



MIDRAS: Viranomaisten luvanvaraiset aineistot tehokkaasti ja turvallisesti tutkimuskäyttöön

CSC – TIETEEN TIETOTEKNIIKAN KESKUS OY

Keilaranta 14 • PL 405 • 02101 Espoo
Puh. (09) 457 2001 • Fax (09) 457 2302 • Y-tunnus 0920632-0 • www.csc.fi

CSC – IT CENTER FOR SCIENCE LTD.

Keilaranta 14 • P.O. BOX 405 • FI-02101 Espoo • Finland
Tel. +358(0)9 457 2001 • Fax +358(0)9 457 2302 • VAT number FI09206320 • www.csc.fi

© CSC – Tieteen tietotekniikan keskus Oy ja Rekisteritutkimuksen tukikeskus (ReTki)

15.4.2011

Esipuhe

Henkilötietojen ja muiden yksikötason tietojen keräämisellä on pitkät perinteet Suomessa. Näistä tiedoista kootut luvanvaraiset rekisterit muodostavat merkittävän aineistopohjan myös tutkimukselle. Rekisteriaineistojen tehokkaaseen ja monipuoliseen tutkimuskäyttöön liittyy tällä hetkellä kuitenkin monia haasteita ja ongelmia. Tämä raportti tuo näihin haasteisiin uudenlaisia ratkaisuja mm. ehdotetun etäkäyttöjärjestelmän avulla. Selvityshankkeen luoman vision mukaan: *Kansallisen etäkäyttöjärjestelmän kautta viranomaisten hallussa olevat, luvanvaraiset, yksikötasoiset aineistot ovat käytettävissä kattavasti, turvallisesti, kustannustehokkaasti ja helposti tutkimukseen.*

Raportissa esitetään tavoitteet toteuttava rekisteriaineistojen etäkäyttöjärjestelmän toimintamalli ja sen toteuttaminen vaiheittain. Raportti on suunnattu ensisijaisesti tietohallinnon organisoinnista päättävälle ja toiminnan rahoittajille, mutta myös rekisteriaineistoja tuottaville ja hallinnoiville organisaatioille sekä tutkijoille.

Raportti perustuu CSC – Tieteen tietotekniikan keskuksen ja Rekisteritutkimuksen tukikeskuksen (ReTkin) yhteistyönä toteuttamaan selvityshankkeeseen, jolle opetus- ja kulttuuriministeriö myönsi avustuksen 22.4.2009. Selvitystyön toteutuksesta on vastannut projektiryhmä, joka on muodostettu CSC:n ja ReTkin yhteistyönä. Ryhmässä ovat toimineet: Marianne Johnson, Panu Kalliokoski (projektipäällikkö 1.9.2010 alkaen), Katariina Peltonen, Terhi Pennanen ja Jussi Tella (projektipäällikkö 31.8.2010 asti). Työn etenemisen on varmistanut hankkeelle perustettu ohjausryhmä, jossa on hanketahojen lisäksi ollut edustettuina rekisteritietoja tuottavia ja hallinnoivia viranomaisia sekä opetus- ja kulttuuriministeriö. Ohjausryhmän ovat muodostaneet: Pirjo-Leena Forsström, CSC; Jaason Haapakoski, Terveystieteiden ja hyvinvoinnin laitos; Kimmo Koivunen, CSC (ohjausryhmän puheenjohtaja); Irma-Leena Notkola, ReTki; Leena Storgårds, Tilastokeskus; Markku Suvanen, opetus- ja kulttuuriministeriö ja Jouko Verho, Kansaneläkelaitos. Lisäksi ohjausryhmän toimintaan osallistuivat Esko Karjala Kansaneläkelaitoksesta ja Antti Katainen Tilastokeskuksesta.

Selvityshankkeeseen on lisäksi osallistunut monin tavoin suuri määrä tutkijoita ja viranomaisten edustajia. Kiitämme opetus- ja kulttuuriministeriötä hankkeen käynnistämisestä sekä kaikkia selvitystyöhön osallistuneita arvokkaasta panoksesta. Toivomme, että raportti kannustaa tutkimusta palvelevan etäkäyttöjärjestelmän rakentamiseen ja käyttöönottoon ja tätä kautta rekisteritutkimuksen mahdollisuuksien merkittävään parantamiseen.

Tiivistelmä

Rekisteriaineistojen etäkäyttöhanke on ollut selvittää, miten rekisteritutkimusta ja rekisteritietoihin perustuvien aineistojen (jatkossa rekisteriaineistojen) saatavuutta voidaan tehostaa etäkäytön avulla. Opetusministeriön rahoittaman hankkeen ovat toteuttaneet CSC – Tieteen tietotekniikan keskus ja Rekisteritutkimuksen tukikeskus (ReTki). Hankkeen ehdottama toimintamalli helpottaa ja monipuolistaa rekisteritutkimuksen tekemistä, tarjoaa tutkimuksen tueksi uuden teknisen ratkaisun ja parantaa julkisin varoin kerätyn tiedon saatavuutta ja uudelleenkäyttömahdollisuuksia. Toimintamalli parantaa erityisesti tietosuojamielessä arkaluontoisten aineistojen käytettävyyttä ja yhdistämistä useammilta viranomaisilta. Tässä raportissa etäkäyttöjärjestelmästä käytetään lyhennettä MIDRAS (Micro Data Remote Access System).

Selvityksessä on hankkeelle asetettujen tavoitteiden mukaisesti keskitytty luvanvaraisten, yksikötasojen rekisteriaineistojen tutkimuskäytön edistämiseen. MIDRAS-järjestelmää voidaan kuitenkin hyödyntää myös muuntotyypisten aineistojen luovuttamisessa sekä yleiskäyttöisempänä tutkimus- ja analyysialustana esimerkiksi viranomaisten selvitystöissä.

MIDRAS-järjestelmä vastauksena rekisteritutkimuksen nykyhaasteisiin

Rekisteriaineistolla tarkoitetaan tässä raportissa viranomaisten hallussa olevia, julkisin varoin hallinnollisiin ja tilastotarkoituksiin rekistereihin kerättyjä yksikötasoisia aineistoja, joiden käyttö on luvanvaraista. (Selvityksessä käytetty sanasto on esitelty tarkemmin liitteessä 1.)

Rekisteriaineistot ovat tärkeä tietovaranto tutkimuksen kannalta, mutta niiden käyttöön tutkimuksessa liittyy nykyisin monia ongelmia ja haasteita. Tutkijoiden näkökulmasta ongelmia ovat erityisesti seuraavat:

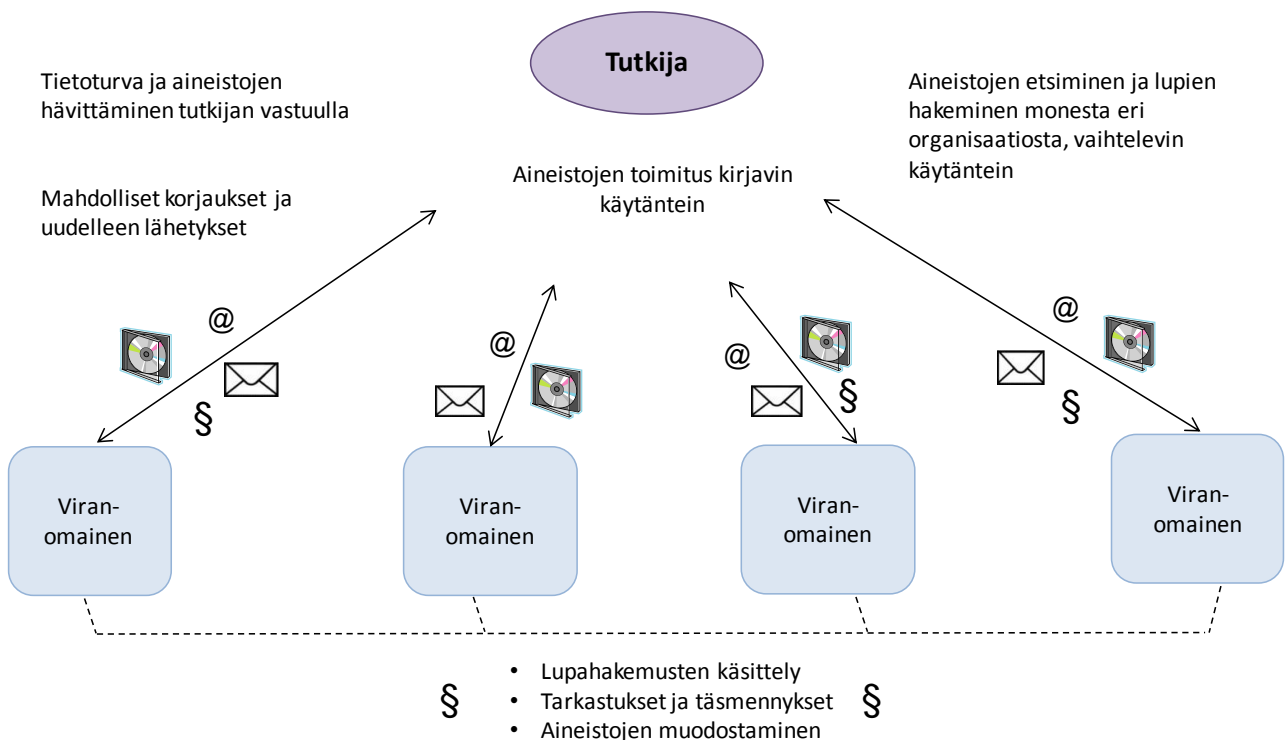
- Käyttölupien myöntämiseen ja tutkimusaineistojen muokkaamiseen kuluu viranomaisilta paljon aikaa.
- Rekistereitä ylläpitäviä viranomaisia on monia ja aineistojen luovutukseen liittyvät käytännöt ja lait vaihtelevat.
- Rekisteriselosteissa ja aineistojen kuvailussa on tutkimuksen kannalta puutteita.
- Tutkimusasetelman kannalta parasta aineistoa ei välttämättä saa tutkittavaksi tilastolain asettaman anonymisointivaatimuksen takia.
- Joidenkin aineistojen hintataso on hyvin korkea.
- Jos aineisto kulkee monen viranomaisen kautta tutkimuskäyttöön, tutkimukseen luovutettavat tiedot eivät ole tuoreita.

Viranomaisten näkökulmasta keskeisiä haasteita rekisteriaineistojen tutkimuskäytön edistämiseen liittyen:

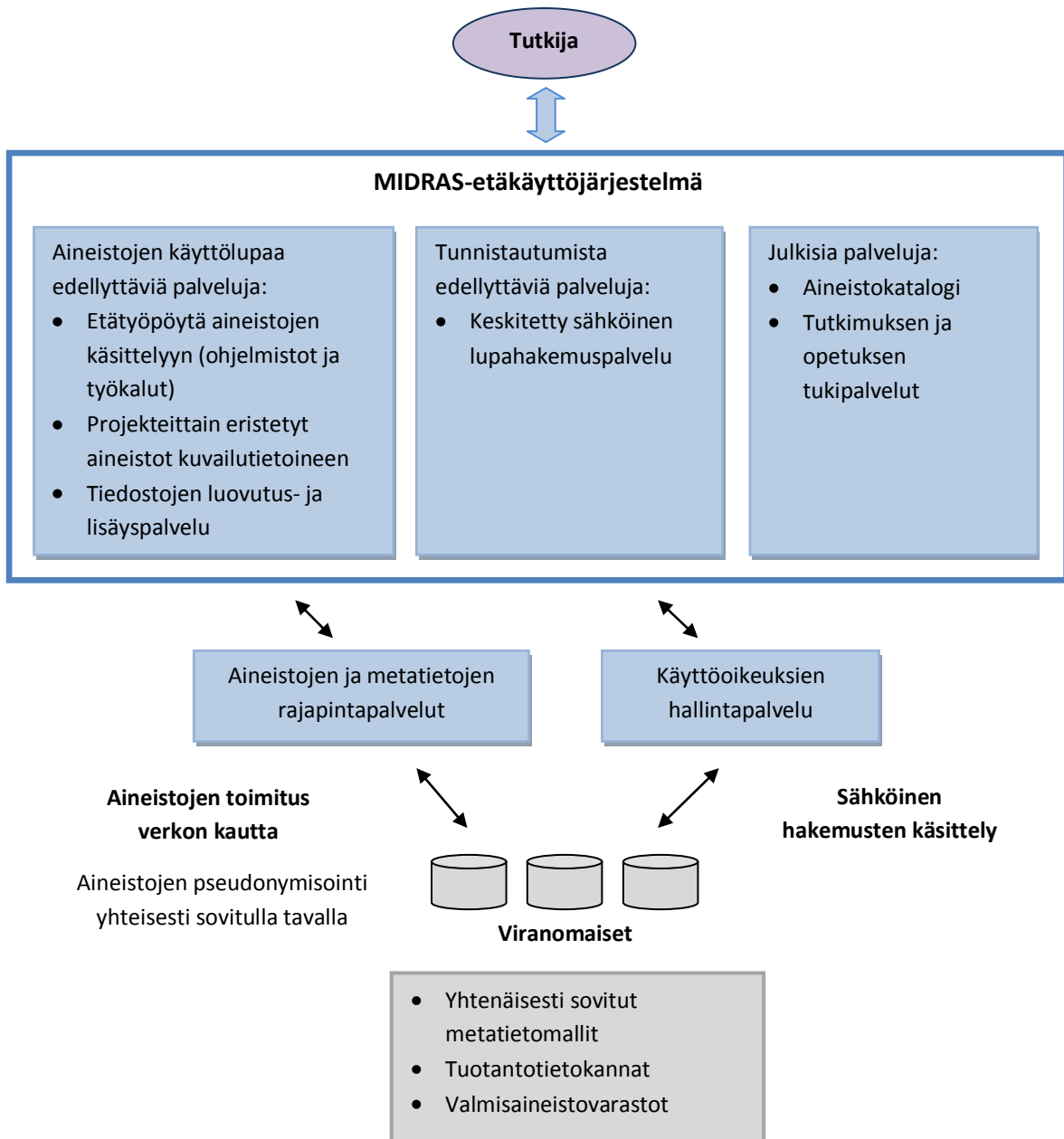
- Tutkijapalvelu ei ole viranomaisten ensisijainen tehtävä ja siksi resurssit tähän ovat vähäiset.
- Tilastolain vaatimus aineistojen anonymisoinnista vaikuttaa aineistojen kokoamisen vaatimaan työmäärään.
- Aineistojen nykyisiin toimitustapoihin (cd tai muistitikku) liittyy tietoturvariskejä, eikä aineistojen käyttöä ja hävittämistä ole mahdollista valvoa.
- Rekisterien dokumentaatiossa, tuottamisessa ja säilytyksessä ei ole otettu huomioon aineiston tutkimuskäyttöä.
- Metatietojen sisältöä ja säilytysmuotoja ei ole standardisoitu.
- Tutkijat eivät tunne aineistoja ja tekevät puutteellisia tutkimussuunnitelmia ja lupahakemuksia.

MIDRAS-toimintamalli tarjoaa selkeitä parannuksia nykyisiin toimintamalleihin verrattuna (kuvat 1 ja 2): se tehostaa ja virtaviivaistaa aineistojen saatavuuteen liittyviä prosesseja, ja säästää näin sekä tutkijoiden että viranomaisten työmäärää, aikaa ja hallinnollista vaivaa. MIDRAS-toimintamalli tukee aineistojen laajempaa ja monipuolisempaa käyttöä, tarjoaa tutkimuksen tueksi uuden teknisen ratkaisun, ja parantaa olennaisesti tietoturvaan ja anonymisoimattomien aineistojen tietosuojaan liittyviä seikkoja.

Kuva 1 Nykytilanne, jossa tutkija hakee lupia ja aineistoja useilta viranomaisilta



Kuva 2 Kuvaus vision mukaisesta MIDRAS-toimintamallista ja tarjottavista palveluista



MIDRAS-järjestelmä toimii kanavana, jonka kautta luovutetaan käyttöluvalla tutkimuskäyttöön viranomaisten rekisteritiedoista muodostettuja rekisteriaineistoja, erityisesti arkaluontoisia yksikötason aineistoja.

Tutkijan kannalta ehdotetun MIDRAS-järjestelmän keskeisimpiä palveluita ovat:

1. Tutkimusaineistojen käsittelyyn ja analysointiin tarkoitettu eristetty etätyöpöytäympäristö palveluineen. Etäkäyttöjärjestelmässä tutkija ei saa tutkimusaineistoa haltuunsa, vaan tutkija saa aineiston käyttöönsä etätyöpöydälle, johon hän voi kirjautua omalta työpisteeltään.

2. Tutkimuslupien hakemista ja tutkimusten suunnittelua helpottava aineistokatalogi ja sille perustuva keskitetty lupahakemuspalvelu. Aineistokatalogi on www-palvelu, johon kerätään mahdollisimman yhteismitallisia metatietoja järjestelmän kautta saatavilla olevista aineistoista.

Aineistoja tuottavan viranomaisen kannalta keskeisimpiä palveluita ovat:

1. Tietojen siirto- ja käyttörajapinta MIDRAS-ylläpito-organisaation ja viranomaisen välillä. Tietojen käyttörajapinnan kautta kulkevat järjestelmän aineistot, metatiedot ja muut tiedot.
2. Käyttölupien ja -oikeuksien hallinnointityökalu. Käyttöoikeuksien hallinnointityökalu sallii tiedon ylläpitäjien myöntää käyttölupia sekä hallinnoida ja tarkastella, millä tutkimusprojekteilla ja keillä tutkijoilla on pääsy mihinkin aineistoon.

Visio

Kansallisen etäkäyttöjärjestelmän kautta viranomaisten hallussa olevat, luvanvaraiset, yksikötasoiset aineistot ovat käytettävissä kattavasti, turvallisesti, kustannustehokkaasti ja helposti tutkimukseen.

MIDRAS-toimintamalli luo yhteiset käytännöt ja suositukset, jotka yhtenäistävät ja virtaviivaistavat tutkimusprosessin, jolloin tutkijat keksivät uusia tapoja hyödyntää vanhaa ja tuottaa uutta tietoa. MIDRAS-järjestelmä tarjoaa hyvän ja monikäyttöisen työskentely-ympäristön, aineistojen löytämistä helpottavan aineistokatalogin ja aiempaa kattavamman aineistovalikoiman yhdessä järjestelmässä. MIDRAS-järjestelmän ratkaisut helpottavat ja nopeuttavat myös viranomaisten omaa selvitys-, suunnittelu- ja raportointityötä. Uudet välineet mahdollistavat sen, että aineistojen tulkinta ja muokkaus tutkimustarkoituksiin voi siirtyä tutkijoiden tehtäväksi ja aineiston tuottajalle jää vain suoraviivainen työ, joka automatisoidaan. Automaatio, aineistojen yhtenäinen kuvailu ja kuvailuihin perustuva lupahakemus vähentävät niin tutkijan kuin aineiston tuottajan työtä.

Ehdotukset etäkäyttöjärjestelmän rakentamiseksi ja toteutus vaiheittain

Projektissa kerättyjen kokemusten perusteella eri toimijoiden mielestä Suomeen tarvitaan tutkimuksen tarpeisiin yksi toimiva rekisteriaineistojen etäkäyttöjärjestelmä. Suositamme, että:

1. Hallintomallista ja rahoituksesta päästään yksimielisyyteen vuoden 2011 aikana.
2. Järjestelmän palveluista aineistokatalogi ja sähköinen lupamenettely toteutetaan ensimmäisinä.
3. Pseudonymisointimalli sovitaan yhdessä sidosryhmien kanssa vuoden 2011 aikana. Pseudonymisointimallia ei voi kuitenkaan käyttää nykyisen tilastolain alaisissa aineistoissa.
4. Aloitetaan aineistojen saattaminen järjestelmän välityksellä käyttöön vaiheittain 2012 alkaen. Sopivia aineistoja tutkijoiden kannalta ovat esimerkiksi: hoitoilmoitusrekisteri, syntymärekisteri,

kuolemansyyrekisteri, rekisteri lääkkeiden erityiskorvaus oikeuksista, työssäkäyntitilaston aineisto ja väestölaskennan pitkittäisaineisto.

MIDRAS-etäkäyttöjärjestelmän toteutusvaiheita on kuvattu tarkemmin liitteessä 2.

Raportin sisältö ja rakenne

Tässä raportissa kuvataan selvityshankkeen tavoitteet ja toteutus sekä esitetään suositettu rekisteriaineistojen etäkäyttöjärjestelmän toimintamalli ja sen edellyttämät toimenpiteet vaiheittain. Raportti on tarkoitettu ensisijaisesti tietohallinnon suunnittelua ja päätöksentekoa varten, mutta se soveltuu myös rekisteriaineistojen tuottavien ja hallinnoivien viranomaisten ja tutkijoiden käytettäväksi. Hankkeessa on toteutettu myös laajempi selvitysraportti, joka on tarkoitettu yksityiskohtaisemmaksi oppaaksi etäkäyttöjärjestelmän rakentamiseen.

Raportin rakenne on seuraava:

- Luvussa 1 esitellään hankkeen taustaa ja tavoitteita, rekisteritutkimuksen nykytilannetta ja haasteita sekä näiden pohjalta muodostettuja MIDRAS-toimintamallin ja etäkäyttöjärjestelmän suunnittelun tavoitteita.
- Luvussa 2 kuvataan hankkeen toteutusmenetelmät ja eteneminen.
- Luvussa 3 käydään läpi hankkeen tuloksia ja eri vaiheissa opittuja asioita, sekä listataan MIDRAS-järjestelmän suunnittelua ohjaavia vaatimuksia ja periaatteita.
- Luvussa 4 esitellään selvityshankkeen suositukset: ehdotetun MIDRAS-järjestelmän rakenne, palvelut ja toiminta, MIDRAS-toimintamallin prosessit ja vastuut, sekä edellytykset ja tarvittavat toimenpiteet toimintamallin ja etäkäyttöjärjestelmän rakentamiseksi.

Sisällysluettelo

Esipuhe.....	2
Tiivistelmä	3
Sisällysluettelo.....	8
1. Johdanto.....	10
1.1. Selvityshankkeen tausta ja tavoitteet	10
1.2. Rekisteritutkimuksen nykytilanne ja haasteet.....	11
1.3. Nykytilanteeseen liittyviä haasteita ja ongelmia	16
1.4. MIDRAS-toimintamallin suunnittelun tavoitteet nykytilanteen perusteella.....	17
2. Selvityshankkeen toteutus	19
2.1. Menetelmät ja aineistot	19
2.2. Selvityshankkeen aikana toteutetun pilotin kuvaus	20
3. Selvityshankkeen aikana opittuja asioita	22
3.1. Toimijoiden tahtotila.....	22
3.2. Aineistojen tuottaminen ja toimittaminen.....	23
3.3. Teknisiä ratkaisuja.....	24
3.4. Lainsäädännölliset kysymykset.....	24
3.5. Kustannukset	25
3.6. Yhteistyön ja koordinaation tärkeys	26
3.7. MIDRAS-järjestelmän laajentamismahdollisuudet	26
3.8. Käyttäjä- ja järjestelmävaatimukset ja suunnittelua ohjaavat periaatteet	26
4. Selvityshankkeen suositukset.....	28
4.1. MIDRAS-visio ja MIDRAS-toimintamallin yleiskuvaus.....	28
4.2. MIDRAS-järjestelmän palvelut.....	29
4.3. MIDRAS-toimintamallin prosessit ja vastuut.....	31
4.4. MIDRAS-hallintomalli ja -ylläpito	34
4.5. MIDRAS-toimintamallin vaatimat toimet.....	35
4.6. Toteutusvaiheet.....	38
4.7. MIDRAS-toimintamallin hyödyt	40
4.8. Yhteenveto	41
Lähteet	42
Liitteet.....	44
Liite 1. Sanasto	44
Liite 2. MIDRAS-toimintamallin toteutusehdotus vaiheittain	47
Liite 3. Kustannusarvio.....	48

Liite 4. Riskit	50
Liite 5. Eräitä keskeisiä tutkimuksessa käytettäviä rekistereitä ja rekistereihin perustuvia aineistoja	52

Kuvat ja taulukot

Kuva 1 Nykytilanne, jossa tutkija hakee lupia ja aineistoja useilta viranomaisilta	4
Kuva 2 Kuvaus vision mukaisesta MIDRAS-toimintamallista ja tarjottavista palveluista	5
Kuva 3 Väitöskirjoissa vuosina 2005-2009 eniten käytetyt henkilörekisterit	12
Kuva 4 Nykytilanne, jossa tutkija hakee lupia ja aineistoja useilta viranomaisilta	17
Kuva 5 Etäkäyttövisio	17
Kuva 6 Kuvaus MIDRAS-toimintamallista ja tarjottavista palveluista	29

Taulukko 1 Rekisteritutkimuksen kannalta keskeisimpien viranomaisten käsittelemät hakemukset ja tutkijapalveluun käyttämät resurssit henkilötyövuosina mitattuna	14
Taulukko 2 Rekisteritutkimukseen liittyviä haasteita	16
Taulukko 3 MIDRAS-toimintamallin suunnittelua ohjanneet tavoitteet	18
Taulukko 4 Kooste etäkäyttäjärjestelmän suunnittelua ohjaavista periaatteista	27
Taulukko 5 MIDRAS-järjestelmän palvelut	30
Taulukko 6 Palvelut ja vastuut hallinnon prosesseissa	31
Taulukko 7 Palvelut ja vastuut valmistelun prosesseissa	32
Taulukko 8 Palvelut ja vastuut lupahakemusprosesseissa	32
Taulukko 9 Palvelut ja vastuut tutkimusaineistoihin liittyvissä prosesseissa	33
Taulukko 10 Palvelut ja vastuut tutkimuksen prosesseissa	34
Taulukko 11 MIDRAS-toimintamallin edellytykset	36
Taulukko 12 MIDRAS-toimintamallin hyödyt eri sidosryhmille	40

Liitteiden taulukot

Taulukko 13 Viranomaisten työt	48
Taulukko 14 Tekninen järjestelmä	49
Taulukko 15 MIDRAS-hallinto	49
Taulukko 16 MIDRAS-toimintamallin riskit ja varautumiskeinot	50

1. Johdanto

1.1. Selvityshankkeen tausta ja tavoitteet

Rekisteriaineistojen etäkäyttöhankkeen taustalla ovat rekisteritutkimuksen toteuttamiseen liittyvät ongelmat sekä pyrkimys lisätä julkisin varoin kerättyjen aineistojen uudelleenkäyttöä tutkimuksessa ja parantaa rekisteritutkimuksen teknisiä edellytyksiä. Rekisteriaineistolla tarkoitetaan tässä raportissa viranomaisten hallussa olevia, julkisin varoin hallinnollisiin ja tilastotarkoituksiin rekistereihin kerättyjä yksikötasoisia aineistoja, joiden käyttö on luvanvaraista. Rekisteritutkimuksen tukikeskus (ReTki) on vienyt rekisterien etäkäyttöideaa eteenpäin joulukuussa 2007 järjestämästään neuvottelupäivästä lähtien. Keväällä 2008 ReTki teki hankkeesta esityksen kansalliseen tutkimusinfrastruktuurikartoitukseen. Hankeesitys menestyi hyvin kansainvälisen arviointipaneelin arvioinnissa ja se katsottiin yhdeksi niistä 13 tiekarttaehdotuksesta, ”joilla voisi olla mahdollisuudet kehittyä merkittäviksi kansallisiksi infrastruktuureiksi” (Opetusministeriö, 2009). Keväällä 2009 ReTki ja CSC – Tieteen tietotekniikan keskus tekivät yhteisen hanke-ehdotuksen silloiselle opetusministeriölle, jossa ehdotettiin rekisteritietojen etäkäytön selvityshankkeen rahoittamista. Opetusministeriö myönsi avustuksen rekisteritiedon etäkäytön valmisteluhankkeelle 22.4.2009 (päätös DNRO 16/525/2009). Tässä raportissa etäkäyttöjärjestelmästä käytetään lyhennettä MIDRAS (Micro Data Remote Access System).

Hankkeen tavoitteena on ollut suunnitella toimintamallia ja pilotoida etäkäyttöjärjestelmää, jonka avulla viranomaisrekistereihin kerättyjä tietoja voidaan käyttää tutkimustarkoituksiin nopeasti, helposti ja tietoturvallisesti, yksiköiden tietosuoja varmistaen.

Hankkeelle asetetut tavoitteet olivat:

- Etäkäyttöjärjestelmän toimintamalliehdotuksen laatiminen
- Vastaavien järjestelmien ja eri toimintamallivaihtoehtojen selvittäminen ja esittely
- Selvitys eri toimijoiden mahdollisuuksista ja halukkuudesta osallistua järjestelmän rakentamiseen ja aineistojen toimittamiseen, sisältäen myös mahdollisten lainsäädännöllisten esteiden ja muutostarpeiden selvittämisen
- Pilottijärjestelmän rakentaminen.

Hankkeen tavoitteisiin kuului lisäksi järjestelmän rakentamis- ja käyttökustannusten arviointi sekä rahoitus- ja hallintomallin ehdottaminen. Näitä tavoitteita toteutettiin lopulta kuitenkin vain osittain, koska hankkeen rinnalle asetettiin kesällä 2010 työryhmä, jonka nimenomaisiin tehtäviin kuuluu ehdotuksen tekeminen etäkäyttöjärjestelmän rahoitus- ja hallintomallista (Valtiovarainministeriö, 2010e).

Hankkeessa lähtökohtana on ollut parantaa viranomaisten hallinnollisiin käyttötarkoituksiin keräämien rekisteritietojen käyttöä tieteellisissä tutkimuksissa. Selvityshanke on rajattu koskemaan ensisijaisesti henkilörekistereitä, joiden luovuttamiseen ja käyttöön liittyy useita tietosuojaan ja tietoturvaan liittyviä haasteita. Hankkeen ehdottama MIDRAS-järjestelmä soveltuu kuitenkin useiden erityyppisten aineistojen luovuttamiseen sekä yleiskäyttöisemmäksi tutkimus- ja analyysialustaksi.

1.2. Rekisteritutkimuksen nykytilanne ja haasteet

1.2.1. Rekisterit ja niitä hallinnoivat viranomaiset

Henkilötietojen ja muiden yksikötason tietojen keräämisellä eri rekistereihin on pitkät perinteet Suomessa. Rekisterit sisältävät tiettyä joukkoa kokonaisuudessaan kuvaavia tietoja sähköisessä muodossa, joita päivitetään tietojen muuttuessa. Useimmiten rekisteritiedon yksikkönä on tapahtuma.

Rekisteritiedot kerätään pääosin hallinnollisiin tarkoituksiin, mutta Suomessa on myös lakisääteisiä rekistereitä, joita kerätään tilastointi- ja tutkimustarkoituksiin (keskeiset rekisterit on lueteltu liitteessä 5). Väestörekisterikeskuksen väestötietojärjestelmä on henkilötietojen perusrekisteri. Useat tahot keräävät koko Suomen kattavia rekisteritietoja omalta hallinnonalaltaan: Verohallitus (verorekisteri), Kela (etusrekisteri), Eläketurvakeskus (eläke- ja ansaintarekisterit), Oikeusrekisterikeskus (rikosrekisteri), työ- ja elinkeinoministeriö (työhallinnon asiakaspalvelurekisteri) ja Opetushallitus (yhteisvalintarekisteri). Terveiden ja hyvinvoinnin laitoksella on oikeus kerätä ja ylläpitää useita rekistereitä tutkimus- ja tilastointitarkoituksiin, muun muassa hoitoilmoitusrekisteri HILMO, syöpärekisteri ja syntyneiden lasten rekisteri (esim. Gissler & Haukka, 2004). Moniin näihin rekistereihin on kerätty tietoja useiden vuosikymmenien ajalta. Myös Tilastokeskus on merkittävä tiedonkerääjä ja perusrekisterien ylläpitäjä. Tilastokeskukseen kerätään tilastolain nojalla laajasti tietoja yhteiskunnallisia oloja kuvaavista rekistereistä tilastointia varten ja Tilastokeskus pitää myös itse yllä useita rekistereitä (Tilastokeskus, 2004). Tässä selvityksessä Tilastokeskuksen yksikötasoisia aineistoja tarkastellaan rekistereinä, vaikka aineistot eivät päivitty jatkuvasti.

1.2.2. Rekisteritutkimuksen merkittävyys tutkimusalana ja tutkimusmäärät

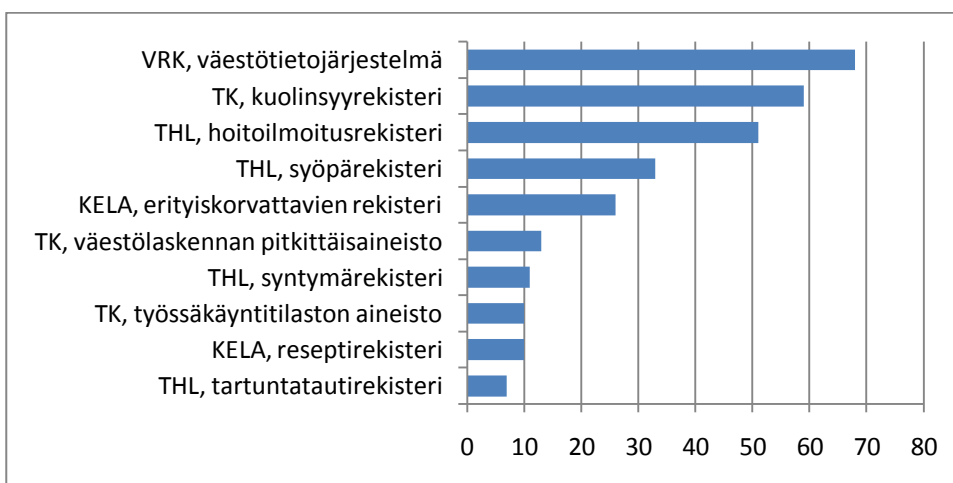
Pohjoismaissa on kansallisiin henkilörekistereihin koottu poikkeuksellisen hyvä tietovaranto, joka muodostaa kansainvälisesti verrattaessa erinomaisen pohjan tutkimukselle. Rekisteritutkimuksen nykykäytänteiden ongelmien vuoksi näitä aineistoja ei tutkimuksessa kuitenkaan ole voitu vielä täysimääräisesti hyödyntää. Suomessa viranomaiset käsittelivät vuonna 2009 yhteensä noin 450 käyttöluvahakemusta rekisteritietoihin tutkimuskäyttöä varten. Erityisesti terveystieteissä rekisteritutkimuksella on vahva perinne. Joka viides vuosina 2008–2009 julkaistu terveystieteellinen

väitöskirja perustui rekistereistä kerättyihin tietoihin (Rekisteritutkimuksen tukikeskus, 2010). Julkaisutietokanta PubMedin haun mukaan suomalaiset tutkijat julkaisevat kansainvälisissä aikakauslehdissä vuosittain noin 200 lääketieteellistä artikkelia, jossa on käytetty tutkimusaineistona suomalaisia rekistereitä.

Rekisteritutkimuksen etu on, että eri rekisterien tietoja voi yhdistää toisiinsa sekä myös esimerkiksi kliinisiin aineistoihin, potilasasiakirjoihin tai kyselyaineistoihin. Esimerkiksi kuolinsyyrekisterin (Tilastokeskus), hoitoilmoitusrekisterin (THL) ja Kelan erityiskorvattavien lääkkeiden rekisterin (osa Kelan etuusrekisteriä) tietoja yhdistelemällä saadaan tarpeellista tietoa tiettyjen sairauksien hoidosta ja hoidon tuloksista. Kuvasta 3 käy ilmi, että juuri näiden rekisterien tietoja käytetään paljon väitöskirjatutkimuksissa. Pääsääntöisesti rekisteritietoja hyödyntävissä väitöskirjatutkimuksissa käytetään sekä rekistereistä saatuja tietoja että muilla tavoilla kerättyjä tietoja (ibid.). Väestörekisterikeskuksen väestötietojärjestelmän tietoja käytetään usein tutkimusjoukon muodostamiseen.

Tilastokeskuksen aineistot muodostaisivat puolestaan erinomaisen aineistopohjan esimerkiksi yhteiskunnalliselle ja taloustieteelliselle tutkimukselle. Tällä hetkellä nämä aineistot ovat vajavaisesti hyödynnettyjä. Tutkimuksen perusteella (ibid.) vain noin viisi prosenttia vuosina 2008–2009 julkaistuista yhteiskuntatieteellisistä väitöskirjoista sisälsi rekistereistä poimittuja tietoja.

Kuva 3 Väitöskirjoissa vuosina 2005–2009 eniten käytetyt henkilörekisterit



(Rekisteritutkimuksen tukikeskus, 2010.)

1.2.3. Rekisteritietojen saatavuus tutkimukseen, lupamenettelyt ja hakuprosessit

Rekistereihin kerättyjen tietojen luovuttaminen ja käyttö on hyvin tarkasti säädeltyä niin viranomaisen kuin tutkijan näkökulmasta. Rekisteritietojen käsittelyä ja luovutusta säätelee kaksi yleislakia, henkilötietolaki (523/1999) ja viranomaisten toiminnan julkisuudesta annettu laki eli julkisuuslaki (621/1999). Tämän lisäksi

eri viranomaisten hallinnoimien rekisteritietojen keräämistä ja luovuttamista säätelevät lukuisat erityislait. Tilastoviranomaisen tilastotarkoituksiin saamien tietojen luovutukseen sovelletaan tilastolakia (280/2004).

Henkilötietolaki on yleislakina aina otettava huomioon henkilöaineistojen eli henkilötietoja sisältävien aineistojen käsittelyssä. Henkilörekisterit sisältävät henkilöitä kuvaavia yksilötasoisia tietoja, jonka seurauksena tiedot on usein määrätty lain nojalla salassa pidettäviksi. Terveys- ja etuusrekisterien sisältämä tieto on lisäksi arkaluonteista. Tilastolain alaisia aineistoja lukuun ottamatta lainsäädäntömme mahdollistaa useimpien yksilötasoisien aineistojen luovuttamisen henkilötunnukseen tieteelliseen tutkimukseen ja THL:n terveydenhuollon valtakunnallisia rekistereitä lukuun ottamatta myös viranomaisten tilastollisiin selvityksiin, jos tutkimuksen tai selvityksen toteuttamisessa henkilötunnukset ovat tarpeen. Tilastolain alaisista aineistoista henkilötunnuksen saa luovuttaa vain muutamiin tietoihin liittyen (ikä, sukupuoli, koulutus, ammatti).

Julkisuuslain mukaan käyttöluvan tutkimusaineistoihin myöntää useimmissa tapauksissa aineistoja tuottava viranomaisen itse. Tilastoviranomaisella on kuitenkin tilastolain mukaan oikeus myöntää käyttöluva tietoihin, jotka muut viranomaiset ovat luovuttaneet tilastoviranomaisen tietopyynnön perusteella tilastoviranomaiselle tietopyynnössä määriteltä tarkoitusta varten (tilastointi-, tutkimustarkoitus).

Käyttöluvan hakeminen voi tapahtua joko vapaamuotoisesti tai viranomaisen tarjoamalla määrämuotoisella käyttöluvahakemuksella. Ennen luvan myöntämistä henkilörekistereitä hallinnoivat viranomaiset varmistavat hakemusasiakirjoista, että henkilötietolain edellytykset tietojen luovuttamiselle tutkimukseen tai selvitykseen täyttyvät. Tutkijat allekirjoittavat lisäksi hakemuksen yhteydessä salassapitositoumuksen.

Hakemusten käsittely vie aikaa (kuukausia tai jopa vuoden), varsinkin jos tutkijan hakemuksessa, tutkimussuunnitelmassa tai tutkimuksen organisoinnissa havaitaan ongelmia. Lisäksi viranomaiset pyytävät lausuntoja rekisteriasiantuntijoilta ja tarvittaessa tietosuojavaltuutetulta.

Tilastolain perusteella luovutetuista tutkimusaineistoista ei saa tunnistaa yksilöitä edes välillisesti, eli aineisto täytyy anonymisoida ennen kuin se luovutetaan tutkijalle. Tilastolain alaisten aineistojen luovutuksissa aika kuluu pitkälti tutkimusaineiston sisällön täsmentämiseen anonymisoinnin vaatimukset huomioon ottaen. Tutkijan näkökulmasta nykytilanteessa on paradoksaalista, että tutkija voi saada käyttöönsä hyvin arkaluonteista henkilötunnuksellista tietoa mm. terveys- ja etuusrekistereistä, mutta tilastolain alaisia tietoja, joista osa on alkuperäisrekistereissä jopa julkisia tietoja, saa tutkimuskäyttöön vain anonymisoituna.

1.2.4. Aineistojen maksullisuus

Koska rekistereihin kerätään tietoja viranomaisten omaan hallinnollisiin ja tilastollisiin tarpeisiin, viranomaisen ei laskuta aineiston käyttäjää (tutkijaa) tietojen keräämisen aiheuttamista kustannuksista. Maksuperustelain (150/1992) mukaan viranomaisen tulee kuitenkin laskuttaa siitä työstä, joka joudutaan tekemään tutkijan toimeksiannosta, kuten tiettyjen muuttujien poimiminen rekisteristä tutkijan tilaamaa rekisteriaineistoa varten (irrottamiskustannukset). Lisäksi aineistojen maksullisuutta säädellään asetuksin (1027/2010; 1212/2010). Asetuksen mukaan Tilastokeskus veloittaa tutkijoita myös tilastolain vaatimasta aineistojen anonymisointiin kuluva työstä.

Tietojen poiminta rekistereistä voi tulla kalliiksi jos rekisteritiedot on jaettu lukuisiin vuosiaineistoihin tai varsinkin siinä tapauksessa, jos viranomaisen on ulkoistanut rekisterin ylläpidon kaupalliselle toimijalle sopimuksella, jossa ei oteta huomioon satunnaiskäyttöä. Rekisteritietojen poimimisesta veloitetaan joskus myös poimittujen tietojen lukumäärän mukaan. Käytäntö ei vaikuta perustellulta, sillä tietotekniikka mahdollistaa kymmenen tai tuhannen henkilön tietojen poimimisen samalla työmäärällä. Tutkimusaineiston hinta voikin muodostua parista sadasta eurosta useisiin kymmeneen tuhansiin euroihin.

1.2.5. Viranomaisten tutkimusaineiston valmistelutyö

Selvityshankkeessa tehtyjen kyselyjen perusteella tutkijapalvelun vaatimat resurssit vaihtelevat viranomaisittain (taulukko 1). Eniten hakemuksia tulee Väestörekisterikeskukselle, jossa on hyvin vakiintunut prosessi rekisteriaineistojen luovuttamiseksi. Muille kuin taulukoissa mainituille viranomaisille tulee vuodessa korkeintaan 20 tutkimuslupahakemusta. Näilläkin viranomaisilla on kuitenkin oltava valmiudet käsitellä lupahakemuksia ja luovuttaa aineistoja. Minkään viranomaisen ensisijainen tehtävä ei ole tutkijoiden palveleminen. Tämän takia resurssit tutkimuspalveluun ovat alimitoitettut, kun työtä tehdään muiden toimien ohella.

Taulukko 1 Rekisteritutkimuksen kannalta keskeisimpien viranomaisten käsittelemät hakemukset ja tutkijapalveluun käyttämät resurssit henkilötyövuosina mitattuna

Viranomaisen	Lupahakemuksia vuosittain, n.	Käyttölupien käsittely	Aineiston poiminta ja muokkaus	Yhteensä
THL	60–70	2 htv	1,5–2 htv	3,5–4 htv
KELA	50–70	0,5 htv	1,5–2 htv	2–2,5 htv
Tilastokeskus - kuolinsyy - muut	60–80 50–70			4,5–5 htv
VRK	100–150			4 htv

Eriyiset tutkimusasetelmat vaativat useimmiten juuri tietynsisältöisen tutkimusaineiston ja tutkija voi saada käyttöönsä vain sellaisia tietoja, joita hän tutkimussuunnitelmansa mukaan katsoo tarvitsevansa. Viranomaiset räätälöivät tutkijalle hänen tarpeidensa mukaisen tutkimusaineiston, mikä vaatii työaika.

Tilastotarkoituksiin kerättyjen aineistojen anonymisointivaatimus ennen niiden luovuttamista tutkijalle aiheuttaa tilastoviranomaiselle lisätyötä muihin rekisteritietoja luovuttaviin viranomaisiin verrattuna. Aineistojen poiminnan helpottamiseksi Tilastokeskuksessa on muodostettu tutkimustarkoitusta varten valmisaineistoja, joista tutkijat voivat tilata itselleen otoksia. Toisinaan myös tilastotaulukkomuotoon kerätyt aggregoidut tiedot voivat olla riittävä tietolähde tutkijalle. Jos tutkijan tarvitsemaa taulukkoa ei ole aikaisemmin tehty, tutkijalle tuotetaan taulukko yksilöaineistosta. Tietosuojan näkökulmasta tilastomuotoinen luovutustapa olisi toivottava. Aineiston muokkaaminen tutkimuskohtaiseksi tilastotaulukoksi on kuitenkin työläämpää kuin yksilöaineiston poiminta ja täten myös kalliimpaa.

1.2.6. Rekisteriselosteet ja aineistojen mukana toimitettavat metatiedot

Rekisteritiedot on yleensä alun perin kerätty muuhun tarkoitukseen kuin tutkimusaineistoksi. Tästä syystä on tärkeää niin suunnittelu- kuin tutkimusvaiheessa, että tutkija saa rekisteriaineistojen kuvailutiedoista tarkan käsityksen siitä, miten tiedot kuvaavat tutkijan tutkimuskohdetta. Tutkijan tulee esimerkiksi tietää, mitkä havaitut muutokset voivat johtua rekisteritietojen ilmoittamistavassa tapahtuneista muutoksista.

Henkilötietolaissa viranomaisia veloitetaan ylläpitämään hallinnoimistaan henkilörekistereistä rekisteriselosteita. Rekisteriselosteet on pidettävä julkisesti esillä. Useat viranomaiset pitävät rekisteriselosteita verkkosivuillaan. Rekisteriselosteesta tulee ilmetä, mitä henkilötietoja rekisterissä on, mutta usein rekisteriselosteissa kuvataan kuitenkin vain yleisellä tasolla rekisterin sisältöä (Tietosuojavaltuutetun toimisto, 2010). Viranomaisilla voi kuitenkin olla omassa käytössään tarkempia tietuekuvauksia.

Aineistoluovutuksen yhteydessä tutkijoille toimitetaan kuvaus muodostetusta tutkimusaineistosta. Nämä voidaan toimittaa aineiston osana (SAS- tai SPSS-tiedostoissa), mutta useimmiten tutkijoille toimitetaan tiedot vapaamuotoisina tekstidokumentteina.

Rekisteriaineistojen kuvaaminen vaihtelee viranomaisittain ja kuvausten hahmottaminen voi olla tutkijalle haastavaa. Viranomaiset ovat myöntäneet, että rekisterien kuvailutasossa on puutteita ja että asialle pitäisi tehdä jotain. Tutkimuskäytön suhteen keskeinen ongelma on, että aineistojen kuvailulla tutkimustarkoituksiin on erilaisia vaatimuksia kuin tietojen kuvailulla hallinnolliseen tai tilastokäyttöön. Viranomaisten kannalta ongelmana on, että dokumentointiin kuluu paljon työaika eikä käytössä ole hyviä, yhtenäisiä työkaluja. Rekistereissä käytetyt laajemmat luokitukset ovat usein standardoituja.

1.3. Nykytilanteeseen liittyviä haasteita ja ongelmia

Taulukkoon 2 on koostettu ne keskeiset ongelmat ja haasteet, joita rekisteritietojen tutkimuskäytössä on tällä hetkellä. Tutkijoiden ongelmat liittyvät ennen kaikkea aineistojen saatavuuteen, viranomaisten ongelmat resurssien puutteeseen.

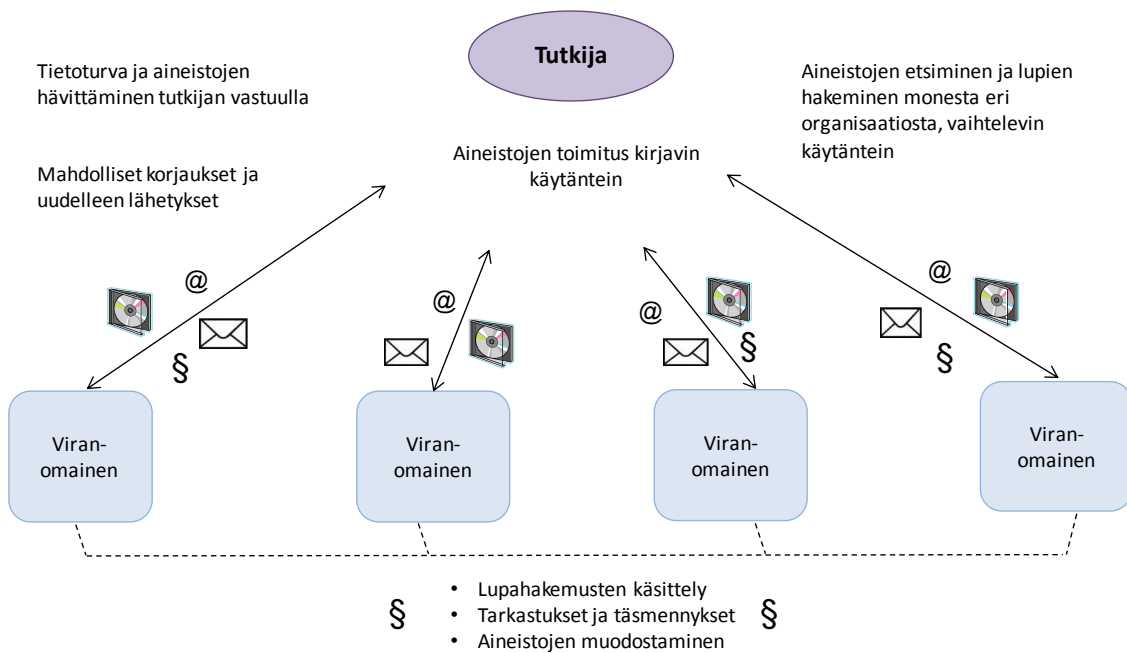
Taulukko 2 Rekisteritutkimukseen liittyviä haasteita

Haaste	Vaikutus
Tutkijan näkökulma	
Käyttölupien myöntämiseen ja aineistojen toimittamiseen kuluva aika	Tutkijalle on vaikeaa suunnitella tutkimuksen ajankäyttöä. Tutkimuksen rahoitusajasta voi kuluu suuri osa aineistojen odotteluun.
Monta eri rekisteritoimijaa ja vaihtelevat käytännöt	Tutkimuslupaprosessi sekä aineistojen käyttöön saaminen on työlästä varsinkin useamman viranomaisen aineistoja yhdistettäessä.
Rekisteriselosteiden ja aineistojen kuvailun puutteet	Tutkijalle on haasteellista tietää saatavilla olevista aineistoista. Aineistojen kuvailuja joudutaan täsmentämään myös aineistojen toimituksen jälkeen.
Tutkimusasetelman kannalta parasta aineistoa ei välttämättä saa tutkittavaksi (liittyen erityisesti tilastolain vaatimukseen aineistojen anonymisoinnista).	Aineistot voivat olla riittämättömiä tutkimusasetelman kannalta tai tutkimusta voi olla mahdotonta suorittaa laadukkaasti.
Joidenkin aineistojen korkea hintataso	Aineistoja voi jäädä käyttämättä korkeiden kustannusten takia. Kustannukset eivät ole aina nopeasti selvillä.
Hidas tuotantoprosessi	Ei voi saada nykytilannetta kuvaavia tietoja.
Viranomaisen näkökulma	
Rajalliset resurssit tutkijoiden palvelemiseen	Rajalliset resurssit tutkimuslupakäsittelyyn ja aineistojen muokkaamiseen vaikuttavat erityisesti prosessien kesto.
Tilastolain vaatimus aineistojen anonymisoinnista	Vaikuttaa aineistojen muodostamisen vaatimaan työmäärään.
Aineistojen toimitustapaan liittyvät tietosuojat ja -turvariskit	Henkilötietoja sisältäviä tutkimusaineistoja luovutetaan CD:llä tai muistitikulla. Mahdollisuudet valvoa aineistojen hävittämistä ovat hyvin rajalliset.
Rekisterien dokumentaatiossa, tuottamisessa ja säilytyksessä ei ole otettu huomioon aineiston tutkimuskäyttöä.	Prosessit vievät paljon aikaa ja resursseja. Pöiminoissa tapahtuu virheitä ja väärinkäsityksiä. Aineiston tallennusmuoto ei välttämättä tue tutkimustarkoituksia (tutkijan tarvitsemien tietojen poimiminen aineistosta voi olla vaikeaa).
Metatietojen standardoinnin puutteet	Kuvailu on työlästä eivätkä aineistojen kuvailut ole yhteismitallisia.
Tutkijoiden vähäinen aineistotuntemus ja puutteelliset tutkimussuunnitelmat ja lupahakemukset	Tutkijoiden tarpeiden selvittämiseen, aineiston muodostamiseen ja tutkimusprosessin tukemiseen kuluu aikaa ja vaivaa.
Tutkimuspalvelun kustannukset	Tutkimusaineistojen kokoamiseen tarvittavan työmäärän vuoksi hinnat eivät yleensä kata kaikkia kustannuksia.

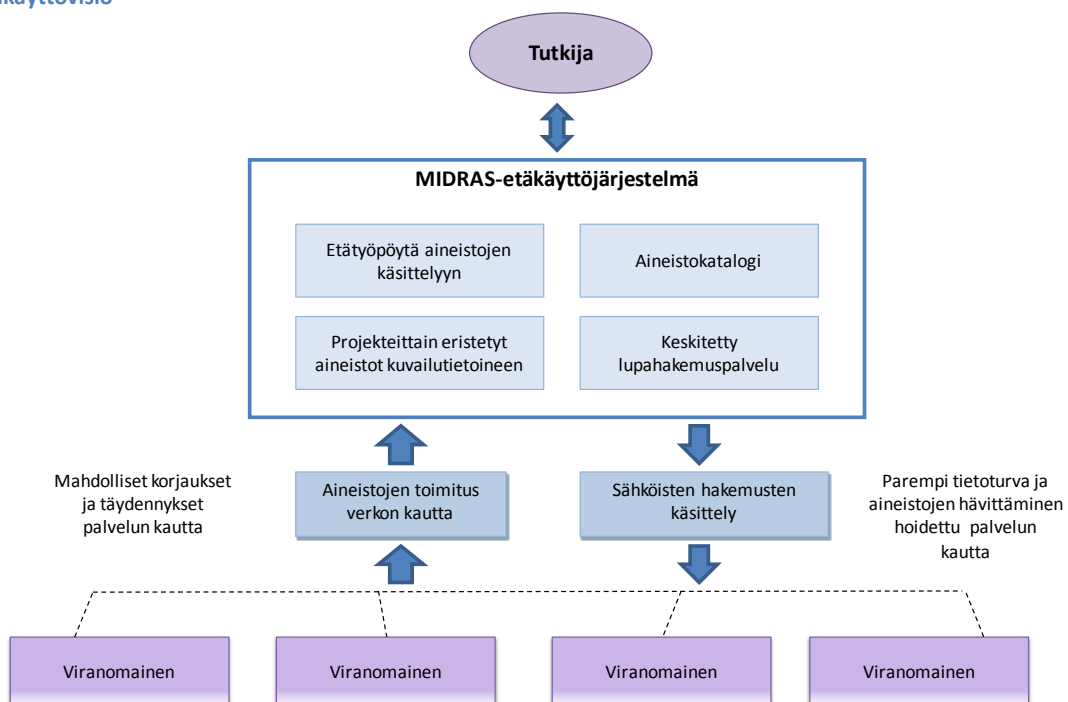
1.4. MIDRAS-toimintamallin suunnittelun tavoitteet nykytilanteen perusteella

MIDRAS-toimintamalli on kehitetty ratkaisemaan nykytilanteen ongelmia. Nykytilannetta, jossa tutkija pyytää aineistoja ja niiden käyttöluvia usealta viranomaiselta on kuvattu kuvassa 4. Kuvassa 5 on puolestaan kuvattu vision mukaista tilannetta, jossa etäkäyttöpalvelun avulla tutkija voi sekä etsiä aineistoja että hakea niiden käyttöluvia helposti, yhden sähköisen palvelun kautta. Vision mukaisesti myös aineistojen toimitus ja mahdolliset korjaukset ja päivitykset hoituvat kätevästi verkon kautta.

Kuva 4 Nykytilanne, jossa tutkija hakee lupia ja aineistoja useilta viranomaisilta



Kuva 5 Etäkäyttövisio



Toimintamallin keskeinen tavoite on se, että aineistojen saatavuus paranee, eri osapuolten työmäärä vähenee eikä mikään asia huononnu nykyhetkestä. MIDRAS-toimintamallia ja etäkäyttöjärjestelmää on suunniteltu taulukossa 3 lueteltujen tavoitteiden pohjalta. Taulukossa esitellään myös mahdollisia toteutustapoja tavoitteisiin vastaamiseksi.

Taulukko 3 MIDRAS-toimintamallin suunnittelua ohjanneet tavoitteet

Tavoite	Mahdollisia toteutustapoja
Rekisteriaineistojen saatavuuden paraneminen	Useiden viranomaisten tiedot voi saada yhden palvelun kautta. Toimintatapojen vakiointi esim. tutkimusluvissa ja tietojen lähetyksessä ja yhdistelyssä nopeuttaa kommunikaatioprosesseja. Aineistojen kuvailuja kokoava aineistokatalogi lisää tietoisuutta käytettävissä olevista tutkimusaineistoista.
Rekisteritutkijoiden aineistojen laadun paraneminen	Ajantasaisemmat aineistot saadaan automatisoimalla tutkimusaineistojen tuotantoprosessia ja tukemalla tietojen hakemista suoraan lähderegistereistä. Tietotekniset aineistojen toimituskanavat helpottavat aineistojen korjausta ja päivitystä. Paremmen tietosuojan turvin voidaan tarjota tutkijoille laajempia aineistoja.
Arkaluonteisten aineistojen käsittelyn tietosuojan paraneminen	Tiedot siirretään kryptografisesti suojattuina. Analysointiprosessista tehdään läpinäkyvää aineistojen tuottajille. Tutkimusaineiston hävittämiskäytäntö vakioidaan. Aineistot lähetetään ilman henkilötunnuksia.
Rekisteritutkimuksen luotettavuuden, toistettavuuden ja säilytyksen paraneminen	Käytetyt menetelmät ja tulokset säilytetään keskitetysti.
Uuden ja uudenlaisen tutkimuksen rohkaisu	Aineistokatalogi kattaa erilaisia aineistoja, joiden yhdistely voi tuottaa uutta tietoa. Aiempien tutkimusten ja menetelmien arkisto tarjoaa resursseja tutkimukseen.
Tämänhetkistä parempi ympäristö aineistojen työstämiseen	Tilasto- ja analyysiohjelmit tarjotaan keskitetysti. Tarjotaan tehokkaat laskentaresurssit. Tutkimusryhmä pääsee eri paikoista samaan tutkimusympäristöön.
Järjestelmä, joka palvelee aineistontuottajien tutkimuspalveluita	Aineistojen luovuttamiseksi luodaan tietoturvallinen, keskitetty ratkaisu.
Järjestelmän laajennettavuus kattamaan uusia tarpeita ja aineistotyyppjä	Järjestelmäarkkitehtuuri suunnitellaan yleiskäyttöiseksi.

2. Selvityshankkeen toteutus

2.1. Menetelmät ja aineistot

Selvityshankkeen toteutuksessa on hyödynnetty useita eri menetelmiä. Raportin tuottamisessa on noudatettu JHS-suosituksia ICT-palvelujen kehittämisen esiselvityksen ja vaatimusmäärittelyn prosessista (JHS 172; JHS 173).

Hanke on toteutettu käyttäen hyväksi aineistoja tuottavien viranomaisten vahvaa asiantuntemusta. Projektiryhmä on järjestänyt useita tapaamisia Tilastokeskuksen (TK), Terveystieteiden tutkimuskeskuksen (THL), Kansaneläkelaitoksen (Kela), työ- ja elinkeinoministeriön (TEM), Väestötietokeskuksen (VRK) sekä Eläketurvakeskuksen (ETK) edustajien kanssa. Tutkijoiden tarpeita on pyritty kartoittamaan haastattelemalla yksittäisiä tutkijoita ja järjestämällä tutkijoille yhteisiä keskustelupaneeleita.

Eri aihealueiden asiantuntijat ovat kommentoineet selvityshankkeeseen liittyviä kysymyksiä hankkeen aikana. CSC:n tietoturvan ja tiedostoylläpidon yksikkö on antanut asiantuntevan panoksensa hankkeeseen. Metatietostandardeihin liittyviä keskusteluita on käyty Yhteiskuntatieteellisen tietoarkiston (FSD), Tilastokeskuksen, Helsinki Region Infosharen ja JHS-jaoston edustajien kanssa. Lisäksi on käyty aktiivista keskustelua eri sidosryhmien ja muiden aihepiiriin liittyvien hankkeiden kanssa.

Lainsäädännöllisiä kysymyksiä käsiteltiin erityisessä lakimiespaneelissa sekä keskustelussa tietosuojavaltuutetun kanssa. Lisäksi projektiryhmä on käynyt sopimusteknisiä neuvotteluita CSC:n sopimusasiantuntijan sekä THL:n sopimuslakimiehen ja tietoturvapäällikön kanssa.

Tutkijapalveluihin vuosittain käytettävien henkilöresurssien arvioimiseksi Väestötietokeskukseen, Tilastokeskukseen, Kelaan ja THL:ään lähetettiin kysely, jossa pyydettiin erikseen arvioimaan käyttöluvakäsittelyyn ja aineistojen muokkaamiseen menevää aikaa. Lisäksi ReTKi on kerännyt vuosittain lukumäärätietoja Kelaan, THL:ään ja Tilastokeskukseen tulleista käyttöluvapahakemuksista.

Selvitystyössä on otettu huomioon myös laajemmat kansalliset ja EU-tason reunaehdot ja uudistushankkeet, kuten julkisen hallinnon kokonaisarkkitehtuurityö, uusi laki julkisen hallinnon tietohallinnon ohjauksesta (246/2010) ja Julkishallinnon perustietovarantojen rajapinnat -työryhmä, keskeiset JHS-suositukset ja standardit, tilastolainsäädännön uudistamishanke (Valtioneuvosto, 2010; valtiovainministeriö, 2010a; 2010b; 2010c ja 2010d) sekä EU:n tilastoasetus (EY 223/2009) ja direktiivi julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (EY 98/2003).

Hankkeessa on hyödynnetty aihealuetta käsitteleviä kirjallisia lähteitä. Toimintamallin selvityksen tueksi on kerätty etäkäyttöjärjestelmän vaatimukseen, hallintoon ja toteutukseen liittyvää tietoa sekä tietoa muiden

maiden järjestelmistä. Projektiryhmä kävi Ruotsissa tutustumassa vastaaviin järjestelmiin paikan päällä ja Tilastokeskus antoi selvityshankkeen käyttöön matka- ja selvitysraportit, jotka on laadittu Tilastokeskuksen Online-projektin yhteydessä.

CSC tutustui Tilastokeskuksen vuonna 2009 toteuttamaan etäkäyttäjärjestelmään (TK Online -järjestelmä), sen toteutukseen ja teknisiin ratkaisuihin. TK Online -järjestelmä on käytössä oleva tutkimuksen infrastruktuuri Suomessa.

Selvitystyön tueksi rakennettiin pilottijärjestelmä, jonka avulla tehtiin kaksi pilottitutkimusta. Pilotin toteutuksesta ja sen avulla kerätyistä tiedoista kerrotaan tarkemmin seuraavassa kappaleessa sekä erillisessä selvitysraportissa.

2.2. Selvityshankkeen aikana toteutetun pilotin kuvaus

2.2.1. Pilotin suunnittelu ja tarkoitus

Pilotissa kerättiin käytännön kokemuksia rekisteritietojen etäkäytöstä MIDRAS-järjestelmän suunnittelun tueksi. Pilotin suunnittelun lähtökohtana käytettiin ulkomaisia rekisteritietojen etäkäyttäjärjestelmiä ja Tilastokeskuksen TK Online -järjestelmää. Pilotin suunnittelussa otettiin huomioon myös tutkijapaneelissa esiin tulleet rekisteritutkijoiden kokemukset ja etäkäyttäjärjestelmään kohdistetut toiveet.

Pilotoinnin tavoitteena oli testata käytännössä ehdotettavia toimintamalleja sekä tietotekniikan että hallinnon osalta. Tarkoituksena oli tuottaa pilottijärjestelmä, joka olennaisilta osin vastaa ehdotettua varsinaista MIDRAS-järjestelmää, ja jonka hallinnollisia ja teknisiä ratkaisuja voitaisiin sellaisenaan soveltaa jatkossa MIDRAS-toimintamallissa. Mikäli pilottivaiheessa valittu ratkaisu osoittautuisi ongelmalliseksi, tavoitteena oli kokemusten perusteella ehdottaa vaihtoehtoisia ratkaisua.

2.2.2. Pilotin toteutus

Pilotti päätettiin toteuttaa kahdessa vaiheessa. Ensimmäisessä vaiheessa testattiin järjestelmän tietointegraatiota ja perustoimintaa pseudodatalla; toisessa vaiheessa pilottijärjestelmää hyödynnettiin varsinaisessa rekisteritutkimuksessa, jossa käytettiin aitoja rekisteriaineistoja.

Jo varhaisessa vaiheessa todettiin, että pilotissa ei voida kokeilla kaikkia suunniteltuja MIDRAS-järjestelmän osia. Varsinainen rekisteriaineistojen tutkimusetäkäyttö katsottiin pääasiaksi, joten pilotissa toteutettiin käsityönä sellaiset taustapalvelut, joiden on tarkoitus toimia MIDRAS-järjestelmässä automaattisin, vakioiduin tavoin. Näitä palveluita ovat määrämuotoisista metatiedoista koottu aineistokatalogi, siihen

perustuva keskitetty lupahakemus, käyttäjähallinnon käyttöliittymät, saatavuus- ja kokeiluaineistot ja tutkimusaineistojen ja -menetelmien säilytyspalvelu.

Pilotin ensimmäinen vaihe, tietointegraatio ja etätyöpöydän peruskäyttö toteutettiin CSC:n sisäisesti. Samaan aikaan etsittiin pilotin tutkimusprojektit ja sovittiin pilotoitikäytännöistä niiden viranomaisten kanssa, jotka toimittivat aineistoja näihin projekteihin. Pilotissa mukana olleita viranomaisia olivat Kela, THL ja Tilastokeskus. Koska rekisteriaineistojen lupaprosessi kestää pitkään ja joka tapauksessa ennustamattoman ajan, pilottia varten etsittiin tutkimusprojekteja, jotka olivat jo saaneet käyttöluvan aineistoihinsa. Toinen vaihe, tutkimuskäyttö, toteutettiin syksyllä 2010.

Varsinaisen tutkimustyön ja sen teknisten edellytysten lisäksi pilotointiin liittyi paljon viestintää sekä aineistojen tuottajien että käyttäjien suuntaan. Viranomaiset tutustutettiin MIDRAS-hankkeeseen ja hankkeen yhteydessä toteutettavaan pilottiin, niiden kanssa sovittiin hallinnolliset toimintatavat ja varsinaisen aineistotoimituksen tekniset yksityiskohdat. Lisäksi tutustuttiin tärkeimpien viranomaisten metatietokäytäntöihin.

2.2.3. Pilottijärjestelmä ja sen käyttö

Pilottijärjestelmä oli käytössä 1.7.2010-14.1.2011. Pilotissa toteutettiin suunnitelluista MIDRAS-järjestelmän palveluista seuraavat:

- aineistojen pseudonymisointi
- aineistojen lähetys ja vastaanotto (osin suunnitellun mallin mukaisesti)
- vahva autentikointi (hyödyntäen mikrosirutunnistautumista)
- etätyöpöytä (Windows remote desktop -ohjelmalla)
- käyttäjien etätyöalueet ja analyysi- ja tilasto-ohjelmistot osittain
- aineistojen käyttäminen pilottiympäristön tietokannasta.

Pilottijärjestelmässä toteutettiin kaksi tutkimusprojektia. Tutkimuksiin käytettiin tietoja yhteensä kahdeksasta rekisteristä, kolmelta eri viranomaiselta sekä yhtä tutkijan omaa aineistoa. Useamman viranomaisen aineistoja käyttäneessä pilottiprojektissa tutkijat yhdistelivät aineistot pseudonyymien perusteella etätyöpöydän työkaluilla. Tutkijat käyttivät aineiston käsittelyssä työkaluina SPSS-, R- ja taulukkolaskentaohjelmia sekä MySQL-tietokantaohjelmistoa.

Tietoturva ja tietosuojat ovat olleet keskeisessä asemassa MIDRAS-ympäristön suunnittelussa. Tietojen päätyminen ulkopuolisten käsiin estettiin hyödyntämällä salakirjoitettuja tietoliikenneyhteyksiä, tunnistamalla käyttäjät vahvalla autentikoinnilla ja eristämällä MIDRAS-palvelu asianmukaisesti.

Pilottiympäristö toteutettiin neljällä virtuaalipalvelimella, jotka olivat omassa aliverkossaan CSC:n testiverkossa. Viranomaiset pseudonymisoivat aineistot algoritmisesti käyttäen projektikohtaista salalauseetta, joka annettiin aineistojen toimittajille pilottijärjestelmän aineistonvastaanottopalvelun kautta. Aineistot annettiin tietokannasta tutkijoiden käyttöön siinä muodossa, jossa viranomaiset olivat lähettäneet ne; aineistoja ei pseudonymisoitu toiseen kertaan. Käyttäjien vahvaan tunnistamiseen käytettiin Väestörekisterikeskuksen HST-organisaatiosirukortteja. Sirukorteissa käyttäjä tunnistetaan varmenteella, joka saadaan käyttöön sirukortilta PIN-koodilla.

Pilottijärjestelmän käyttäjäkokemus ei ollut siinä mielessä täysin autenttinen, että käyttäjät olivat hakeneet luvat aineistoihin normaalin nykykäytännön mukaisesti ja aineistojen vastaanotto järjestelmään sekä tulosten siirto pois järjestelmästä toteutettiin hieman eri tavoin kuin varsinaisessa etäkäyttöjärjestelmässä on suunniteltu. Käyttäjille tarjottiin myös erittäin paljon henkilökohtaista neuvontaa, apua ja opastusta; kirjallisia ohjeita ei juuri ollut. Pilottikäyttäjiltä saatiin kuitenkin arvokasta palautetta ja ideoita järjestelmän jatkokehitystä varten ja kokemukset pilottijärjestelmän käytöstä olivat pääosin positiivisia.

3. Selvityshankkeen aikana opittuja asioita

Tässä luvussa vedetään yhteen palaute, jota on kerätty hankkeen aikana MIDRAS-toimintamallin ja -järjestelmän kehittämisen tueksi. Tutkijoiden kommentteja on kerätty muun muassa tutkijapaneeleissa. Pilottijärjestelmän käyttäjiltä on kerätty palautetta käyttäjätutkimuksen avulla. Aineistoja ylläpitävien viranomaisten edustajat ovat myös kommentoineet etäkäyttöjärjestelmän toteutustapaa hankkeen edetessä.

3.1. Toimijoiden tahtotila

”Olisi kauhean kätevää jos yhdestä palvelusta pääsisi moniin aineistoihin... Jos etäkäyttösteemiin saisi yhdistettyä useita rekisterejä se olisi mahtavaa! Jos kaikki tiedot olisi saatavilla samasta paikasta ja jos jotenkin pääsisi käsiksi ilman kamalan pitkää luparumbaa. Siinä olisi todella hienot mahdollisuudet tutkijalle!” (pilottijärjestelmän käyttäjä)

Projektissa kerättyjen kokemusten perusteella eri toimijoiden mielestä Suomeen halutaan saada yksi toimiva rekisteritietojen etäkäyttöjärjestelmä. Tutkijoiden mielestä etäkäyttöjärjestelmän tulee tarjota kokonaisvaltainen ratkaisu rekisteritutkimuksen tekemiseen. Viranomaiset ovat suhtautuneet varovaisen myönteisesti siihen, että aineistojen välitys ja käyttö MIDRAS-järjestelmän kautta toteutuisivat, sillä se voisi helpottaa myös viranomaisten työtä ja parantaa palvelun laatua. Viranomaiset ovat ymmärtäneet, että teknisten käytäntöjen lisäksi pitää muuttaa toimintamalleja ja prosesseja, jotta rekisteritietojen saatavuus

tutkimukseen paranee. Tunnisteellisia henkilötietoja tutkimustarkoituksiin luovuttavat viranomaiset ovat kiinnostuneita järjestelmästä myös tietosuojaparannusten takia.

MIDRAS-järjestelmää kommentoineet tutkijat ovat osoittaneet vahvaa kiinnostusta järjestelmää kohtaan ja innostusta sen mahdollisuuksista. Pilottikäyttäjät kokivat järjestelmän tutkijoita hyödyttäväksi palveluksi, joka voisi helpottaa ja monipuolistaa rekisteritutkimuksen tekoa ja parantaa sen laatua. Tutkijat ovat kuitenkin myös ilmaisseet huolensa järjestelmän mahdollisista kielteisistä vaikutuksista, kuten siitä, lisäkö järjestelmä kustannuksia, tutkijan työtä ja viranomaisten työkuormaa ja toimitusaikoja.

Tutkijat ovat toistuvasti korostaneet, että oheispalvelut (aineistokatalogi, keskitetty lupahakemus) ovat vähintään yhtä tärkeitä etäkäyttöjärjestelmän osia kuin etätyöpöytä. Tutkijat ovat erityisesti toivoneet, että aineistojen hakemisprosessi suoraviivaistuu ja helpottuu etäkäyttöpalvelun avulla. Avoin selailu- ja tutustumismahdollisuus rekisteriaineistoihin taas tukisi tutkimuksen suunnitteluvaihetta.

Tutkijoiden mukaan ideaalinen etäkäyttöpalvelu olisi helppokäyttöinen, joustava, kattava (koskien niin tarjottuja aineistoja, metatietoja kuin ohjelmia), vuorovaikutteinen, nopea ja tehokas. Tutkijat ovat painottaneet erityisesti toimintavarmuutta ja sitä, että aineistojen käsittely ei saa hankaloitua etätyöpöydän, käyttöliittymän, suorituskyvyn tai yhteysongelmien vuoksi.

3.2. Aineistojen tuottaminen ja toimittaminen

Jotta MIDRAS-toimintamalli saadaan toteutetuksi ja käyttöön, viranomaisten pitää analysoida ja muuttaa prosessejaan ja toimintatapojaan. Tämä koskee niin aineistojen tuottamista, toimenkuvia ja vastuita kuin tutkijoiden kanssa toimimistakin. Rekisteriasiantuntijoiden ja tutkijoiden välisen yhteyden ylläpitäminen on keskeistä etäkäyttöjärjestelmän kehittämisessä, jotta olemassa oleva tietotaito aineistojen tuottamisessa saadaan hyödynnettyä. Viranomaiset painottavat, että heidän aineistoasiantuntijoillaan on sellaista osaamista aineistojen tulkinnan ja muodostamisen suhteen, johon tutkijoiden tiedot tai taidot eivät välttämättä riitä.

Aineistojen tuottamisesta ja toimittamisesta etäkäyttöjärjestelmään on saatu monenlaisia näkemyksiä. Osa viranomaisista olisi kiinnostuneita toimittamaan rekisterinsä kokonaisuudessaan etäkäyttöjärjestelmään, kun taas osa viranomaisista toimittaisi ainoastaan tutkimuskohtaisia tutkimusaineistoja käytettäväksi etäkäyttöjärjestelmän kautta. Tutkijat toivovat mahdollisimman laajojen aineistojen saamista MIDRAS-järjestelmään.

Koska tutkimusaineistojen esikäsittely on työläs prosessi, tutkijat ovat ehdottaneet, että MIDRAS-järjestelmässä tarjottaisiin uudelleen käytettäväksi eri tutkimusprojektien muodostamia ja muokkaamia

tutkimusaineistoja. Vähintään toivotaan, että aineistojen esikäsittelyssä käytetyt työkalut ja ohjelmat tallennettaisiin MIDRAS-järjestelmään.

3.3. Teknisiä ratkaisuja

Projektissa ei mm. tietoturvasyistä selvitetty yksityiskohtaisesti rekisterien teknisiä tallennusmuotoja ja tietokanta-arkkitehtuuria. On kuitenkin selvää, että viranomaisten valmiustaso liittyä rajapintojen kautta MIDRAS-järjestelmään on vielä hyvin vähäinen. Siksi MIDRAS-järjestelmän aineistojen rajapintapalveluiden käyttöönotossa on edettävä vaiheittain. Tutkijalle luovutettavien aineistojen siirtäminen järjestelmään etukäteen (ns. push-malli) nähtiin toteuttamiskelpoiseksi vaihtoehdoksi jo nykytilanteessa, sen sijaan tietojen kysely rajapintojen kautta (ns. pull-malli) viranomaisten tietokannoista vaatii viranomaisilta kehittämistyötä.

Projektin lähtökohta oli, että MIDRAS-järjestelmästä rakennetaan federoitu ratkaisu käyttäen myös federoituja tietokantoja (esim. IBM InfoSphere, Oracle DB link). Federoidut eli virtuaaliset tietokannat ovat kuitenkin tällä hetkellä vielä kypsymättömiä ratkaisuja, joiden rajapinnat ovat vakiintumattomat ja suorituskyky vaihtelevaa.

3.4. Lainsäädännölliset kysymykset

Viranomaisten lakiasiantuntijoiden sekä tietosuojavaltuutetun mukaan etäkäyttöjärjestelmä ei muuta rekisteriaineistojen luovutuksen lakisääteisiä vaatimuksia. Tutkijoille annettu käyttöoikeus yksikkötasoiseen aineistoon rinnastetaan aineiston perinteiseen luovutukseen. Tutkijan katsotaan olevan tutkimusaineistonsa rekisterinpitäjä siitä huolimatta, että hän ei saa siirrettyä aineistoa etäkäyttöjärjestelmästä omaan haltuunsa. Aineiston ylläpidon ja säilytyksen tietoturvallisuus ovat tutkijan vastuulla. Myöskään aineistojen pseudonymisointi (eli suorien tunnistetietojen korvaaminen surrogaateilla) ei sinällään muuta aineistojen luovutuksen lakisääteisiä vaatimuksia.

Rekisteriaineistoja luovuttavien viranomaisten tulee aineistojen luovutus päätöksiä tehdessään varmistaa, että tutkimuksen henkilötietojen käsittely täyttää henkilötietolain ja muiden lakien vaatimukset. Luovutus päätökseen voidaan liittää ehtoja, mutta päätökseen perustuvan aineiston luovutuksen jälkeen vastuu aineiston asianmukaisesta käsittelystä siirtyy aineiston saajalle, eli tutkijalle. Koska tutkija on MIDRAS-järjestelmässä tutkimusaineistonsa rekisterinpitäjä ja vastuussa aineistonsa käsittelyn tietoturvallisuudesta, etäkäyttöjärjestelmässä on organisoitava tutkijoiden ja etäkäyttöjärjestelmän ylläpitäjän väliset oikeudet ja vastuut joko sopimuksin tai uuden lainsäädännön kautta. Lisäksi järjestelmässä tulee varata viranomaisille mahdollisuus tarkistaa, että tutkimusten tulokset eivät sisällä salassapitosopimusta loukkaavia tietoja.

Jotta etäkäyttäjärjestelmä voisi palvella tutkimusta täysimääräisesti, tilastolain alaisia aineistoja olisi hyvä saada tutkimuskäyttöön nykyistä astetta lievemmällä suojaustasolla. Jos kansallisen tilastolain luovutussäännös yhdenmukaistettaisiin EU-tilastoasetuksen (EY 223/2009) 23 artiklan kanssa, käyttöluopa voitaisiin tieteellisiä tarkoituksia varten myöntää myös sellaisiin salassa pidettäviin tietoihin, joiden perusteella tilastoyksiköt ovat välillisesti mutteivät suoraan tunnistettavissa. Erityisesti yhteiskuntatieteilijät ovat painottaneet tilastolain muuttamisen tarpeellisuutta, jotta etäkäyttäjärjestelmän avulla voitaisiin käyttää laadukkaita ja laajoja aineistoja. Tutkijat ovat myös ehdottaneet, että tilastolakiin lisättäisiin Tilastokeskuksen tehtäväksi/velvoitteeksi tuottaa ja ylläpitää tutkimusaineistoja.

Viranomaisten on lähivuosien kuluessa varauduttava nykyistä tiukempiin tietoturva- ja tietosuojavaatimuksiin luovuttaessaan tietoja henkilörekistereistä. Lokakuussa 2010 voimaan tullut valtionhallinnon tietoturva-asetus (681/2010) edellyttää, että viiden vuoden siirtymäkauden jälkeen valtakunnalliseen henkilörekisteriin talletettuja arkaluonteisia henkilötietoja voidaan siirtää vain sellaisessa viranomaisen tietoverkossa, jonka käyttö on rajoitettu, ja viranomainen on varmistanut, että tietoverkko ja tietojenkäsittely kokonaisuudessaan täyttävät tavanomaisesti sovellettavan korotetun tietoturvallisuuden tason vaatimukset. Myös valmisteilla olevat EU-tietosuojadirektiivin muutokset tuonevat lähivuosina viranomaisille lisää henkilötietojen luovutukseen liittyviä velvotteita.

3.5. Kustannukset

Sekä tutkijat että rekistereitä ylläpitävät viranomaiset ovat esittäneet huolensa etäkäyttäjärjestelmän kustannusvaikutuksista. Tutkijat ovat olleet huolissaan, että tutkimusaineistojen hinnat voivat nousta etäkäyttäjärjestelmän käyttöönoton seurauksena. Viranomaiset puolestaan ovat nostaneet esille aineistojen kuvaamisen ja muokkaamisen vaatimat resurssit, joihin ei tällä hetkellä ole rahoitusta tai tarvittavaa työvoimaa.

Tutkijat ovat korostaneet, että olisi kaikkien kannalta edullisempaa, jos rekistereistä tehtäisiin moneen tutkimukseen soveltuvia valmisaineistoja. Tutkijat kokevat, että jos tutkija tekee järjestelmän avulla enemmän itse, sen tulisi näkyä myös aineistojen hinnassa.

Monissa etäkäyttäjärjestelmää ylläpitävissä maissa on tarjottu valtion tukea etäkäyttäjärjestelmän kaltaiselle tutkimusinfrastruktuurille ja sen kehittämiseksi. Esimerkiksi Ruotsissa ja Tanskassa tukea on saatu rekisteriaineistojen metatietojen kuvailuun. (Kansainvälisen selvityksen tuloksia on esitelty tarkemmin hankkeen yhteydessä julkaistavassa selvitysraportissa).

MIDRAS-järjestelmän kustannusarvio on esitelty liitteessä 3.

3.6. Yhteistyön ja koordinaation tärkeys

Yhteistyön vahvistaminen kaikilla tasoilla, myös kansainvälisesti, on tärkeää niin MIDRAS-toimintamallin toteuttamiseksi kuin päällekkäisen työn estämiseksi. Koska MIDRAS toteutetaan yhteistyöhön perustuvana järjestelmänä, viranomaisten mielestä on tärkeää määrittää, miten tehtävät MIDRAS-toimintamallissa on jaettu ja miten eri toimijoiden välinen yhteistoiminta on järjestetty.

Metatietoja tulee kehittää yhteistyössä erilaisten sidosryhmien kanssa. Tärkeitä kehitysprojekteja, joiden kanssa metatietotyötä pitäisi koordinoita, ovat JHS 143 -suosituksen päivitys, arkistolaitoksen tietokantakuvausmäärittely ja DDI3-standardi. Tavoitteena on sopia tavat, joilla metatietoja voidaan kerätä, yhdistää ja replikoida.

Rajapintojen määrittelyssä valtiovarainministeriön asettaman Julkishallinnon perustietovarantojen rajapinnat (Pera) -työryhmän tavoitteena on yhtenäistää käytännöt, joilla julkisen sektorin perustietovarannot tarjotaan tietojen tarvitsijoille. Työryhmän linjaukset tulevat vaikuttamaan MIDRAS-toimintamallin toteutukseen ja aikatauluun.

3.7. MIDRAS-järjestelmän laajentamismahdollisuudet

Tulevaisuudessa MIDRAS-järjestelmällä on mahdollisuus toimia paitsi tutkimusaineistojen jakelu- ja analysointialustana, myös aineistojen tai työmenetelmien talletusalustana ja ohjelmistojen ylläpidon keskitysalustana. Viranomaiset ovat esittäneet kiinnostustaan järjestelmän käyttöön esimerkiksi viranomaisten omien selvitysten teossa. Pilottikäyttäjät näkivät jatkokehityksen mahdollisuuksia erityisesti järjestelmän vuorovaikutteisuuden vahvistamisen kautta.

Selvityshankkeen aikana käydyissä keskusteluissa nostettiin painokkaasti esiin myös eri tietoinfrastruktuurien yhteentoimivuus. Esimerkiksi biopankkitutkimuksen aineistoja ja paikkatietoaineistoja on tärkeää päästä yhdistämään rekisteriaineistoihin.

3.8. Käyttäjä- ja järjestelmävaatimukset ja suunnittelua ohjaavat periaatteet

Seuraavassa taulukossa on tiivistelmä selvityshankkeen aikana opituista ja korostuneista MIDRAS-järjestelmän suunnittelua ohjaavista periaatteista. Perusvaatimukset on luokiteltu kolmitasoisella tärkeysluokituksella:

- 1: järjestelmä ei toteuta tarkoitustaan, mikäli vaatimusta ei täytetä.
- 2: järjestelmän hyödyllisyys vaarantuu, mikäli vaatimusta ei täytetä.
- 3: vaatimus tekee järjestelmästä hyödyllisemmän.

Tutkijoiden keskeinen vaatimus on, että järjestelmä helpottaa rekisteritutkimuksen tekemistä. Viranomaisten keskeisimmät vaatimukset ovat, ettei järjestelmä hankaloita resurssitilannetta ja etteivät tietoturva ja -suoja heikkene.

Taulukko 4 Kooste etäkäyttöjärjestelmän suunnittelua ohjaavista periaatteista

Vaatimuksia tutkijoiden näkökulmasta		Tärkeys
Toimintavarmuus	Järjestelmän ja yhteyksien tulee toimia varmasti ja luotettavasti. Työtä ei saa hukkaa eikä työskentely vaikeutua järjestelmän ongelmien takia.	1
Kattavuus aineistojen ja niiden kuvailujen suhteen	Järjestelmän kautta pitää saada mahdollisimman kattavasti aineistoja. Aineistojen kuvailutietojen tulee olla kattavasti, helposti ja nopeasti saatavilla.	1
Lupahakemusprosessin selkeys	Järjestelmän tulee sisältää myös keskitetty sähköinen lupahakemuspalvelu, joka helpottaa ja tehostaa lupahakemusprosesseja.	2
Joustavuus	Järjestelmän tulee olla käytettävissä helposti, useammalla koneella, ajasta ja paikasta riippumatta.	2
Helppokäyttöisyys	Järjestelmä ei saa vaatia useita tai hankalia asennuksia ja hankintoja. Mahdolliset yhteensopivuus- ja palomuuriongelmat tulee selvittää ja tiedottaa etukäteen. Järjestelmän tukee tukea ja opastaa käyttäjää niin käyttöönotossa kuin käytön eri vaiheissa.	2
Kattavuus käytössä olevien ohjelmistojen ja työkalujen suhteen	Koska aineistoja ei saa ulos järjestelmästä ja kaikki käsittely tapahtuu järjestelmän sisällä, järjestelmän tulee sisältää kaikki aineistojen analysoinnissa tarvittavat ohjelmistot ja työkalut, sekä tarjota keino niiden ja tutkijoiden omien aineistojen lisäämiseksi.	2
Tehokkuus ja tuloksellisuus	Työskentelyn tulee olla sujuvaa ja palvelinten tulee olla tehokkaita ja nopeita.	2
Hinnoittelun kilpailukykyisyys	Järjestelmän käytön hinnoittelun tulee olla perusteltua, eikä se saa johtaa ”tuplamaksuihin”. Hinnoittelun pitää rohkaista käyttämään järjestelmää.	2
Vuorovaikutteisuus	Järjestelmän tulee tarjota lisäarvoa nykyiseen työskentelyyn verrattuna edistämällä sekä projektin sisäistä että laajemminkin tutkijoiden ja viranomaisten välistä vuorovaikutusta.	3
Vaatimuksia rekistereitä ylläpitävien viranomaisten näkökulmasta		Tärkeys
Tietoturvallisuus	Rekisteritietojen tulee olla vain käyttöluvan saaneiden hallussa.	1
Mukautuvuus	Eri lakien erityisvaatimukset aineistonluovutuksiin on otettava huomioon.	1
Yhteensopivuus	Järjestelmän pitää pystyä vastaanottamaan aineistoa riippumatta viranomaisen tietoteknisistä ratkaisuista ja tasosta.	2
Vuorovaikutteisuus	Järjestelmän tulee edistää tutkijoiden ja viranomaisten vuorovaikutusta.	2
Kustannustehokkuus	Aineistojen toimittaminen järjestelmän välityksellä ei saa olla vaikeampaa tai enemmän kustannuksia vaativaa kuin tällä hetkellä. Pitkällä tähtäimellä järjestelmän tulee johtaa kustannussäästöihin.	2
Rahoitus	Järjestelmän toteuttamisesta, ylläpidosta sekä oheispalveluiden tuottamisesta ei saa koitua lisäkustannuksia rekistereitä ylläpitäville viranomaisille.	2

4. Selvityshankkeen suositukset

Selvityshankkeen tuloksena suositamme, että rekisteritutkimuksen haasteita ratkotaan rakentamalla rekisteriaineistojen tai muiden luvanvaraisten aineistojen tutkimuskäyttöön yhtenäinen MIDRAS-toimintamalli, joka perustuu toimintamallia tukevan MIDRAS-etäkäyttöjärjestelmän palveluille. MIDRAS-toimintamallissa tutkimusprosessin eri vaiheissa käytetään MIDRAS-järjestelmän tarjoamia tietoteknisiä ratkaisuja ja kaikissa aineistojen ja aineistokuvailujen käsittelyn vaiheissa noudatetaan vakioituja toimintatapoja, jotta tutkimusprosessi ja sen oheisprosessit ovat mahdollisimman nopeita, sujuvia ja kustannustehokkaita. Suositukset ovat vahvasti liitoksissa laajempiin julkisen hallinnon tietohallinnon kehitystavoitteisiin. Tässä osiossa kuvataan rekisteritietojen etäkäyttöä tukeva MIDRAS-toimintamalli sellaisena, kuin se suositetaan toteutettavaksi.

4.1. MIDRAS-visio ja MIDRAS-toimintamallin yleiskuvaus

Visio

Kansallisen etäkäyttöjärjestelmän kautta viranomaisten hallussa olevat, luvanvaraiset, yksikötasoiset aineistot ovat käytettävissä kattavasti, turvallisesti, kustannustehokkaasti ja helposti tutkimukseen.

MIDRAS-toimintamallin keskiössä on etäkäyttöjärjestelmä, jonka kautta viranomaisrekistereistä luovutetaan tutkimuskäyttöön yksikötason tutkimusaineistoja ja jossa näiden aineistojen analyysit suoritetaan. Käytännössä järjestelmä soveltuu myös yleiskäyttöiseksi alustaksi, jolla käyttöluvan saaneet voivat tutkia ja analysoida käyttöluvan mukaisia aineistoja saamatta niitä omaan haltuunsa.

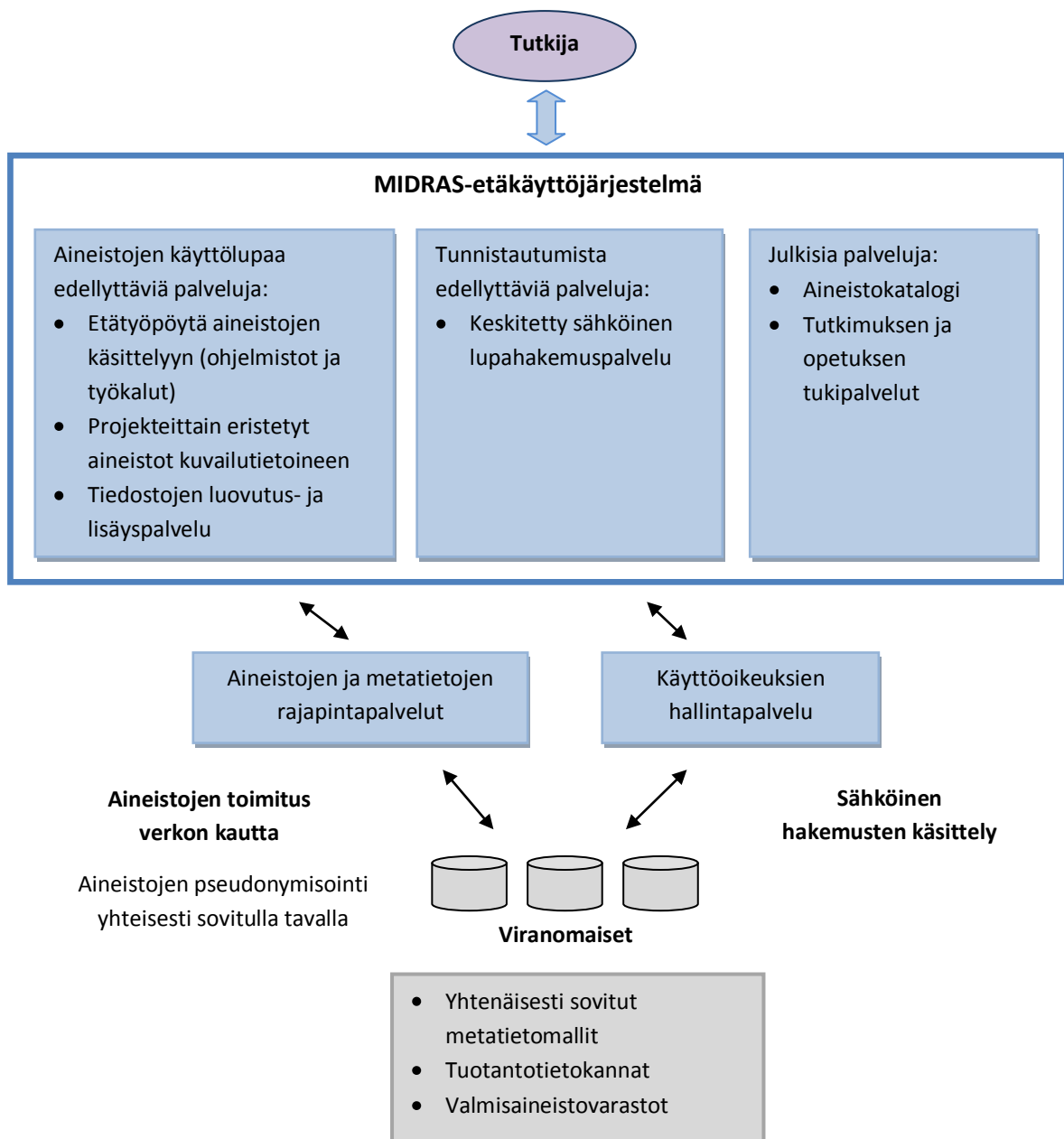
Tutkijan kannalta ehdotetun MIDRAS-järjestelmän keskeisimpiä palveluita ovat:

1. Tutkimusaineistojen käsittelyyn ja analysointiin tarkoitettu eristetty etätyöpöytäympäristö palveluineen. Etäkäyttöjärjestelmässä tutkija ei saa tutkimusaineistoa haltuunsa, vaan tutkija saa aineiston käyttöönsä etätyöpöydälle, johon hän voi kirjautua omalta työpisteeltään.
2. Tutkimuslupien hakemista ja tutkimusten suunnittelua helpottava aineistokatalogi ja sille perustuva keskitetty lupahakemuspalvelu. Aineistokatalogi on www-palvelu, johon kerätään mahdollisimman yhteismitallisia metatietoja järjestelmän kautta saatavilla olevista aineistoista.

Aineistoja tuottavan viranomaisen kannalta keskeisimpiä palveluita ovat:

1. Tietojen siirto- ja käyttörajapinta MIDRAS-ylläpito-organisaation ja viranomaisen välillä. Tietojen käyttörajapinnan kautta kulkevat järjestelmän aineistot, metatiedot ja muut tiedot.
2. Käyttöoikeuksien hallinnointityökalu. Käyttöoikeuksien hallinnointityökalu sallii aineistojen ylläpitäjien myöntää käyttölupia sekä hallinnoida ja tarkastella, millä tutkimusprojekteilla ja keillä tutkijoilla on pääsy mihinkin aineistoon.

Kuva 6 Kuvaus MIDRAS-toimintamallista ja tarjottavista palveluista



4.2. MIDRAS-järjestelmän palvelut

MIDRAS-järjestelmässä on kuusi toisistaan erottuvaa palvelukokonaisuutta, jotka on kuvattu taulukossa 5. Kaikki MIDRAS-järjestelmän palvelut ovat automatisoituja tietoteknisiä palveluita, joissa MIDRAS-hallinto ylläpitää palveluita toteuttavia järjestelmiä ja antaa ohjeistusta, tukea ja työvälineitä palveluiden käyttöön. Palvelut on ryhmitelty kokonaisuuksiksi, jotka ovat tutkimuksen suunnitteluvaiheeseen kuuluvat aineistokatalogi ja käyttöluopapalvelut, tutkimuksen toteutusvaiheeseen kuuluvat ohjelmistopalvelu (etätyöpöytä) ja tarkastettu tiedostonsiirtopalvelu sekä näitä kaikkia kokonaisuutena tukevat rajapintapalvelut ja tukipalvelut.

Taulukko 5 MIDRAS-järjestelmän palvelut

Palvelukokonaisuus	Keskeinen sisältö	Käyttäjät	Hyödyt	Käyttötapa	MIDRAS-hallinnon rooli
Aineistokatalogi	MIDRAS-järjestelmän kautta saatavilla olevien aineistojen kuvaukset julkaisu- ja ohjeviitteineen	Tutkijat	Tutkimusten suunnittelu, laadukkaat käyttölujarahakemukset, aineistojen käytön tuki	Selailu ja tekstihaut WWW-selaimella	Katalogin tuottaminen viranomaisten aineistokuvailuista
Käyttölupien haku- ja hallintapalvelut	Aineistojen käyttöoikeuksien haku, myöntö ja tarkastelu	Tutkijat, viranomaiset	Hakuprosessin helppous, nopeus ja läpinäkyvyys; käyttölupien seuranta	WWW-lomakkeet	Lupapalvelun ylläpito ja kehittäminen; Käyttövaltuustietokannan varastointi ja käyttöliittymät
Aineistojen ja metatietojen rajapintapalvelut	Ohjeet ja menetelmät tietojen pseudonymisoinniseksi ja välittämiseksi MIDRAS-järjestelmään	Viranomaiset	Aineistojen ja metatietojen yhteensopivuus, tietotoimitusten nopeus ja vaivattomuus	Ohjeet ja työkalut haettavissa WWW-sivuilla, web service -rajapinnat	Rajapintojen ohjeistus ja aputyökalut, aineistojen ja metatietojen välitys
Tutkimus-, raportointi- ja analyysiohjelmistopalvelu	Aineistojen käsittely ja tutkimus	Tutkijat	Monipuolinen tutkimusympäristö, yhteistyöalusta, hyvät tietojenkäsittelyresurssit, tietoturva, tietosuojat	Etätyöpöytä, jolla aineiston käsittelyyn sopivat ohjelmistot	Etätyöpöydän toteutus, yhteisten ohjelmistojen asennukset ja päivitykset
Tutkimuksen ja opetuksen tukipalvelut	Käyttäjätuki, tiedotus- ja keskustelualueet, dokumentti- ja välinevarastot	Tutkijat, viranomaiset	Tietojen, menetelmien ja kokemusten jakaminen ja varastointi, palaute	WWW-keskustelualue, puhelintuki	Tiedotteiden ja ohjeiden julkaisu, yhteydenottoihin vastaaminen, palvelun ylläpito ja kehittäminen
Omien tiedostojen lisäys- ja luovutuspalvelu tarkastuksen kautta	Tutkimustulosten luovutus etätyöpöydältä tutkijalle, omien tutkimusaineistojen liittäminen rekisteriaineistoihin, työkalut ja päivitykset	Tutkijat	Tutkimuksen tulosten raportointi; tutkimusaineistojen, tulosten ja työvälineiden hyödyntäminen useissa ympäristöissä	Hakemisto etätyöpöydällä, WWW-palvelu	Tiedostojen välitys ja ohjeistus

4.3. MIDRAS-toimintamallin prosessit ja vastuut

MIDRAS-toimintamallissa pitää määrittää selkeästi, kenen vastuulle kuuluu mikäkin tehtävä ja miten eri toimijoiden välinen yhteistoiminta on järjestetty. Vastuiden jakautumista suunniteltaessa on noudatettu seuraavia periaatteita:

1. Tehtävät jaetaan karkeasti toimijoiden ydintehtävän mukaan.
2. Kehitystehtävät jaetaan tarvittavan osaamisen mukaan.
3. Lain asettamat velvoitteet eri osapuolille täyttyvät.
4. Osapuolet pystyvät useimmissa tapauksissa muuttamaan sisäistä toimintaprosessiaan tarvitsematta muuttaa yhteistyökäytäntöjä.

Seuraavassa kuvataan yleisellä tasolla ehdotetun MIDRAS-toimintamallin vakioidut toimintatavat eri prosesseissa. Prosessit on luokiteltu aihealueittain ja niihin liittyvät vastuut on eritelty, kun mukana on useampia toimijoita. MIDRAS-hallinnolla viitataan MIDRAS-järjestelmää ylläpitävään tahoon. Viranomaisten lisäksi prosessit voivat tulevaisuudessa koskea myös muita rekisteriaineistoja tuottavia tai hallinnoivia organisaatioita.

4.3.1. Hallinnon prosessit

Viranomaisen liittyminen MIDRAS:iin

MIDRAS-hallinto ja viranomainen määrittävät yhteyshenkilöt, tekevät yhteistyösopimuksen ja antavat toisilleen rajapintapalveluiden käyttötunnukset.

Taulukko 6 Palvelut ja vastuut hallinnon prosesseissa

Käytetyt palvelut	Viranomaisen vastuut	MIDRAS-hallinnon vastuut
Tukifoorumi	Vastuuhenkilöiden nimeäminen, palveluiden käyttöönotto, aineistokuvailut	Sopimus pohjat, palveluiden ja aineistokuvailujen ohjeistus ja työkalut

4.3.2. Valmistelun prosessit

Valmisaineistojen tuottaminen

Viranomainen pitää yllä valmisaineistovarastoa, jonka päivitystä tuotantorekisteristä jatkossa automatisoidaan julkisen hallinnon tietohallinnon kehittämislinjausten mukaisesti. Tietopyynnöt toteutetaan mahdollisuuksien mukaan valmisaineistoista.

Aineistojen kuvailu

Viranomainen tuottaa aineistoistaan MIDRAS-toimintamallin yhtenäisen metatietostandardin mukaiset kuvailut ja lähettää ne MIDRAS-järjestelmään, jossa ne toimivat osana aineistokatalogia ja lupahakemustyökalua.

Taulukko 7 Palvelut ja vastuut valmistelun prosesseissa

Käytetyt palvelut	Viranomaisen vastuut	MIDRAS-hallinnon vastuut
Metatietojen rajapintapalvelut	Aineistojen kuvailu, kuvailujen lähettäminen ja päivittäminen	Kuvailun ohjeistus ja työkalut, kuvailujen vastaanotto ja säilöntä, metatietostandardin kehittäminen yhteistyössä eri osapuolten kanssa

4.3.3. Lupahakemusprosessit

Uuden tutkimuksen perustaminen

Tutkija hakee tutkimuslupaa MIDRAS-käyttölupahakemustyökalulla. Lupahakemuksen perusteella viranomaiset myöntävät luvat ja muodostavat valmisaineistoista tai muusta aineistosta tutkijalle lähetettävät aineistot. Aineistojen muodostamista voidaan automatisoida rakenteisen lupahakemuksen perusteella.

Lupahakemuksen päivittäminen

Tutkija päivittää lupahakemustaan MIDRAS-käyttölupahakemustyökalulla. Viranomaiset käsittelevät muutokset kuten uudet hakemukset.

Taulukko 8 Palvelut ja vastuut lupahakemusproesseissa

Käytetyt palvelut	Viranomaisen vastuut	MIDRAS-hallinnon vastuut	Tutkijan vastuut
Aineistokatalogi, käyttöoikeuksien haku- ja hallintapalvelut, tukifoorumi	Hakemusten tarkastaminen ja käsittely, aineistojen poiminta	Tutkimusprojektin työtilan perustaminen, palveluiden ylläpito, käyttäjätuki	Tutkimuksen suunnittelu, aineistotarpeen ja tutkimuksen kuvailu

4.3.4. Tutkimusaineistoihin liittyvät prosessit

Aineiston pseudonymisointi

Viranomainen korvaa muodostamissaan tutkimusaineistoissa suorat tunnisteet (esim. henkilötunnukset) pseudonyymeilla yhteisesti sovitun mallin mukaisesti ennen aineiston lähettämistä MIDRAS-järjestelmään. Pseudonymisointimallia ei kuitenkaan voi käyttää nykyisen tilastolain alaisiin aineistoihin, vaan tilastolain edellyttämässä tapauksissa kaikki tutkimusaineistot lähetetään tilastoviranomaiselle yhdisteltäviksi ennen MIDRAS-järjestelmään lähettämistä.

Tutkimusjoukon määrittäminen

Jos tutkittavien asioiden joukon määrittämiskriteeri on monimutkainen, tutkija määrittää etätyöpöydällä erillisen poiminta-aineiston perusteella ne yksilöt/yksiköt, joiden tietoja tarvitaan

tutkimuksessa. Tällaisessa tapauksessa tutkimusjoukko annetaan pseudonyymilistana viranomaisille poimintaa varten rajapintojen kautta.

Aineiston toimittaminen

Viranomaisen toimittaa tutkimukselle aineistot, niiden kuvailut ja molempien päivitykset joko lähettämällä aineistot MIDRAS-järjestelmän rajapinnalla (ns. push-malli) tai tarjoamalla rajapinnan, jolla MIDRAS-järjestelmä voi hakea aineiston tarpeen vaatiessa (ns. pull-malli).

Omien aineistojen lisäys

Jos tutkijalla on oma aineisto, jossa tutkituilta on saatu lupa tietojen yhdistelemiseksi rekisteritietoihin ja jonka käyttö on kuvailtu lupahakemuksessa, MIDRAS-hallinto lisää tutkijan aineiston pseudonymisoinnina etätyöpöydälle. Kuitenkin tilastolain edellyttämässä tapauksissa kaikki tutkimusaineistot lähetetään tilastoviranomaiselle yhdisteltäviksi ennen MIDRAS-järjestelmään lähettämistä.

Taulukko 9 Palvelut ja vastuut tutkimusaineistoihin liittyvissä prosesseissa

Käytetyt palvelut	Viranomaisen vastuut	MIDRAS-hallinnon vastuut	Tutkijan vastuut
Aineistojen rajapintapalvelut, tutkimuspalvelut	Pseudonymisointi, aineistojen valmistelu ja tarjoaminen MIDRAS-järjestelmään	Pseudonymisoinnin ja aineistotoimitusten ohjeistus, työkalut ja rajapinnat, tutkijan omien aineistojen pseudonymisointi, aineiston tarjoaminen tutkimuksen työtilaan	Tutkimusjoukon määrittäminen Omien aineistojen toimittaminen

4.3.5. Tutkimuksen prosessit

Aineistojen käsittely

Tutkija kirjautuu etätyöpöydälle, yhdistelee eri viranomaisilta saamansa aineistot pseudonyymien perusteella ja käsittelee ja analysoi aineistojaan työpöydällä olevilla ohjelmistoilla.

Tulosten luovutus

Tutkija siirtää tai kopioi tutkimuksensa tulokset etätyöpöydällä tulostenluovutusalueelle ja hakee ne tulostenluovutuspalvelun kautta omalle koneelleen.

Tutkimuksen jälkeiset toimet

Tutkija kuvailee aineistonsa ja työkalunsa ja julkaisee työkalut kuvailuineen MIDRAS-tukifoorumilla. Tutkimusaineistot säilytetään jatkokäyttöä varten (käyttöluvalla) ja lisätään aineistokatalogiin.

Taulukko 10 Palvelut ja vastuut tutkimuksen prosesseissa

Käytetyt palvelut	Viranomaisen vastuut	MIDRAS-hallinnon vastuut	Tutkijan vastuut
Tutkimus-, raportointi- ja analyysiohjelmanpalvelu, tukipalvelut, aineistokatalogi, omien tiedostojen lisäys- ja luovutuspalvelu	Aineistokuvailujen tarkistus tarpeiden mukaan, aineistojen korjaukset ja päivitykset	Etätyöpöydän ylläpito, tietoturva ja käyttäjätuki	Tutkimus, tutkimusetiikan noudattaminen, ohjelmistojen käyttö, tutkimuksen aineistojen ja työvälineiden kuvailu jatkokäyttöä varten

4.4. MIDRAS-hallintomalli ja -ylläpito

Selvityshankkeen tavoitteena on esitellä hallintomallivaihtoehtoja, jotka eri rahoittajat ja viranomaiset voivat hyväksyä. Hallintomalliksi ehdotetaan alustavasti kaksitasoista mallia, jossa toimintaa ohjaava strateginen taso ja operatiivinen taso erotetaan toisistaan.

Strategisella tasolla MIDRAS-hallinnosta vastaa johtoryhmä, joka linjaa kehittämistä ja laatii ehdotukset ja suosituksen rekisteritietojen tutkimuskäytön edistämiseksi. Johtoryhmässä ovat edustettuina valtiovarainministeriö, opetus- ja kulttuuriministeriö ja muita ministeriöitä, Tutkimus- ja innovaationeuvosto, tutkimuslaitokset, yliopistot ja rekisteriaineistoja hallinnoivat viranomaiset sekä operatiivisesta toiminnasta vastaavat tahot. Johtoryhmä asettaa toimintansa tueksi sidosryhmäkonsortion, joka seuraa järjestelmän toimintaa, tekee kehitysehdotuksia ja edistää yhteistoimintaa. Sidoryhmäkonsortion jäseniä ovat järjestelmän käyttäjät eli tutkijat ja aineistoja toimittavien viranomaisten edustajat.

Operatiivisella tasolla MIDRAS-hallinto on selkeintä jakaa yleisestä hallinnosta ja tietoteknisistä palveluista vastaaviin tahoihin. Hallinnollinen ylläpitotaho kantaa kokonaisvastuun järjestelmän kehittämisestä ja toiminnasta ja sopimuksista. Tietotekninen ylläpitotaho vastaa tietojärjestelmän toimivuudesta, kehitystyöstä ja käyttäjätuesta. Etäkäyttöjärjestelmän infrastruktuurin rakentaminen ja ylläpito vaativat korkeaa tietotekniikan asiantuntemusta. Järjestelmän toimivuuden kannalta toimivat yhteydet, tarvittavat työskentely- ja tallennustilat sekä asiantunteva tietotekninen opastus ongelmatilanteissa ovat ensisijaisen tärkeitä.

Hallinnollinen ylläpito voidaan järjestää esimerkiksi jommallakummalla seuraavista tavoista:

1. *Erillinen uusi viranomainen / virasto:*

Perustetaan uusi erillinen organisaatio, jonka lakisääteinen tehtävä olisi tieteellistä tutkimusta ja viranomaisten selvitystyötä palvelevan etäkäyttöjärjestelmän ylläpito. Erillisen viraston malli edellyttää vähintään asetustasoista säädöstä sen perustamisesta. Uusi virasto voisi toimia

ministeriön tai jonkin olemassa olevan viraston yhteydessä ja tukeutua tämän isäntäorganisaation hallintopalveluihin (tukiprosessit, henkilöstö- ja taloushallinto).

2. *Olemassa olevan viranomaisen yhteyteen perustettu yksikkö:*

Sisäinen palvelukeskusmalli, jossa MIDRAS-palvelu toimisi jonkun olemassa olevan viranomaisen sisällä omana hallinnollisena yksikkönään. Isäntäviranomaisen tulisi olla tehtäviltään laaja-alainen ja kuulua MIDRAS-toimintamallin sidosryhmiin.

Tietosuojavaltuutetun kanta on, että MIDRAS-hallinnolla on hyvä olla erityistehtävä hallinnoida pseudonymisointiavaimia. Näin toiminnan oikeutuksesta ei ole epäselvyyttä. Suositamme, että hallintomallista tehdään tarkemmat ehdotukset Tilastokeskuksen ja Midras-projektin etäkäyttöjärjestelmien yhteiskehittämistä valmisteleivassa työryhmässä (Valtiovarainministeriö, 2010e).

4.5. MIDRAS-toimintamallin vaatimat toimet

Taulukossa 11 kuvataan ne edellytykset, joita MIDRAS-toimintamallin ja etäkäyttöjärjestelmän toteuttaminen vaatii. Edellytykset on luokiteltu vaativuuden mukaan: välttämättömät edellytykset on täytettävä, jotta MIDRAS-toimintamalli voidaan ottaa käyttöön; muut edellytykset ovat tärkeitä jollekin MIDRAS-toimintamallin osa-alueelle, mutta MIDRAS-toimintamalli voidaan ottaa käyttöön, vaikkei edellytys täytyisikään. Hallinnollisia edellytyksiä ovat rahoitus, hallinnon organisointi ja lupahakemusten käsittelyprosessi; teknisiä edellytyksiä ovat MIDRAS-järjestelmän, valmisaineistojen ja rajapintapalveluiden toteutus sekä ohjeistus; ja lainsäädännöllisiä edellytyksiä ovat tilastolain muutos, tutkijat huomioon ottava laintulkinta ja MIDRAS-hallinnon toimivalta.

Taulukko 11 MIDRAS-toimintamallin edellytykset

Edellytys	Tavoitetila	Suositus	Käytännön toimet
Hallinnolliset edellytykset			
MIDRAS-järjestelmän rahoitus (välttämätön)	MIDRAS-järjestelmällä on vakaat toimintaedellytykset, jotka rohkaisevat käyttämään järjestelmää.	MIDRAS-järjestelmästä tehdään osa kansallista tutkimusinfrastruktuuria ja sen keskeiset osat pidetään käyttäjille ilmaisina.	Budjetoidaan MIDRAS-järjestelmälle tarvittavat resurssit. Viranomaisille myönnetään resursseja aineistohallinnan saamiseksi MIDRAS-järjestelmän tasolle.
MIDRAS-hallinnon organisointi (välttämätön)	MIDRAS-järjestelmällä on hallinnollinen ja tietotekninen vastuutaho, joilla on asiantuntemukseltaan valmius tehtäviensä hoitamiseen.	Valtuutetaan sopivat toimijat vastaamaan MIDRAS-hallinnosta ja MIDRAS-järjestelmän tietoteknisestä toteutuksesta ja ylläpidosta.	Rahoittaja(t) sopivat MIDRAS-järjestelmän hallinnoinnista. MIDRAS-hallinnon toimivallasta säädetään lailla tai asetuksella.
Lupahakemusten tehokas käsittely (välttämätön)	Tutkijat tietävät, mikä on tarjolla olevien aineistojen rakenne. Lupahakemuksissa on konekäsitteltävä kuvaus aineistotarpeesta. Useimpien lupahakemusten aineistotoimitus pystytään toteuttamaan automaattisesti tarkistetun lupahakemuksen perusteella.	Keskitetty lupahakemus rakennetaan sellaiseksi, että se perustuu aineistokatalogissa oleviin, tarkkoihin aineistojen rakennekuvailuihin.	Viranomaiset muodostavat prosessin lupahakemusten käsittelyyn, jossa lupahakemuksen tarkistuksen jälkeen käsittely on automaattista. MIDRAS-hallinto tarjoaa tähän apu työkaluja.
Järjestelmän kautta tarjottavien aineistojen kartoittaminen	Järjestelmässä on alusta alkaen mukana tutkimuksen kannalta hyödyllinen aineistovalikoima. Järjestelmään rakentuu kattava ja monipuoliseen tutkimukseen soveltuva aineistotarjonta.	Järjestelmän aineistovalikoimaa kehitetään aktiivisesti sekä tutkijoiden tarpeiden että viranomaisten tarjonnan pohjalta.	Ministeriöt ohjeistavat viranomaisia kartoittamaan omat aineistonsa. MIDRAS-hallinto neuvottelee uusien aineistojen lisäämisestä.
Ohjaus ja tuki viranomaisille aineistonmuodostus-prosessista	Viranomaisten tietohallinnon prosessi ja automaatio vastaavat MIDRAS-toimintamallin suosituksia.	Tietohallinnon automatisointia vahvistetaan yhteisten julkishallinnon kehittämisperiaatteiden mukaisesti.	Ministeriöt ohjaavat viranomaisia eri ohjausmekanismein.

Edellytys	Tavoitetilä	Suositus	Käytännön toimet
Tekniset edellytykset			
MIDRAS-järjestelmän määrittely, toteutus ja ylläpito (välttämätön)	MIDRAS-järjestelmässä on selkeät, hyvin dokumentoidut ja helposti käyttöön otettavat rajapinnat ja standardit. MIDRAS-järjestelmä kattaa tutkimusprosessin kaikki alueet ja on tietoturvallinen, luotettava järjestelmä.	MIDRAS-järjestelmän rajapinnat, metatietomuodot ja tiedostomuodot määritellään ja toteutetaan. MIDRAS-järjestelmää ja sen käyttöä seurataan ja auditoidaan aktiivisesti.	Toimijat asettavat riittävät henkilöresurssit määrittelyyn, toteutukseen ja ylläpitoon.
Valmisaineistojen tuottaminen ja kuvailu (välttämätön)	Tutkijoille on olemassa selkeä, hyvin dokumentoitu aineistotarjonta, josta on helppoa löytää omaa tutkimusta tukevat aineistot. Viranomaisten valmisaineistojen päivittyminen ja kerääminen on automaattista.	Tuotantorekistereiden sisältö kerätään valmisaineistovarastoon, jossa aineistot on valmiiksi yhdistelty muotoon, josta tietopyynnöt on helppo toteuttaa. Valmisaineistoista tehdään yhteisen metatietostandardin mukaiset kuvailut.	Viranomaiset kartoittavat tutkimusta palvelevat valmisaineistot, rakentavat automaation niiden tuottamiseen ja kuvailevat aineistot. Viranomaisille annetaan rahoitusta ja tukea tarvittavien aineistomuutosten toteuttamiseen.
Viranomaisten rajapintapalveluiden toteutus	Viranomaiset pystyvät vastaamaan aineistokyselyihin automatisoidun, käyttäjät tunnistavan rajapinnan kautta. Tutkijat voivat MIDRAS-etätyöpöydältä tehdä kyselyitä, jotka hakevat ja yhdistelevät eri viranomaisten aineistoja.	Rajapintapalvelut rakennetaan ja niitä pidetään yllä.	Viranomaiset resursoivat rajapintapalveluiden pystytyksen ja ylläpidon. MIDRAS-hallinto tarjoaa tukea ja työkaluja rajapintapalveluiden toteuttamiseen.
Ohjeiden, esimerkkien ja työkalujen tuottaminen	MIDRAS-järjestelmän palveluita on helppo käyttää ja MIDRAS-järjestelmän rajapintoihin ja standardeihin on valmiit työkalut.	MIDRAS-järjestelmään tuotetaan ohjeet, esimerkit ja työkalut.	Osoitetaan riittävät henkilöresurssit ohjeistukseen sekä esimerkkien ja työkalujen tuottamiseen.

Edellytys	Tavoitetila	Suositus	Käytännön toimet
Lainsäädännölliset edellytykset			
Tilastolain muutos	Tilastokeskuksen ja muiden tilastoviranomaisten yksikköaineistoja voidaan käyttää MIDRAS-järjestelmän kautta niin kuin muitakin aineistoja.	Ehdotetaan tilastolain uudistamista valmistelevalle työryhmälle tilastolain muuttamista EU-asetuksen mukaiseksi. Tutkimusta helpottaisi erityisesti, jos aineistoja ei tarvitsisi anonymisoida tutkimuskäyttöön luovutettaessa	Otetaan muutostarve huomioon tilastolain uudistusta valmistelevalle työryhmässä.
Tutkijoiden tarpeet huomioon ottava laintulkinta (henkilötietolaki, erityislait)	Tutkijat saavat nopeasti riittävät tiedot tutkimuksen suunnitteluun ja toteuttamiseen. Aikaa vievät työvaiheet jätetään tutkijoiden tehtäväksi.	Tutkijoille luovutetaan riittävän laaja tutkimusaineisto lainsäädännön antamin valtuuksin.	Tutkijat ja viranomaiset keskustelevat avoimesti toteutettavien aineistopyyntöjen kriteereistä ja muodostavat yhtenäiset käytännöt.
MIDRAS-hallinnon toimivalta	MIDRAS-hallinnolle on säädetty oikeudet ja vastuut	MIDRAS-hallinnolle annetaan toimivalta etäkäyttöjärjestelmän ylläpitoon.	Vastuuministeriö antaa säädöksen MIDRAS-hallinnon toimivallasta.

4.6. Toteutusvaiheet

MIDRAS-toimintamallin toteutuminen on monivaiheinen prosessi, koska viranomaisten lähtötilanteet ovat erilaiset. MIDRAS-toimintamalli hioutuu ja laajenee ajan myötä, ja tutkimusinfrastruktuuria valmistelevat toimet (palveluiden rakentaminen, valmisaineistojen muodostaminen ja kuvailu) vievät oman aikansa. MIDRAS-toimintamallin toteutusvaiheita on kuvattu tarkemmin liitteessä 2.

4.6.1. Perustamisvaihe (2011-2012)

Perustamisvaiheeseen kuuluvat kaikki ne toimet, jotka pitää tehdä, jotta MIDRAS-järjestelmä saadaan käyttöön. Perustamisvaiheessa MIDRAS toimii ennen kaikkea uutena rekisteritietojen lupahakemus- ja luovutuskanavana. Aineistojen lähetysvaihe aikaisemmassa tietojenluovutusprosessissa on korvattu tietoteknisellä tiedonsiirtoyhteydellä aineiston ylläpitäjän ja MIDRAS-järjestelmän välillä. MIDRAS tarjoaa myös keskitetyn paikan, jossa viranomaiset ja rekisteritutkijat voivat kommunikoida ja johon voi kerryttää tietoa rekisteriaineistoista ja rekisteritutkimuksesta.

- MIDRAS-järjestelmän ensimmäinen tekninen toteutus otetaan tuotantokäyttöön.
- Keskeiset viranomaiset liittyvät MIDRAS:iin ja alkavat tarjota aineistonsa tutkimuskäyttöön MIDRAS-lupaprosessin ja -etätyöpöydän kautta.

- MIDRAS-järjestelmään alkaa kertyä aineistojen kuvailuja ja arkistoituja tutkimusaineistoja.

4.6.2. Palveluvaihe (2012-2014)

Palveluvaihetta ovat ne kehitysaskleet, joita MIDRAS-järjestelmässä toteutetaan yhä useampien palveluiden ollessa käytössä. Palveluvaiheessa MIDRAS-järjestelmä sisältää koko ajan laajenevan valikoiman tietoa tarjolla olevista aineistoista, niiden käsittelystä, niillä tehdystä tutkimuksesta ja etäjärjestelmän käytöstä. Viranomaiset ja tutkijat voivat korvata osan omista tietojenkäsittelytarpeistaan MIDRAS-etätyöpöydän tarjoamilla palveluilla. Suuri osa viranomaisista toteuttaa valmisaineistovaraston, johon tiedot tulevat automaattisesti tuotantotietokannoista.

- Aineistojen tuottajat muodostavat valmisaineistoja ja automatisoivat aineistoluovutusten prosessia.
- Järjestelmään liittyy uusia aineistojen tuottajia omalähtöisesti tai yhdistelmäaineistojen yhtenä lähteenä.
- Aineistojen kuvailuja kerätään ja muunnetaan yhteiseen muotoon.
- Tietojen vaihto MIDRAS-järjestelmän ja muiden aineistokatalogien välillä käynnistyy.
- Keskitetty etätyöpöytä kehittyy käyttäjäpalautteen perusteella ja alkaa vähentää erillisten tiedonkäsittely-ympäristöjen tarvetta.
- Palveluun aletaan kerätä eri rekistereistä yhdisteltyä saatavuusaineistoa.

4.6.3. Jatkokehitysvaihe (2014-)

Jatkokehitysvaiheeseen kuuluvat ne toimet, joissa MIDRAS-järjestelmä ja –toimintamalli kehittyvät uusiin ja ennalta arvaamattomiin suuntiin. Jatkokehitysvaiheessa MIDRAS-järjestelmä on yleiskäyttöinen tietojen keruu- ja käsittelyalusta, joka soveltuu erityisesti arkaluontoisen tiedon käsittelyyn. MIDRAS-järjestelmän oheispalvelut helpottavat kommunikaatiota tiedon tuottajien ja käyttäjien kesken. MIDRAS-järjestelmässä tehtyä vakiointityötä hyödynnetään muissa palveluissa. Automatisointi lisää jatkuvasti aineistojen ajantasaisuutta ja nopeuttaa lupa- ja aineistonluovutusprosesseja.

- MIDRAS-järjestelmän teknisiä ratkaisuja päivitetään.
- Viranomaiset rakentavat rajapintoja, joilla voi tehdä tietokyselyitä tutkimuskäyttöön tarkoitettuihin aineistoihin.
- Tietopyyntöjen tarkistusta (esim. otoskoko) automatisoidaan.
- Aineistojen kuvailuja linkitetään kasvavassa määrin kansallisiin ja semanttisen webin käsitteistöihin.
- MIDRAS-järjestelmästä saatua rajapintakokemusta hyödynnetään muiden palveluiden kehittämisessä.
- MIDRAS-järjestelmää käytetään sovelluskehitysalustana, kun käsiteltävä tieto on arkaluontoista.

4.7. MIDRAS-toimintamallin hyödyt

MIDRAS-toimintamalli tarjoaa selkeitä parannuksia nykyisiin toimintamalleihin verrattuna: se tehostaa ja virtaviivaistaa aineistojen saatavuuteen liittyviä prosesseja, ja säästää näin sekä tutkijoiden että viranomaisten työtarvetta, aikaa ja hallinnollista vaivaa. Järjestelmä tukee aineistojen laajempaa ja monipuolisempaa käyttöä, parantaa tutkimuspalveluiden laatua ja tasalaatuisuutta, helpottaa rekisteritutkimuksen tekemistä ja parantaa olennaisesti tietoturvaan ja anonymisoimattomien aineistojen tietosuojaan liittyviä seikkoja. Taulukossa 12 on eritelty MIDRAS-toimintamallin hyötyjä eri toimijoiden näkökulmista.

Taulukko 12 MIDRAS-toimintamallin hyödyt eri sidosryhmille

Hyödyt rahoittajille:	Hyödyt yhteiskunnalle:
<ul style="list-style-type: none">• Tutkimus lisääntyy.• Tutkimuksen laatu paranee ajantasaisempien aineistojen, laajemman aineistopohjan ja parantuneiden metatietojen myötä.• Uuden tutkimuksen ja innovaatiotoiminnan edellytykset paranevat.• Arkaluontoisten aineistojen käsittely saadaan kustannustehokkaaksi keskittämällä.• Kokonaiskustannukset alenevat pitkällä aikajänteellä.• MIDRAS koordinoi sidosryhmien yhteistyötä järjestelmien yhteentoimivuuden parantamiseksi.	<ul style="list-style-type: none">• Tutkimuksen kansainvälinen kilpailukyky turvataan: rekisterit ovat ainutlaatuinen tietovaranto, jonka tutkimuskäyttöä ei ole varaa jättää tehostamatta.• Rekisteritietoja voidaan hyödyntää tehokkaammin ja paremmin päätöksenteossa.• Yksityishenkilöiden tietosuoja paranee tutkimusprosessissa.• Järjestelmä antaa esimerkin avoimesta, yhteisiin toimintatapoihin perustuvasta uudesta palvelusta.• Rekisteritutkimuksen avoimuus lisää kansalaisten luottamusta.
Hyödyt tutkijoille:	Hyödyt aineistoja hallinnoiville viranomaisille:
<ul style="list-style-type: none">• Avoimuus ja tietoisuus lisääntyvät; rekisteritietojen olemassaolon ja tietosisältöjen tunnettuus paranee.• Laajempiin aineistoihin pääsee vaivattomasti käsiksi; rekisterien käytön kynnyks madaltuu ja käyttö monipuolistuu.• Vakioituneet sähköiset ratkaisut tekevät lupaprosesseista ja aineistojen saamisesta nopeampaa ja läpinäkyvämpää.• Tietoturva paranee.• Datojen päivitys ja korjaaminen nopeutuu.• Etätyöpöydällä on monipuolinen työvälineistö keskitetyksi ja helposti saatavilla.• Keskitetyllä ympäristöllä voidaan tarjota paremmat laskenta- ja tallennusresurssit.• Käyttäjäfoorumien ja tutkimusarkiston kautta pystyy hyödyntämään aiempien tutkimusten menetelmiä, työkaluja ja (käyttöluvalla) aineistoja.• Työtapojen ja -välineiden arkistointi parantaa tutkimusten toistettavuutta ja siten luotettavuutta.	<ul style="list-style-type: none">• Tietoturva ja anonymisoimattomien aineistojen tietosuoja paranevat.• Tietoisuus kansallisesti tärkeistä aineistoista paranee.• Tulee mahdolliseksi seurata, miten aineistoja käytetään.• Aineistot hävitetään aina asianmukaisesti käyttöluvan loputtua.• Keskitetyn ratkaisun vuoksi viranomaisen ei tarvitse kehittää omia etäkäyttöratkaisuja.• Vakioitu, aineistojen metatietoihin pohjautuva lupahakemusprosessi vähentää lupahakemusten käsittelytyötä.• Aineistoluovutusten työtä voi vähentää kehittämällä ja tarjoamalla valmisaineistoja.• Järjestelmä tarjoaa hyvän käsittelyalustan myös tietojen viranomaiskäytölle.• Viranomainen voi omassa tutkimus- ja selvitystoiminnassaan hyödyntää MIDRAS-järjestelmää.• Viranomainen saa näkyvyyttä ja meriittiä aineistojen lisääntyneestä käytöstä.

4.8. Yhteenveto

Projektissa kerättyjen kokemusten perusteella eri toimijoiden mielestä Suomeen tarvitaan tutkimuksen tarpeisiin yksi toimiva rekisteriaineistojen etäkäyttöjärjestelmä. Suositamme, että:

1. Hallintomallista ja rahoituksesta päästään yksimielisyyteen vuoden 2011 aikana.
2. Järjestelmän palveluista aineistokatalogi ja sähköinen lupamenettely toteutetaan ensimmäisinä.
3. Pseudonymisointimalli sovitaan yhdessä sidosryhmien kanssa vuoden 2011 aikana. Pseudonymisointimallia ei voi kuitenkaan käyttää nykyisen tilastolain alaisissa aineistoissa.
4. Aloitetaan aineistojen saattaminen järjestelmän käyttöön vaiheittain 2012 alkaen. Sopivia aineistoja tutkijoiden kannalta ovat esimerkiksi: hoitoilmoitusrekisteri, syntymärekisteri, kuolemansyyrekisteri, rekisteri lääkkeiden erityiskorvausoikeuksista, työssäkäyntitilaston aineisto ja väestölaskennan pitkittäisaineisto.

Lähteet

Kirjallisuus, artikkelit ja selvitykset

Gissler M, Haukka J: Finnish health and social welfare registers in epidemiological research. Norsk Epidemiologi 14: (1), 113-120, 2004.

Opetusministeriö, 2009. Kansallisen tason tutkimusinfrastruktuurit: Nykytila ja tiekartta. Opetusministeriön julkaisu 2009:3. <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2009/liitteet/opm03.pdf?lang=fi>.

Rekisteritutkimuksen tukikeskus, 2010. Rekisterien käyttö väitöskirjoissa 2005-2010.

Tilastokeskus, 2004. Use of Register and Administrative Data Sources for Statistical Purposes. Käsikirjoja 45. Helsinki 2004.

Lait

Euroopan parlamentin ja neuvoston asetus Euroopan tilastoista (EY 223/2009)

Euroopan parlamentin ja neuvoston direktiivi julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (2003/98/EY)

Henkilötietolaki (523/1999)

Laki viranomaisten toiminnan julkisuudesta (621/1999)

Tilastolaki (280/2004)

Valtion maksuperustelaki (150/1992)

Valtiovarainministeriön asetus Tilastokeskuksen suoritteiden maksullisuudesta (1027/2010)

Valtiovarainministeriön asetus Väestörekisterikeskuksen suoritteiden maksuista (1212/2010)

Suosituks, esitykset ja hankkeet

DDI Alliance, 2009. DDI-Lifecycle 3.1. <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1>

Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA), 2006. JHS 143 Asiakirjojen kuvailun ja hallinnan metatiedot. <http://docs.jhs-suositukset.fi/jhs-suositukset/JHS143/JHS143.pdf>

Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA), 2010. JHS 172 ICT-palvelujen kehittäminen: Esiselvitys. <http://docs.jhs-suositukset.fi/jhs-suositukset/JHS172/JHS172.pdf>.

Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA), 2010. JHS 173 ICT-palvelujen kehittäminen: Vaatimusmäärittely. <http://docs.jhs-suositukset.fi/jhs-suositukset/JHS173/JHS173.pdf>.

Rekisteritutkimuksen tukikeskus, 2007. Rekisteritutkimuksen tukikeskuksen neuvottelupäivien hanke-esitys etäkäyttöjärjestelmäksi. <http://retki.stakes.fi/FI/toiminta/koulutus/esitysarkisto/neuvottelupaiva.htm>

Tietosuojavaltuutetun toimisto, 2010. Tietosuojavaltuutetun ohjeet rekisteriselosteen sisällöstä. <http://www.tietosuoja.fi/2584.htm>

Valtioneuvosto, 2010. Hallituksen esitys Eduskunnalle laeiksi julkisen hallinnon tietohallinnon ohjauksesta sekä viranomaisten toiminnan julkisuudesta annetun lain 18 ja 36 §:n muuttamisesta. HE 246/2010.
<http://www.finlex.fi/fi/esitykset/he/2010/20100246>

Valtiovarainministeriö, 2010a. Julkishallinnon perustietovarantojen rajapinnat –työryhmä.
http://www.hare.vn.fi/mHankePerusSelaus.asp?h_ild=15360

Valtiovarainministeriö, 2010b. Julkishallinnon tietovarantojen saatavuuden ja käytön edistäminen –työryhmä.
http://www.hare.vn.fi/mHankePerusSelaus.asp?tVNo=1&h_ild=14220

Valtiovarainministeriö, 2010c. Valtiotason arkkitehtuurit –hanke. Julkisen hallinnon tietoarkkitehtuuri. Versio 1.00. 12.11.2010. Valtiovarainministeriö, hallinnon kehittämisosasto, valtion IT-toiminnan johtamisyksikkö.
http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/20100105Valtio/05_Julkishallinnon_tietoarkkit_ehtuuri.pdf

Valtiovarainministeriö, 2010d. Tilastolain uudistamista valmisteleva työryhmä. Asettamispäätös 30.8.2010.
http://www.hare.vn.fi/mHankePerusSelaus.asp?h_ild=165

Valtiovarainministeriö, 2010e. Tilastokeskuksen ja Midras-projektin etäkäyttäjärjestelmien yhteiskehittämistä valmisteleva työryhmä. Asettamispäätös 31.5.2010.

Liitteet

Liite 1. Sanasto

Aggregointi

Yhden aineiston yksikkötietojen yhdistäminen toisiinsa siten, että tuloksena oleva aineisto sisältää summattua tietoa (frekvensseinä tai tunnuslukuina) lähdeaineiston tiedoista

Anonymisointi

Aineiston käsitteleminen siten, että aineiston yksiköiden sekä suora että välillinen tunnistaminen ovat mahdottomia

Federaatio

Usean organisaation tai tietojärjestelmän yhteenliittymä

Federoitu tietokanta (virtuaalinen tietokanta)

Tietokanta, jonka tietosisällöt säilytetään useammassa taustatietokannoissa

Henkilörekisteri

Käyttötarkoituksensa vuoksi yhteenkuuluvista merkinnöistä muodostuva henkilötietoja sisältävä tietojoukko, jota käsitellään osin tai kokonaan automaattisen tietojenkäsittelyn avulla taikka joka on järjestetty kortistoksi, luetteloksi tai muulla näihin verrattavalla tavalla siten, että tiettyä henkilöä koskevat tiedot voidaan löytää helposti ja kohtuuttomitta kustannuksitta

- henkilötiedolla tarkoitetaan kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi.

Käyttölupa (hallinnollinen termi)

Anomuksen käsittelyn tuloksena syntyvä sopimus, jonka perusteella käyttäjällä (MIDRAS-järjestelmässä tutkijalla) on oikeus käyttää jotain aineistoa tai sen osaa

Käyttöoikeus (tekninen termi)

Käyttäjätunnukseen liitetty tieto siitä, mitä resursseja kyseistä tunnusta käyttävä käyttäjä saa käyttää; käyttöluvan tekninen toteutus

Luokitus

Lista tiettyjä arvoja saavan muuttujan arvojen merkityksistä

Metatieto

Tiedon kontekstia, sisältöä ja rakennetta sekä niiden hallintaa ja käsittelyä koko elinkaaren ajan kuvaavaa tietoa. Tätä tietoa voidaan käyttää mm. aineiston hakuun, paikallistamiseen ja tunnistamiseen. Metatiedot ovat olennaisia aineistojen löytämisen, luetteloinnin ja käytön kannalta. Metatiedoista saatetaan käyttää myös termiä metadata. Metatiedot sisältävät sekä aineiston kuvailutietoja että teknisiä, järjestelmän metatietoja.

Muuttuja

Määrämuotoisen aineiston yksi sarake eli tieto, joka annetaan aineiston jokaiselle riville eli havainnolle

Otos

Yksikköaineistosta tuotettu versio, jossa on vain osa yksiköistä (henkilöistä, yrityksistä tai muista rekisteriyksiköistä) mukana. Käytetään, jos esimerkiksi rekisteriin liittyvät lait tai käytännöt estävät kokonaisaineiston tutkimuskäytön.

Primäärirekisteri

Hallinnollinen alkuperäinen tietokanta, jossa rekisteridata sijaitsee

Pseudonymisointi

Tunnisteiden korvaaminen yksiselitteisillä tunnisteilla, joista alkuperäistä tunnistetta ei voi päätellä

Pull-malli

Tiedonsiirtokäytäntö, jossa tiedon lähettäjä tarjoaa tiedonsiirtorajapinnan ja vastaanottaja avaa tiedonsiirtoyhteyden

Push-malli

Tiedonsiirtokäytäntö, jossa tiedon vastaanottaja tarjoaa tiedonsiirtorajapinnan ja lähettäjä avaa tiedonsiirtoyhteyden

Rekisteri

Yksikkötasoista aineistoa sisältävä tietokanta, jota tuottaa ja ylläpitää rekisterinpitäjä, ja joka on talletettu rekisterinpitäjän tietojärjestelmään tai tietojärjestelmiin

- Rekisteri sisältää yksilöivän tunnistein, esimerkiksi henkilötunnus tai Y-tunnus, jolla rekisterin yksi yksikkö voidaan tunnistaa.
- Yleensä rekisteri on jatkuvasti päivittyvä hallinnollinen tietokanta, ei kertaalleen kerätty tai tilastotarkoitukseen kerätty hetkellinen tietokanta.

Rekisteriaineisto

Rekisteriin pohjautuva eli rekisteritiedoista muodostettu aineisto. Viranomaisten hallussa olevat, julkisin varoin hallinnollisiin ja tilastotarkoituksiin rekistereihin kerätyt yksikkötasoiset aineistot, joiden käyttö on luvanvaraista.

Rekisterinpitäjä

Yksi tai useampi henkilö, yhteisö, laitos tai säätiö, jonka käyttöä varten henkilörekisteri perustetaan ja jolla on oikeus määrätä henkilörekisterin käytöstä tai jonka tehtäväksi rekisterinpito on lailla säädetty. Lainsäädännöllisesti myös henkilötietoja sisältävän tutkimusaineiston saanut tutkija on rekisterinpitäjä.

Rekisteriseloste

Henkilötietolain 10 § edellyttämä julkinen metatieto henkilörekisteristä

Saatavuusaineisto

Yksikkötason aineisto (mikrodata), joka sisältää tietoa lähinnä siitä, mitä tietoja on saatavissa mistäkin yksiköstä. Voidaan käyttää esim. tutkimusta tukevien biopankkiaineistojen löytämiseen.

Tietovaranto

Hallinnolliselta tai sisällölliseltä kannalta yhtenäinen kokonaisuus hyödynnettävää tietoa

Tutkimusaineisto

Tiettyä tutkimusta varten tuotettu (kerätty, muokattu) aineisto, jonka perusteella tutkimus tehdään

Vahva tunnistaminen

Kuten lakitekstissä: autentikointi, jossa käyttäjän varmistaminen käyttäjätunnuksen haltijaksi perustuu ainakin kahteen näistä kolmesta: (1) käyttäjän tietämään asiaan, (2) käyttäjän hallinnoimaan esineeseen, (3) käyttäjän yksilöivään ominaisuuteen

Valmisaineisto

Aineisto, josta organisaatio tuottaa tutkijoille luovutettavia aineistoja.

Valmisaineistovarasto

Tietojärjestelmä, joka sisältää kaikki valmisaineistot

Yhdistely, linkkaus

Kahden aineiston yhdistäminen käyttäen avaimena aineistoista löytyviä tunnisteita (esim. henkilötunnus)

Yksikköaineisto (mikrodata)

Aineisto, joka sisältää erillistä tietoa kustakin havaintoyksikön edustajasta, esim. henkilöstä, yrityksestä tai tapahtumasta

Yksilöaineisto

Yksikköaineisto, jossa havaintoyksikkönä on henkilö

Liite 2. MIDRAS-toimintamallin toteutusehdotus vaiheittain

	2011	2012	2013	2014
Edellytykset	Päätös hallintomallista ja rahoituksesta	Lakimuutokset ja sopimukset		Asenteiden, tietotekniikan ja julkisen tietohallinnon kehitys
MIDRAS-järjestelmän toteutus	Suunnittelu	Ensimmäinen tekninen toteutus tuotantokäyttöön	Kehitys käyttäjien palautteen myötä	
MIDRAS-tukikeskus (hallinto)	Tutkijapalvelu; Suunnittelu		Lakisääteiset tehtävät	
Datakatalogi ja metatiedot	Aineistokatalogin suunnittelu ja katalogien yhteisfoorumi	Järjestelmään alkaa kertyä yhtenäisiä kuvailuja; JHS-suositus	Tiedonvaihdon käynnistyminen katalogien välillä	Aineistojen kuvailujen linkitys semanttisen webin käsitteistöihin
Sähköinen lupamenettely	Yhteisen lupahakemuksen suunnittelu	Yhteinen lupahakemus	Kehitys käyttäjien palautteen myötä	
Etätyöpöytä ja tukipalvelut		Etätyöpöytä ja tukifoorumi tutkimuskäytössä	Etätyöpöytä tekee tarpeettomaksi osan muista ympäristöistä	Teknisten ratkaisujen päivitys; Sovelluskehitysalusta
Viranomaisten liittyminen MIDRAS-järjestelmään	Keskeiset viranomaiset mukaan		Uusia aineiston tuottajia mukaan	
Aineistot		Aineistoja aletaan toimittaa; Valmisaineistojen muodostaminen	Saatavuusaineistoja eri rekistereistä	
Aineistojen toimittaminen		Aineistoja toimitetaan järjestelmään push-mallin mukaisesti	Aineistoluovutusten automatisoinnin kehitys; Pull-mallin testausta	Tietoluovutuksissa yhteiset rajapinnat

Liite 3. Kustannusarvio

Kustannusarvio perustuu viranomaisilta saatuihin työmääräarvioihin sekä tunnettujen komponenttien kustannusanalyysiin.

Palvelun infrastruktuurin kustannusten toteutuminen riippuu siitä, missä määrin palvelut ovat toteutettavissa jo olemassa olevilla järjestelmillä. Esimerkiksi tutkijoiden tunnistamisessa, sähköisessä käyttöluopakemuksessa ja järjestelmän käytön seurannassa on mahdollista saavuttaa huomattavia säästöjä, jos ne voidaan toteuttaa täysin tai osittain valmiiksi olemassa olevilla järjestelmillä. Suurimmat säästöt syntyvät kuitenkin työskentelytapojen muutoksista.

Tällä hetkellä Suomessa viranomaisten henkilötietojen käyttölupakäsittelyihin ja aineistojen poimintaan sekä toimittamiseen käytetään karkean arvion mukaan vuodessa noin 17 henkilötyövuotta. Rekisteritutkimukseen liittyvät palvelut maksavat viranomaisille yhteensä noin 2 miljoonaa euroa vuodessa. Tämän kustannuksen rahoittavat ennen kaikkea tutkijat (tutkimushankkeiden rahoittajat) ja rekisterinpitäjät.

Seuraavassa esitetään arvio siitä, millaisia kustannuseriä etäkäyttöjärjestelmään sisältyy.

Taulukko 13 Viranomaisten työt

Kustannuserä	Aloitus (€)	Ensimmäinen vuosi noin 150 käyttäjää (€)	Viides vuosi noin 700 käyttäjää (€)
Tutkimuskohtaisten aineistojen muokkaus		900 000	500 000
Valmisaineistojen teko	200 000	400 000	200 000
Aineistohallinnon kehitys	200 000	100 000	
Dokumentaatio	300 000	100 000	50 000
Käyttöluvut	50 000	300 000	200 000
<i>Yhteensä</i>	<i>750 000</i>	<i>1 800 000</i>	<i>950 000</i>

Taulukko 14 Tekninen järjestelmä

Kustannuserä	Aloitus (€)	Ensimmäinen vuosi noin 150 käyttäjää (€)	Viides vuosi noin 700 käyttäjää (€)
Palvelinalustan ylläpitotyö	15 000	30 000	30 000
Ohjelmistojen ylläpitotyö	10 000	15 000	15 000
Kehitystyö (oheispalvelut, määrittelyt, aputyökalut)	30 000	150 000	50 000
Virtualisointiohjelmistot		33 000	33 000
Palvelinlaitteistot	30 000	2 000	2 000
Etätyöpöytä ja vahva tunnistaminen	4 000	6 000	20 000
Levytila ja varmuuskopiot	2 000	500	1 000
SPSS	20 000	15 000 (25 käyttäjää)	25 000 (75 käyttäjää)
SAS		50 000 (25 käyttäjää)	100 000 (75 käyttäjää)
Muut ohjelmistot (sis. Stata)		15 000	45 000
<i>Yhteensä</i>	<i>111 000</i>	<i>316 500</i>	<i>321 000</i>

Taulukko 15 MIDRAS-hallinto

Kustannuserä	Aloitus (€)	Ensimmäinen vuosi noin 150 käyttäjää (€)	Viides vuosi noin 700 käyttäjää (€)
Tukipalvelut		350 000	400 000
Sähköinen käyttöluva ja käyttäjätunnushallinto	50 000	50 000	10 000
Aineistohallinto	45 000	50 000	40 000
Auditointi ja seuranta	5 000	5 000	5 000
<i>Yhteensä</i>	<i>100 000</i>	<i>455 000</i>	<i>455 000</i>

Liite 4. Riskit

Uuden toimintamallin käyttöönottoon liittyy riskejä, mutta niin myös siihen, ettei uudistuksia tehdä. Tärkeitä ja arvokkaita mahdollisuuksia jää hyödyntämättä, mikäli rekisteritutkimuksen edellytykset pysyvät ennallaan. Tämänhetkiseen tilanteeseen liittyy riskejä, joihin MIDRAS-toimintamalli tarjoaa ratkaisun:

- Viranomaisten työmäärä tutkimuspalveluissa kasvaa, mutta resursseja ei tule lisää eikä prosessia pystytä uudistamaan; seurauksena on kaikille huono tietopalvelu.
- Rekisteriaineistojen vähäinen dokumentointi ja byrokraattinen hakukäytäntö herättää kansalaisissa ennakkoluuloja.
- Rekisteriaineistojen vähäinen dokumentointi lisää kynnystä rekisteritutkijaksi ryhtymiseen ja tekee rekisteritutkimuksesta pienen eliitin erityisalan.
- Aineistojen sisältöön ja muutoksiin liittyvä hiljainen tieto menetetään aineistoja tuntevien työntekijöiden poistuessa, ennen kuin kuvailut saadaan kirjatuksi muistiin.
- Suomeen perustetaan useita, keskenään epäyhteensopivia aineistopalveluita tai etätyöpöytiä, joiden päällekkäinen ylläpito tuottaa paljon vaivaa ja sirpaleisuus hankaloittaa entisestään useiden viranomaisten aineistojen yhdistämistä.

MIDRAS-toimintamallin toteuttamiseen liittyy riskejä, joihin vastataan taulukossa 16 esitellyillä keinoilla.

Taulukko 16 MIDRAS-toimintamallin riskit ja varautumiskeinot

Riski	Varautumiskeino
Kansalaiset pelkäävät superrekisteriä.	Kansalaisille esitetään, mitä hyödyllistä tutkimusta on tehty käyttäen rekistereitä, ja rekistereiden tietosisältö kuvaillaan mahdollisimman hyvin. MIDRAS-järjestelmään toteutetaan tutkimuskatalogi ja -arkisto, jossa luetellaan MIDRAS:n kautta tehtyjä tutkimuksia.
Järjestelmälle ei ole kustannuksia vastaavaa tarvetta.	Toteutetaan MIDRAS-järjestelmästä ensin hyödyllisimmät osiot. Suunnitellaan järjestelmä siten, että sitä voi soveltaa laajasti arkaluontoisten tietojen käsittelytarpeisiin. Tiedotetaan järjestelmästä tehokkaasti ja sovelletaan sitä kaikkien sidosryhmien käyttöön.
Ulkopuoliset murtautujat pääsevät MIDRAS-järjestelmään.	Rakennetaan VAHTI- ja muiden tietoturvamääräysten mukainen järjestelmä suojaamaan ulkopuolisilta. Noudatetaan järjestelmässä korotettua tietoturvasoa (suurin osa MIDRAS-järjestelmässä käsiteltävistä aineistoista on turvallisuusluokituksen III mukaisia ja siksi VAHTI 2/2010 edellyttää tätä). Seurataan aktiivisesti etäjärjestelmän käyttöä.

Riski	Varautumiskeino
Tietoa päätyy väärin käsiin.	Käytetään salakirjoitettuja yhteyksiä kaikessa tietoliikenteessä ja vahvaa tunnistusta etätyöpöydälle kirjautumisessa. Seurataan tietojen käyttöä ja kirjautumisia työpöydälle.
Tutkijat käyttävät väärin aineistoja ja tekevät väärää johtopäätöksiä, koska eivät ole aineistojen luovutuksessa tekemisessä aineistoasiantuntijoiden kanssa.	Viranomaiset kuvaavat aineistoja tarkasti ja erityisesti aineistoissa ja muuttujissa tapahtuneita muutoksia vuosien varrella. Säilytetään keskusteluyhteys tutkijoiden ja viranomaisten aineistoasiantuntijoiden välillä.
Tutkijat tunnistavat yksilöitä aineistosta ja julkaisevat tutkimustietonsa niin, että yksilön voi tunnistaa.	Tutkijat allekirjoittavat salassapitositoumuksen. Tutkijoita informoidaan salassapidon rikkomuksesta sekä tutkijalle että tutkimusorganisaatiolle tulevista seuraamuksia.
Keskeinen viranomainen jättäytyy ulkopuolelle, eikä salli aineistonsa luovutusta MIDRAS-järjestelmän kautta.	Järjestelmä varmistetaan luotettavaksi (auditointi). Järjestetään MIDRAS-järjestelmän rahoitus siten, että käyttö on kannattavaa.
Tilastolakiin ei tule muutosta ja tilastoaineistoja on vielä anonymisoitava ennen luovutusta.	Esitellään MIDRAS-järjestelmä järjestelmänä jonka kautta voi ottaa aineistoja käyttöön EU-asetuksen mukaisesti. Tutkijat hankkivat aineistonsa suoraan primääirekistereistä.
Viranomaisilla ei resursseja käyttöluvahakemusten käsittelyyn ja tutkimusaineistojen ja valmisaineistojen luomiseen.	Viranomaisilla on julkisuuslain mukainen velvoite palvella aineistojen saannissa. MIDRAS-järjestelmän kustantamiseen kuuluu myös rahoittaa viranomaisten tutkimusaineistojen luovuttamiseen liittyvät työt.
Uudet aineistojen toimitustavat lisäävät viranomaisten työtä.	MIDRAS-järjestelmän käyttöönotossa järkeistetään myös viranomaisten aineistohallintaa.
Järjestelmästä tulee kallista ylläpitää.	Automatisoinnilla ja valmisaineistoilla vältetään toistuvaa työtä.
Tutkijat eivät halua käyttää järjestelmää.	Tarjotaan hyvät resurssit tietojen käsittelyyn ja etätyöpöydällä laaja ohjelmistovalikoima. Suunnitellaan järjestelmä yhdessä käyttäjien kanssa ja varmistetaan sen käytettävyys.
Suomessa ylläpidetään kahta (tai useampaa) samojen aineistojen etäkäyttöjärjestelmää.	VM:n asettaman yhteistyöryhmän tehtävä on katsoa ettei Tilastokeskuksen ja MIDRAS-hankkeen välillä tehdä päällekkäistä työtä.
MIDRAS-järjestelmä edellyttää viranomaisilta työvuota, joka on hankala järjestää.	Tarjotaan viranomaisille useampaa toimintamallia, kun on syytä olettaa yhden toimintatavan rajoittavan liiaksi viranomaisen toimintatapoja.

Liite 5. Eräitä keskeisiä tutkimuksessa käytettäviä rekistereitä ja rekistereihin perustuvia aineistoja

Rekisterit rekisterinpitäjän mukaan
Eläketurvakeskus (ETK)
Ansaintarekisteri (työsuhderekisteri)
Eläkerekisteri
Kansaneläkelaitos (KELA)
Etuusrekisteri (tallennettu lukuisiin erillisiin eri etuuksia koskeviin osarekistereihin)
Maa- ja metsätalousministeriö
Maatilarekisteri
Maatilayritysrekisteri
Oikeusrekisterikeskus
Rikosrekisteri
Opetushallitus
Opiskelijavalintarekisterit
Säteilyturvakeskus (STUK)
Annosrekisteri
Terveyden ja hyvinvoinnin laitos (THL)
Hoitoilmoitusrekisteri (Hilmo)
Lastensuojelurekisteri
Syntyneiden lasten rekisteri
Syöpärekisteri (ylläpidetään Suomen Syöpäyhdistyksen kanssa)
Tartuntatautirekisteri
Toimeentulotukirekisteri
Tilastokeskus
FLEED-aineisto (Finnish Longitudinal Employer-Employee Data)
Kuolemansyytilaston aineisto (kuolemansyyrekisteri)
Tulonjakotilaston aineisto
Tutkintorekisteri
Työssäkäyntitilaston aineisto
Yritysrekisteri
Väestölaskentojen pitkäaikaisaineisto
Työ- ja elinkeinoministeriö
Työhallinnon asiakaspalvelurekisteri (URA)
Työterveyslaitos (TTL)
Työperäisten sairauksien rekisteri (ammattitautirekisteri)
Verohallitus
Verotietokanta
Väestörekisterikeskus (VRK)
Väestötietojärjestelmä