

Datalukutaito ja vastuullinen päätöksenteko (DATALIT)

Tilannekuvaraportti 2021



Tilannekuvaraportti

DataLit – Datalukutaito ja vastuullinen päätöksenteko

1. Johdanto

DataLit on yhteiskunta-, oikeus- ja tietojenkäsittelytieteitä yhdistävä hanke, jonka tutkimusryhmät ovat Helsingin yliopistosta, Aalto yliopistosta ja Itä-Suomen yliopistosta. Hankkeen tarkoituksena on kehittää ymmärrettäviä ja luotettavia käytäntöjä ja työkaluja hyödyntää suomalaista sosiaali-, terveys- ja hyvinvointidataa sekä edistää datalukutaitoa. Tavoitteena on myös luoda uusi monitieteinen ja tieteidenvälinen datalukutaitoon keskittyvä tutkimusalue.

Datalukutaito muodostuu datan ja siihen liittyvien keskeisten sosiaalisten, oikeudellisten ja teknisten kysymysten ymmärryksestä, ja se on perusta eettiselle ja näyttöön perustuvalla päätöksenteolla. Datan hallinta ja käyttö päätöksenteossa ovat keskeisiä haasteita nyky-yhteiskunnassa. Data-analytiikan kuten koneoppimisen ja tekoälyn kehitys on luonut uusia mahdollisuuksia käyttää dataa. Mutta näyttöön ja tietoon perustuva päätöksenteko edellyttää aikaisempaa parempaa ymmärrystä sekä datan luonteesta että sen käyttöä ja säilyttämistä koskevista teknisistä, yhteiskunnallisista ja oikeudellisista kysymyksistä.

Hankkeessa kehitetään yhdessä sidosryhmien kanssa ratkaisuja ja työkaluja tietoon perustuvaan päätöksenteon ja hyvän datalukutaidon tueksi. Näitä ovat esimerkiksi menetelmät tuottaa anonymisoitua synteettistä dataa ja hyväksyttäviä koneoppimismalleja rekisteridatan analysointiin sekä interaktiiviset välineet datan kommunikointiin ja visualisointiin. Samalla hankkeessa tunnistetaan datan hyödyntämisen keskeiset haasteet ja edistetään kansalaisten ja päätöksentekijöiden datalukutaitoa osallistumalla julkisiin keskusteluihin datan käytöstä.

2. Ilmiöpohjaisuus

Datan hallinta ja käyttö päätöksenteossa ovat keskeisiä haasteita nyky-yhteiskunnassa niin yksilöille kuin organisaatioille. Data-analytiikan kuten koneoppimisen ja tekoälyn kehitys on luonut uusia mahdollisuuksia käyttää dataa sekä julkisella että yksityisellä sektorilla. Erityisesti terveyteen ja hyvinvointiin liittyvässä datassa ja analytiikassa on ennakoitu olevan suuria mahdollisuuksia – niiden avulla voisi olla mahdollista kehittää terveys- ja hyvinvointipalveluita, lisätä kustannustehokkuutta ja edistää ihmisten terveyttä. Esimerkiksi Suomen nykyisessä hallitusohjelmassa on monia strategisia tavoitteita, missä datalla ja digitalisaatiolla on suuri rooli. Hallitus haluaa edistää eettisesti, taloudellisesti ja sosiaalisesti kestävästä datapolitiikan ja tekoälypoliittisen sääntelykehikon laatimista ja avata julkisen sektorin dataa hyödynnettäväksi joustavasti ja laajamittaisesti ottaen kuitenkin huomioon korkean tietosuojan. Myös ajatus näyttöön tai tietoon perustuvasta päätöksenteosta nähdään kasvavassa määrin perustuvan erilaisille datavarannoille ja päätöksenteon tukityökalujen käyttöön. Uudet käytännöt ja tavoitteiden saavuttaminen asettavat kuitenkin kasvavia vaatimuksia datan hallinnalle ja sen käytölle päätöksenteossa. Vaatimuksiin vastaaminen edellyttää aikaisempaa parempaa ymmärrystä sekä datan luonteesta että sen käyttöä ja säilyttämistä koskevista teknisistä, yhteiskunnallisista ja oikeudellisista kysymyksistä. Tarvitaan **datalukutaitoa**.

Datalukutaito muodostuu datan ja siihen liittyvien keskeisten sosiaalisten, oikeudellisten ja teknisten kysymysten ymmärryksestä. Sen perustana on ymmärrys siitä, kuinka dataa kerätään, prosessoidaan, analysoidaan, tulkitaan ja esitetään. Nykyaikana on myös entistä tärkeämpää ymmärtää, kuinka dataa hyödyntävät mallit toimivat. Tämän päälle rakentuu tiedollinen ymmärrys siitä, kuinka data voi toimia näyttönä ja kuinka datasta tehtävät päätelmät perustuvat datan ominaisuuksiin. Samalle perustalle rakentuu myös ymmärrys datan käyttöä ohjaavista lainsäädännöllisistä ja eettisistä periaatteista. Datalukutaito vaatii näiden eri ulottuvuuksien kokonaisvaltaista tarkastelua. Esimerkiksi käy datan anonymisointi: motivaatio datan anonymisoinnille on eettinen, henkilötietojen suojaaminen, mutta sillä on merkittäviä seurauksia datan käytettävyydelle ja viime kädessä molempien tavoitteiden menestys riippuu anonymisoinnin teknisen toteutuksen yksityiskohdista. Datalukutaito mahdollistaa datan käyttöön liittyvien mahdollisuuksien ja rajoitusten tunnistamisen. Ilman sitä yhteiskunnallisesti kestävä datan käyttö ei ole mahdollista.

Datalukutaidon eri ulottuvuuksia on luontevaa tarkastella suhteessa datan käytön koko kaareen. Kuten oheinen taulukko 1. osoittaa, kaikkiin vaiheisiin datan keräämisestä sen käyttöön päätöksenteossa liittyy tiedollisia, teknisiä, eettisiä ja oikeudellisia kysymyksiä, joita ei aina ole helppo erottaa toisistaan.

Taulukko 1. **Datalukutaidon kysymyksiä**

datan tuottaminen tai kerääminen	Millä tavoin datan tuottamisen tai keräämisen tapa vaikuttaa sen luotettavuuteen? Millä tavoin varmistetaan relevantin metadatan säilyminen? Mistä asioista dataa ei kerry? Mitä edellytyksiä oikeusjärjestelmämme asettaa datan keräämiselle? Millaista dataa on eettistä kerätä?
datan säilyttäminen ja käsittely	Vaikuttaako datan säilytystapa tai muuttuvat määritelmät datan käytettävyyteen? Millä tavoin voidaan turvata datan säilytys ja henkilötietojen yksityisyys? Mitä lainsäädäntö vaatii datan säilyttämiseltä?
datan analysointi	Millaisia vaatimuksia koneoppimismenetelmät asettavat datalle? Kuinka paljon sisältöasiantuntemusta datan tulkinta edellyttää? Voiko datan perusteella ennustaa tai tehdä kausaalipäätelmiä? Millaisia taustaoletuksia malli edellyttää? Kenellä on oikeus käyttää dataa?
datan esittäminen	Millaisia ovat esitystapojen tuottamat väärinymmärrykset ja vinoumat? Kuinka esittää data ymmärrettävästi ja johtamatta harhaan? Kuinka mahdollistaa käyttäjän vuorovaikutus visuaalisen datan kanssa?
datan käyttö päätöksenteossa	Kuinka vahvaa näyttöä data antaa? Kuinka vahvaa näytön pitäisi olla? Mitkä ovat käytettävissä olevan näytön rajoitukset? Kuinka läpinäkyviä päätöksentekoprosessissa käytettävän datan ja päätelmien tulisi olla?

DataLit-hankkeen lähtökohtana on ajatus, että datalukutaito on edellytys vastuulliselle ja tietoon perustuvalla päätöksenteolla. Se mahdollistaa sekä datan oivaltavan käytön että siitä käytävän kriittisen yhteiskunnallisen keskustelun. Datalukutaidon vaatimus ei rajoitu vain rekistereitä ylläpitäviin viranomaisiin, dataa käyttäviin tutkijoihin tai poliittisiin päättäjiin. Sitä tarvitsevat myös yritysjohtajat, terveydenhuollon ammattilaiset ja kunnalliset päättäjät. Viime kädessä datalukutaito on kaikkia kansalaisia koskeva kansalaistaito.

Hankkeen tavoite on luoda välineitä datalukutaidon kehittämiseen suomalaisessa yhteiskunnassa. Erityisenä sovelluskohteena on suomalainen sosiaali- ja terveysalan rekisteridata. Hanke tarkastelee datan koko käytön kaaren sen keräämisestä sen käyttöön päätöksenteossa. Tavoitteena on luoda kokonaisvaltainen käsitys datalukutaidosta, jonka perusteella voidaan kritisoida esimerkiksi big dataan ja tekoälyyn liittyviä epärealistisia odotuksia. Ensisijaisena tavoitteena on kuitenkin luoda konkreettisia välineitä datan analysointiin, anonymisointiin ja esittämiseen sekä luoda ymmärrystä datan käytön eettisistä, yhteiskunnallisista ja juridisista taustaehtoista.

3. Tutkimuksen state of the art

Datalukutaito on monitieteinen tutkimusalue, joka yhdistää yhteiskuntatieteitä, oikeustiedettä, kognitiotiedettä, tietojenkäsittelytiedettä, datatiedettä ja filosofiaa. Luonteenomaista tutkimusalueelle on sekä osa-alueiden nopea kasvu että tutkimuskentän hajaannus. Tietojenkäsittelytieteen ja datatieteen menetelmäkehitys on nopeaa ja siitä puuttuu usein yhteydet ihmistieteelliseen tutkimukseen. Vastaavasti tekoälyn ja datan käytön yhteiskuntatieteellinen tutkimus on uusi ja nopeasti kasvava tutkimusalue, jolle on tyypillistä lähestymistapojen moninaisuus ja yhdistävän kokonaisjäsennyksen puuttuminen. Sosiologit, antropologit, oikeustieteilijät ja kulttuuritutkijat tutkivat ilmiötä omista (hajanaisista) teoreettisista lähtökohdistaan ja useimmiten vailla läheistä yhteyttä toisiinsa tai tietojenkäsittelytieteen tutkimukseen. Filosofissa (tieteenfilosofia ja etiikka) datan ja tekoälyn hallinnan tutkimus on vasta alkamassa.

DataLit-hankkeen lähtötilanne on lupaava. Se on osa Suomen Akatemia rahoittamaa Suomen tekoälykeskus (FCAI) lippulaivahanketta. Hanke yhdistää kolmea FCAIn tutkimusohjelmaa (Privacy-preserving and secure AI, AI in society sekä Applications of AI in healthcare) ja on ensimmäinen laaja suomalainen hanke, joka pyrkii yhdistämään tietojenkäsittelytieteellisen ja yhteiskuntatieteellisen tutkimuksen. Myös kansainvälisesti katsottuna tällainen yhteistyö on harvinaista. Hankkeeseen osallistuvien yhteiskuntatieteilijöiden tausta on tieteen- ja teknologiantutkimuksessa ja heillä on jo aiempaa kokemusta dataan liittyvästä tutkimuksesta.

Hankkeen tavoitteena on muodostaa oman ja eri tieteenaloilla tapahtuvan tutkimuksen pohjalta kokonaisjäsenitys datalukutaidosta ja sen edellytyksistä. Tavoitteena on luoda monitieteellisen tutkimuksen kokonaisuudesta tieteidenvälinen tutkimuskenttä. Olennainen osa hanketta on myös osallistuminen ajankohtaisiin keskusteluihin, esimerkiksi differentiaalisesta tietosuojasta, koneoppimismenetelmien rajoista ja mahdollisuuksista, dataa ja tekoälyä koskevan eurooppalaisen oikeudellisen säätelyn kehityksestä sekä kriittisen datatutkimuksen esiin nostamista ongelmakohdista. Seuraavassa lyhyesti joistakin hankkeen kannalta keskeisistä tutkimusaiheista.

Yksityisyyden suojan osalta differentiaalinen tietosuojatuki on vakiintunut tutkimuskentässä keskeiseksi menetelmäksi tietosuojan varmistamiseen arkaluonteisten tietojen käsittelyssä. Tähän pohjautuvia menetelmiä on esitetty esimerkiksi datan anonymisointiin synteettistä dataa käyttäen, mutta näiden tarkkuus tiukan muodollisten tietosuojatulkintojen alla on usein heikko. Tällaiseen dataan liittyviä todellisia yksityisyyssriskejä ja sen käytön vaikutuksia muihin analyyseihin tunnetaan huonosti.

Tekoälyllä ja koneoppimismenetelmillä on mahdollista vahvistaa päätöksentekoa terveydenhuollossa. Uusilla menetelmillä on mahdollista esimerkiksi tehostaa diagnostiikkaa automatisoimalla lääketieteellisen kuvantamisdatan analyysi, monitoroida reaaliaikaisesti potilaiden terveydentilaa, sujuvoittaa sairaaloiden hallinnollisia prosesseja tai vaikkapa pyrkiä ymmärtämään erilaisia hoitotapoja ja muokkaamaan niistä parhaita käytäntöjä. Tekoälyn käyttöön lääketieteessä liittyy kuitenkin myös haasteita, kuten mallien opettaminen valikoituneella tai harhaisella aineistolla, suuren yleisön vähäinen luottamus, puutteellinen regulaatio ja rajoittunut ymmärrys siitä, minkälaisen kysymysten ratkaisuun tekoäly soveltuu ja mihin ei.

Suomessa on nyt ensimmäistä vuotta voimassa oleva sosiaali- ja terveystietojen toissijaista käyttöä koskeva lainsäädäntö, jota EU:ssa pidetään esikuvana terveysdatan hyödyntämiselle uudella eurooppalaisella terveysdata-alueella (European Health Data Space). Uuden lainsäädännön vaikutuksista ei ole vielä tehty tieteellistä tutkimusta. Tekoälyn käyttö sote-tietojen hyödyntämisessä nostaa myös esille eettisiä ja yksityisyyden suojaan ja syrjinnän ehkäisyyn liittyviä kysymyksiä, joihin hankkeessa pyritään vastaamaan monitieteistä tutkimusasetelmaa hyväksi käyttäen. EU:n tasolla aihepiiriin suoraan liittyvä lainsäädäntö on myös voimakkaassa muutostilassa, mikä antaa tutkimuksellisesti dynaamisen näkökulman hankkeeseen.

Kriittinen yhteiskunnallinen datatutkimus (critical data studies) korostaa näkemystä, että big dataa ja siihen liittyviä odotuksia tulee tarkastella kriittisesti suhteessa sekä sosiaaliseen ja kulttuuriseen ympäristöön että eri toimijoihin. Tutkimuksen keskeisenä tavoitteena on tunnistaa datan käyttöön liittyviä yhteiskunnallisia ja eettisiä rajoitteita, valtasuhteita ja vaikutuksia ihmisten arkeen. Tutkimussuuntausta on kritisoitu siitä, että kriittistä datatutkimusta on usein toteutettu datahankkeiden ulkopuolella, mikä tässä hankkeessa tullaan välttämään.

DataLit-hanke jakautuu viiteen työpakettiin (WP1-5), jotka ovat tiiviissä vuorovaikutuksessa keskenään. Kukin työpaketti lähestyy datan keräämisen, analysoinnin, käytön ja tulkinnan haasteita yhdistäen tietoteknistä ja yhteiskuntatieteellistä osaamista. Tutkimuskohteeksi on tässä vaiheessa määritelty seitsemän eri tapaustutkimusta, mutta sekä kohteet että niiden määrä saattavat muuttua tutkimuksen edetessä. Tapaustutkimusten kohteeksi valitut hankkeet eivät välttämättä toteudu aiotulla tavalla ja nopeasti liikkuvalla alueella voi nousta esiin uusia tärkeitä tutkimuskohteita. Valitut tapaustutkimukset ovat rekisterinpitäjien ja -viranomaisten sekä terveys- ja hyvinvointipalveluja tarjoavien organisaatioiden konkreettisia meneillään tai kehitteillä olevia hankkeita. Mukana on mm. THL:n hankkeita, joissa käsitellään rekisteridataa ja terveys- ja hyvinvointikeskusten tietojohdantamista sekä Valtiovarainministeriön johtama Aurora AI-hanke, jossa kehitetään kohdennettuja elämäntilanteisiin liittyviä palveluväyliä. Muita yhteistyökumppaneita tapaustutkimuksissa ovat Findata, Kela, Tilastokeskus, Digi- ja väestötietovirasto sekä joitakin sairaanhoitopiirejä.

WP1: Luotettavat päätelmät ja niiden taustaehdot. Työpaketissa kehitetään koneoppimismalleja rekisteridatan analysointiin ja arvioidaan minkälaisia vaatimuksia eri mallinnus- ja päättelypäämäärät kuten kausaalinen mallintaminen, algoritmeihin perustuva päätöksenteko, yksilökohtainen ennustaminen, tai

populaatiotason prosessien mallintaminen asettavat datalle. Tieteen- ja teknologiantutkimus ja filosofia tuovat työpakettiin refleksiivisyyttä kysymällä mm. minkälaista taustatietoa, asiantuntemusta ja datalukutaitoa mallien rakentaminen ja hyödyntäminen edellyttävät? Kuinka datan ja rekisteriaineistojen puutteellisuus, virheet, vinoumat tai esimerkiksi kirjauskäytäntöjen muutokset vaikuttavat malleihin ja niistä tehtäviin päätelmiin?

WP2: Yksityisyyteen liittyvät haasteet ja ratkaisut rekisteritutkimuksessa. Tämän työpaketin aiheena ovat yksityisyyden suojaamiseen liittyvät menetelmät kuten differentiaalinen tietosuoja ja synteettinen data. Työpaketissa kehitetään menetelmiä arvioida alkuperäisen ja synteettisen datan välisiä eroja ja synteettisen datan käytön vaikutuksia. Lisäksi tutkitaan eri yksityisyyttä suojaavien menetelmien ymmärrettävyyttä, hyväksyttävyyttä sekä soveltuvuutta eri käyttö- ja päätöksentekotilanteissa. Työpaketissa kysytään myös minkälaisia vaatimuksia eri sidosryhmät asettavat yksityisyydelle eri asiayhteyksissä.

WP3: Oikeudelliset ja eettiset vaatimukset vastuulliselle datan käytölle. Uudistuvat ja monimutkaiset säännökset tuottavat datan käyttäjille ja hallinnoijille jatkuvia haasteita. Tässä työpaketissa keskitytään erittelemään keskeisiä datan hallintaan ja datalukutaitoon liittyviä käsitteitä kuten yksityisyys, läpinäkyvyys, reiluus ja ymmärrettävyys ja kuinka ne näyttäytyvät eri konteksteissa (esim. lainsäädäntöteksteissä, strategioissa tai tieteellisissä keskusteluissa) ja avaamaan lakien tulkinnessa ilmeneviä solmukohtia. Tässä työpaketissa tietojenkäsittelytieteet tuovat tietoa teknisistä mahdollisuuksista ja niiden yhteensovittamisesta eettisiin ja oikeudelliseen sääntelyyn. Työpaketissa tutkitaan myös julkisen sektorin datavarantojen käytön läpinäkyvyyttä, muun muassa tutkimalla tekoälyn käyttöä sote-tietojen hyödyntämisessä.

WP4: Datan esittämisen ja visualisoinnin vaikutukset datalukutaidolle. Monimuotoinen data ja usein läpinäkymättömät algoritmit tuottavat haasteita datan ja mallien käytölle päätöksenteossa. Työpaketissa kehitetään interaktiivisia välineitä datan kommunikointiin ja visualisointiin, jotka auttavat käyttäjää ymmärtämään ja myös jäsentämään dataa oman päätöksenteon tueksi. Datan käyttötarpeet vaihtelevat, jolloin myös datan esittämisen vaatimukset ovat erilaisia. Työpaketissa tutkitaan eri käyttäjäryhmien ja organisaatioiden tiedon tarpeita yhteiskuntatieteiden, kognitiotieteen ja tietojenkäsittelytieteen näkökulmista.

WP5: Julkinen sektori vastuullisena data hyödyntäjänä ja päätöksentekijänä. Työpaketissa keskitytään datan ja tekoälymallien käyttöön julkisen sektorin sosiaali- ja terveysalan päätöksenteossa. Tutkimuksen kohteena ovat toisaalta nykyiset käytännöt ja kehityshankkeet ja toisaalta niiden institutionaaliset ja yhteiskunnalliset sidokset ja vaikutukset. Yhtenä lähtökohtana on tutkia ja arvioida muiden kuin yksilön profilointiin tähtäävien työkalujen ja mallien hyödyntämismahdollisuuksia päätöksenteossa. Työpaketti tarkastelee myös, kuinka dataa ja tekoälymalleja voidaan käyttää marginaalisten ryhmien aseman parantamiseen, hallinnon läpinäkyvyyden lisäämiseen ja poliittisten keskustelujen asiasisällön vahvistamiseen.

4. Monitieteisyys

Hanke yhdistää uudella tavalla useita tutkimusaloja: tieteen- ja teknologiantutkimus, kriittinen datatutkimus, sosiologia, oikeustiede, tieteenfilosofia, kognitiotiede, yhteiskuntatiede ja tietojenkäsittelytiede. Tavoitteena on luoda uusi monitieteinen ja tieteidenvälinen datalukutaitoon keskittyvä tutkimusalue. DataLit

on kansainvälisestikin katsottuna laaja ja kunnianhimoinen yritys yhdistää tietojenkäsittelytieteellistä ja yhteiskuntatieteellistä tutkimusta.

Mielekäs monitieteinen tutkimus edellyttää pitkäjänteisyyttä ja panostamista. Hankkeen kesto luo tälle hyvän perustan ja rekrytoinneissa on pyritty siihen, että hankkeen tutkijoilla on mahdollisuus pitkäjänteisen yhteistyön toteuttamiseen. Hankkeen työpaketit on muodostettu siten, että niissä on mukana tutkijoita useammalta eri tieteenalalta. Erityisesti on pyritty yhteiskuntatieteiden ja tietojenkäsittelytieteen yhteistyörajapintojen rakentamiseen. Näin on mahdollista kokeilla **refleksiivistä yhteisluomista** (reflexive co-creation), jossa yhteistyötahojen kanssa toteutettavat kehittämishankkeet ovat samanaikaisesti yhteiskuntatieteellisen tutkimuksen kohteita. Tämä mahdollistaa sekä hankkeiden paremman mukauttamisen muuttuviin olosuhteisiin että vahvistaa yhteisluomisessa käytettävää tietopohjaa. Monitieteisyyden vahvistaminen otetaan huomioon myös hankkeen sisäisissä toimintatavoissa: tieteidenvälistä ymmärrystä tuetaan sisäisten työpajojen, tapaamisten ja retriittien avulla. Ensimmäisinä vuorossa ovat työpajat ”Mitä ja miten yhteiskuntatieteilijät tutkivat?” sekä ”Mitä on synteettinen data?”.

Hankkeen tutkimuksellinen päähuomio on sosiaali-, terveys, ja hyvinvointidatassa. Tämä alue kattaa paljon eri toimijoita yhteiskunnan eri sektoreilta: yksityisiä ja julkisia, kansallisia ja kansainvälisiä, eri ministeriöiden hallinnonaloihin kuuluvia, erilaisia yrityksiä, viranomaisia ja tutkimusorganisaatiota. Tutkimustulosten käytettävyys ei kuitenkaan rajoitu tälle alueelle, vaan keskeiset datalukutaitoon liittyvät käsitteet ja työkalut tulevat soveltumaan myös laajemmin, mikä otetaan huomioon hankkeen viestinnässä.

5. Vaikuttavuus

Hankkeen tutkijat ovat hyvin verkostoituneita ja heitä kuuluu moniin kotimaisiin ja kansainvälisiin työryhmiin ja verkostoihin. Tutkijat toimivat asiantuntijoina esimerkiksi hankkeeseen liittyvissä lainsäädäntöhankkeissa ja osallistuvat aktiivisesti julkiseen keskusteluun. DataLit-hankkeen osaamista hyödynnetään esimerkiksi sosiaaliturvan uudistamisen työryhmässä, AuroraAI:n etiikkatyöryhmän toiminnassa sekä laajassa Horizon2020 URBANAGE hankkeessa. Vaikuttavuus perustuu jo olemassa oleville yhteistyöverkostoille ja asiantuntijuudelle, mutta hankkeen myötä vaikuttavuutta rakennetaan järjestelmällisesti sekä ulotetaan uusille sidosryhmille ja laajemmalle yleisölle.

Tutkimuksen keskeinen tavoite on tuottaa yhdessä sidosryhmien kanssa ratkaisuja ja työkaluja tietoon perustuvan päätöksenteon ja hyvän datalukutaidon tueksi. Tämä tavoite on mahdollista saavuttaa vain tiiviissä vuorovaikutuksessa sidosryhmien ja hyödyntäjätahojen kanssa. Vuorovaikutuksen keinot voidaan jakaa viiteen ryhmään, jotka käyttävät eri kanavia ja joiden kohteena on erilaiset sidosryhmät. Näiden lisäksi hanke viestii verkkosivujen, sosiaalisen median, STN:n, yliopistojen ja yhteistyökumppaneiden viestintäkanavien kautta sekä edistää tieteellistä vaikuttavuutta osallistumalla aktiivisesti kotimaisissa ja kansainvälisissä tiedeyhteisöissä.

- **Työpajat** ovat eri sidosryhmille tarkoitettuja kohtaamistilanteita, joissa tietoa välitetään kaksisuuntaisesti hankkeen tutkijoiden ja sidosryhmien välillä. Ensimmäinen työpaja on suunnattu koko hankkeen sidosryhmille, jotka sitoutuivat hankkeeseen jo hakuvaiheessa. Työpajan tarkoituksena on toisaalta kertoa hankkeen lähtökohdista ja tavoitteista ja toisaalta selvittää millaista tietoa sidosryhmät tarvitsevat. Työpajoja järjestetään säännöllisesti siten, että osa työpajoista keskittyy esimerkiksi yksittäiseen tapaustutkimukseen tai työpakettiin.

- **Roadshows eli kiertueet** ovat jonkin teeman ympärille rakennettuja tilaisuuksia, joissa hankkeen tuloksia esitellään eri käyttäjäryhmille. Alustaviksi teemoiksi on suunniteltu esimerkiksi synteettisen datan käyttömahdollisuuksia ja datan visualisointia. Kiertue tarkoittaa, että saman teeman tilaisuuksia pidetään useissa eri organisaatioissa ja eri käyttäjäryhmille (esim. sairaanhoitopiirit ja terveyspalveluiden tarjoajat). Kiertueita voidaan toteuttaa sekä virtuaalisesti että vierailemalla eri organisaatioissa.
- **Politiikkaraportit** ovat keskeinen vaikuttamisen keino. Niitä tuotetaan myös yhdessä muiden STN-hankkeiden ja muiden toimijoiden kanssa (esim. FCAI).
- **Datamyytien murtaajat -kampanja** on suunnattu sekä päättäjille että laajemmalle yleisölle. Kampanjan tarkoituksena on kumota yleisiä väärinkäsityksiä ja sekaannuksia datasta ja sen käyttömahdollisuuksista. Esimerkki kampanjan aiheesta on big dataan liittyvät myytit. Kampanjaa viedään mm. sosiaaliseen mediaan, Helsingin yliopiston Tiedekulmaan ja siitä tuotetaan podcast- tai radio-ohjelmasarja.
- Hanke pyrkii myös tuottamaan erilaisia **koulutusmateriaaleja** yhteistyökumppaneiden kanssa. Suunnitelmissa on esimerkiksi monitieteiden yliopistokurssi datasta yhteiskunnassa, oppimateriaaliksi soveltuva Handbook of Data Literacy -kirja, sekä mahdollinen verkkokurssi datalukutaidosta.

Hankkeen vaikuttavuuden tavoitteet jakautuvat **neljään käytännön tavoitteeseen**. DataLit-hankkeen tavoitteena on:

1. Herättää tietoisuutta datalukutaidon tarpeesta.
2. Edistää datalukutaitoa vastuullisen päätöksenteon tueksi.
3. Kehittää ratkaisuja ja työkaluja datan käyttöön, esittämiseen ja hallintaan, jotka pohjautuvat tutkimukseen, yhteiskehittämiseen ja ajatukseen datalukutaidon edistämisestä.
4. Kouluttaa tulevaisuuden asiantuntijoita joilla on vahva ja monitieteellinen ymmärrys datasta ja datalukutaidosta sekä edistää monitieteellistä keskustelua.