studentų ir diplomų, *Sodros* įdarbinimų / socialinio draudimo įmokų, Valstybinės mokesčių inspekcijos mokesčių, Valstybinės akreditavimo Asmens sveikatos priežiūrai tarnybos prie SAM licencijuotų specialistų bei įstaigų, *Sveidros*, Kompetencijų platformos duomenys.
- Bandomojoje aplinkoje **Savižudybių ir savižudybių grėsmės atvejų rodiklių stebėseną** sujungti Mirties atvejų ir jų priežasčių valstybės registro, *Sveidros*, ESPBI, *Sodros*, Gyventojų registro, Socialinės paramos šeimai informacinės sistemos (SPIS), Studentų registro ir Mokinių registro duomenys.

Nors pats duomenų ežeras prieinamas tik registruotiems vartotojams, jo viduje sukurti nuasmeninti ar agreguoti produktai gali būti skelbiami viešai ir pasiekti geroai platesnę auditoriją. Tokiu principu veikia VLK asmens sveikatos priežiūros įstaigų veiklos vertinimo rodiklių švieslentė, skiepavimo nuo žmogaus papilomos viruso apimčių švieslentė, naudojama Nacionalinio visuomenės sveikatos centro (NVSC) veikloje (2 pav.), COVID-19 pandemijos valdymo įrankiai, rengiami Lietuvos atvirų duomenų portale skelbiami duomenų rinkiniai ir daugelis kitų produktų. Taigi ežere atliekama analizė gali virsti ne tik vidiniu darbo rezultatu, bet ir visuomenės naudojamu įrankiu.

Dar viena, kol kas mažai išnaudota galimybė – platformoje kurti ir administruoti registrus. Įsivaizduokime, kaip galėtų veikti hipotetinis ligos X registras. Kiekvieną kartą jis automatiškai gautų informaciją apie naujus ligos X atvejus iš *E. sveikatos* sistemos. Dirbtinio intelekto agentas iš klinikinių tekstų atrinktų registrai reikalingą informaciją ir ją struktūrizuotų. Automatinė kokybės patikros sistema aptiktų trūkstamus ar prieštaringsus duomenis ir apie juos informuotų registro tvarkytoją. Šis prisijungęs prie registro patvirtintų ar pataisytų informaciją specialioje formoje. Tuo pat metu sprendimų priėmėjai realiuoju laiku atsinaujinančiose ataskaitose matytų šios ligos epidemiologinę situaciją, gydymo rezultatus ar kitus



2 pav. Skiepijimo nuo žmogaus papilomos viruso apimčių švieslentė, sukurta ežero duomenų pagrindu

aktuales rodiklius.

Tai nėra futuristinis scenarijus. Jam reikalingi technologiniai įrankiai VDA platformoje jau egzistuoja, o kai kurie projektai jau yra nuėję nemažą šio kelio dalį. Pavyzdžiui, bendradarbiaujant NVSC ir VDA buvo modernizuota Užkrečiamųjų ligų ir jų sukėlėjų valstybės informacinė sistema (ULSVIS), automatizuojant dalį anksčiau rankiniu būdu atliekamų duomenų surinkimo ir apdorojimo procesų. Panašiai buvo skaitmenizuota miokardo infarkto klasterio stebėsena Lietuvoje ir įsitraukta į tarptautinę *EuroHeart* registrą. Kitaip tariant, jau neklausime, ar tokius registrus techniškai galime kurti, klausime, kiek procesų vis dar norime atlikti rankomis.

RYTOJAUS IŠŠŪKIAI

Ar šiandien jau išnaudojame visas galimybes ir visus sprendimus priimame remdamiesi duomenimis? Tikrai ne. Turime visą reikalingą infrastruktūrą ir daug sėkmės istorijų, tačiau lieka ir keli ribojantys veiksniai.

Duomenų gylis. Administraciniai duomenų šaltiniai turi savo ribas. Dalis reikalingos informacijos glūdi laisvame klinikiniam tekste. Jį jau mokame panaudoti – VDA reguliariai atliekama laisvo teksto analizė, pavyzdžiui, siekiant nustatyti laboratorinių tyrimų ar gyvybinių funkcijų matavimų reikšmes. Tačiau sudėtingesniems klausimams reikalinga informacija dažnai apskritai nėra renkama, ypač apie paciento elgseną ir savijautą už gydymo įstaigos ribų. Pavyzdžiui, iš administracinių

duomenų sunku pasakyti, ar pacientas rūko, kaip kinta jo fizinis aktyvumas. Be to, vis daugiau duomenų apie žmogaus sveikatą sukuriami ne gydymo įstaigoje, o paties žmogaus naudojamuose išmaniuosiuose nešiojamuosiuose įrenginiuose (angl. *wearable devices*). Kad tokie duomenys taptų sisteminga sveikatos informacijos dalimi, reikės nemažų pokyčių elektroninės sveikatos sistemų organizavime.

Metaduomenys. Net ir gavus duomenis, juos suprasti dažnai nėra paprasta. Kiekvienas duomenų šaltinis turi savų niuansų: kas ir kada pildoma, ko nepildoma, ką reiškia konkrečios reikšmės. Šios žinios neretai yra sukaupotos vos kelių tos srities specialistų, registro tvarkytojų ar IT tiekėjų galvose. Norint duomenis patikimai naudoti, šias žinias būtina paversti prieinama dokumentacija – aprašyti registrų struktūrą, turinį, lyginti juos su kitais šaltiniais. Net į tokį bazinį klausimą, kokia dalis *Sveidros* ir *E. sveikatos* įrašų sutampa, šiandien neturime vieno atsakymo. VDA jau skelbia publikacijas, padedančias geriau suprasti valstybės duomenų šaltinius, tačiau reikalingas ir pačių duomenų valdytojų įsitraukimas – juk būtent jie geriausiai žino duomenų atsiradimo aplinkybes ir niuansus. Dar vienas svarstomas kelias – plačiau taikyti gausiai dokumentuotus tarptautinius duomenų standartus, pavyzdžiui, klinikiniam duomenims skirtą *openEHR*.

Kompetencijos. Net ir turint gerai aprašytus bei prieinamus duomenis reikia mokėti jais naudotis. Tam reikalingi ne tik techniniai analizės įgūdžiai, bet ir sveikatos

sistemos žinios. Šiandien tokių unikalių specialistų ieškome tarp sveikatos priežiūros specialistų, rezidentų ar gyvybės mokslų tyrėjų, kurie savarankiškai persikvalifikavo į duomenų analizės sritį. Tačiau pavienių ekspertų nepakaks, jei norime, kad duomenimis grįsta sveikatos analizė taptų įprasta praktika. Poreikiui atliepti Vilniaus ir Kauno universitetai šiuo metu inicijuoja naujus studijų dalykus ir programas, skirtas tokioms tarpdisciplininėms kompetencijoms ugdyti.

Galiausiai, turime pripažinti, kad pats duomenų ežeras – tik vienas žingsnis skaitmeninės sveikatos sistemos ke-

lyje. Didžiausia vertė atsiranda tada, kai turime išsamius duomenis, suprantame jų prasmę ir turime žmonių, galinčių užduoti tinkamus klausimus bei gautus atsakymus panaudoti sprendimams. Techninės kliūtis Lietuva jau įveikė ir žengia pirmaujančiu Europoje žingsniu. Dabar svarbiausia išmokti šia galimybe naudotis.

Julius Juodakis,

Valstybės duomenų agentūra

El. p. julius.juodakis@stat.gov.lt

Literatūra

1. European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. COM(2018) 233 final. Brussels: European Commission; 2018. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0233>
2. Lietuvos Respublikos pakartotinio sveikatos duomenų naudojimo įstatymas. Nr. XIV-789, 2021 m. gruodžio 16 d. TAR, 2021-12-22, Nr. 26680. <https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/72c77a52626411ecb2fe9975f8a9e52e>
3. Lietuvos Respublikos oficialiosios statistikos ir valstybės duomenų valdysenos įstatymas. <https://www.e-tar.lt/portal/lt/legalAct/f79dd2d00c1711edb4cae1b158f98ea5>
4. IQVIA. Analytics Research Accelerator. IQVIA Real World Evidence Solutions. <https://www.iqvia.com/solutions/real-world-evidence/health-data-apps-and-ai/analytics-research-accelerator>
5. Valstybės duomenų agentūra. Valstybės duomenų surašymas. <https://experience.arcgis.com/experience/c45584925bb94df2a5948253c6685c1e/page/Puslapis?views=VALSTYB%C4%96S-DUOMEN%C5%B2-SURA%C5%A0YMAS>
6. Valstybės duomenų agentūra. Oficialiosios statistikos portalas – e. paslaugos. <https://osp.stat.gov.lt/e.-paslaugos>
7. Palantir Technologies. Palantir Foundry. <https://www.palantir.com/platforms/foundry/>
8. Europos Sąjungos investicijų administravimo informacinė sistema. Medicininių klasterių duomenų mainų ir stebėsenos platforma. <https://www.esinvesticijos.lt/sutartys/medicininiu-klasteriu-duomenu-mainu-ir-stebesenos-platforma>
9. Valstybinė ligonių kasa. VLK rodiklių švieslė. <https://rodikliai.vlk.lt/>
10. Valstybės duomenų agentūra. Skiepai nuo ŽPV [interaktyvi švieslė]. Oficialiosios statistikos portalas. <https://osp.stat.gov.lt/skiepai-nuo-zpv>
11. Lietuvos atvirų duomenų portalas. <https://data.gov.lt/>
12. Dapkutė A, Niakšų O, Juodakis J, et al. Lithuania's National Digital Health System Architecture, Evolution, and Lessons From a Centralized Approach Using Longitudinal Document Analysis: Descriptive National Case Study. JMIR Formative Research. In press.
13. Juodakis J, Zubavičiūtė S, Dapkutė A, et al. Data Resource Profile: The Lithuanian health data reuse pathway. Int J Epidemiol. 2025;54(4):dyaf118. DOI:10.1093/ije/dyaf118.
14. openEHR Foundation. openEHR specifications. <https://specifications.openehr.org/>

A HEALTH (AND BEYOND) DATA LAKE IN LITHUANIA

Julius Juodakis

State Data Agency, e-mail: julius.juodakis@stat.gov.lt

Every day, large quantities of data are collected in various state information systems. Historically, the rules governing access and use of this data for decision making, especially in the health area, were unclear and varied between institutions and countries. Lithuania has tackled these challenges, and now features a national data lake with an EU-leading technical infrastructure, >800 data sources, and legal pathways to use these safely for research, innovations and public policy. Hundreds of

projects are using state data each year, from internal analyses to national dashboards on hospital service quality or vaccination coverage, open data releases for the public, or even high-level registry automatization. Nonetheless, effective use of this resource will require good metadata, expanding data depth, and increasing data literacy and analysis competences throughout the health sector.