



KATSAUS

Ilkka Jokipii ja Sami Outinen

Suomalaiset Venäjällä 1917–1964 -tietokanta historian- tutkimuksen apuvälineenä

Kansallisarkiston Suomalaiset Venäjällä 1917–1964 -tutkimushankkeen tehtävänä on muun muassa laatia interaktiivinen tietokanta Venäjällä, Neuvosto-Venäjällä ja Neuvostoliitossa vuosina 1917–1964 oleskelleista Suomen suuriruhtinaskunnan alamaaisista, itsenäisen Suomen kansalaisista ja eräissä tapauksissa heidän jälkeläisistään (esimerkiksi amerikansuomalaiset). Tässä artikkelissa esitellään tutkimushankkeen interaktiivinen tietokanta, tietokannan tekninen toteutus, tietokantatyön vaiheet, tietojen manuaalisesta syöttämisestä joukkoistamiseen ja tekoälyyn ulottuvat tiedonkeruumenetelmät, lähdeaineiston luonne, tietokantatyössä käytettävät historiatieteelliset menetelmät ja kokonaisuuteen liittyvä viestintäsuunnitelma. Tarkastelun kohteena on myös se, miten käytettävät lähteet on valittu, mitkä tekijät rajoittavat tietojen julkaisemista ja mitä mahdollisuuksia digivälitteisen tutkimusympäristön huima tekninen kehitys avaa lähitulevaisuuden tietokantavälitteiselle historianitutkimukselle. Suomalaiset Venäjällä 1917–1964 -tutkimushankkeen interaktiivinen tietokanta tuo internetiin avoimesti käytettäväksi elämäkertatietoja suomalaisista Venäjällä, Neuvosto-Venäjällä ja Neuvostoliitossa 1917–1964. Tiedot esitetään lähdeviitotettuna sähköisenä henkilökortistona ja osittain linkkeinä digitoituihin asiakirjoihin. Kansalaiset pääsevät osallistumaan tutkimusprosessiin taulukoimalla digitalisoituja lähdetietoja hankkeen joukkoistamisalustalla. Tietojen vapaata saatavuutta rajoittavat tietosuojalainsäädäntö ja osaan arkistoaineistoa liittyvät aineistonluovutussopimukset. Kymmeniätuhansia henkilöjä ja joitakin satoja tuhansia henkilötietorivejä sisältävä tietokantahanke vaatii taakseen riittävästi henkilöresursseja ja toimivan teknisen toimintaympäristön. Digitointi- ja joukkoistamistekniikan kehittyminen on mahdollistanut yhtenäisemmän hajautetun tiedonkeruutoiminnan hallinnoimisen ja tutkijoiden keskittymisen henkilötietoja koskevaan tutkimustyöhön. Hankkeessa on myös tehty kokeiluja tekoälyavusteisen arkistoaineistojen sisällöntunnistuksen kanssa, ja tulokset ovat todella lupaavia. Tulevaisuuden tietokantatutkimushankkeissa näillä HTR-teknologiaan perustuvilla menetelmillä voidaan säästää merkittävä osa hankkeiden manuaalisesta tietojen syöttötyöstä.

tietokanta, joukkoistaminen, tekoäly, digitalisaatio, historianitutkimus, Venäjä

Ilkka Jokipii, Kansallisarkisto (ilkka.jokipii@kansallisarkisto.fi); Sami Outinen, Kansallisarkisto (sami.outinen@kansallisarkisto.fi)

Suomalaiset venäjällä 1917–1964 -tutkimushanke

Suomalaiset Venäjällä 1917–1964 -tutkimushanke käynnistyi Kansallisarkistossa 1. syyskuuta 2020 ja se päättyy 31. joulukuuta 2025. Valtioneuvoston kanslian rahoittamassa hankkeessa toteutetaan interaktiivinen tietokanta vuosina 1917–1964 Venäjällä, Neuvosto-Venäjällä ja Neuvostoliitossa oleskelleista suomalaisista. Tietokanta toimii myös useamman hankkeen tutkijan voimin toteutettavan historiatieteellisen tutkimuksen apuvälineenä. Sen on tarkoitus palvella tutkijoiden lisäksi myös sukulaisistaan tietoja etsiviä tai muuten aiheesta kiinnostuneita kansalaisia.

Interaktiiviseen tietokantaan viedään tietoja Suomen suuriruhtinaskunnan alamaista, itsenäisen Suomen kansalaisista ja heidän ensimmäisen polven jälkeläisistään. Tutkimuksen kohteena ovat etenkin vuonna 1917 Venäjällä olleet suomalaiset, punapakolaiset, ”loikkaarit”, amerikansuomalaiset ja vainojen uhrit.

Vuonna 1917 Venäjällä eli arviolta 15 000–25 000 Suomen alamaista. Useimmat heistä liittyivät ensimmäisen maailmansodan aikana käynnistyneeseen paluumuuttajien virtaan. Suomen ja Neuvosto-Venäjän Tarton rauhansopimus käynnisti vielä yhden suuren paluumuuttoaallon: virallista tietä Suomeen palasi noin 10 000 Suomen kansalaista, joiden henkilötiedot on kirjattu Suomen Pietarin-konsulaatin arkistoon.

Neuvosto-Venäjälle siirtyi 10 000–13 000 punapakolaista Suomen sisällissodan lopulla ja sen jälkeisinä vuosina. 1920- ja 1930-lukujen aikana Neuvostoliittoon siirtyi vielä kaksi sellaista suomalaisryhmää, joilla oli – tai joiden vanhemmilla oli ollut – Suomen tai Suomen suuriruhtinaskunnan kansalaisuus. Arviolta 6 000–7 000 amerikansuomalaista muutti Neuvostoliittoon 1920-luvun ja etenkin 1930-luvun alun mittaan. Heidän vanavedessään liikkeelle lähtivät myös niin sanotut ”loikkaarit”, lähinnä salatein Suomesta Neuvostoliittoon siirtyneet Suomen kansalaiset, joiden määräksi on arvioitu 15 000–20 000.

Neuvostoliitossa toteutettiin 1920-luvulta lähtien mutta etenkin 1930-luvulla laajoja väestönsiirtoja. Stalinin ajan poliittinen terrori ja pakkotoimet kohdistuivat voimakkaina myös moniin etnisiin tai muihin vähemmistöihin. Surmaamisten ja pakkotoimien huippu saavutettiin vuosina 1937–1938. Surmatujen tai tuomittujen Suomen kansalaisten lukumäärä asettunee 10 000–15 000 ihmisen välimaastoon. Monet kuolivat väestönsiirtojen aikana tai pakkotyössä nälkiintymisen tai heikon terveydenhuollon vuoksi. Interaktiivisen tietokannan yksi tehtävistä on tarkentaa, kuinka monta suomalaista oleskeli Neuvostoliitossa vuosina 1917–1964.

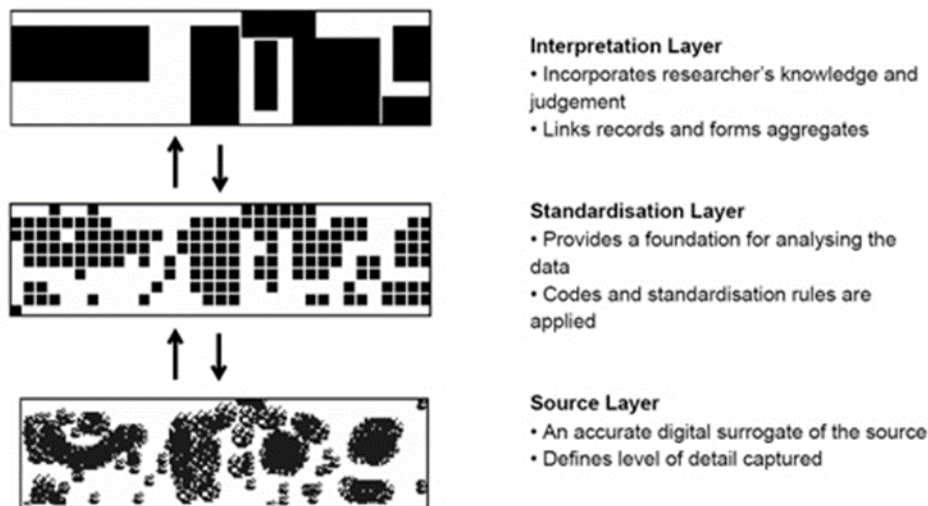
Maantieteellisesti tarkasteltuna tutkimus kattaa kaikki ne alueet, jotka kuuluivat 7.11.1917 perustettuun Neuvosto-Venäjään ja ne alueet, jotka eri vaiheissa liitettiin 30.12.1922 perustettuun Neuvostoliittoon. Tämä merkitsee sitä, että esimerkiksi nykyisen Ukrainan alueella oleskelleiden suomalaisten vaiheiden selvittäminen kuuluu hankkeen tehtäviin. Koronakriisi ja Venäjän hyökkäys Ukrainaan helmikuussa 2022 ovat siirtäneet arkistotyön painopistettä nykyisen Venäjän valtialueen ulkopuolelle. Useita tietokantatyön kannalta tärkeitä suomalaisia ja venäläisiä arkistokokonaisuuksia tai niiden kopioita on käytettävissä erityisesti Kansallisarkistossa.

Tietokannan kolmikerroksinen toteutus

Suomalaiset Venäjällä 1917–1964 -tutkimushankkeen interaktiivisen tietokannan toteuttamiseen liittyy useita sisällöllisiä ja teknisiä ulottuvuuksia. Suomalaiset Venäjällä -tietokannan eräänlaisena teoreettisena lähestymistapana ja pohjana on käytetty University of Londonin Mark Merryn formuloimaa historiallisen tietokannan kolmikerroksista mallia. Mallissa tietokantaa pitää ajatella teknisiä ratkaisujaan laajemmin, ja

mallin kolme eri kerrosta ottavat huomioon historiantutkimuksen lähdekriittiset vaatimukset siirrettäessä merkintöjä vanhoilta papereilta biteiksi tietokantajärjestelmään¹.

Mallin mukaisesti hankkeen tietokantatyö alkaa lähdetietojen tarkasta ja kirjaimellisesta syöttämisestä, seuraavassa kerroksessa syötetyt tiedot saatetaan standardimuotoon (mm. kirjoitusasu) ja tulkinta-kerroksessa esitetään lopullinen tutkimustulos (kooste/yhdistelmä syötetyistä tiedoista). Mallia voi tulkita molempiin suuntiin, ja näin ollen tulkinnasta tulee olla mahdollista palata myös yksittäiseen kirjaimellisesti lähdemateriaalista kopioituun tietokantariviin. Artikkelin on jaettu näitä kolmea kerrosta ja niihin liittyviä työvaiheita mukailleen.



58

Kuva 1. "Kolmikerroksinen" tietokantamalli (Mark Merry).

Lähdeaineistoa löytyy ympäri maailmaa

Koska lähdeaineistoa suomalaisista Venäjällä 1917–1964 löytyy selvästi enemmän kuin viidessä vuodessa tietokantaan on mahdollista syöttää, työ toteutetaan tärkeysjärjestyksessä. Henkilöryhmistä tärkeimmät ovat 1930-luvun jälkimmäisen puoliskon vainojen uhrin, 1930-luvulla Neuvostoliittoon laittomasti menneet ns. loikkarit, punapakolaiset, amerikansuomalaiset ja vuoden 1917 alussa Venäjällä olleet suomalaiset.

Vainojen uhreista kerätään tietoja useista tuomittuja suomalaisia sisältävistä aineistoista. Ensimmäiset tiedot ovat julkaistavissa sen jälkeen, kun meneillään oleva tuomittujen tunnistamistyö viranomaisrekistereistä ja eri lähteistä löytyvien tietojen vertailu on tehty. Vainojen yhteydessä tuomituista (suuruusluokka 10 000) ja teloitetuista (suuruusluokka 5000) hankkeella on hallussa tutkija Mirko Harjulan laatima pohjatiedosto (6672 henkilöä/4472 teloitettua).

Lähteinä on käytetty Eila Lahti-Argutinan matrikkeliä *Olimme joukko vieras vaan. Venäjänsuomalaiset vainonuhrit Neuvostoliitossa 1930-luvun alusta 1950-luvun alkun²*, L. Gildin ja M. M. Braudzen *Kniga pamjati finnov* ("Suomalaisten muistokirja")³ ja eräitä muita yksittäisiä lähteitä, kuten venäläisen kansalaisjärjestön Memorialin keräämiä tietoja. Täydentäviä tietoja saamme etenkin kymmenien vuosien työn yksittäisten suomalaisten kohtaloiden selvittämisessä tehneen Aimo Ruususen tutkimuksista. Tutkimushankkeessa käytetään henkilöiden tunnistamistyössä avuksi myös tutkija Tauno Tukkinen asiantuntemusta. Vuonna 1917 Venäjällä olleista suomalaisista saadaan pohjatiedot vuoden 1917 henkikirjoista. Vertailu Pietarin passiviraston ja mahdollisesti myös kirkonkirjojen tietoihin tehdään ennen tietojen julkaisemista.

Näiden kahden pääryhmän tiedot ovat julkaistavissa ensimmäisiksi. Talkoolaiset ovat syöttäneet taulukkomuotoon henkikirjatiedot Venäjällä olleista suomalaisista vuonna 1917 tutkimushankkeen joukoistamisalustalla. Vuotta 1917 koskevien henkilöjen seulominen passinhaltijain kortistosta talkoolaisten avulla käynnistyi syksyllä 2022.

Vuoden 1918 sodassa hävinneitä punaisia pakeni Neuvosto-Venäjälle etenkin loppukeväästä 1918. Punapakolaisten lukumäärän on arvioitu nousseen kymmeneen tuhanteen. Osa heistä jäi Neuvostoliittoon vuosikymmeniksi. Punapakolaisia koskevat aineistot ovat ensisijaisia suhteessa muihin tässä artikkelissa esiteltyihin kokonaisuuksiin lukuun ottamatta muiden pääryhmien pohja-aineistoja.

Tärkein lähdekokonaisuus punapakolaisten vaiheiden selvittämiseksi on Suomen Kommunistisen Puolueen (SKP) arkistojen se osa, joka on saatu kopioituna Moskovasta Kansallisarkistoon. Muita punapakolaisia koskevia arkistoaineistoja löytyy Pietarin arkistoista (mm. entinen puoluearkisto, TsGAIPD SPb), Venäjän sota-arkistosta (RGVA), Etsivän keskuspoliisin/valtiollisen poliisin arkistosta Kansallisarkistosta, Karjalan tasavallan kansallisarkistosta, valtionrikosylioikeuden arkistosta, turvallisuuspoliisi Säpon arkistosta Ruotsista (Riksarkivet) ja Kustaa Vilkun Aänislinnan arkistosta (Urho Kekkosen arkisto).

Arviolta 5 000–7 000 amerikansuomalaisia saapui Pohjois-Amerikasta Neuvostoliittoon lähinnä rakentamaan Neuvosto-Karjalaa etenkin 1920-luvun ja 1930-luvun vaihteessa. Varpu Lindströmin koamat Missing in Karelia -luettelot ovat pohjatiedosto, samoin Karjalan kansallisarkiston tiedot. Suomen sukututkimusseuran amerikansuomalaisten luettelo, ns. Laiho list, pohjoisamerikkalaiset arkistotiedot ja mahdollisesti Mayme Sevanderin arkistot ovat hyviä täydentäviä henkilötietolähteitä.

Vuosina 1917–1929 rajan ylittäneiden tiedoista (muut kuin punapakolaiset) viedään tietokantaan tärkeimmät. Niitä ovat etenkin rajaviranomaisten ja etsivän keskuspoliisin keräämät arkistoissa säilyneet tiedot. Etenkin vuosia 1918–1920 koskeva aineisto on hajanaista ja hajallaan eri viranomaisten arkistoissa. Ainakin lääninhallitusten passiluettelot ovat täydennettävää aineistoa.

Neuvostoliittoon pakeni 1930-luvun ensimmäisen puoliskon aikana Suomesta jopa 15 000–20 000 suurta lamaa ja oikeistoradikalismia, osin sosialismiin viehtynyttä ns. loikkaria. Pohjatiedostoksi tietokantaan viedään Pietarin pääkonsulinviraston ja ulkoasiainministeriön loikkarikortistojen tiedot. Koska loikkarikortistoista löytyy ainoastaan 4147 henkilöä, osan heistä tiedot löytyvät vasta vainojen joukosta ja Suomeen palanneita tutkittaessa. Neuvostoliitosta palasi Suomeen jo 1930-luvulla vähintään 2600 suomalaista.⁴

EK-Valpon henkilökortistot saadaan tietokantaan hyödyntämällä Transkribus-ohjelmiston tekoälyä (ks. tarkemmin alla). Nämä tiedot täydentävät kaikkiin pääryhmiin kuuluvien henkilötietoja. Tarkoitus on viedä tietokantaan myös hankkeen arkistoon kansalaisilta saadut tiedot ja hankkeen tutkijoiden keräämät henkilöluettelot. Venäjän tilanteen vuoksi korostuvat myös lähialueiden (Ukraina, Georgia, Kazakstan, Baltia) arkistoista saatavat henkilötiedot, joita julkaistaan myös ensimmäisten tietojen joukossa. Tietokannasta rajataan pois yksinomaan sotilastehtävissä itärajan vuosina 1917–1964 ylittäneet suomalaiset kuten heimosoturit ja toisen maailmansodan Suomen armeijan sotilaat.

Haasteena tietokantatyön kannalta on se, miten olennaisia elämänkaaren aikaisia tietoja kymmenistä tuhansista henkilöistä onnistutaan löytämään ympäri maailmaa sijaitsevista arkistoista. Tiedot ovat hajallaan kymmenissä arkistokokonaisuuksissa, sadoissa arkistosarjoissa, tuhansissa dokumenteissa ja satoja tuhansia henkilötietoja sisältävissä rekistereissä. Venäläisissä asiakirjoissa henkilöiden nimet on saatettu kirjoittaa monella eri tavalla (lisäksi on käytetty salanimiä), henkilöluetteloiden tietosisältö saattaa olla hyvinkin niukka ja henkilöiden tunnistaminen suomalaisista viranomaisrekistereistä on näin ollen paikoin vaivalloista. Kun vielä tarkoituksena on löytää tietoja lähes viidenkymmenen vuoden aikana itänaapurissa

oleskelleista suomalaisista, tietokantatyön organisoimiseen on kiinnitettävä erityistä huomiota. Näin ollen hankkeessa on laadittu jatkuvasti päivittyvä tietokannan tietojenhallintasuunnitelma.

Henkilöllisyydeltään epäselväksi jääneiden tietoja etsitään hankkeen loppupuolella kirkonkirjoista ja väestörekisterikeskuksen aineistoista. Myös ulkoasiainministeriön yleisestä henkilökortistosta voi hakea lisätietoja yksittäisistä henkilöistä, samoin vuosien 1939–1958 kelakorteista, Sisäasiainministeriön jatkosodan aikaiset väestönsiirtoasiakirjat -arkistosta, vuoden 1950 ja sitä seuraavien väestönlaskentojen aineistoista, sanomalehtitietokantojen kuolinilmoituksista ja sukututkimussivustoilta.

Tietojen kerääminen on suuri palapeli

Kolmivaiheisen tietokantatyön ensimmäinen vaihe on lähdetietojen kerääminen taulukkomuodossa hankkeen henkilökunnan, hankkeen avustajien ja joukkoistamisprojektin sähköisille lomakkeille keräämistä tiedoista. Teknisenä toimintaympäristönä tietojen keräämiselle toimivat Microsoft Excel ja Microsoft Access -sovellukset.

Näillä tavoin hanke on onnistunut keräämään syksyyn 2022 mennessä yli 100 000 henkilötietoa, joista on kerätty joukkoistamalla lähes 40 000 tietoa. Hankkeen työntekijöiden lisäksi merkittävän panoksen tähän työhön ovat tarjonneet aiemmin mainittujen lisäksi monet muutkin avustajat. Osa avustajien töistä ovat erillistutkimuksia ja osa avustajista on syöttänyt yksittäisistä lähteistä henkilötietoja sähköiseen muotoon. Yksi henkilö saattaa esiintyä näissä tiedoissa useampaan kertaan. Kaikista tiedoista kerätään myös lähdetiedot tietokantatyön toisen vaiheen vertailuja ja luokittelua varten. Näin myös kaikki käyttäjät saavat tiedon käytetyistä lähteistä tietokantatyön kolmannen vaiheen eli tietokannan julkistamisen kautta.

Tietoja kerätään ja viedään tietokantaan tietojenhallintasuunnitelmassa hahmotellussa tärkeysjärjestyksessä. Henkilöiden vaiheista koskevista tiedoista keskitytään tietokantatyössä erityisesti seuraavien tietojen etsimiseen ja todistusvoimaisuuden punnitsemiseen: nimi, syntymäkuunta, syntymäaika, kansalaisuus, rajanylitys (aika, tapa ja suunta), viimeisin kirjoillaolokunta Suomessa, olinpaikat Neuvostoliitossa, kuolinaika, kuolinkuunta, kuolintapa ja linkitykset henkilöitä koskeviin digitaalisiin dokumentteihin.

Ensimmäisenä tietokantaan viedään vainojen uhrien tiedot. Seuraavat ryhmät ovat järjestyksessä vuonna 1917 Venäjällä olleet suomalaiset, Venäjän nykyisillä lähialueilla oleskelleet suomalaiset, punapakolaiset, amerikansuomalaiset, rajanylittäjät 1918–1929, loikkaarit, Suomeen palanneet ja muut mahdolliset ryhmät/aiempia tietoja täydentävä aineisto. Tietokanta sisältää henkilöiden elämäkertatietojen lisäksi linkkejä tutkimuksen kohteena olevien vaiheista kertoviin asiakirjoihin ja valokuviiin.

Joukkoistaminen ja käsialantunnistus uusinta tekniikkaa

Merkittävän osan tietojen keräystyöstä tekevät vapaaehtoiset avustajat hankkeen joukkoistamissivustolla⁵. Joukkoistettava aineisto koostuu tietosuojasyistä aikaisintaan vuonna 1921 syntyneistä tai yli 50 vuotta sitten kuolleista. Vuonna 1917 Venäjällä olleiden suomalaisten tiedot välittyvät tietokantaan suurimmalta osin joukkoistamistoimintaan osallistuvien talkoolaisten avulla. Aineistona toimivat tässä tapauksessa aikakauden henkikirjat ja Pietarin pääkonsulinviraston ylläpitämä passinhaltijain kortisto. Henkikirjoista löytyy kattavia tietoja Venäjällä henkirahan eli henkilökohtaisen veron maksaneista suomalaisista. Passinhaltijain kortistosta löytyvät tiedot Venäjällä passinsa leimaamaan velvoitetuista Suomen suuriruhtinaskunnan alamaaisista.

Joukkoistettavana ovat syksyyn 2022 mennessä käyneet myös Neuvosto-Venäjältä Suomeen vuonna 1921 evakuoitujen henkilötietolomakkeet, Pietarissa toimineen suomalaisen Pyhän Marian seurakunnan kuolleiden ja haudattujen luettelot, etsivän keskuspoliisin rajanylittäjäaineistoja, ja eräitä pienempiä arkistokokonaisuuksia. Näin saamme tuotua tietokantaan itärajan takana oleskelleiden henkilötietoja ja tarkempia tietoja rajanylittäjistä.

Hankkeen medianäkyvyys on tuonut sivustolle yli 400 käyttäjää. Noin 40 000 henkilötiedon muuttaminen sähköiseen muotoon on ollut talkoolaisille kova urakka etenkin siitä syystä, että eri käyttäjät ovat syöttäneet tiedot 2–3 kertaa laadunvarmennuksen vuoksi. Joukkoistaminen toimii siten, että talkoolaiset voivat syöttää tietoja sähköiseen muotoon sivustolle ladatuista digitoiduista henkilöasiakirjoista. Syötetyt tiedot on mahdollista muuntaa yleisimpien toimisto- ja taulukko-ohjelmistojen tunnistamaan csv-muotoon.

Merkittävää on ollut myös vuorovaikutus talkoolaisten ja hankkeen tutkijoiden välillä joukkoistamisalustan keskustelualueen välityksellä. Keskustelualueella on vaihdettu vuoden 2022 lokakuun puoliväliin mennessä 1100 viestiä, joista 900 on liittynyt vapaaehtoisten yksittäisistä dokumenteista tekemiin havaintoihin ja 200 yleisluonteisempiin teemoihin. Suurin osa keskustelusta koskee lähdeaineistosta esiin nousseita tulkintakysymyksiä. Keskustelupalsta toimii myös talkoolaisten keskinäisenä vertaistukena eli vapaaehtoiset neuvovat toisiaan lähteiden tulkinnessa ja palvelun käyttämisessä. Hanke pyrkii lisäämään palautteen perusteella joukkoistamissivuston helppokäyttöisyyttä ja ohjeiden selkeyttä. Talkoolaisilta on saatu myös paljon hyödyllisiä tutkimuksellisia huomioita.

Sivustolla tehtävästä toiminnasta on tässä artikkelissa käytetty nimeä ”joukkoistaminen”. Tällöin tarkoitetaan toimintaa, ”jossa ongelma annetaan ratkaistavaksi tai tehtävä suoritettavaksi ennalta määrittelemättömälle joukolle avoimen kutsun avulla”. Samasta asiasta voidaan käyttää myös ilmaisuja joukkouttaminen, yleisön osallistaminen, joukolle ulkoistaminen, talkoistaminen (kun toiminnasta ei makseta rahallista korvausta). Esimerkiksi yritykset ja julkisen sektorin organisaatiot voivat osallistaa ongelman ratkaisemiseen tai tehtävän suorittamiseen alan ammattilaisten lisäksi harrastajia ja muuten alasta kiinnostuneita.⁶

Hankkeen joukkoistamistoiminta nivoutuu kansalaistieteen (citizen science) perinteeseen. Jo 1800-luvulla sanakirja Oxford English Dictionary syntyi maallikoiden avustamana. Joukkoistaminen (crowdsourcing) alettiin tieteellisesti ymmärtää vuonna 2008 verkkovälitteisenä ongelmanratkaisu- ja tuotantomallina. Luonnontieteistä ja eri alojen digitointihankkeista (kuten Suomen Sukuhistoriallisen Yhdistyksen digiarkisto) joukkoistaminen on edennyt 2010-luvun aikana osaksi historiatieteellisiä tutkimushankkeita.⁷

Hankkeemme toteuttamassa laajuudessa ei joukkoistamista ole tietävästi harjoitettu Suomessa aikaisemmin historiatieteissä. Toiminta on ollut teettämämme kyselytutkimuksen mukaan vapaaehtoisille mielekkästä. Käyttäjät mainitsivat toimintansa suurimmiksi motiiveiksi auttamishalun, historiaan kohdistuvan kiinnostuksen, sukulaisten kohtalojen selvittämisen ja inkeriläistaustan. Vaikeimmiksi asioiksi osoittautuivat käsiala, lomakkeissa olevat epäselvät merkinnät ja lomakkeissa kerrottujen paikkakuntien selvittäminen.

Suurin osa vastaajista piti ohjeita riittävinä, mutta osa vastaajista toivoi tarkempia ohjeita yksittäisistä asioista. Vastaajilla ei ole ollut suurempia teknisiä ongelmia tai ongelmiin löytyi ratkaisu. Yksittäisissä vastauksissa pidettiin käyttämämme zooniverse-alustaa hankalana, koska lomaketta joutuu pienentämään/suurentamaan tai näytössä piti liikkua välillä moneen suuntaan. Puolet vastaajista osallistui keskusteluun zooniversen keskustelupalstalla. Yksittäiset vastaajat toivoivat raportointia tutkimuksen etenemisestä tai nopeampaa vastaamista kysymyksiinsä.

Käyttäjien mukaan tutkimus tulee tällä tavalla lähelle tavallista kansalaista ja [voi] olla osa yhteistä kokemusta. Kiinnostus osallistua joukkoistamiseen tulevaisuudessa oli myös suuri. Yhteensä 43 vastaajasta 67 % on erittäin kiinnostunut osallistumaan toimintaan tulevaisuudessa, 30 % melko kiinnostunut ja 2 % hieman kiinnostunut. Useammille joukkoistaminen toimi korona-ajan henkireikänä, mahdollisti työnteon jatkamisen eläkkeellä, tutustutti historiaan, tuki muihinkin joukkoistamishankkeisiin osallistumista ja tarjosi mielekästä tekemistä.

Joukkoistaminen on yhteydessä myös tutkimustoiminnan pelillistämiseen⁸. Tätä ulottuvuutta kannattaa kehittää mahdollisissa tulevissa vastaavissa hankkeissa. Kirjapalkintoja käytimme hankkeen alkuvaiheessa yhtenä keinona houkuttaa sivustolle käyttäjiä. Palkinnoissa ei tule edetä kuitenkaan sellaisiin tapoihin, joilla käyttäjät korostaisivat nopeutta tarkkuuden sijaan.

Lähdeaineistojen läpikäymisessä on hyödynnetty myös uusinta tekoälyteknologiaa. Etsivän keskuspoliisin henkilökortistoa⁹ varten on luotu oma käsialamalli Transkribus-sovellusta¹⁰ käyttäen ja mallin avulla kortistosta on mahdollista luoda haut mahdollistava data-aineisto. Tekoälyn tarkkuus on varsin hyvä myös merkinnöiltään vaihtelevassa kortistossa. Käsialamalli tunnistaa tekstin jopa lähes 95 prosenttisesti oikein, näin toimien yli 200 000 henkilön kortistosta voi helposti poimia kaikkia Neuvostoliitossa tutkimuksen aikarajauksen puitteissa oleskelleet henkilöt. Käsialantunnistusteknologia säästää huomattavasti aikaa aineiston läpikäymisestä.

Tutkijakäyttöliittymä tietojen standardoinnin apuvälineenä

Tutkimushankkeen tutkijakäyttöliittymässä vertaillaan, standardoidaan ja luokitellaan suomalaisista Venäjällä kerättyjä tietoja. Tutkijakäyttöliittymään ladataan henkilöistä kerättyjä taulukkomuotoisia tiedostoja. Tiedostot sisältävät yksittäisistä tutkimuksen kohteena olevista henkilöistä yhdessä lähteessä (yleensä arkistosarja) tai osatutkimuksessa kerrotut tiedot.

Tietoja kerätään 36 asiasta, jotka on ryhmitelty tietokannassa 140 kenttään. Kysymyksessä ovat henkilö-, asema-, sijainti-, sotilas-, tiedustelija-, tuomio-, vankeus- ja kuolintiedot suomalaisista Venäjällä 1917–1964. Varsinaisten tietokenttien lisäksi tietokanta sisältää kentät lähdeviitteille, alaviitteelle, viittauksiin kortistossa esiintyviin sukulaisiin ja linkit digitoituihin dokumentteihin.

Henkilöiden vaiheista koskevista tiedoista keskitytään tietokantatyössä erityisesti seuraavien tietojen etsimiseen ja todistusvoimaisuuden punnitsemiseen: nimi, syntymäkunta, syntymäaika, kansalaisuus, rajanylitys (aika, tapa ja suunta), viimeisin kirjoillaolokunta Suomessa, olinpaikat Neuvostoliitossa, kuolinaika, kuolinkunta ja kuolintapa. Muista tiedoista vähintään Venäjällä oleskelleiden suomalaisten ammatit, kansallisuudet, kuolinpaikat, järjestöjen jäsenyydet (mukaan lukien sotilasjärjestöt ja tiedusteluorganisaatiot), vankeuspaikat ja tuomiot pyritään mahdollisuuksien mukaan selvittämään.

Tutkijanäkymään on valmius syöttää satojen taulukkojen sisältämiä tietoja vähintään 70 000 henkilöstä siten, että yhden henkilön tiedot saattavat löytyä useasta, jopa kymmenestä lähteestä. Yhdessä ladattavassa taulukossa voi olla tuhansia rivejä. Tiedot on kerätty kymmenistä arkistosarjoista. Tämä merkitsee sitä, että tietokantaan ladataan tutkimushankkeen aikana kymmeniä taulukoita, joissa on satoja tuhansia rivejä henkilötietoja.

Koska tietoja syötetään tutkijanäkymään taulukko kerrallaan, ja yksi taulukko sisältää tietoja yhdestä tai korkeintaan muutamasta lähteestä, tutkijakäyttöliittymässä useasta lähteestä löytyneiden henkilöiden tietoja voidaan yhdistellä. Vain yhdestä syöttölomakkeesta löytyneet henkilöt ja heidän tietonsa siirretään sellaisenaan sähköiseen henkilökortistoon. Sen sijaan silloin kun sama henkilö löytyy kahdesta tai useammasta syöttölomakkeesta, henkilötiedot yhdistetään toisiinsa ja sijoitetaan samaan sähköiseen henkilökorttiin.

Tutkijakäyttöliittymä ehdottaa tietojen yhdistämistä, jos perustiedot (sukunimi, etunimet, syntymäkunta, syntymäpäivä, syntymävuosi, kirjoillaolokunta, kuolinpäivä, kuolinvuosi, kuolinkunta ja kuolintapa) ovat samoja tai muistuttavat riittävästi toisiaan. Tiedot näkyvät tutkijakäyttöliittymässä allekkain. Perustiedot kahden tai useamman lähteen sisältämistä henkilötiedoista ovat samaan aikaan näkyvissä. Samaan henkilöön viittaavien eri lähteistä syötettyjen tietojen yhdistäminen tehdään yhdellä komennolla.

Silloin kun yhdistettävien henkilöiden yksittäiset tiedot ovat täsmälleen samoja (esimerkiksi sama kuolinpäivä), yhdistämiskomento liittää kaikki tietoon liittyvät lähdetunnukset samaan tietoon. Muussa tapauksessa tieto ja siihen liittyvä lähdetunnus merkitään yhdistettävien henkilöiden henkilökorttiin vaihtoehdoiksi tiedoiksi. Tutkijakäyttöliittymän tutkimuksellista luonnetta vahvistaa lähdeviitteiden lisäksi alaviitekenttä, johon voidaan kirjoittaa huomioita tietokannassa esiintyvistä henkilöistä.

Tutkijakäyttöliittymän Henkilöiden tietojen muokkaus -kohdassa henkilötiedot muokataan julkaistaviksi henkilö kerrallaan. Tässä vaiheessa tietyt tiedot merkitään kenttäkohtaisesti todistusvoimaisimmiksi. Tietoja on mahdollista muokata myös siten, että useamman lähteen kertoman tiedon perusteella voidaan päätellä esimerkiksi henkilön tarkka sijaintipaikkatieto, kirjoitusasuun tehdä muutoksia, väärin kirjattuja tietoja poistaa, kirjoitusvirheitä korjata ja tietorakennetta yhtenäistää. Myös epäselviksi jääneiden henkilöllisyyksien selvittely tehdään osittain tässä vaiheessa.

Jokaisessa kentässä esiintyvän tiedon alkuperä todennetaan lähdeviitteen avulla. Tämä mahdollistaa tietojen todistusvoimaisuuden puntaroinnin eli historiatieteellisen lähdekriittisyyden vaalimisen. Tietokantatyön tutkijakäyttöliittymässä kerätyt tiedot lajitellaan todistusvoimaisimpiin tietoihin ja epätarkempiin tietoihin. Tulkinta tehdään avoimesti eli tietokannan loppukäyttäjät näkevät sekä tutkimushankkeen tulkinnan kaikkien henkilöiden vaiheista että kaikki nekin tiedot, jotka hankkeen tutkijat ovat arvioineet vähemmän luotettaviksi tai epätarkemmiksi. Tästä syystä Suomalaiset Venäjällä 1917–1964 -tutkimushankkeen interaktiivinen tietokanta on luonteeltaan sekä tutkimustulos että lähdetietopankki.

Tietosuoja rajoittaa tietojen julkaisemista

Tutkimushanke julkaisee interaktiivisessa tietokannassa olevia tietoja tietosuojasäädösten sallimissa rajoissa. Peruslähdekohtana on tällöin se, että tietoja saa julkistaa yli 50 vuotta sitten kuolleista tai yli sata vuotta sitten laadituista asiakirjoista löytyneistä. Tutkimushankkeen jälkeen tietokannan julkaisematta jäävät tiedot on mahdollista julkaista Kansallisarkistossa siinä tahdissa kuin edellä mainitut aikarajoitukset ajan myötä poistuvat.

Tätä kirjoitettaessa interaktiivisen tietokannan käyttöliittymäsuunnitelma on valmistunut. Käyttöliittymän teknisestä toteutuksesta vastaa Kansallisarkiston sovelluskehitystiimi. Näin varmistetaan, että kehitettävä verkkopalvelu on yhteensopiva Kansallisarkiston arkkitehtuuriratkaisujen ja teknologiavaatiusten kanssa. Tämä takaa syntyvälle palvelulle pidemmän elinkaaren, paremman toimintavarmuuden ja jatkuvan ylläpidon.

Internetin välityksellä käytettäväksi tulevassa julkisessa käyttöliittymässä näkyvät kaikki tietosuojan sallimat tutkimushankkeen keräämät tiedot suomalaisista Venäjällä, Neuvosto-Venäjällä ja Neuvostoliitossa vuosina 1917–1964. Käyttöliittymä sisältää teksti- ja karttapohjaisia hakutoimintoja, joiden avulla voi tehdä hakuja ja hakuyhdistelmiä interaktiivisen tietokannan tietosisällöstä. Sivustolta löytyy myös taustatietoja hankkeesta ja interaktiivisesta tietokannasta.

Tekstihakujen tuloksena syntyy luettelo haun kohteena olevista henkilöistä (tiedot: sukunimi, etunimet, syntymäaika, viimeisin kirjoillaolokunta Suomessa, kuolinaika, kuolinkunta ja kuolintapa). Luette-

losta löytyy linkki jokaisen hakuehdot täyttävän henkilön sähköiseen henkilökorttiin. Henkilökortista löytyvät kaikki henkilöä koskevat tiedot ja linkkejä henkilön vaiheista kertoviin dokumentteihin. Karttahaute kohdistuvat yksittäisiin henkilöihin tai kartalla näkyviin henkilöjen oleskelupaikkoihin.

Medianäkyvyys lisää kansalaisten kiinnostusta

Talkoolaisten merkittävä osuus Suomalaiset Venäjällä 1917–1964 -tutkimushankkeen tiedonkeruutyössä varmistui sillä, että tieto joukkoistamisalusta levisi helmikuussa 2021 STT:n välityksellä¹¹ suuren yleisön keskuuteen. *Helsingin Sanomien* jatkojuttu¹² aiheesta varmisti joukkoistamiselle lentävän lähdön.

Tämän jälkeen talkoolaisia on värvätty etenkin sosiaalisen median välityksellä. Näin on tapahtunut tavallisesti uusien joukkoistamisaineistojen lisäämisen yhteydessä. Tietoa on levitetty Kansallisarkiston sosiaalisen median kanavissa¹³ ja yhteistyökumppanien välityksellä¹⁴.

Näkyvyys mediassa auttaa myös siinä, että kansalaiset osaavat ottaa yhteyttä. Kysymykset ja kerkotukset suomalaisista Venäjällä, Neuvosto-Venäjällä ja Neuvostoliitossa vuosina 1917–1964 tuottavat tutkijoille ja tietokantaan uusia tietoja ja selvityspyynnöjä henkilöistä, jotka tietokantaan kuuluvat. Tutkimushankkeen kahden ensimmäisen toimintavuoden aikana yhteenottoja kansalaisilta on tullut arviolta kaksi tuhatta. Ensimmäiset tietokantatiedot julkaistaan hankkeen toiminta-aikana, jotta kansalaiset pääsevät tuoreeltaan täydentämään ja kommentoimaan julkaistuja tietoja.

Yhteenveto

Suomalaiset Venäjällä 1917–1964 -tutkimushankkeen interaktiivinen tietokanta tuo internetiin avoimesti käytettäväksi elämäkertatietoja suomalaisista Venäjällä, Neuvosto-Venäjällä ja Neuvostoliitossa 1917–1964. Tiedot esitetään lähdeviitotettuna sähköisenä henkilökortistona ja osittain linkkeinä digitoituihin asiakirjoihin. Kansalaiset pääsevät osallistumaan myös itse tutkimusprosessiin taulukoimalla digitalisoituja lähdetietoja hankkeen joukkoistamisalustalla.

Kaikki tutkimushankkeen käyttämät lähteet ovat käytettävissä Kansallisarkistossa, ja hanke vastaa kansalaisten tiedusteluihin jo tutkimushankkeen aikana. Ainoa tietojen vapaata saatavuutta rajoittava tekijä on tietosuojalainsäädäntö. Vuonna 2018 käyttöön otetun eurooppalaisen tietosuojasetuksen¹⁵ jälkeen organisaatioiden lainopillinen tulkinta ohjaa tietosuojatulkintoja siinä missä aiemmin lupaa julkaista arkaluonteisia henkilötietoja kysyttiin tietosuojalautakunnalta¹⁶. Osaa arkistoaineistosta voi käyttää vain Kansallisarkiston toimipisteissä aineistonluovutus sopimusten määräysten seurauksena.

Kymmeniätuhansia henkilöjä ja joitakin satoja tuhansia henkilötietorivejä sisältävä tietokantahanke vaatii taakseen riittävästi henkilöresursseja ja toimivan teknisen toimintaympäristön. Aikaisempiin vastaaviin valtiollisiin tutkimushankkeisiin (Suomen sotaturmat 1914–1922/Suomi, sotavangit ja ihmisluovutukset 1939–1955) verrattuna tiedonkeruun painopistettä on siirretty palkatuille ja vapaaehtoisille avustajille.

Aiempien hankkeiden tapaan mediajulkisuus on auttanut siihen, että vuorovaikutus kansalaisten kanssa suomalaisten vaiheiden selvittämisestä itärajan takana on sujuvaa. Tällä kertaa myös tiedotuskampanjat sosiaalisessa mediassa ovat ylläpitäneet talkootoimintaa hankkeen joukkoistamisalustalla. Tämä johtuu laajemman ja hajanaisemman lähdeaineiston lisäksi siitä, että digitointi- ja joukkoistamistekniikan kehittyminen on mahdollistanut yhtenäisemmän hajautetun tiedonkeruutoiminnan hallinnoimisen.

Näin myös hankkeen tutkijoille jää enemmän aikaa tutkimuksellisten johtopäätösten tekemiseen, joskin käsiteltävän aineiston laajuus asettaa toiminnalle omat rajoituksensa. Samaa henkilöä koskevien

henkilötietojen yhdistäminen onnistuu nykyisessä hankkeessa kuitenkin tietoteknisten ratkaisujen ansiosta sujuvammin kuin aiemmissa hankkeissa.

Tekoäly on ottanut viime vuosina suuria harppauksia, ja arkistoaineistojen sisällön tunnistamiseen liittyvät teknologiat ovat jo tavallistenkin tutkijoiden saatavilla. Tällä hetkellä ollaan tasolla, missä tekstin tunnistus onnistuu riittävän hyvällä tarkkuudella vähintään aineiston haettavaksi tekemiseksi, mutta usein voidaan tuottaa myös lähes täydellistä vaikkapa koneellisille tutkimusmenetelmille sopivaa tekstiaineistoa digitaalisista arkistoaineistokuvista. Seuraava suurempi harppaus teknologiassa on taulukko- ja lomakemuotoisten aineistojen tunnistuksen kehitys. Teknologia lienee kypsää lähivuosina, ja sen jälkeen jää hankkeemme kaltaisilta tutkimuksilta kenties aikaa vievin manuaalista aineistojen syöttämistä sisältävä vaihe parhaimmillaan kokonaan väliin. Tämä tuonee historiantutkimuksen huomattavasti lähemmäs muita yhteiskuntatieteitä monilta menetelmiltään, mutta samalla vaikka tutkimusdatan saatavuus helpottuu, tulee sen laatua tarkastella jatkossakin lähdekriittisesti. Tähän soveltuu esimerkiksi esittelemämme kolmikeroksinen tietokantamalli.

- ¹ Mark Merry, *Designing Databases for Historical Research* (University of London), <https://port.sas.ac.uk/mod/book/tool/print/index.php?id=75#ch134>.
- ² Eila Lahti-Argutina, *Olimme joukko vieras vaan. Venäjänsuomalaiset vainonuhrit Neuvostoliitossa 1930-luvun alusta 1950-luvun alkuun* (Vammala: Siirtolaisuusinstituutti 2001).
- ³ L. Gildi & M. M. Braudze, *Kniga pamjati finnov* ("Suomalaisten muistokirja", Gjolj, 2010).
- ⁴ Pekka Nevalainen, *Punaisen myrskyn suomalaiset* (Vammala: SKS 2002), 281–287; Kostianen, Auvo: *Loikkaarit. Suuren lamakauden laitton siirtolaisuus Neuvostoliittoon* (Helsinki: Otava 1988), 205–209.
- ⁵ <https://www.zooniverse.org/projects/ijokipii/suomalaiset-venajalla-1917-1964>.
- ⁶ Sanastokeskus, TEPA-termipankki, <https://termipankki.fi/tepa/fi/>.
- ⁷ Tiina Onikki-Rantajääskö & Kaarina Pitkänen-Heikkilä, "Joukkoistamisen haasteita ja mahdollisuuksia Tieteen termipankin wikitalloissa," *Tieteessä Tapabtuu* 33, 6 (2015): 14–19; Daren Brabham, "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases," *Convergence: The International Journal of Research into New Media Technologies*, 14, 1 (2008): 75–90; Zooniverse: 10 years of people-powered research. News 12.12.2019. University of Oxford: <https://www.ox.ac.uk/news/2019-12-12-zooniverse-10-years-people-powered-research>.
- ⁸ Vrt. myyräpelinä toteutetut Kansalliskirjaston digitalkoot 2011–2012, ks. Onikki-Rantajääskö & Pitkänen-Heikkilä, "Joukkoistamisen haasteita," 14.
- ⁹ http://wiki.narc.fi/portti/index.php/Valtiollinen_poliisi#Henkil.C3.B6kortisto.
- ¹⁰ <https://readcoop.eu/transkribus>.
- ¹¹ Esimerkki: <https://www.uusimaa.fi/uutissuomalainen/3595639>.
- ¹² <https://www.hs.fi/kotimaa/art-2000007838380.html>.
- ¹³ Esimerkiksi <https://www.facebook.com/Kansallisarkisto>; <https://twitter.com/Kansallisarkist/status/1564241650158903296>), erikseen muun muassa pääjohtaja Jussi Nuortevan toimesta (<https://twitter.com/JussiNuorteva/status/1450497152904540165>).
- ¹⁴ Suomi-Venäjä-seuran uusin päivitys Pietarin passiviraston passinhaltijain kortiston joukkoistamisesta 20.9.2022: <https://www.facebook.com/SuomiVenajaSeura/>.
- ¹⁵ <https://tietosuoja.fi/gdpr>.
- ¹⁶ <https://www.finlex.fi/fi/viranomaiset/ftie/>.