

Arkaluonteisen tiedon tutkimuskäytön viitearkkitehtuuri

Viitearkkitehtuurikuvaus

Satu Tissari, Nestori Syynimaa, Ari Lukkarinen

Julkaisu Arkaluonteisen tiedon tutkimuskäytön viitearkkitehtuuri	
Julkaisija Avoin tiede ja tutkimus -hanke	Julkaisuajankohta 10.2.2016
Tekijät Satu Tissari, Nestori Syynimaa, Ari Lukkarinen	
Lisenssin antaja: Opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus -hanke	
Lisenssi  Tämä teos on lisensoitu Creative Commons Nimeä 4.0 Kansainvälinen –lisenssillä .	
Julkaisun jakelu PDF-tiedosto ladattavissa sivuilla avointiede.fi/selvitykset	
Yhteystiedot http://avointiede.fi avointiede@postit.csc.fi	

12.2.2016

Versio: 1.0

Sisällys

1. Johdanto	4
1.1. Dokumentin tarkoitus.....	4
1.2. Kenelle tämä dokumentti on tarkoitettu.....	4
1.3. Tämän kuvauksen rajaukset ja reunaehdot	4
2. Kokonaisarkkitehtuurimenetelmän hyödyntäminen	6
3. Periaatetason arkkitehtuurikuvaukset.....	6
3.1. Strategia, kehittämisvaatimukset ja tavoitteet	7
3.2. Visio arkaluonteisen tiedon tutkimuskäytön infrastruktuurista	8
3.3. Sidosarkkitehtuurit, -hankkeet ja -ratkaisut	9
3.4. Arkkitehtuuriperiaatteet	13
3.5. Tietoturvaperaatteet	14
4. Toiminta-arkkitehtuurin arkkitehtuurikuvaukset	18
4.1. Sidosryhmät	18
4.2. Substanssipalvelut	19
4.3. Toimijat ja roolit	21
4.4. Prosessit	22
4.4.1. Prosessikartta.....	23
4.4.2. Prosessikuvaukset.....	24
4.4.3. Keskeisimmät tehtävänmukaiset huomioitavat asiat.....	32
5. Tietoarkkitehtuurin arkkitehtuurikuvaukset.....	34
5.1. Sanasto	34
5.2. Tiedot ja tietomallit	40
5.3. Loogiset tietovarannot	41
6. Lähtötilanteen havainnot.....	43
6.1. Palveluiden ja hankkeiden nykytilanne	43
6.2. Palveluiden tavoitetila.....	46
6.3. IT-infrastruktuuripalvelujen tavoitetila	47
6.4. Tietoturvallisuuden nykytilasta.....	47



1. Johdanto

1.1. Dokumentin tarkoitus

Tämä dokumentti on CSC - Tieteen tietotekniikan keskukselta tilattu selvitystyö. Dokumentissa kuvataan ehdotus arkaluonteisen tiedon tutkimuskäytön viitearkkitehtuurityön pohjaksi. Viitearkkitehtuuriehdotus on tuotettu opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus -hankkeelle (ATT), ja se julkaistaan hankkeen avoimella lisenssillä.

Tilatun selvitystyön tavoitteena oli laatia arkaluonteisten aineistojen hyvän hallinnan pohjaksi arkkitehtoninen kehys, joka edesauttaa keskustelua aiheesta sekä toimii mahdollisen jatkotyön pohjana. Pyrkimyksenä oli löytää tärkeimmät tekijät, periaatteet, prosessit, yms., jotka olisi huomioitava, sekä tunnistaa tarvittavien yhteisten palveluiden kirjo ja toteutuksen periaatteet.

Dokumentissa ei tuoteta yksityiskohtaisia kuvauksia, eikä oteta kantaa tekniisiin ratkaisuihin.

ATT-hankkeen tavoitteena on, että vuoteen 2017 mennessä Suomi nousee yhdeksi johtavaksi maaksi tieteen ja tutkimuksen avoimuudessa ja että avoimen tieteen mahdollisuudet hyödynnetään laajasti yhteiskunnassamme. Lisäämällä tutkimuksen avoimuutta, parannamme samalla tieteen luotettavuutta ja läpinäkyvyyttä. Avoimuus tuo kaikille mahdollisuuksia olla osallisena tieteen edistämisessä ja mahdollistaa tutkimustulosten tehokkaamman hyödyntämisen. Tähän päästään toteuttamalla avoimuuden periaatteita niin, että tutkimustulosten hyödyntäminen helpottuu ja nopeutuu ja tutkimuksen vaikuttavuus kasvaa. Tähän tarvitaan paitsi tutkijayhteisön laajaa osallistumista, myös koko tutkimusympäristöjen, tutkijapalveluiden ja tutkimusinfrastruktuurien kehittämistä. Tämä puolestaan vaati tutkimusorganisaatioiden, tutkimusinfrastruktuurien ja tukipalveluiden laajamittaista yhteistyötä ja koordinaatiota sekä uusien toimintatapojen omaksumista.

1.2. Kenelle tämä dokumentti on tarkoitettu

Tämä arkkitehtuurikuvaus on tarkoitettu kaikille sidosryhmille, jotka ovat tekemisissä arkaluonteisen tiedon tutkimuskäytön kanssa.

Erityisesti seuraavien sidosryhmien tulee huomioida viitearkkitehtuuri:

- Aineiston tuottajat
- Rekisterinpitäjät
- Palveluntuottajat
- Palveluiden omistajat
- Tietojärjestelmien toteuttajat

1.3. Tämän kuvauksen rajaukset ja reunaehdot

Viitearkkitehtuurissa kuvataan arkaluonteisen tiedon tutkimuskäyttöön liittyvät sidosryhmät ja toimijat, sekä prosessit ja käsitteet. Viitearkkitehtuuri listaa aiheeseen liittyvät palvelut ja tarpeet ja luo vision yhteentoimivasta sähköisestä infrastruktuurista.



Viitearkkitehtuurin ulkopuolelle on rajattu muun kuin arkaluonteisen tiedon tutkimuskäyttö, kuten tieto, joka ei ole salassa pidettävää tai se on luokiteltu suojaustasoille I ja II (FINLEX 681/2010). On huomattava, että arkaluonteisella tiedolla tässä dokumentissa ei tarkoiteta vain henkilötietolain määritelmän mukaista tietoa vaan kaikkea muutakin salassa pidettävää tietoa myös muuta kuin henkilötietoa. Henkilötietolain määritelmä koskee vain lakitekstiä, yleiskielessä määritelmä on toinen. Viitearkkitehtuuri ei ota kantaa tekniseen toteutukseen.

Arkaluonteisella tiedolla tarkoitetaan esimerkiksi tietoa, joka on:

- asetuksen (681/2010) [9](#) § mukaan luokiteltu korkeintaan suojaustasoille III ja IV,
- lain (621/1999) [24](#) § tarkoitettu salassa pidettävä viranomaisen asiakirja,
- henkilötietolain (523/1999) [11](#) § tarkoitettua arkaluonteista tietoa,
- biopankkilain (688/2012) [3](#) § tarkoitettuihin näytteisiin ja niiden käsittelyyn liittyvää tietoa,
- potilaan asemaa ja oikeuksia käsittelevän lain (785/1992) [13](#) § perusteella potilasasiakirjoista koottu tieto
- tilastokeskuksen keräämä tieto yhteisöistä ja säätiöistä, pois lukien tilastolain (280/2004) [18](#) § mainittu julkinen tieto

On huomioitavaa, että arkaluonteista tutkimustietoa ei sinällään lueta lain (621/1999) [5](#) § tarkoittamaksi viranomaisen asiakirjaksi.

Viitearkkitehtuuri sisältää tutkimuskäyttöön kerätyn datan lisäksi myös muuhun kuin tutkimuskäyttöön alun perin kerätyn datan käsittelyn tutkimuskäyttöä varten. Tällaista tutkimusta kiinnostavaa dataa on esimerkiksi julkishallinnon rekistereissä ja terveydenhuollon järjestelmissä tai tutkittavien omassa hallinnassa.

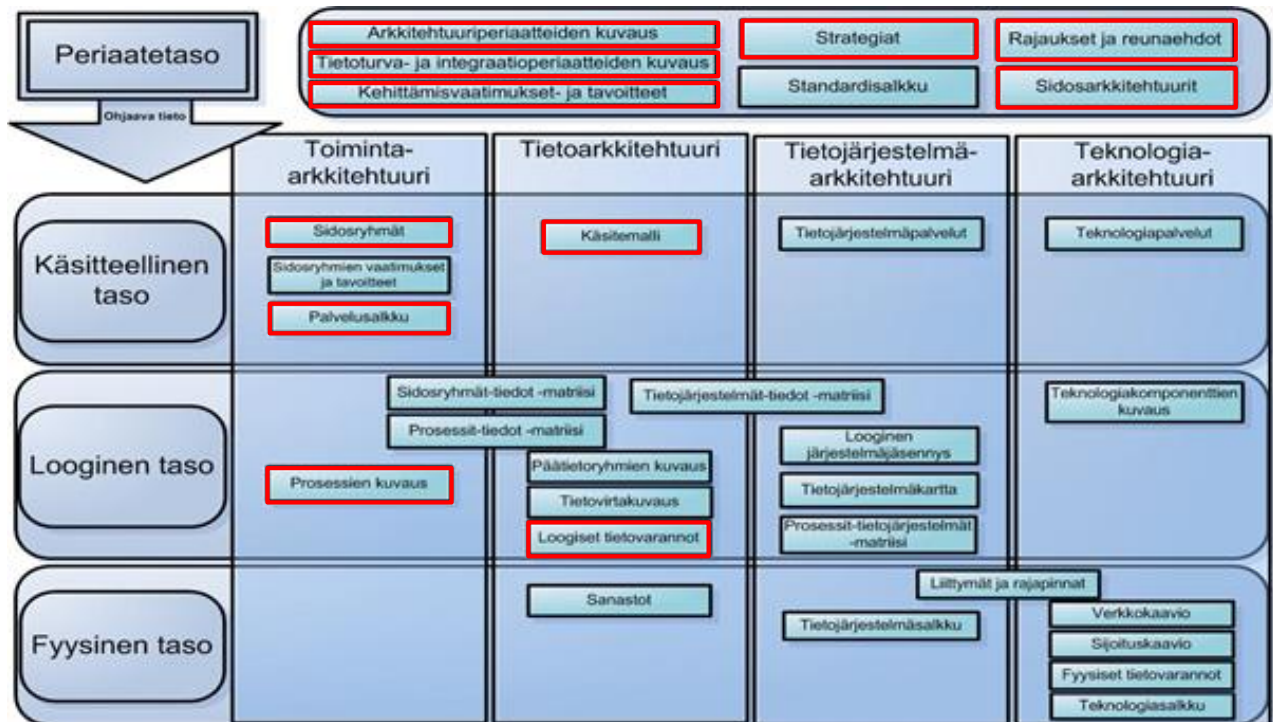
Viitearkkitehtuuri käsittelee arkaluonteisen tiedon tutkimuskäyttöä, joka tässä ymmärretään sisältävän tieteellisen tutkimuksen lisäksi viranomaisten päätöksentekoa ja johtamista tukevat selvitykset ja yritysten tutkimus- ja tuotekehityksen uusien tuotteiden ja palveluiden kehittämiseksi.

Henkilötietolaki sallii arkaluonteisen tiedon käsittelyn erityisesti, kun henkilö on antanut siihen nimenomaisen suostumuksensa tai kun on kyse historiallisesta tai tieteellisestä tutkimuksesta taikka tilastoinnista.

Viitearkkitehtuurin ulkopuolella ovat terveydenhuollon, sosiaalihuollon tai muun palvelutuotanto, joihin liittyy olennaisesti erilaisia vaatimuksia kuin tutkimuskäyttöön. Sosiaali- ja terveydenhuollon tietojen toissijainen käyttö eli ko. tiedon käyttö tutkimukseen, kehittämiseen, tietojohtamiseen ja innovaatio-toimintaan sisältyy viitearkkitehtuuriin, mutta viitearkkitehtuuri ei ota kantaa ko. tiedon luovutuksen säädäntöön, vaan lähtee siitä, että tieto tieteelliseen tutkimukseen ja muuhun toissijaiseen käyttöön on luovutettu voimassa olevan lainsäädännön mukaisesti.

2. Kokonaisarkkitehtuurimenetelmän hyödyntäminen

Tässä työssä on käsitelty seuraavia **tavoitetilan** kokonaisarkkitehtuurin osakuvauksia:



Kuva 1: Viitearkkitehtuurissa tuotetut kokonaisarkkitehtuurikuvaukset on merkitty punaisella

3. Periaatetason arkkitehtuurikuvaukset

Periaatetason keskeisimmät viitearkkitehtuurilinjaukset ovat:

- Lakien noudattaminen ja yhtenäinen tulkinta
- Tutkimustiedon, menetelmien ja tulosten avoimuus
- Tutkimustiedon jatkokäyttö
- Tietosuojaan ja tietoturvan ajantasaisuus
- Tietosuojaan huomiointi jo suunnittelussa (Privacy by Design)
- Hyvä tiedonhallintatapa
- Tiedon arkaluonteisuuden pienentäminen
- Vain tarpeellisen tiedon luovuttaminen
- Käyttöoikeudet vain tarpeellisille tiedon käsittelijöille
- Käyttöoikeuksien elinkaaren hallinta tutkimuksen aikana
- Tiedon koko elinkaaren huomiointi
- Tietojen luokittelu
- Palvelujen skaalautuvuus, laajennettavuus ja muovautuvuus

3.1. Strategia, kehittämisvaatimukset ja tavoitteet

Keskeisimpänä strategisena tavoitteena on avoimen tieteen ja tutkimuksen hengessä^{1,2,3} mahdollistaa arkaluonteisen tiedon turvallinen, tehokas ja yhtenäinen tutkimuskäyttö, mukaan lukien käyttö jatkotutkimuksiin ja monitieteisiin tutkimuksiin, sekä turvata aineistojen asianmukainen pitkäaikaissäilytys.

Avoimuus ei tässä arkkitehtuurissa tarkoita tietojen jakamista internetissä kehen tahansa hyödynnettäväksi, vaan avoimuus tarkoittaa tässä arvokkaiden aineistojen saatavuutta tutkimuskäyttöön siten, että erityisesti panostetaan tutkittavien tietosuojan varmistamiseen ja tietoturvaan. Selvennyksenä todettakoon, että tavoitteena ei ole muuntaa arkaluonteisia tietoja avoimeksi tiedoksi eikä edes kaikkien tutkijoiden vapaasti käytettäväksi vaan arkaluonteinen tieto on aina käytettävissä vain tutkijakohtaisin asianmukaisin luvuin. Samoin pitkäaikaissäilytyksessä otetaan huomioon arkaluonteisten aineistojen asianmukainen säilytysaika.

Tavoitteena on yhtenäisillä toimintatavoilla ja menetelmillä lakien tulkinnanvaraisuuksien vähentäminen ja aineistojen jatkotutkimuskäytön lisääminen lain sallimissa rajoissa. Tavoitteena on myös lisätä tutkimusresursseista saatavaa hyötyä kumuloituvina laadukkaina tutkimusaineistoina, jotka mahdollistavat uudenlaisten tutkimuskysymysten asettamisen ja uusia tutkimustuloksia. Kaiken ohessa tutkittavien tietosuoja paranee ja motivaatio osallistua tutkimuksiin kasvaa.

Arkaluonteisen tiedon tutkimuskäytön viitearkkitehtuuri pyrkii vastaamaan dataintensiivisen tutkimuksen trendin tuomiin kehittämisvaatimuksiin yhden osa-alueen osalta. Henkilötietoa sisältävää tutkimusaineistoa syntyy ja on tarvetta hyödyntää hyvin erilaisilla tieteenaloilla kuten lääketieteessä, yhteiskuntatieteessä, oikeustieteessä, kielitieteessä sekä monitieteisessä tutkimuksessa⁴. Tämän viitearkkitehtuurin tavoite on kattaa kaikki tutkimusalat, joissa käsitellään arkaluonteista tietoa. Aineistoissa piilee valtava potentiaali, erityisesti aineistojen jatkokäytössä siten, että aineistoihin kohdistetut panostukset kumuloituvat pitkäaikaisesti säilytettäviin aineistoihin.

Erityisesti Suomessa muiden pohjoismaiden lisäksi viranomaisten rekisterit muodostavat merkittävän tietovarannon, joiden nykyistä laajempi hyödyntäminen on toivottavaa. Suomen vuonna 2013 voimaantullut edistyksellinen biopankkilaki edistää osaltaan samoja avoimeen tieteen ja tutkimuksen tavoitteita. Suomalaisten vahva luottamus tutkimukseen tukee tutkittavien antaman panoksen mahdollisimman laajaa hyödyntämistä tehokkaasti. Tietosuojasta huolehtiminen on edellytys tämän luottamuksen säilyttämiselle ja vahvistamiselle.^{5,6,7}

¹ [Tutkimuksen tietoaaineistot Olennaisen käsikirja päättäjille](#) 2010 www.avointiede.fi

² [Tieto käyttöön Tiekartta tutkimuksen sähköisten tietoaaineistojen hyödyntämiseksi](#) 2011 www.minedu.fi

³ [Avoin tiede ja tutkimus -hanke](#) 2014 www.avointiede.fi

⁴ [Ihmistieteellisten tutkimusaineistojen jatkokäyttö ja tietosuoja](#) 2014 www.helsinki.fi

⁵ [MIDRAS: Viranomaisten luvanvaraiset aineistot tehokkaasti ja turvallisesti tutkimuskäyttöön](#) 2011

⁶ [Rekisteriaineistojen etäkäyttöjärjestelmien yhteiskehittämisen työryhmä TURE loppuraportti](#) 2011 www.valtioneuvosto.fi

⁷ [Joint Nordic Registers and Biobanks, A goldmine for health and welfare research](#) 2014 www.nordforsk.org

My Data -trendissä⁸ yksilöt hallitsevat kaikkea itseensä kohdistuvaa tietoa riippumatta sen syntytavasta. Nämä aineistot voivat myös olla arvokkaita tutkimukselle ja epäilemättä joukoittain terveydestään kiinnostuneita yksilöitä ovat kiinnostuneita luovuttamaan aineistojaan tutkimukseen. Nämä aineistot laajentavat esim. terveyteen kohdistuvan tutkimuksen näkökulmaa ja siten tuovat erilaiset palveluntuottajat hyödyntämään tutkimustuloksia.

Terveysalan tutkimus- ja innovaatiotoiminnan kasvustrategia^{9,10} ja tekeillä oleva genomistrategia¹¹ korostavat arkaluonteisen tiedon tutkimuskäyttöä ja hyödyntämistä terveydenhuollossa. Sote-tieto hyötykäyttöön -strategia tuo esille sote-tiedon toissijaisen käytön tutkimukseen, johtamiseen, tilastointiin ja kaupalliseen käyttöön.¹² Tilastolainsäädännön tuore uudistus edistää arkaluonteisen tiedon tutkimuskäyttöä ja henkilörekisterien lainsäädännön tulevilla uudistuksella varmistetaan, että julkisen hallinnon rekistereihin tallennettuja sote-tietoja voidaan käyttää myös toissijaisiin käyttötarkoituksiin.

Yhteiskunnan rakentama infrastruktuuri tulee olla yhteiskunnan kaikkien toimijoiden – niin yksilöiden, julkisten organisaatioiden kuin yritystenkin – käytettävissä. Yritykset ovat keskeisessä asemassa yksilöitä hyödyttävien tuotteiden ja palvelujen kehittämisessä ja tuottamisessa sekä yhteiskunnan verotulojen kerryttämisessä tutkimuksen ja yhteisten palvelujen rahoittamiseksi. Sama lainsäädäntö velvoittaa kaikkia toimijoita. Yhteiskunnan etu on saada tutkimustulokset hyötykäyttöön tuotteina ja palveluina.

Arkaluonteista tietoa sisältävien aineistojen täysimääräiseksi hyödyntämiseksi tutkimuksessa tarvitaan tietosuojan ja tietoturvaan liittyviä kansallisia yhteisiä palveluja heterogeenisen tutkimusyhteisön yhteiseen käyttöön resurssien tehokkaan käytön varmistamiseksi. Tällaiset usean hallinnonalan ja rahoittajatahon yhteisten tavoitteiden saavuttamista edistävät palvelut luovat Suomelle ainutlaatuisen kilpailuvaltin tutkimusympäristöissä.

Tämän viitearkkitehtuurin tavoitteena on:

1. Kuvata arkaluonteisen tiedon tutkimuskäyttöön liittyvät tärkeimmät toimijat, periaatteet ja prosessit,
2. Kuvata toiminnallinen tavoitetilä arkaluonteisen tiedon tutkimuskäytölle,
3. Hahmottaa kansallisia yhteisiä palveluja.

3.2. Visio arkaluonteisen tiedon tutkimuskäytön infrastruktuurista

Suomessa on arkaluonteisen tiedon tutkimuskäyttöön ja siihen liittyvään tiedon hallintaan luotettavat ja helppokäyttöiset kansalliset palvelut, joiden rakentumista tämä viitearkkitehtuuri ohjaa. Yhteentoimivuutta silmällä pitäen rakennetut palvelut ovat tieteellisen tutkimuksen, julkishallinnon selvitysten sekä yritysten tuotekehityksen tekijöiden käytettävissä. Palvelukokonaisuus

⁸ My Data – johdatus ihmiskeskeiseen henkilötiedon hyödyntämiseen 2014 www.lvm.fi

⁹ Terveystieteen tutkimus- ja innovaatiotoiminnan kasvustrategia 2014 www.tem.fi

¹⁰ Biopankkien liiketoimintamahdollisuudet 2014 www.tekes.fi

¹¹ Parempaa terveyttä genomitiedon avulla, Kansallinen genomistrategia: Työryhmän ehdotus 2015 www.sitra.fi

¹² Tieto hyvinvoinnin ja uudistuvien palvelujen tukena, Sote-tieto hyötykäyttöön -strategia 2015 www.stm.fi

tarjoaa palvelut myös arkaluonteisia tietoja tutkimus- ja kehityskäyttöön luovuttaville organisaatioille sekä tutkittaville. Yhteiset prosessit yhtenäistävät arkaluonteisen tiedon käsittelyä ja varmistavat tutkittavien tietosuojan säilymisen. Tämä takaa eri alojen tutkimuksen kukoistuksen sekä yksilöitä hyödyttävien tuotteiden ja palveluiden kehittämisen.

3.3. Sidosarkkitehtuurit, -hankkeet ja -ratkaisut

Seuraavaan taulukkoon on koottu keskeiset arkaluonteisen tiedon tutkimuskäytön viitearkkitehtuuriin liittyvät sidosarkkitehtuurit. Velvoittavuus -sarake kuvaa, onko kyseisessä kehittämisessä velvoittava, huomioitava vai ohjaava määrittäminen.

Alla on myös lueteltu viitearkkitehtuuriin liittyvät hankkeet, joiden vuorovaikutus tähän viitearkkitehtuuriin tulee selvittää.

Taulukko 1: Sidosarkkitehtuurit

Nimi	Velvoittavuus	Kuvaus
Suomen perustuslaki (FINLEX 731/1999)	Velvoittava	Yksityisyyden suoja määritellään perustuslaissa perusoikeudeksi.
Laki viranomaisten toiminnan julkisuudesta (FINLEX 621/1999)	Velvoittava	Viranomaisten asiakirjat ovat julkisia, jollei tässä tai muussa laissa erikseen toisin säädetä.
Henkilötietolaki (FINLEX 523/1999)	Velvoittava	Lain tarkoituksena on toteuttaa yksityiselämän suojaa ja muita yksityisyyden suojaa turvaavia perusoikeuksia henkilötietoja käsiteltäessä sekä edistää hyvän tietojenkäsittelytavan kehittämistä ja noudattamista.
Laki potilaan asemasta ja oikeuksista (FINLEX 785/1992)	Velvoittava	Laki sisältää säädökset potilaan asemasta ja oikeuksista mm. tietojen käsittelyssä.
Valtioneuvoston asetus tietoturvasuudesta valtioneuvoston asetuksella (FINLEX 681/2010)	Velvoittava	Asetuksessa säädetään valtionhallinnon viranomaisten asiakirjojen käsittelyä koskevista yleisistä tietoturvasuodatuksista sekä asiakirjojen luokittelun perusteista ja luokittelua vastaavista asiakirjojen käsittelyssä noudatettavista tietoturvasuodatuksista.
Biopankkilaki (FINLEX 688/2012)	Velvoittava	Lain tarkoituksena on tukea tutkimusta, jossa hyödynnetään ihmisperäisiä näytteitä, edistää näytteiden käytön avoimuutta sekä turvata yksityisyyden suoja ja itsemääräämisoikeus näytteitä käsiteltäessä.
Tilastolaki (FINLEX 280/2004)	Velvoittava	Lakia sovelletaan tietojen keräämiseen, luovuttamiseen, suojaamiseen ja muuhun käsittelyyn tilastoja laadittaessa.

Nimi	Velvoittavuus	Kuvaus
Tietohallintolaki (FINLEX 634/2011)	Velvoittava/ Hyödyllinen riippuen orga- nisaatiosta	Lain tarkoituksena on tehostaa julkisen hallinnon toimintaa sekä parantaa julkisia palveluja ja niiden saatavuutta säätämällä julkisen hallinnon tietohallinnon ohjauksesta ja tietojärjestelmien yhteentoimivuuden edistämistä ja varmistamisesta.
Asetus viranomaisen toiminnan julkisuudesta ja hyvästä tiedonhallintatavasta (FINLEX 1030/1999)	Velvoittava	Hyvän tiedonhallintatavan ja tiedonsaantioikeuksien toteuttaminen
Laki lääketieteellisestä tutkimuksesta (FINLEX 488/1999)	Velvoittava	Laki sisältää ihmiseen ja ihmisen alkioon ja sikiöön kohdistuvan lääketieteellisen tutkimuksen määräykset.
Laki terveydenhuollon valtakunnallisista henkilörekistereistä (FINLEX 556/1989)	Velvoittava	Laissa määrätään keskusrekisterit, jotka sisältävät henkilön terveydentilaa, sairautta tai vammaisuutta taikka häneen kohdistettuja hoitotoimenpiteitä tai niihin verrattavia toimia koskevia tietoja.
Asetus terveydenhuollon valtakunnallisista rekistereistä (FINLEX 774/1989)	Velvoittava	Asetus sisältää poistoilmoitusrekisterin, synyneiden lasten rekisterin, lääkkeiden sivuvaikutusrekisterin, raskauden keskeyttämis- ja sterilöimisrekisterin, syöpärekisterin, epämuodostumarekisterin, näkövammarekisterin ja implanttirekisterin.
Asetus potilasasiakirjoista (FINLEX 298/2009)	Velvoittava	Asetus koskee potilaan hoidon järjestämisessä ja toteuttamisessa käytettävien asiakirjojen laatimista sekä niiden ja muun hoitoon liittyvän materiaalin säilyttämisestä.
Laki sosiaali- ja terveydenhuollon tietojen sähköisestä käsittelystä (FINLEX 159/2007)	Velvoittava	Sähköinen potilastietojen käsittely- ja arkistointijärjestelmä terveydenhuollossa
Arkistolaki (FINLEX 831/1994)	Velvoittava	Määrittää arkistonmuodostajat ja arkistolaitoksen tehtävät.
Laki sosiaali- ja terveystieteen tutkimus- ja kehittämiskeskusten tilastotoimesta (FINLEX 409/2001)	Velvoittava	Tässä laissa tarkoitettuja tilastoja tuottaa ja julkaisee sosiaali- ja terveystieteen tutkimus- ja kehittämiskeskus (tutkimus- ja kehittämiskeskus), joka toimii tilastolain (62/1994) 3 §:n 1 momentissa tarkoitettuna tilastoviranomaisena.
Asetus ELIXIR tutkimusinfrastruktuurista (FINLEX 701/2014)	Velvoittava	Asetuksessa säädetään Eurooppalaisen luonnontieteiden tutkimusinfrastruktuurin biologiselle tiedolle (ELIXIR) perustavan konsortiosopimuksen voimaansaattamisesta ja sopimuksen lainsäädännön alaan kuuluvien määräysten voimaansaattamisesta annetun lain voimaantulosta.
EU-tietosuoja-asetus	Velvoittava	Asetus on tavoitteena hyväksyä 2015 aikana



Nimi	Velvoittavuus	Kuvaus
Suomen Akatemian suositukset tutkimusinfrastruktuuripalvelujen käytöstä	Velvoittava (kevät 2015 lähtien)	Suomen Akatemia suosittelee tutkimusaineistojen avaamista ja siihen sopivien palvelujen käyttämistä
BBMRI ERIC	Velvoittava	Suomen jäsenyys ERICissä asettaa vastaavanlaisia velvollisuuksia kuin ELIXIR asetus
EATRIS ERIC	Velvoittava	Suomen jäsenyys ERICissä asettaa vastaavanlaisia velvollisuuksia kuin ELIXIR asetus
VAHTI 2/2010 Ohje tietoturvallisuudesta valtionhallinnossa annetun asetuksen täytäntöönpanosta	Velvoittava/Hyödyllinen riippuen organisaatiosta	Ohje korvaa aiemmat ohjeet Valtionhallinnon tietoineistojen käsittelyn tietoturvasuosohje (VAHTI 2/2000), Salassa pidettävien tietojen ja asiakirjojen turvaluokittelu- ja merkintä-ohje (VM 5/01/2000) sekä Arkaluonteiset kansainväliset tietoineistot (VAHTI 4/2002).
VAHTI 3/2012 Teknisen ICT-ympäristön tietoturvaso-ohje	Velvoittava/Hyödyllinen riippuen organisaatiosta	Ohje valtionhallinnon tietohallinto- ja tietoturvahenkilöstön tietoturvasojen teknisen toteuttamisen tuki.
VAHTI 1/2013 Sovelluskehityksen tietoturvaohje	Velvoittava/Hyödyllinen riippuen organisaatiosta	Ohjeen tavoitteena on ohjata sovelluskehittäjiä valtionhallinnossa ottamaan huomioon tietoturva-vaatimukset alusta lähtien.
Avoin tiede ja tutkimus -hanke	Ohjaava	Tiedon, menetelmien ja julkaisujen avoimuutta edistävä OKM:n hanke 2014-2017
Valtakunnallinen biopankkien kokonaisarkkitehtuuri	Ohjaava	Biologisten näytteiden ja niihin liittyvien tietojen hallinta biopankkitutkimuksessa
Valtakunnallinen sote-kokonaisarkkitehtuuri	Ohjaava	Työnalla oleva sosiaali- ja terveydenhuollon kokonaisarkkitehtuuri sote-tiedon hyötykäytöstä
Sähköisen asioinnin viitearkkitehtuuri	Ohjaava	Julkisen hallinnon sähköisen asioinnin viitearkkitehtuuri, jossa on lähtökohtana asiakaslähtöisyys ja palvelukeskeisyys
JHKA yhteiset ICT palvelut arkkitehtuuri	Ohjaava	Useammalla kohdealueella käytössä olevien palvelujen arkkitehtuuri
Perustietovarantojen viitearkkitehtuuri	Ohjaava	Perustietovarantojen suunnittelua ja kehittämistä ohjaava viitearkkitehtuuri
Arkistolaitoksen SÄHKE metatietomalli	Ohjaava	Asiakirjojen metatietomalli
Tutkimuksen kokonaisarkkitehtuuri	Ohjaava	OKM:n Avoin tiede ja tutkimus -ohjelman kokonaisarkkitehtuuri: valmisteilla
Tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuri	Ohjaava	Tieteellisen laskennan kokonaisarkkitehtuuri

Nimi	Velvoittavuus	Kuvaus
Suomen tutkimusinfrastruktuurien strategia ja tiekartta 2014-2020	Ohjaava	Kokonaiskuva tutkimusinfrastruktuurien ekosysteemin kehitysvaiheesta tukee ministeriöiden ja elinkeinoelämän toimintaa.
Tutkimuslaitosten ja korkeakoulujen yhteistyön kehittämisprosessi KOTUMO	Ohjaava	KOTUMOn, kahdeksan ministeriön, yliopistojen, ammattikorkeakoulujen ja tutkimuslaitosten työryhmä tuottaa tiekartan helmikuun 2015 loppuun mennessä.
Terveysalan tutkimus- ja innovaatio-toiminnan kasvustrategia	Ohjaava	TEM, STM ja OKM sekä Tekesin ja Suomen Akatemian yhteistyössä tutkimus- ja innovaatio-rahottajien sekä terveysalan toimijoiden kanssa laatima kasvustrategia
SOTE-tieto hyötykäyttöön strategia 2020	Velvoittava	Strategiassa on olennaista saada sote-tieto hyötykäyttöön ja jalostaa sitä tietämykseksi, joka auttaa niin palvelujärjestelmää kuin yksittäistä kansalaistakin.
Genomistrategia	Ohjaava	Strategian alustavat toimenpide-ehdotukset
Kansallisen palveluväylän viitearkkitehtuuri	Huomioitava	Viitearkkitehtuuri ohjaa kansallisesti tietojen ja palvelujen yhdistämisen ratkaisukokonaisuuksia ja tiedonvälityksen välineiden kehittämistä
Sähköinen henkilöiden tunnistus	Huomioitava	Henkilöt voidaan tunnistaa erilaisilla sähköisillä menetelmillä kuten HAKA, TUPAS tai henkilövarmenne.

Taulukko 2: Sidoshankkeet

Nimi	Kuvaus
Sosiaali- ja terveydenhuollon asiakas- ja potilastiedon toissijaista käyttöä koskevaa lainsäädäntöä valmisteleva työryhmä	SOTE-tieto hyötykäyttöön strategian toimeenpanoon liittyvä lakihanke. Lisäksi tarvitaan muutoksia voimassa olevaan lainsäädäntöön koskien sähköisiä asiakas- ja potilasasiakirjoja.
Tryggve-projekti	Pohjoismaisen arkaluonteisen tiedon käsittelyn yhteistyöhanke
Euro-Bioimaging-hanke	Eurooppalainen bioalojen ja lääketieteen kuvantamisteknologioiden tutkimusinfrastruktuuri
FMAS-hanke	Palvelunkehityshanke helpottamaan rekisteritietojen ja tilastoaineistojen tutkimuskäyttöä
CESSDA-hanke	Yhteiskuntatieteellisen tutkimuksen aineistojen arkistoija on CESSDA-hankkeen palveluntuottaja
Akusti-hanke	Sote-tiedon toissijaisen käytön arkkitehtuuri

ECRIN	Eurooppalainen klinisen tutkimuksen verkosto, joka erityisesti tukee monikansallisia klinisiä tutkimushankkeita.
-------	--

3.4. Arkkitehtuuriperiaatteet

Viitearkkitehtuurin ja siihen liittyvien palvelujen keskeiset suunnittelun ja toteutuksen peruskivinä toimivat linjaukset on seuraavassa koottu arkkitehtuuriperiaatteiksi. Seuraavassa taulukossa on kuvattu periaatteet toiminta- (TO), tieto- (TI), järjestelmä- (JA) ja teknologia-arkkitehtuureille (TE).

Taulukko 3: Arkkitehtuuriperiaatteet

	Nimi	Kuvaus	Tärkeys (1-5)
TO-1	Lakien noudattaminen	Viitearkkitehtuurissa noudatetaan lakeja ja säännöksiä. Erityisesti kiinnitetään huomiota palveluja tuotettaessa siinä käsiteltäviin aineistoihin liittyvä lainsäädäntö.	5
TO-2	Tietosuojasta ja tietoturvasta huolehtiminen kuuluu kaikille	Jokainen osallinen on vastuussa omalta osaltaan tietosuojasta ja tietoturvasta mukaan lukien palveluntuottajan ja asiakasorganisaatioiden johto ja koko henkilöstö sekä loppukäyttäjät.	5
TO-3	Koko hyödyntämisketjun huomiointi	Tietosuojan ja tietoturvan tulee säilyä koko hyödyntämisketjun	5
TO-4	Elinkaaren huomiointi	Huomioidaan tutkimuksen ja tiedon koko elinkaari.	4
TO-5	Läpinäkyvyys	Toiminta on avointa.	4
TO-6	Palvelukeskeisyys	Kokonaisuus muodostuu yhteentoimivista palveluista.	3
TI-1	Tietojen luottamuksellisuus	Tietosuojan säilyminen on välttämättömyys tutkimukselle.	5
TI-2	Tietojen luokittelu	Tiedot luokitellaan arkaluonteisuuden, laadun ja elinkaaren mukaan. Arkaluonteisuus luokitellaan tapauskohtaisesti aineiston tunnistettavuuden perusteella.	4
TI-3	Tietojen hyödynnettävyys	Arkaluonteinen tieto on hyödynnettävissä, kun tiedon luottamuksellisuus ja tutkittavan tietosuojan varmistetaan.	4
TI-4	Tietojen pysyvyys	Tutkimustiedon validointi ja verifiointi edellyttää tiedon pitkäaikaissäilytystä.	4



Nimi		Kuvaus	Tärkeys (1-5)
TI-5	Keskitettyjen tietovarantojen käyttäminen	Kaikelle tiedolle pyritään määrittämään autoritatiivinen päälähte ja huomioimaan se tiedon käytössä.	3
TI-6	Tiedon kohteiden edunvalvonta	Kahden tai useamman tietoaineiston yhdistäminen tutkimustarkoitusta varten täytyy tehdä tavalla, joka ei vaaranna tiedon luottamuksellisuutta ja tutkittavien tietosuojaa. Yhdistämisen voi tehdä esimerkiksi luotettava kolmas osapuoli.	3
JA-1	Tietoturvan ja riskien arvioinnin ajantasaisuus	Tietoturvan toteutuksessa hyödynnetään jatkuvasti kehityksen tuomia mahdollisuuksia ja uusiin uhkiin reagoidaan.	5
JA-2	Tietomurtoon varautuminen	Tietomurtoon varautuminen ja vahinkojen minimointi ovat keskeisiä suunnitteluperiaatteita.	4
JA-3	Yhden luukun -periaate	Pyritään tuomaan palvelut kunkin toimijaryhmän saatavilla yhdestä paikasta.	4
JA-4	Modulaarisen suunnittelun suosiminen	Kokonaisuus tulee jakaa itsenäisesti toimiviin, loogisiin osiin, jotka kommunikoivat keskenään standardilla tavalla.	3
JA-5	Teknologiariippumattomuus	Palvelut toteutetaan teknologiariippumattomiksi.	2
TE-1	Yhteentoimivuus	Tietojen käsittely ja siirtäminen toteutetaan tavalla, joka takaa palveluiden välisen yhteentoimivuuden ja organisaation ulkopuolisen tutkimuskäytön.	4
TE-2	Skaalautuvuus, laajennettavuus, muovautuvuus	Toimintaympäristön merkittävään kasvuun, laajenemiseen ja muutoksiin tulee varautua.	3

Kyseisen kohdealueen toiminnan ja tietojärjestelmien kehittäjien sekä toteuttajien tulee huomioida kaikessa kehittämisessä yllä kuvatut arkkitehtuuriperiaatteet.

3.5. Tietoturvaperiaatteet

Seuraavassa on listattu viitearkkitehtuurin keskeiset tietoturvaperiaatteet.

**Taulukko 4: Hallinnollinen turvallisuus**

Nimi	Kriittisyys	Kuvaus
Luokittelu ja riskiarviointi	Kriittinen	Tiedot luokitellaan arkaluonteisuuden mukaan ja tietovuodon vaikutukset arvioidaan.
Tietoturvapoikkeamista raportointi	Kriittinen	Tietoturvapoikkeamien käsittelyyn on asianmukaiset prosessit kuten raportointi johdolle ja sidosryhmille sekä asianmukaiset toimenpiteet poikkeamista toipumiseen.
Toimijoiden tunnistaminen ja auktorisointi	Kriittinen	Kaikki toimijat on tunnistettava ja heidän käyttöoikeutensa on hallittava asianmukaisella luotettavuudella.
Roolien asettaminen	Tärkeä	Palvelun käyttäjille on määritelty selkeät roolit (esim. omistaja, vastuullinen vetäjä, jäsen).
Luotettavat ja sertifioidut toimijat	Tärkeä	Toimijoiden on täytettävä palvelun elinkaaren ajan ennalta määritellyt turvallisuusvaatimukset siten, että toteutus on myös arvioitu luotettavan arvioijan toimesta. Suositeltavaa tasoa edustaa hallintajärjestelmän sertifiointi ISO 27001 –standardin mukaisesti. Toimijalla on oltava ajantasainen suunnitelma arvioinneissa havaittujen puutteiden korjaamisesta. Korotettua tietoturvasoaa edellytetään arkaluonteisen tiedon turvallisuuden varmistamiseksi.
Palvelutasosopimukset	Tärkeä	Palvelun tuottajan tehtävät on määritelty palvelutasosopimuksessa.
Turvallisuussopimukset organisaatioiden välillä	Tärkeä	Palvelua käyttävien organisaatioiden välille on laadittu turvallisuussopimukset, joissa määritellään eri osapuolien vastuut, velvollisuudet ja sanktiot.

Taulukko 5: Henkilöstöturvallisuus

Nimi	Kriittisyys	Kuvaus
Palvelun käyttäjien koulutus	Tärkeä	Käyttäjien ja henkilöstön on ymmärrettävä ympäristön käyttöpolitiikka ja oma roolinsa tietoturvallisuuden varmistamisessa, sekä sitouduttava sopimuksiin.

Nimi	Kriittisyys	Kuvaus
Luotettavat ja sitoutuneet palvelun ylläpitäjät	Tärkeä	Ylläpitäjien on ymmärrettävä, että kyseessä on arkaluonteinen aineisto ja toimittava sen mukaisesti. Ylläpitäjät sitoutuvat salassapitosopimukseen. Sitoutumista voidaan tukea osaamissertifiointien avulla.
Organisaation vastuu henkilöstönsä osaamisesta ja tehtävistä	Tärkeä	Henkilöstön rekrytointi- ja irtisanoutumisprosesseissa huomioidaan turvallisuusvaatimukset asianmukaisesti ja henkilöt ovat päteviä ja sitoutuneita tehtäviinsä.
Vaaralliset työyhdistelmät	Kriittinen	Vaaralliset työyhdistelmät on tunnistettava ja niiden muodostuminen estetään.
Virheiden välttäminen	Kriittinen	Kriittisissä toiminnoissa suunnittelu- ja toteutusvirheiden riskiä pienennetään muutoshallinnan sekä teknisten kontrollien (esim. eheystestaus) avulla.

Taulukko 6: Fyysinen turvallisuus

Nimi	Kriittisyys	Kuvaus
Palvelun tuottajan fyysinen turvallisuus oltava riittävä	Kriittinen	Pääsy laittiloihin ja ylläpitotiloihin on suojattu vaatimusten ja luokittelun mukaisesti.

Taulukko 7: Laitteistoturvallisuus

Nimi	Kriittisyys	Kuvaus
Tiedonsiirto	Tärkeä	Arkaluonteista aineistoa voidaan siirtää ainoastaan luotettujen ympäristöjen välillä. Luottamusketju pidetään eheänä.
Luottamusketjun varmistus	Kriittinen	Yhteydenotot sallitaan vain luotettavaksi osoitetuista järjestelmistä
Laitteiden turvallisuus	Tärkeä	Organisaatioiden on varmistettava laitteidensa tietoturvallisuuden ajantasaisuus ja kattavuus.
Laitteiden poisto	Tärkeä	Kaikkien laitteiden käytöstä poisto on tehtävä tietoturvallisesti ja valvotusti.

**Taulukko 8: Ohjelmistoturvallisuus**

Nimi	Kriittisyys	Kuvaus
Ohjelmistojen tietoturva	Kriittinen	Uudet ohjelmistot testataan testiympäristössä ennen käyttöönottoa ja kaikki tietoturvapäivitykset tehdään viivyttämättä.
Käyttäjän ohjelmistot	Kriittinen	Käyttäjän toimesta ympäristöön tuotavien ohjelmistojen ja alustojen riskejä on pyrittävä arvioimaan kattavasti.
Järjestelmäpäivitykset	Kriittinen	Käyttöjärjestelmät ja ohjelmistot päivitetään säännöllisesti ja hallitusti. Vastuutaho on määritelty.

Taulukko 9: Tietoliikenneturvallisuus

Nimi	Kriittisyys	Kuvaus
Arkaluonteisen tiedon siirron esto pois suojatusta ympäristöstä	Kriittinen	Teknisesti estetään luvattomat tietojen siirrot suojatun ympäristön ulkopuolelle. Jäännösriskin hallintaan käytetään teknistä valvontaa ja turvallisuussopimuksia ja -sitoumuksia.
Salatut yhteydet ja tiedonsiirtokanavat	Kriittinen	Kaikki tietoliikenneyhteydet ovat asianmukaisesti salatut.
Tietoliikenteen ja palvelun käytön kirjanpito	Tärkeä	IT-järjestelmien ylläpitoon ja tietoturvaan liittyviä tavanomaisia lokeja kerätään ja analysoidaan riittävästi. Lokien pitkäaikais säilytyksestä on huolehdittava.
Yhteydenotto palveluihin	Tärkeä	Käyttö sallittu vain määritellyistä IP-osoitteista. Riskiarvion perusteella voidaan käyttäjälle antaa rajattu pääsy suojattuun ympäristöön.
Tietoturva-anomalioiden reaaliaikainen havainnointi	Tärkeä	Reaaliaikainen mahdollisia tietoturvaongelmia valvova järjestelmä on käytössä.

Taulukko 10: Tietoaineistoturvallisuus

Nimi	Kriittisyys	Kuvaus
Jatkuvuus- ja toipumissuunnitelmat	Tärkeä	Jatkuvuus- ja toipumissuunnitelmissa huomioidaan aineistojen arkaluonteisuus. Aineistot suojataan sekä normaali- että poikkeusoloissa.

Nimi	Kriittisyys	Kuvaus
Luovutettujen aineistojen kirjanpito	Tärkeä	Luovuttajan käsiteltävä kirjanpitoa arkaluonteisena aineistona, joka tallennetaan luotettavasti.
Tietojen luovutus	Tärkeä	Tutkimukseen luovutetaan vain tarvittava tieto siten, että luottamuksellisen tiedon epäsuora tunnistaminen ei ole käytännössä mahdollista.

Taulukko 11: Käyttöturvallisuus

Nimi	Kriittisyys	Kuvaus
Vahva tunnistaminen	Kriittinen	Henkilöiden identiteetin tunnistuksessa käytetään useaa yhtäaikaista todentamistapaa.
Rooliperusteinen pääsy- ja käyttöoikeuksien hallinta	Kriittinen	Kaikki käyttövaltuudet myönnetään roolien perusteella. Käyttäjä liittyy rooliin henkilökohtaisen tunnuksen perusteella.
Virus ja haittaohjelmatusunnistukset	Kriittinen	Ympäristön palvelimilla tehdään automaattisesti virus- ja haittaohjelmakan-nausta.
Pääsy- ja käyttöoikeuksien peruutus	Tärkeä	Organisaatiolla on mahdollisuus nopeasti estää oman organisaationsa henkilöltä pääsy tietoa-aineistoon.
Etäkäyttö	Tärkeä	Sallitaan vain organisaatioiden varmistamista päätelaitteista.

4. Toiminta-arkkitehtuurin arkkitehtuurikuvaukset

Toiminta-arkkitehtuurin keskeisimmät viitearkkitehtuurilinjaukset ovat: - Kuvataan oleelliset toimijat ja prosessit - Toimitaan valtakunnallisesti ja tieteenalariippumattomasti yhdenmu-kaisesti - Edistetään kansallista ja kansainvälistä yhteistyötä - Toteutetaan yhden luukun periaate tutkittavalle, tutkijoille ja yrityksille
--

4.1. Sidosryhmät

Viitearkkitehtuuria laadittaessa on tunnistettu seuraavassa taulukossa listatut sidosryhmät.

Taulukko 12. Sidosryhmät

Sidosryhmä	Kuvaus
Tietojen kohde	Arkaluonteisen tiedon kohde. Kohde voi olla luonnollinen tai oikeudellinen henkilö, kuten yritys tai yhteisö.
Edustaja	Osapuoli, joka on valtuutettu edustamaan tietojen kohdetta.
Aineiston käyttäjä	Henkilö tai organisaatio, joka käyttää arkaluonteista tietoa tutkimustarkoituksessa. Esimerkiksi viranomainen, tutkimuslaitos, tai yksittäinen tutkija.
Aineiston tuottaja	Henkilö ja organisaatio, joka tuottaa arkaluonteista tietoa. Esimerkiksi viranomainen, tutkimuslaitos, rekisterin ylläpitäjä, tutkija, tai tietojen kohde.
Taustaorganisaatio	Tietojen käyttäjän tai tuottajan organisaatio. Esimerkiksi yliopisto tai tutkimuslaitos, jossa aineistoa käyttävä tutkija työskentelee.
Palveluntarjoaja	Osapuoli, joka tuottaa arkaluonteisin tiedon käsittelyyn, tallentamiseen, tms. liittyviä palveluita. Tutkimusinfrastruktuurit ja biopankit ovat palveluntarjoajia.
Rekisterinpitäjä	Organisaatio tai henkilö, joka ylläpitää henkilörekisteriä ja luovuttaa siitä tietoja tutkimukseen. Henkilörekisterillä tarkoitetaan tässä sekä luonnollisia että oikeushenkilöitä koskevaa tietoa. Esimerkiksi tutkimuslaitokset, sairaalat, biopankit, sekä erilaiset viranomaiset ja tutkijat.
Tietojärjestelmien toteuttaja	Osapuoli, joka toteuttaa tietojärjestelmiä, joilla käsitellään arkaluonteista tietoa.
Valvoja	Osapuoli, joka valvoo tai auditoi arkaluonteisen tiedon käsittelyn muita osapuolia. Esimerkiksi tietosuojavaltuutettu, Valvira sekä tutkimuseettiset lautakunnat.
Tiedon omistaja	Osapuoli, joka omistaa arkaluonteisen tiedon. Esimerkiksi viranomainen, tietojen kohde tai tietojen kohteen edustaja.
Tutkimusinfrastruktuuri	Osapuoli, jolla on keskeistä osaamista edustamallaan tutkimusalueella ja/tai tuottaa tutkimusyhteisölle keskeistä palvelua.

4.2. Substanssipalvelut

Viitearkkitehtuuria laadittaessa on tunnistettu seuraavat substanssitoiminnan palvelut. Palvelujen toteuttaminen fokuoituihin käyttötarkoituksiin edellyttää

kussakin käyttötapauksessa käsiteltäviin aineistoihin liittyvän lainsäädännön selvittämisen ja huomioimisen teknisessä toteutuksessa. Kaikissa käyttötapauksissa ei välttämättä tarvita kaikkia substanssipalveluita saman sisältöisinä.

Taulukko 13. Substanssitoiminnan palvelut

Palvelu	Kuvaus
IT-infrastruktuuripalvelu	Palvelu, jonka avulla tutkimusorganisaatio pystyy hallitsemaan siihen kohdistettujen vaatimusten mukaisesti hallinnoimaansa infrastruktuurinomaista aineistoa ja tarjoamaan dataan liittyviä palveluja tutkimukselle ja tarvittaessa muillekin toimialoille.
Tiedonhakupalvelu	Palvelu tietojen saatavuuden selvittämiseen julkisten metatietojen avulla, tai tarkan arkaluonteisen tiedon avulla. Palvelu voi olla tietotekninen palvelu tai tietyn palveluntarjoajan toimittama palvelu.
Tutkimuslupapalvelu	Kaikkien osapuolien yhteinen tutkimuslupien käsittelypalvelu tutkimussuunnitelman luomisesta, tutkimuslupien, eettiset lupien, aineistojen käyttölupien ym. myöntämiseen.
Suostumustenhallintapalvelu	Tutkittavien suostumusten pyytämiseen, antamiseen, peruuttamiseen ja muuhun hallintaan tarkoitettu palvelu.
Luovutustietopalvelu	Aineistojen luovutusten kirjaamiseen, seurantaan ja raportointiin tarkoitettu palvelu.
Rekisteriselostepalvelu	Rekisteriselosteiden hallintaan tarkoitettu palvelu.
Esikäsittelypalvelu	Tietojen esikäsittelyyn, kuten usean aineiston yhdistämiseen, koodaamiseen, aineiston tunnistettavuuden heikentämiseen ja muuhun vastaavaan tarkoitettu palvelu.
Tiedonsiirtopalvelu	Palvelu tietojen turvalliseen siirtoa varten tutkimusympäristöön infrastruktuurin sisällä.
Turvallinen tutkimusympäristö	Palvelu, joka tarjoaa tutkijoiden käyttöön arkaluonteisen tiedon käsittelyyn soveltuvan tutkimusympäristön.
Analysointipalvelu	Aineiston analysointiin tarkoitettu palvelu, joka voi olla esim. ohjelmistopalvelu tai asiantuntijapalvelu.
Saatavuuspalvelu	Tutkimustuotosten kuten sähköisten tietoaaineistojen säilytykseen tarkoitettu palvelu.
Julkaisupalvelu	Tutkimustulosten julkaisupalvelu.
Neuvontapalvelu	Arkaluonteisen tiedon tutkimuksen neuvontapalvelu.

Palvelu	Kuvaus
Koulutuspalvelu	Arkaluonteisen tiedon tutkimuskäytön koulutuspalvelu.
Pätevöitymispalvelu	Palvelu tutkijoiden ja muiden sidosryhmien kouluttamiseksi arkaluonteisen tiedon käsittelyyn ja osaamisen testaamiseksi.
Tutkittavien rekrytointipalvelu	Palvelu, jonka avulla voidaan rekrytoida tutkittavia henkilöitä tai yhteisöjä.

Tutkimusinfrastruktuuri voi tuottaa yhtä tai useampaa yllä luetelluista palveluista.

4.3. Toimijat ja roolit

Viitearkkitehtuurin toimijat ja roolit ovat pääsääntöisesti samat kuin sidosryhmät. Seuraavassa toimijat ja roolit on kuvattu.

Taulukko 14. Toimijat

Toimija	Kuvaus
Tutkija	Henkilö, joka käyttävät arkaluonteista tietoa tutkimustarkoituksessa.
Tutkittava	Henkilö, jolta arkaluonteista tietoa kerätään hänen suostumuksellaan.
Taustaorganisaatio	Kuten edellä (Taulukko 12).
Palveluntarjoaja	Kuten edellä (Taulukko 12).
Rekisterinpitäjä	Kuten edellä (Taulukko 12).
Tutkimusinfrastruktuuri	Kuten edellä (Taulukko 12).
Valvoja	Kuten edellä (Taulukko 12).

Taulukko 15. Roolit

Rooli	Kuvaus
Tietojen kohde	Kuten edellä (Taulukko 12).
Luvanmyöntäjä	Valvoja, joka voi myöntää luvan tutkimuksen tekemiseen.
Aineiston käyttäjä	Kuten edellä (Taulukko 12).
Aineiston tuottaja	Kuten edellä (Taulukko 12).
Aineiston luovuttaja	Osapuoli, joka luovuttaa aineistoa tutkimukseen.

Rooli	Kuvaus
Esikäsittelijä	Aineiston käyttäjä ja mahdollinen tuottaja. Esikäsittelee aineistoa joko yhdistäen useita aineistoja, heikentämällä tietojen kohteen tunnistamista jne.
Vertaisarvioija	Henkilö, joka suorittaa tieteelliseen julkaisuun liittyvää vertaisarviointia.
Palveluntarjoaja	Kuten edellä (Taulukko 12).

4.4. Prosessit

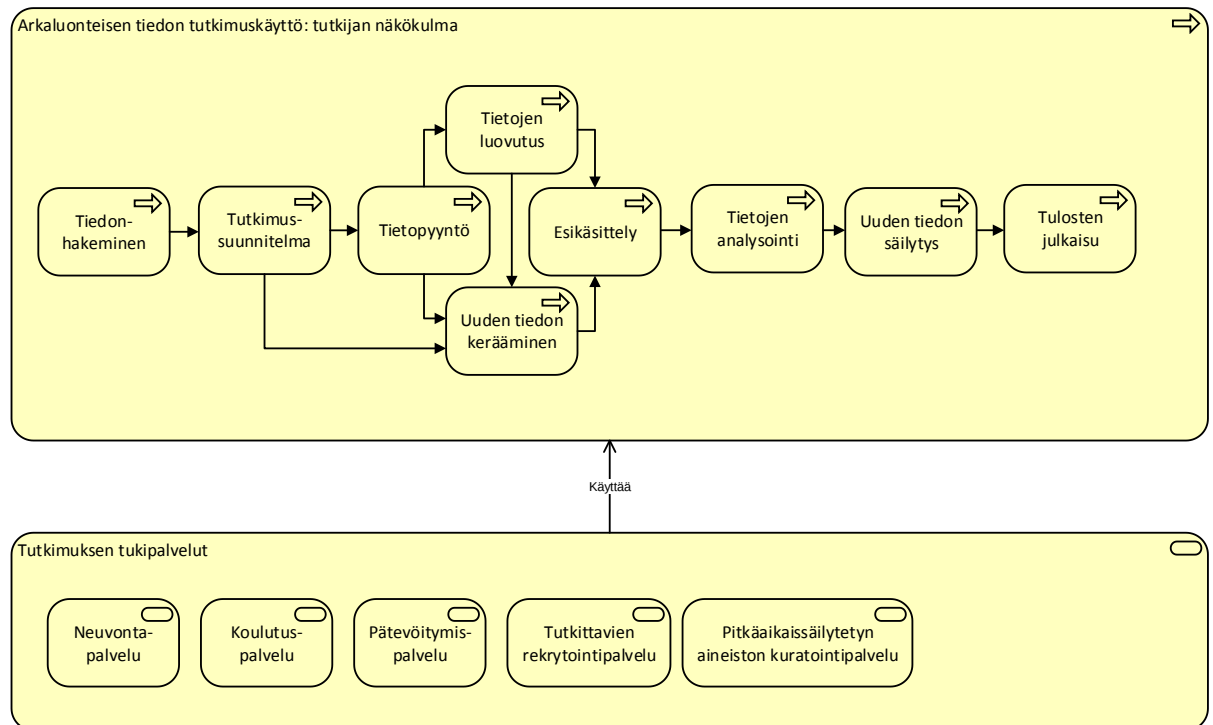
Seuraavaan on kuvattu arkaluonteisen tiedon tutkimuskäytön keskeisimmät prosessit, joita alueen palveluiden ja ratkaisujen tulee tukea.

Taulukko 16. Arkaluonteisen tiedon tutkimuskäytön keskeisimmät prosessit

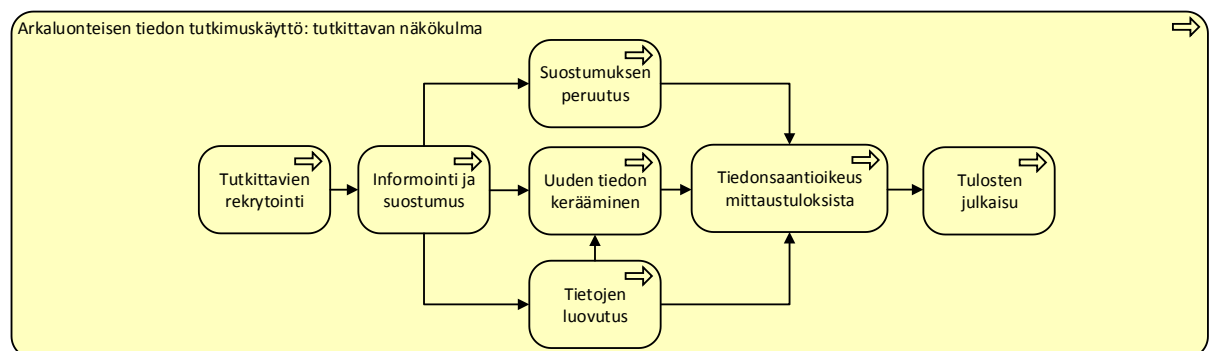
Prosessi	Selite
Tiedon hakeminen	Tutkimukseen tarvittavien olemassa olevien tietojen etsiminen tiedonhakupalvelujen avulla, minkä tarkoituksena on selvittää tutkimussuunnitelman laatimista varten, millaista tietoa on saatavilla.
Tutkimussuunnitelma	Arkaluonteista aineistoa sisältävän tutkimuksen suunnitelman laadinta ja tutkimuslupien haku sekä suunnitelma tutkimusdatan hallinnasta.
Tietopyyntö	Tutkimussuunnitelman mukaisen arkaluonteisen aineiston hankkiminen rekisterinpitäjältä.
Uuden tiedon kerääminen	Arkaluonteisen tiedon kerääminen tutkittavilta suostumuksella tai olemassa olevista tietovarannoista kokoamalla.
Tietojen luovutus	Rekisterinpitäjän muodostaman tutkimussuunnitelman ja -lupien mukaisen aineiston luovutus tutkijalle.
Esikäsittely	Arkaluonteisten tietojen valmistelu tutkimuskäyttöön: yhdistely, koodaus, tunnistettavuuden heikentäminen, anonymisointi tmv. Esikäsittely voi poistaa tiedon arkaluonteisuuden.
Tietojen analysointi	Muodostetun tutkimusrekisterin käsittely ja analysointi yms.
Uuden tiedon säilytys	Tutkimuksessa syntyneen uuden tiedon säilytys asianmukaisissa järjestelmissä huomioiden asianmukainen säilytysaika.
Tulosten julkaisu	Tutkimustulosten julkaiseminen ja tutkimustuotosten kuten sähköisten tietoaineistojen ja menetelmien kuten ohjelmistojen saatavuuden määrittely ja julkinen kuvailu.
Tiedon kuratointi eli huoltaminen	Pitkäaikaissäilytyksessä tiedon arvon, käytettävyyden ja ymmärrettävyyden säilyttäminen.

4.4.1. Prosessikartta

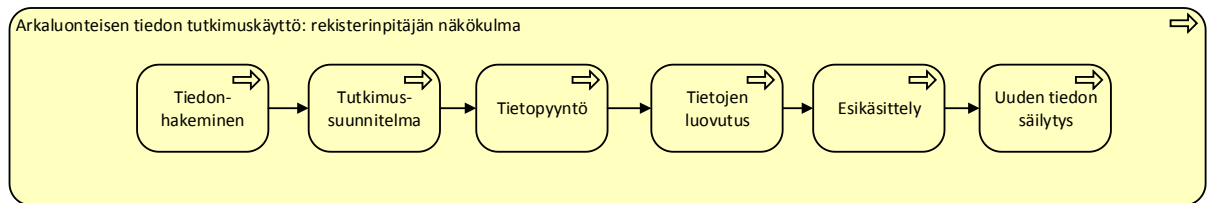
Seuraavassa prosessikartassa on kuvattu viitearkkitehtuurin prosessit ja niiden välinen siirtyminen.



Kuva 2: Arkaluonteisen tiedon tutkimuskäytön prosessikartta: tutkijan näkökulma



Kuva 3: Arkaluonteisen tiedon tutkimuskäytön prosessikartta: tutkittavan näkökulma



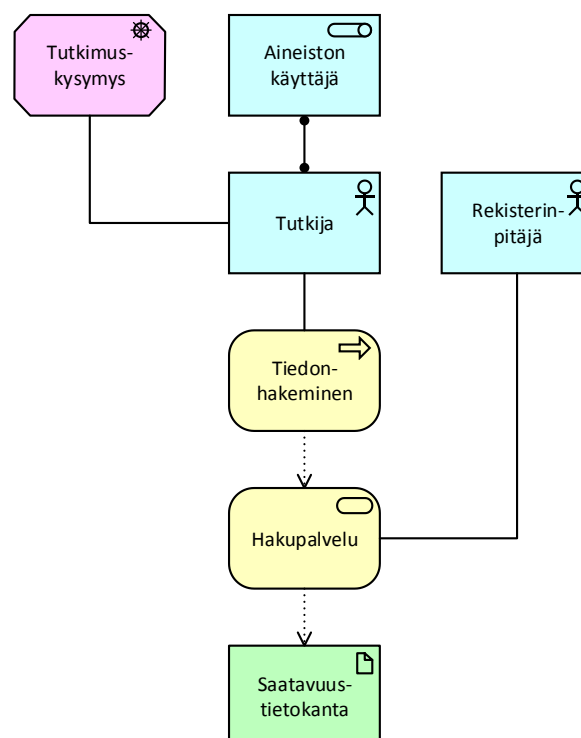
Kuva 4: Arkaluonteisen tiedon tutkimuskäytön prosessikartta: rekisterinpitäjän näkökulma

4.4.2. Prosessikuvaukset

Seuraavassa on kuvattu prosessikartan eri prosessit sekä kaavioina että tekstimuodossa.

Tiedon hakeminen

Arkaluonteisen aineiston tutkimusprosessi alkaa tiedon hakemisella. Prosessissa **toimijana** ja **aineiston käyttäjän roolissa** on **tutkija**, jonka ajurina on **tutkimuskysymys**. **Tiedonhakupalvelun** avulla tutkija selvittää millaista tietoa tutkimuskysymyksen selvittämiseen on saatavilla, ja mistä sitä on saatavilla. Tiedonhakupalvelu hyödyntää yhtä tai useampaa **saatavuustietokantaa**.

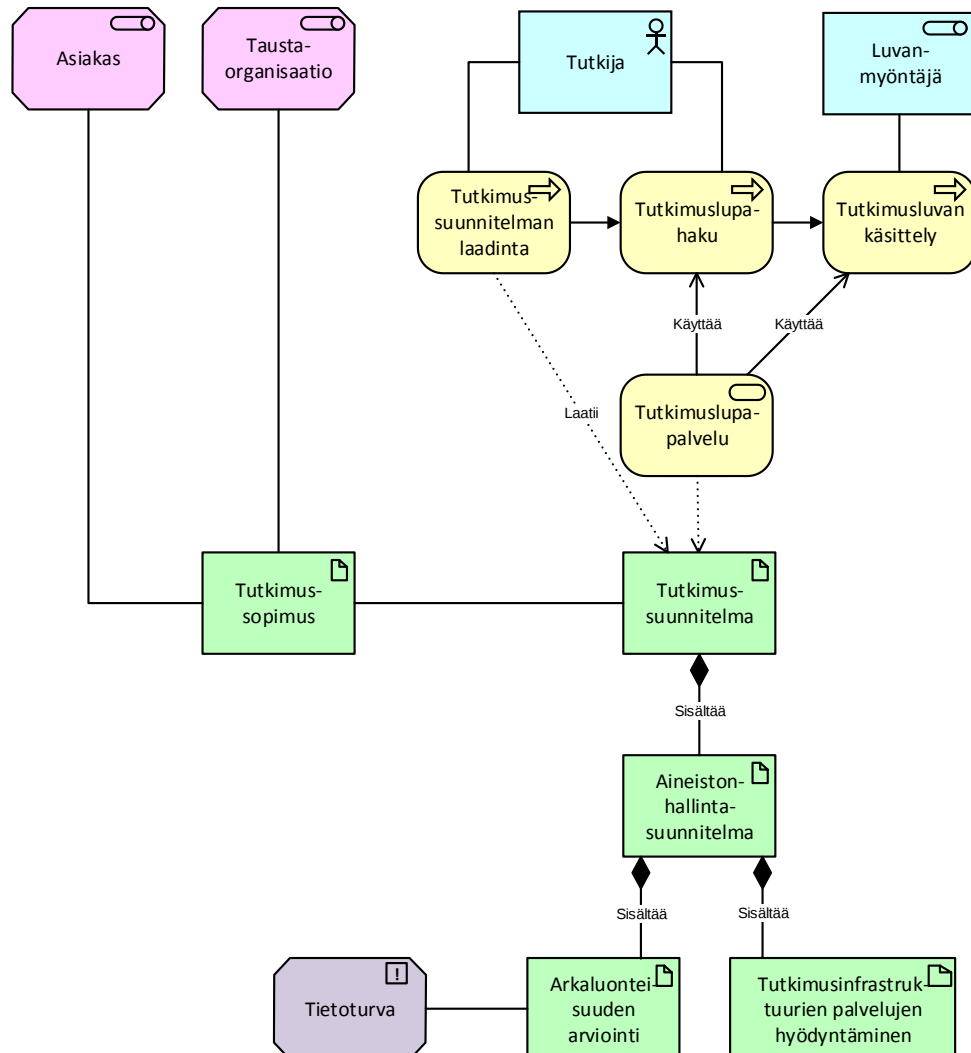


Prosessikaavio 1: Tiedon hakeminen



Tutkimussuunnitelma

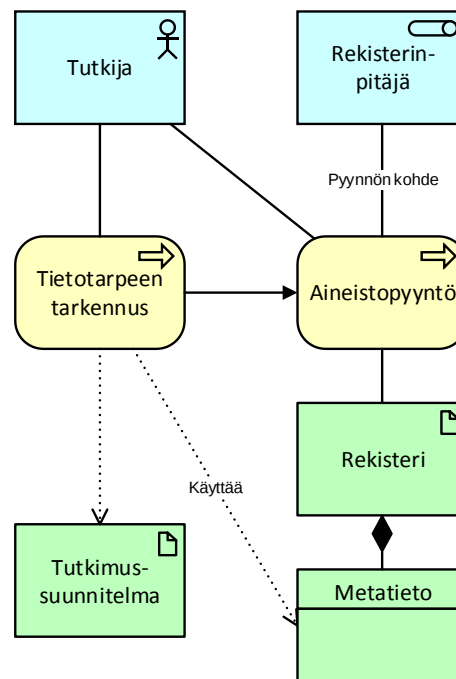
Saatavuuden selvittämisen jälkeen, **tutkija** laatii **tutkimussuunnitelman**. Suunnitelmaan voi liittyä erillinen **tutkimussopimus**, joka on tehty **taustaorganisaation** ja **asiakkaan** välille. Suunnitelman täytyy sisältää **aineistonhallintasuunnitelma**, jonka puolestaan täytyy sisältää aineiston **arkaluonteisuuden arvioinnin tutkimusinfrastruktuurien palvelujen hyödyntämisen lisäksi**. Tutkija hakee suunnitelman laadinnan jälkeen **tutkimuslupaa**. Lupa-haku osoitetaan yhdelle tai useammalle **luvanmyöntäjä** -taholle tutkimussuunnitelman mukaisesti. **Tutkimuslupapalvelussa** luvanmyöntäjä tekee päätöksen tutkimuslupan myöntämisestä käyttäen tutkimussuunnitelmaa päätöksen perusteena.



Prosessikaavio 2: Tutkimussuunnitelma

Tietopyyntö

Tutkimusluvan saamisen jälkeen **tutkija** tekee **tietotarpeen tarkennuksen** **tiedonhakupalvelun** avulla. Tarkoituksena on yksilöidä **tutkimussuunnitelman** toteuttamiseksi tarvittavat tutkimusluvan mukaiset tiedot. Tarkennuksen jälkeen tutkija tekee **rekisterinpitäjälle** kohdistetun **aineistopyynnön**.

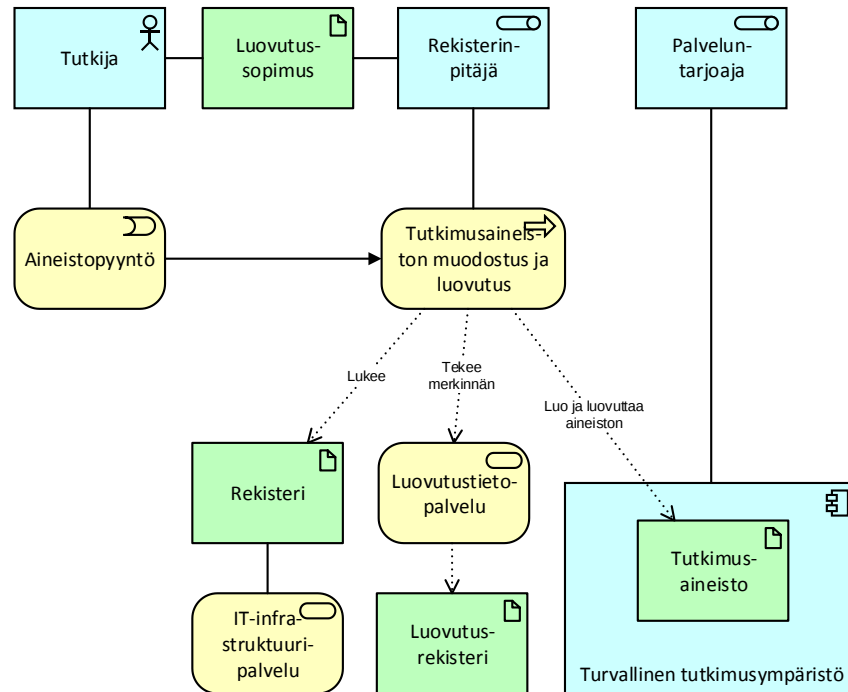


Prosessikaavio 3: Tietopyyntö

Tietojen luovutus

Tietojen luovutus -prosessi alkaa **tutkijan** tekemästä **aineistopyynnöstä**. **Rekisterinpitäjä** ja tutkijan välille on solmittu **luovutussopimus**. Rekisterinpitäjä lukee **rekisteriä**, tekee merkinnän **luovutusrekisteriin** käyttämällä **luovutustietopalvelua**, sekä luo/muodostaa **tutkimusaineiston** turvalliseen **tutkimusympäristöön** ja **asettaa asiaankuuluvat käyttöoikeudet**. Turvallisella tutkimusympäristöllä tarkoitetaan teknistä ympäristöä, joka on tutkimuksen suorittamisen näkökulmasta asianmukaisesti suojattu. Tutkimusympäristöä ylläpitää **palveluntarjoaja**.

Mikäli tiedot on tarkoitettu esimerkiksi tutkimusryhmälle, täytyy tutkijan, jolle tiedot on luovutettu, noudattaa tätä prosessia luovuttaessaan tietoja muille tutkimusryhmän jäsenille.

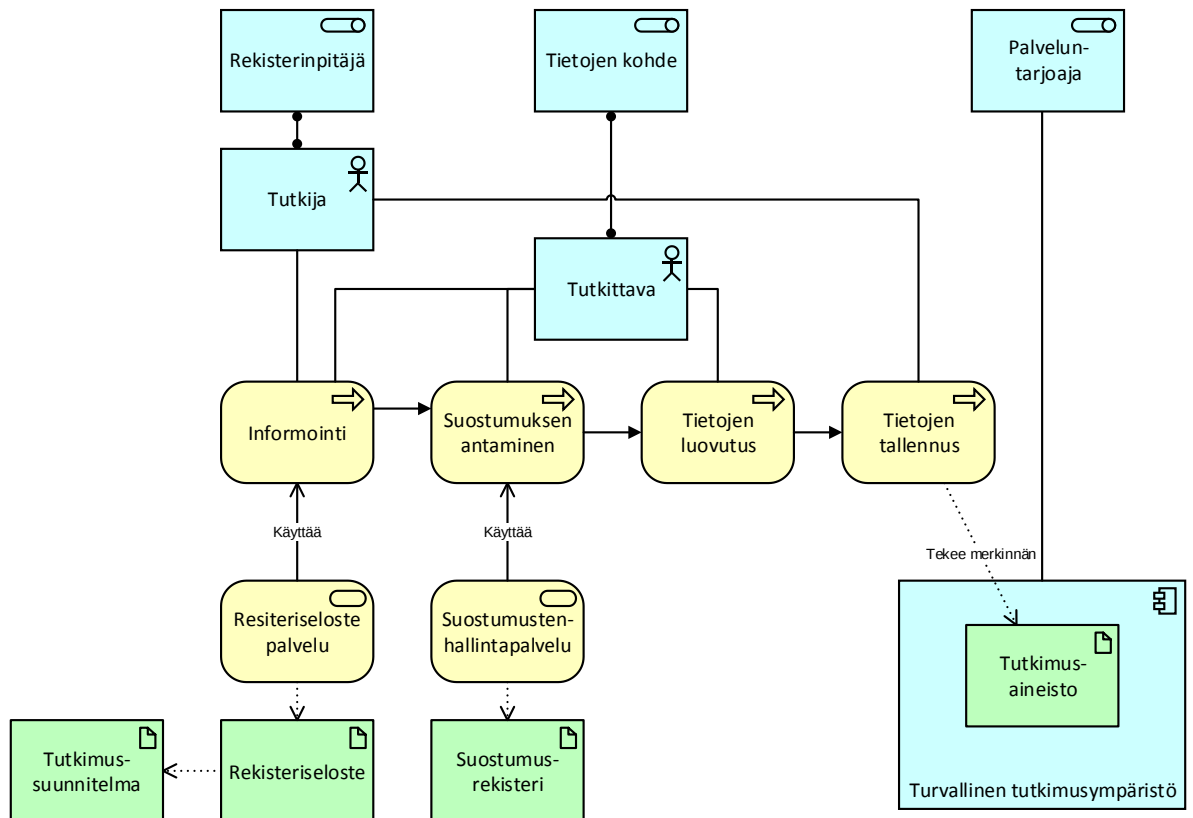


Prosessikaavio 4: Tietojen luovutus

Uuden tiedon kerääminen

Kaikissa tutkimuksissa olemassa olevaa aineistoa ei ole, sitä täytyy täydentää, tai on muutoin tutkimussuunnitelman mukaisesti kerättävä lisää. Uuden tiedon kerääminen -prosessi voi alkaa tutkimussuunnitelma, tietopyyntö ja tietojen luovutus -prosessien jälkeen.

Tutkija informoi **tutkittavaa** sen mukaan kuinka **tutkimussuunnitelmassa** on määritelty. Tutkittava **antaa suostumuksen** käyttäen **suostumustenhallintapalvelua**, joka käyttää **suostumusrekisteriä**. Suostumuksen antamisen jälkeen tutkittava **luovuttaa tiedot** tutkijalle, joka **tallentaa tiedot tutkimusaineistoon**. Muodostettua tutkimusaineistoa kutsutaan **tutkimusrekisteriksi** ja se saattaa olla henkilörekisteri, jonka **rekisterinpitäjä** on tutkija. Tämä tieto on kuvattu **rekisteriselosteessa**.



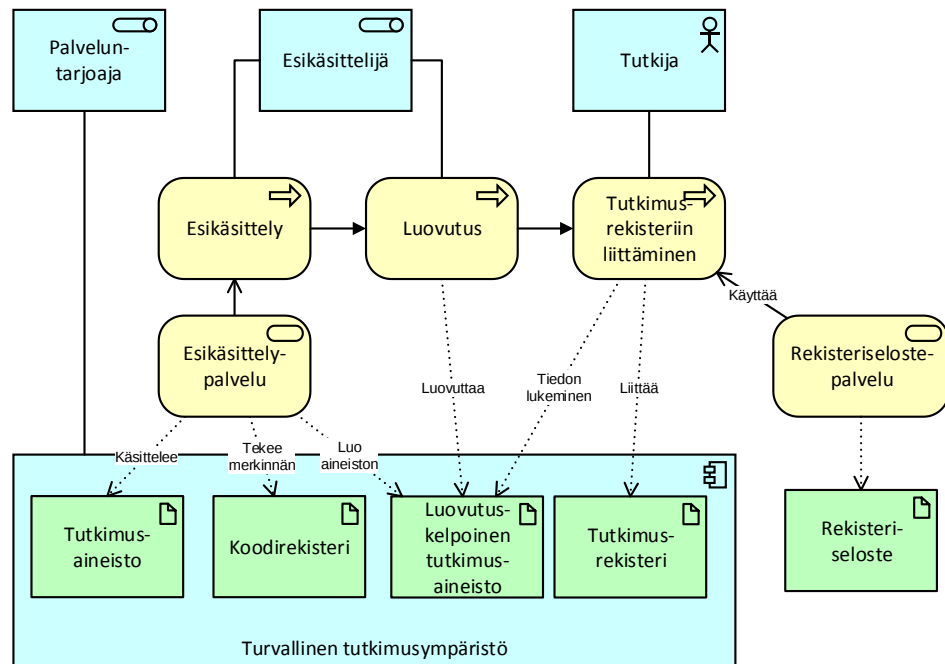
Prosessikaavio 5: Uuden tiedon kerääminen

Esikäsittely

Tutkimusaineiston luovutuksen ja/tai uuden aineiston keräämisen jälkeen aineistoa on esikäsiteltävä. Esikäsittelijä on rooli, joka suorittaa **esikäsittely** prosessin. Prosessissa käytetään **esikäsittelypalvelua**, joka käsittelee tutkimusaineistoa, tekee tarvittaessa merkinnän **koodirekisteriin** ja luo **luovutuskelpoisen aineiston**.

Esikäsittely voi pitää sisällään esimerkiksi usean tunnisteellisen aineiston tietojen yhdistämistä, tutkijan tutkimusta varten keräämien tietojen yhdistämistä muihin aineistoihin, tietojen tunnistettavuuden heikennystä, tiedon anonyymisointia, yms.

Luovutuskelpoinen aineisto luovutetaan/asetetaan käyttöoikeudet **tutkijalle**, joka **liittää tutkimuksen** tiedot omaan **tutkimusrekisteriinsä**. Tarvittaessa laaditaan myös **rekisteriseloste**.



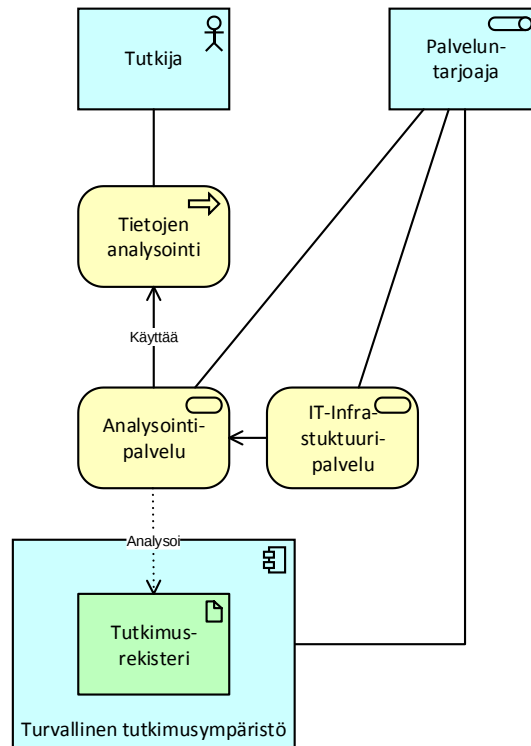
Prosessikaavio 6: Esikäsittely

Esikäsittelyn jälkeinen luovutuskelpoinen aineisto ei enää välttämättä ole arkaluonteista tietoa.



Tietojen analysointi

Tutkimusrekisterin laatimisen jälkeen **tiedot analysoidaan tutkijan** toimesta. Tutkija käyttää **analysointipalvelua** tutkimusaineiston analysointiin.

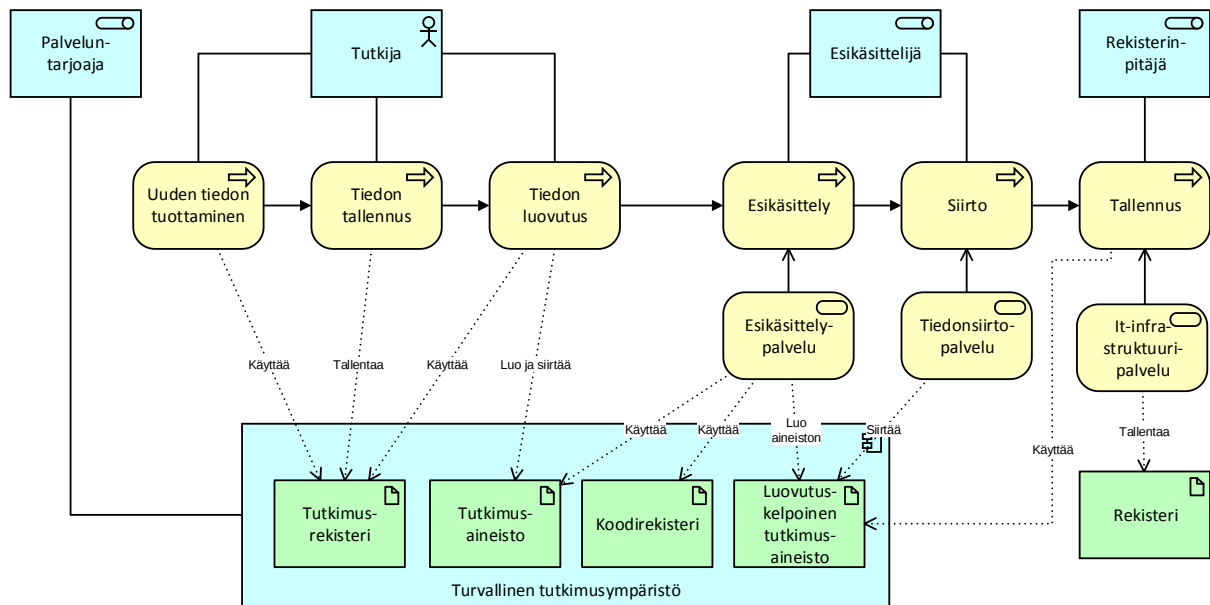


Prosessikaavio 7: Tietojen analysointi

Uuden tiedon säilytys

Analysoinnin jälkeen **tutkija tuottaa uutta tietoa** käyttäen **tutkimusrekisteriä**. Uusi tutkimuksella tuotettu **tieto tallennetaan** tutkimusrekisteriin.

Joissain tutkimuksissa osa tuloksista täytyy palauttaa rekisterinpitäjälle. Silloin tutkija luo ja **luovuttaa tutkimusaineiston esikäsittelyyn**, jossa **esikäsittelypalvelun** avulla aineistosta tehdään **luovutuskelpoista**. **Koodirekisterin** avulla esikäsittelyssä voidaan yhdistää tutkimustulokset aiemmin koodattuun alkuperäiseen tietoon. Esikäsittelijä **luovuttaa** aineiston **rekisterinpitäjälle**, joka **tallentaa** tiedot **rekisteriin**.

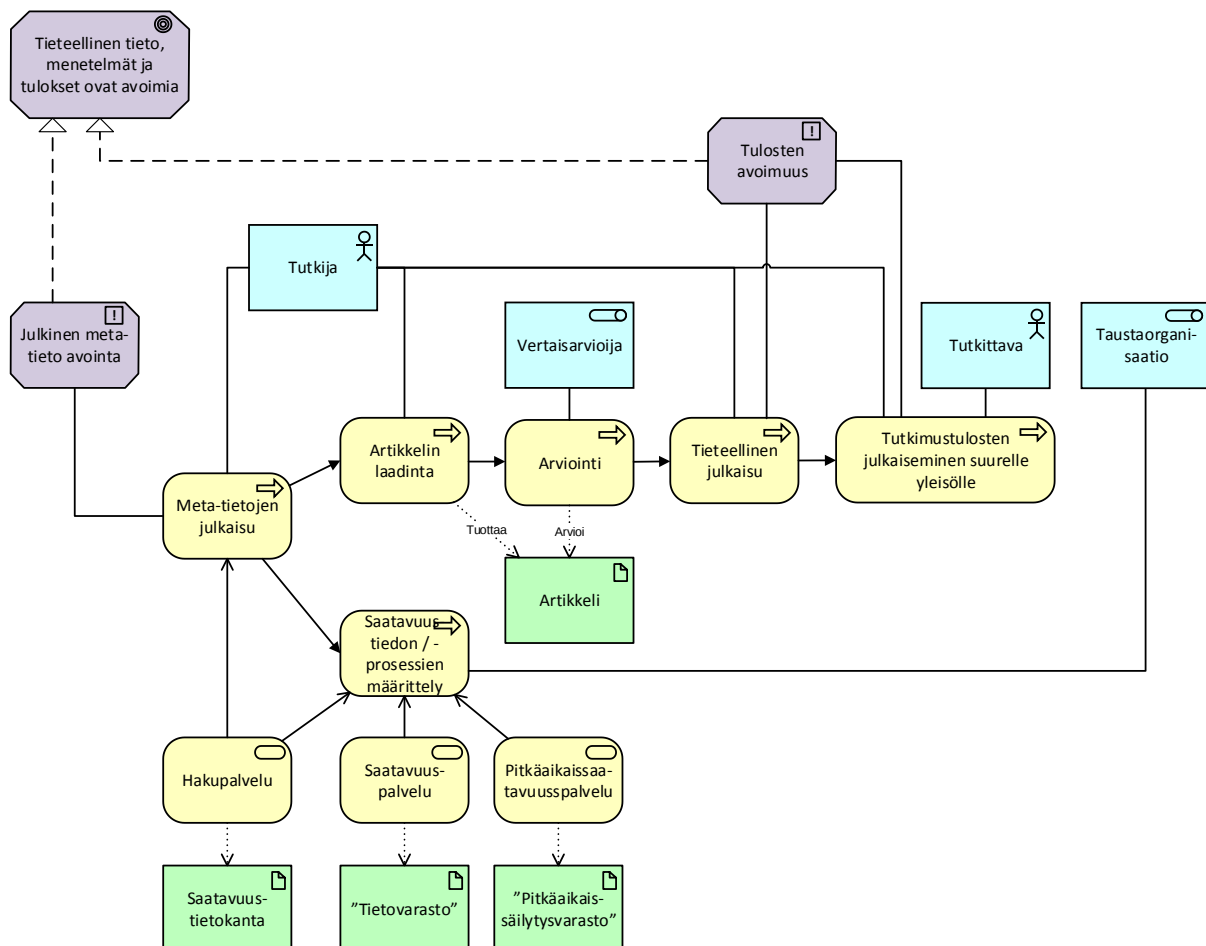


Prosessikaavio 8: Uuden tiedon säilytys

Tulosten julkaisu

Tutkimustiedon tuottamisen ja mahdollisen säilyttämisen jälkeen tulokset julkaistaan. **Tutkija** julkaisee tutkimuksen aineistojen ja menetelmien **metatiedot** käyttäen **tiedonhakupalvelua**, joka tallentaa tiedot yhteen tai useampaan **saatavuustietokantaan**. Tämän jälkeen tutkija pyrkii julkaisemaan tuloksensa tavanomaisen **tieteellisen julkaisu** -prosessin kautta. Tähän liittyy tavanomaisesti **vertaisarviointi**. Kun tieteellinen julkaisu on hyväksytty, tutkija julkaisee tutkimustulokset **yleistajuisesti suurelle yleisölle**.

Metatietojen julkaisun jälkeen **taustaorganisaatio määrittelee tiedon saatavuuden, pitkäaikaissäilytyksen** ja näihin liittyvät **prosessit** tiedon jatkokäytön mahdollistamiseksi. Tässä vaiheessa hyödynnetään **esikäsittelypalvelua** ja **saatavuuspalvelua** tietojen tallennuksessa ”**tietovarastoon**” tai ”**pitkäaikaissäilytysvarasto**”.



Prosessikaavio 9: Tulosten julkaisu

4.4.3. Keskeisimmät tehtävänmukaiset huomioitavat asiat

Käsiteltävät aineistot ja niiden käyttötarkoitukset vaikuttavat siihen, mitä säännöksiä tietojen käyttöön sovelletaan. Toimijoiden on kunkin osaltaan kiinnitettävä huomiota aineisto- ja käytötapauskohtaisesti, että tietojen luovuttajalla on lainmukainen peruste luovuttaa tietoja,



tietojen vastaanottajalla on lainmukainen peruste käsitellä henkilötietoja eikä luovutus ole vastoin käyttötarkoitussidonnaisuuden periaatetta. Tietojen luovutuksensaajan on tunnistettava asemansa rekisterinpitäjänä.

Taulukko 17. Arkaluonteisen tiedon tutkimuskäytössä tehtävämukaisesti huomioitavat asiat

Huomioitava asia	Aineiston käsittely	Aineiston tuottaminen	Aineiston luovutus
Tutkittavien rekrytointi		X	
Tutkittavan informointi		X	
Tutkittavan suostumuksen hankinta		X	
Tutkittavan suostumuksen peruutuksen käsittely	X	X	X
Tutkimus- ja aineistohallintasuunnitelman laatiminen huomioiden tietojen luottamuksellisuus	X	X	X
Tutkimuslupien haku ja myöntö	X	X	X
Aineistojen luovutusperusteiden tarkastelu			X
Aineiston esikäsittely	X	X	X
Tutkimuskuvauksen julkaisu			X
Rekisteriselosteen laadinta	X	X	
Aineiston luovutusten kirjapito			X
Aineiston tuhoaminen	X	X	X
Säilytettävän aineiston arkaluonteisuuden määrittely	X	X	X
Uuden datan säilytys arkaluonteisuuden mukaisesti	X	X	
Aineiston kattavien julkisten metatietojentuotto	X	X	X
Tutkimuksen tulosten julkaisu	X		
Tutkimuksen menetelmien julkaisu	X		
Aineiston kuratointi pitkäaikaissäilytyksessä		X	
Tutkimustulosten julkaiseminen suurelle yleisölle	X	X	X
Tietojen luokittelu ja asianmukaisen tietoturvallisuuden sekä tutkittavien tietosuojan varmistaminen	X	X	X
Palveluntarjoajien arviointi	X	X	X

Palvelun kehittäjien tulee selvittää käsiteltäviin aineistoihin ja niiden käyttötarkoituksen liittyvä lainsäädäntö ja huomioida se palvelun toteutuksessa. Palveluntarjoajan huomioitavat asiat riippuvat sen tuottamasta palvelusta. Kaikkien palveluntarjoajien tulee olla tietoisia arkaluonteisesta aineistosta palvelussaan ja varmistettava asianmukaisesti ja hyvien tietohallintatapojen mukaisesti asianmukainen tietosuoja ja tietoturva.



5. Tietoarkkitehtuurin arkkitehtuurikuvaukset

Tietoarkkitehtuurin keskeisimmät viitearkkitehtuurilinjaukset ovat:

- Tutkimustiedon kumuloituminen jatkotutkimuksista laajoilla suostumuksilla
- Tiedon pitkäaikaissäilytyskelpoisuuden varmistus noudattamalla tutkimusaineistoille määriteltäviä vaatimuksia tallennusmuotoihin, metatietoihin sekä tuotanto- ja tallennusprosesseihin liittyen
- Tiedon kuratoinnin varmistus määrittelemällä ja ylläpitämällä jokaiselle tiedolle vastuullinen omistajataho
- Tietomallinnuksen hyödyntäminen palveluinfrastruktuurin kehittämisessä eri järjestelmien sisältämien tietojen yhteismitallisuuden taakamiseksi
- Yhteisten ja/tai yhtenäisten sanastojen, tietomallien, suostumusmallien, tutkimuslupahakemusten luonti ja hyödyntäminen
- Tiedon luokittelu arkaluonteisuuden, käyttötavan, laadun ja elinkaarin mukaan
- Julkiset ja kattavat metatiedot tutkimusaineistoista
- Tiedon jäljitettävyys
- Autoratiivinen tietolähteen (master data) määrittely
- Käyttöoikeuksien määrittely aineistokohtaisesti
- Tutkimuksen ja tuotannollisten palveluiden kuten terveydenhuollon yhteisten tietovarantojen tunnistus

5.1. Sanasto

Seuraavassa on listattu viitearkkitehtuuriin liittyviä sanoja ja niiden lähteitä.

Taulukko 18. Keskeisin sanasto

Käsite	Selite	Tarkennuksia
Aineisto	Tiettyä tarkoitusta varten koottu kokoelma sähköisessä muodossa olevaa dataa, fyysisiä näytteitä, julkaisuja tai menetelmiä kuten ohjelmisto, jota käsittelemällä ja käytämällä saadaan uutta tietoa.	Biopankkitutkimuksessa aineistolla tarkoitetaan kokoelmaa näytteitä ja niihin liittyvää tietoa. Biopankkisanasto



Käsite	Selite	Tarkennuksia
Anonymisaatio Anonymisointi	Aineiston yksittäisten henkilöiden tunnistamisen tekeminen mahdottomaksi muokkaamalla dataa.	Henkilön tunnistetietojen ja näytteen tai siitä määritetyn tiedon välisen yhteyden pysyvä katkaiseminen. Biopankkisanasto
Anonymisointitieto	Anonymisoinnissa käytettyjen menetelmien kuten yhdistelyn, muuttujien poistamisen ja/tai karkeistuksen kuvaus.	



Käsite	Selite	Tarkennuksia
Arkaluonteinen tieto	Tässä viitearkkitehtuurissa arkaluonteisia tietoja ovat sekä henkilöön että muuhun kuin henkilöön liittyvät tiedot seuraavasti. Henkilöön liittyvät mitkä tahansa tiedot, joilla hänet voidaan tunnistaa suorasti tai epäsuorasti, ovat arkaluonteisia tietoja. Jos henkilöä ei voida tunnistaa, moni ominaisuus ei enää tässä tarkastelussa olekaan arkaluonteinen. Tiedossa mukana olevat ominaisuudet voivat puolestaan mahdollistaa henkilön suoran tunnistamisen, tai joukko sinällään ei-arkaluonteisia ominaisuuksia voi mahdollistaa henkilön epäsuoran tunnistamisen. Tiedon arkaluonteisuus on arvioitava tapauskohtaisesti. Muu kuin henkilöön kohdistuva tieto esimerkiksi organisaatioon liittyvä tieto on arkaluonteista, jos se on luokiteltu tai on luonteeltaan luokiteltavissa vähintään luottamukselliseksi tai salassa pidettäväksi. Tässä viitearkkitehtuurissa arkaluonteisen tiedon määritelmä on merkittävästi laajempi kuin henkilötietolain käyttämä määritelmä, jonka vaikutusalue jää henkilötietolain tulkintaan.	Henkilötietolaissa arkaluonteisina tietoina pidetään henkilötietoja, jotka kuvaavat tai on tarkoitettu kuvaamaan: 1) rotua tai etnistä alkuperää; 2) henkilön yhteiskunnallista, poliittista tai uskonnollista vakaumusta tai ammattiliittoon kuulumista; 3) rikollista tekoa, rangaistusta tai muuta rikoksen seuraamusta; 4) henkilön terveydentilaa, sairautta tai vammaisuutta taikka häneen kohdistettuja hoitotoimenpiteitä tai niihin verrattavia toimia; 5) henkilön seksuaalista suuntautumista tai käyttäytymistä; taikka 6) henkilön sosiaalihuollon tarvetta tai hänen saamiaan sosiaalihuollon palveluja, tukitoimia ja muita sosiaalihuollon etuuksia. Henkilötietolaki



Käsite	Selite	Tarkennuksia
Avoin tiede ja tutkimus	Avoimen tieteen ja tutkimuksen tavoitteena on mahdollistaa arkaluonteisen tiedon turvallinen, tehokas, yhtenäinen ja lainmukainen tutkimuskäyttö, mukaan lukien käyttö jatkotutkimuksiin ja monitieteisiin tutkimuksiin, sekä turvata aineistojen asianmukainen pitkäaikaissäilytys erityisesti siten, että varmistetaan lainmukainen tutkittavien tietosuojan ja arkaluonteisten aineistojen käsittelyjärjestelmien tietoturva. Samoin pitkäaikaissäilytyksessä otetaan huomioon arkaluonteisten aineistojen lain- ja asianmukainen säilytysaika.	
Avoin tieto	Tieto on vapaasti saatavilla ja kaikkien hyödynnettävissä internetissä.	
Henkilörekisteri	Henkilötietoja sisältävä tietokokoelma.	
Henkilötiedot	Kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä, hänen perhettään tai hänen kanssaan yhteisessä taloudessa asuvia koskeviksi tiedoiksi.	Biopankkisanasto Henkilötietolaki
Julkinen tieto	Tieto on julkista, mutta ei välttämättä sellaisenaan saatavana internetissä.	
Koodattu näyte	Näyte, jonka yhteys luonnolliseen henkilöön, sekä näytettä koskevaan tai niihin liitettävään tietoon on merkitty yksilöllisellä tunnisteella, johon ei liity luonnollisen henkilön yksilöivää tietoa.	Biopankkisanasto
Koodaus	Tietoon liittyvä koodi, jonka avulla se pystytään yhdistämään muuhun tietoon esim. henkilön yksilöivään tietoon.	
Käyttölupa	Antaa oikeuden tai pääsyn käsitellä esim. lukemalla tai analysoimalla tai muulla tavoin käyttölupaa alaista aineistoa.	
Laaja suostumus	Antaa tutkimukselle mahdollisuuden käyttää k.o. laajan suostumuksen alaista aineistoa useissa erilaisissa suostumuksessa määritellyn teeman alaisissa tutkimuksissa ilman erillistä henkilötietolain määrittelemään nimenomaista suostumusta. Laaja suostumus on mahdollinen biopankkitutkimuksessa.	



Käsite	Selite	Tarkennuksia
Muuttuja	Muuttuja on tilastotieteen keskeisiä käsitteitä, jolla viitataan johonkin mitattavan kohteen ominaisuuteen, esim. pituus, paino, poliittinen kanta. Muuttujien arvot syntyvät erilaisten mittausoperaatioiden tuloksena. Arvojen vaihtelusta syntyy muuttujan jakauma. Jos x tarkoittaa suomalaisten miesten pituusmuuttujaa, sen arvo voi olla mikä tahansa 120-220 cm:n välillä, ja suurehkoista otoksesta saatu pituusjakauma on Gaussin kellokäyrän kaltainen.	Tilastokeskus
Pitkäaikaissäilytys	Aineisto on tulevien sukupolvien käytettävissä ja ymmärrettävissä.	
Privacy by Design	Periaate, jossa yksityisyys ja tietosuoja ovat mukana teknologioiden koko elinkaareissa varhaisesta suunnittelutasosta käyttöönottoon, käyttöön ja lopulliseen hävitykseen.	
Pseudonymisointi	Henkilön yksilöivän tunnisteiden korvaaminen väliaikaisella tunnuksella. Vrt. koodaus.	
Rekisterinpitäjä	Yhtä tai useampaa henkilöä, yhteisöä, laitosta tai säätiötä, jonka käyttöä varten henkilörekisteri perustetaan ja jolla on oikeus määrätä henkilörekisterin käytöstä tai jonka tehtäväksi rekisterinpito on lailla säädetty. Tutkija on usein tutkimusrekisterinsä rekisterinpitäjä.	
Rekisteröity	Henkilö, jota henkilörekisterissä oleva henkilötieto koskee.	Henkilö, jonka tunnisteellinen näyte on otettu. Biopankkisanasto



Käsite	Selite	Tarkennuksia
Suojaustasoluokitus	Salassa pidettävien asiakirjojen luokittelussa käytetään seuraavia luokkia: 1) suojautaso I, jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa erityisen suurta vahinkoa salassapitosäännöksessä tarkoitettulle yleiselle edulle; 2) suojautaso II, jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa merkittävää vahinkoa salassapitosäännöksessä tarkoitettulle yleiselle edulle; 3) suojautaso III, jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa vahinkoa salassapitosäännöksessä tarkoitettulle yleiselle tai yksityiselle edulle; 4) suojautaso IV, jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa haittaa salassapitosäännöksessä tarkoitettulle yleiselle tai yksityiselle edulle.	Asetus (681/2010)
Suostumus	Kaikenlaista vapaaehtoista, yksilöityä ja tietoista tahdon ilmaisua, jolla rekisteröity hyväksyy henkilötietojensa käsittelyn.	Henkilötietolaki
Tilastoyksikkö	1. Tilastoyksikkö on se tilastollinen perusyksikkö, jota tilastointi koskee, esim. Suomen kansalainen, kotitalous tai kunta, mutta se voi olla myös tapahtuma, kuten rikos tai liikenneonnettomuus. 2. Tilastoyksikkö on tilastollinen perusyksikkö, jota tilastointi koskee eli yksikkö, jolta tiedot kerätään. Yritysten rakennetilastoissa tilastoyksikkö on joko itsenäinen yritys tai yrityksen toimipaikka.	Tilastokeskus
Turvallisuusluokitus	Salassa pidettävät asiakirjat tai niihin sisältyvät tiedot voidaan luokitella sen mukaan, minkälaisia tietoturvallisuutta koskevia vaatimuksia niiden käsittelyssä on tarpeen noudattaa.	Asetus (681/2010)
Tutkimusaineisto	Tutkimuksessa käytetty aineisto.	

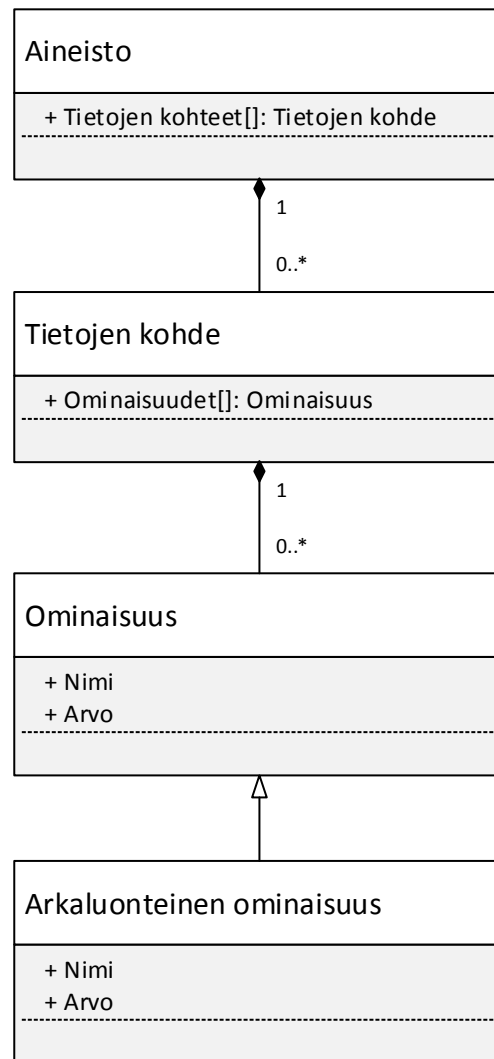
Käsite	Selite	Tarkennuksia
Tutkimuslupa	Tutkijan taustaorganisaation myöntämä lupa tutkimukselle.	
Tutkimusrekisteri	Tutkijan eri lähteistä tiettyä tutkimusta varten koama rekisteri, joka voi olla myös henkilörekisteri.	
Tutkimustulos	Tutkimuksen tuloksena saatu uusi tieto	
Tutkimustuotos	Tutkimuksen lopputuotoksena syntynyt asia, joka voi olla esim. julkaisu, jossa kuvataan saatu uusi tieto, tai sähköinen tietoaaineisto, joka on kuvattu metatiedoin jatkokäyttöä ja tutkimuksessa saatujen uusien tietojen todentamista varten tai tutkimuksessa kehitetty analyysimenetelmä kuten ohjelmisto.	
Tutkittava	Tutkimuksen kohteena oleva henkilö.	Henkilö, joka on antanut tietojaan tai näytteitään biopankille (kts. 'rekisteröity'). Biopankkisanasto

5.2. Tiedot ja tietomallit

Seuraavassa on kuvattu karkealla tasolla viitearkkitehtuurin kannalta keskeisimmän, eli arkaluonteisen tiedon, rakenne.

Arkaluonteinen tieto

Kappaleessa 1.3 on listattu viitearkkitehtuurin piiriin kuuluva tieto. Yleinen aineistojen tietorakenne (Kuva 5) ei ota kantaa siihen, onko tieto arkaluonteista. Yksittäinen aineisto koostuu **tietojen kohteista**, joita voi olla rajoittamaton määrä. Kukin tietojen kohde puolestaan sisältää rajoittamattoman määrän **ominaisuuksia**, joilla on nimi ja arvo. Ominaisuuden erikoistapaus on arkaluonteinen ominaisuus.



Kuva 5 Yleinen aineistojen tietorakenne

Esimerkiksi henkilörekisteri (aineisto) sisältää henkilöitä (tietojen kohteita), joista on tallennettu henkilöä kuvailevia tietoja (ominaisuuksia), kuten nimi ja osoitetiedot. Ominaisuudet voivat olla mitä tahansa kuvailevaa tietoa, kuten vapaata tekstiä, kuvaa, ääntä, tai bionäyte. Näin esimerkiksi asiakirjojen tietyt osat voivat olla luokiteltu salaisiksi ja loput julkisia.

Kuten kohdassa 0 on määritelty, henkilötiedoksi luetaan kaikki sellainen kuvaileva tieto, jolla voidaan tunnistaa henkilö joko suoraan tai epäsuorasti tai välillisesti. Henkilötietolaissa (523/1999 [13 §](#)) ainoastaan henkilötunnus on yksilöity henkilötiedoksi. Tästä seuraa se, että mikään tieto ei lähtökohtaisesti ole arkaluonteista (pl. luokiteltu tieto), vaan se arvioidaan tapauskohtaista.

5.3. Loogiset tietovarannot

Seuraavassa on kuvattu keskeiset loogiset tietovarannot.

**Taulukko 19. Loogiset tietovarannot**

Tietovaranto	Selite
Rekisteri	Julkisen hallinnon tai viranomaisten keräämiä tietoja sisältävä rekisteri. Tieto voi olla arkaluonteista.
Tilastorekisteri	Tilastotarkoituksiin kerätty aineisto. Tieto voi olla arkaluonteista.
Biopankkien näyte- ja tietorekisteri	Tutkimukseen luovutetut näytteet ja niihin liittyvät tiedot
Tutkimustietokanta	Tutkimuksen tuottama tietokanta
Omat tiedot	Yksilön itse itsestään keräämät tiedot
Tutkimusrekisteri	Yhdestä tai useammasta rekisteristä tutkimuskäyttöön muodostettu aineisto tai tutkijan itse keräämä tai tuottama tieto. Aineisto voi muodostaa henkilörekisterin.
Anonymisointitieto	Anonymisoinnissa käytetyn yhdistelyn tms. varanto
Koodirekisteri	Tutkimukseen luovutetun tiedon koodaustiedot
Luovutustiedot	Tutkijalle luovutetun tiedon rekisteri
Suostumusrekisteri	Rekisteri, jossa säilytetään tietojen kohteiden antamat suostumukset.
Rekisteriselosterekisteri	Rekisteri, jossa säilytetään rekisteriselosteet
Edustajarekisteri	Rekisteri, jossa säilytetään tietojen kohteiden edustajia.
Saatavuustiedot	Jatkotutkimukseen saatavilla olevien tutkimuksen tuottamien aineistojen julkiset metatiedot sekä näytteiden julkiset saatavuustiedot
Tutkimustoimintatiedot	Tutkimuksen suunnitelmat ja lupatiedot



6. Lähtötilanteen havainnot

6.1. Palveluiden ja hankkeiden nykytilanne

Viitearkkitehtuuria laadittaessa keväällä 2015 on tutustuttu viitearkkitehtuurissa mahdollisesti hyödynnettävissä oleviin palveluihin tai niitä kehittäviin hankkeisiin nykytilanteen hahmottamiseksi. Luetteloon ei ole otettu mukaan yleisiä tutkimukseen liittyviä palveluita. Alla on lueteltu palveluaihioita eri substanssipalveluille. Todettakoon, että valmiita palveluita ei ole vaan kaikkia aihioita täytyy kehittää enemmän tai vähemmän kaikki viitearkkitehtuurin sovellusalueet huomioiviksi. Osassa palveluita voi olla myös päällekkäisyyksiä, jotka lisäävät kokonaisuuden kustannuksia.

Taulukko 20. Mahdollisia palveluaihioita viitearkkitehtuurin substanssipalveluiden toteuttamiseen.

Tarvittava palvelu	Mahdollisia palveluaihioita
IT-infrastruktuuripalvelu	Turvallinen pilvipalvelu ePouta – Suomen ELIXIR-noodi tarjoaa infrastruktuuripalvelua arkaluonteisen biolääketieteellisen datan tallennukseen ja laskentaan. Palvelua tarjotaan organisaatioasiakkaille, jotka toteuttavat sen avulla omia palveluitaan. Palvelussa eri asiakkaat ovat täysin eristettyjä toisistaan. Palvelun tietoturva on auditoitu korotetulle tietoturvasolulle 2015 lopussa.

Tarvittava palvelu	Mahdollisia palveluaihioita
Tiedonhakupalvelu	- Tutkimuksen tietoaaineistojen hakupalvelu Etsin – OKM:n ATT-hankkeessa kehitetty kansallinen palvelu tutkimuskäyttöön saatavilla olevien aineistojen löytämiseksi. Palvelu edistää tutkimuksen tuottamien aineistojen uudelleen käyttöä. Palvelu sisältää julkisia metatietoja aineistoista. Palvelu on avattu tuotantoon 2014 ja sen lähdekoodi on avoimesti saatavilla. - Aila – Yhteiskuntatieteellisen tietoarkiston FSD:n saatavuuspalvelu yhteiskuntatieteellisille tutkimuksen tuottamille aineistoille. Palvelussa voi tehdä hakuja aineistojen metatiedoista. - KITE – FIMMissä kehitetty ohjelmisto biopankkien saatavuustietojen käsittelyyn. KITE:n lähdekoodi tultaneen julkaisemaan avoimena lähdekoodina. - Arkistojen Portti ja Astia – Arkistolaitoksen palveluja tutkijoiden käytettävissä olevien aineistojen etsimiseen sekä käyttöoikeuksien hakuun. - FMAS hankkeessa on tavoitteena tuottaa metatietokatalogi käytettävissä olevista tietovarannoista vuoden 2015 loppuun mennessä.
Tutkimuslupapalvelu	REMS – ELIXIR-hankkeessa alun perin kehitetty tietoaaineistojen käyttövaltuuksien hakuprosessin hallintaohjelmisto, joka on tuotantokäytössä myös mm. Etsin-tiedonhakupalvelussa. REMSistä on jatkokehitetty ohjelmisto THL biopankin aineistopyyntöjen lupakäsittelyyn. REMS ja sen THL biopankille räätälöity versio on julkaistu avoimen lähdekoodin lisenssillä ja niitä voidaan käyttää eri fokusalueiden omien lupien hakupalveluiden kehittämisessä. FMAS hankkeessa on tavoitteena tuottaa käyttölu-pien sähköinen hakupalvelu vuoden 2015 loppuun mennessä.
Suostumustenhallintapalvelu	Suostumusten hallintapalvelu tutkittavien biopankkisuostumusten hallintaan – Kelan OmaKanta-järjestelmään kehitteillä oleva palvelu tutkittavien biopankkisuostumusten antamiseen ja perumiseen.
Luovutustietopalvelu	-
Rekisteriselostepalvelu	-



Tarvittava palvelu	Mahdollisia palveluaihioita
Esikäsittelypalvelu	-
Tiedonsiirtopalvelu	Palveluväylä – kansallinen palveluväylä on tiedonvälityskonsepti, jossa kukin palveluväylään liitetty järjestelmä hallitsee omia tietojaan sekä vastaa siitä, että muiden tarvitsemat tiedot ovat saatavissa välitysalustan kautta ottaen huomioon tietojen käyttöön liittyvät mahdolliset rajoitukset. Palveluväyläympäristössä kulkevien viestien käsittelyssä tulisi erityisesti varmistaa, että sekä palvelimet että palvelua käyttävät henkilöt ovat oikeutettuja tekemään kyseessä olevia toimenpiteitä
Turvallinen tutkimusympäristö	Kansallinen rekisteritutkijan etätyöpöytä – FMAS-hanke, jossa Tilastokeskus ja Kansallisarkisto yhdessä kehittävät kansallista rekisteri- ja mikroaineistojen tutkijapalvelua, jonka tavoitteena on valmistua vuoden 2015 loppuun mennessä. Etätyöpöytä perustuu Tilastokeskuksen etätyöpöytäpalveluun, joka on ollut tuotannossa 2010 alkaen.
Analysointipalvelu	FIN-CLARIN – tuotannossa kauan ollut aineistopalvelu kieltentutkijoille
Saatavuuspalvelu	- Tutkimus-PAS – OKM:n ATT-hankkeessa suunnitellaan ja kehitetään tutkimusaineistojen pitkäaikaissaatavuuspalvelua, joka tulee nojaamaan tietoturvasertifioituun pitkäaikaissäilytyspalveluun. - Samwise – THL:ssä kehitetty näytetietojen ja niihin liittyvien terveystietojen hallintaohjelmisto, jota kehitetään hyödynnettäväksi biopankeissa. Samwisen lähdekoodi tultaneen julkaisemaan avoimena lähdekoodina. - Aila – Yhteiskuntatieteellisen tietoarkiston FSD:n saatavuuspalvelu yhteiskuntatieteellisille tutkimuksen tuottamille aineistoille. - FIN-CLARIN – tuotannossa kauan ollut aineistopalvelu kieltentutkijoille
Julkaisupalvelu	Eri alojen omat käytössä olevat tutkimusten julkaisupalvelut

Tarvittava palvelu	Mahdollisia palveluaihioita
Neuvontapalvelu	FSD:n tietopalvelut neuvoa ja antaa ohjeita sähköisen tutkimusaineiston hallintaan sen elinkaaren kaikissa vaiheissa. FMAS hankkeessa on tavoitteena tuottaa tukipalvelu rekisteritutkimukseen vuoden 2015 loppuun mennessä.
Koulutuspalvelu	-
Pätevyitymispalvelu	-
Tutkittavien rekrytointipalvelu	-

Yllä olevan lisäksi tehtiin seuraavat havainnot:

- Esimerkiksi tiedonlouhintaan tarvittavaa suurteholaskentakapasiteettia Suomessa on saatavilla CSC – tieteen tietotekniikan keskuksessa, mutta tämä resurssi ei kuitenkaan suuren asiakaskunnan tarkoituksiin suunniteltuna voi täyttää arkaluonteisen tiedon käsittely-ympäristöltä vaadittavaa tietoturvaa ilman huomattavia erityisjärjestelyjä. Kyseinen suurteholaskentakapasiteetti on erillinen yllä mainitusta turvallisesta ePouta-palvelusta.
- Tietojen salaus ja salauksen purku edellyttävät laskentatehoa ja erityisesti suurten aineistojen tapauksessa tietojen salaus ei ole käytännössä mahdollinen toimintamalli. Salaukseen liittyy myös salausavainten pitkäaikainen hallinta, johon tarvittaisiin erillinen hyvin hallittu palvelu. Tätä ei ole otettu huomioon substanssipalveluluettelossa. Jos salaus nähdään tarpeelliseksi, viitearkkitehtuuria tulee kehittää tämän osalta.

6.2. Palveluiden tavoitetilä

Tavoitteena on luoda Suomeen kansallisia palveluita erilaisen arkaluonteisen tiedon tutkimuskäyttöön erillisten fokuoitujen tutkimusalueiden palvelujen sijaan. Substanssipalveluiden, joita on käsitelty aiemmin tässä dokumentissa, kehittämisessä voidaan hyödyntää yllä esitettyjä kehitteillä olevia palveluja tai ohjelmistoja. Palveluita voitaisiin sopivilta osin laajentaa kansallisiksi palveluiksi hyödyttämään kaikkia arkaluonteista tietoa tutkimuksessaan käsitteleviä tahoja.

Substanssipalveluista useimmat olisi tehokasta tuottaa kansallisina yhteisinä palveluina, joissa kuitenkin kukin taho hallitsee vain omaa organisaatiotaan koskevia tietoja. Tällaisia kaikille samanlaisia toimintoja ovat suostumusten, rekisteriselosteiden ja tietojen luovutusten hallinta. Tutkimuslupienhakupalvelu ja aineistojen esikäsittelypalvelu kansallisina palveluina yhtenäistävivät toimintaa. Tiedonhakupalvelut ovat myös luonteeltaan kansallisia. Osa näistä palveluista voisi olla yhteisiä tutkimukselle ja viranomaisille kuten luovutus-tietopalvelu, jolloin tiedot voisivat olla vain yhdessä järjestelmässä.

Erityisesti yhteentoimivuuden kehittäminen mm. määrittelemällä yhteisiä rajapintoja ja liittymiä, sekä hyödyntämällä tietomallinnusta, edistäisi merkittävästi tämän viitearkkitehtuurin toteuttamista. Lisäksi arkaluonteisen tiedon käsittelyyn tarvitaan muita yleisiä tutkimusprosessin palveluja, kuten aineiston hyvään hallintaan liittyviä sanastopalveluja ja aineistojen tunnistepalveluja.

Yksittäisten palvelujen ja tietovarantojen lisäksi tarvitaan yhteisiä toimintatapoja, joilla voidaan luoda tehokas ja yhtenäinen kansallinen toimintakulttuuri. Näillä toimenpiteillä parannetaan myös yksilön tietosuojaa kaikkien toimijoiden luottamuksen kasvun lisäksi, sekä ansaitaan yksilöiden luottamus ja tataan tarvittavien tietojen saatavuus tutkimukseen.

6.3. IT-infrastruktuuripalvelujen tavoitetilä

Suuret tietoaaineistot ovat kasvavassa määrin muodostumassa tutkimuksen merkittäviksi infrastruktuureiksi ja näiden hyödyntämiseen tarvitaan erilaisia palveluita tutkijoille. Tässä viitearkkitehtuurissa nähdään erityisen tarpeelliseksi IT-infrastruktuuripalvelut, joilla tutkimusorganisaatiot pystyvät helposti täyttämään tuottamiinsa tai hallitsemiinsa arkaluonteisiin aineistoihin kohdistuvat tarpeet ja muodostamaan näistä tietoaaineistoinfrastruktuureja laajempaan käyttöön osaksi yhteentoimivaa tutkimuksen infrastruktuuria.

Esimerkki edellä mainittujen palvelujen tarpeesta on biopankkien lakisäätöisten rekisterien hallintaa toteuttava palvelu, jota kaikki biopankit voisivat hyödyntää. Tämä yhtenäistäisi ja tehostaisi biopankkien hallinnoimien vain tutkimuskäyttöön tarkoitettujen datojen käsittelyä ja saatavuutta tutkimukseen. Toinen selkeä tarve on Big data -tyyppisten arkaluonteisten tutkimusaineistojen kuten genomidatan tai kuvadatan säilytys- ja hallintapalvelulle. Tällaisten Big data -palvelujen ominaispiirteisiin kuuluvat suuren turvallisen tallennustilan lisäksi datan käsittelyn tarvitsema poikkeuksellisen suuri prosessointiteho esim. vertailutiedon hakuun tai tiedonlouhintaan.

Ihmisiin liittyvässä tutkimuksessa käytettävää aineistoa käsittelevät infrastruktuurinomaiset Big data -palvelut voivat olla myös samanaikaisesti keskeisiä palvelutuotannon infrastruktuuripalveluita: esimerkiksi referenssitietovarantoja terveydenhuollon päätöksentekojärjestelmille yksilöllisen lääketieteen rantaustuessa arkipäivän terveydenhuoltoon. Palveluita suunniteltaessa ja toteutettaessa on alusta lähtien otettava huomioon arkaluonteisten aineistojen laatu ja sen myötä mahdollisuudet tutkimusta laajempaan hyödyntämiseen, jolloin voidaan saavuttaa merkittäviä kustannussäästöjä tutkimuksen ja palvelutuotannon yhteisillä tietoturvalisillä infrastruktuureilla.

6.4. Tietoturvallisuuden nykytilasta

Kuva tietoturvallisuudesta on muuttunut merkittävästi muutaman viimeisen vuoden aikana mm. tietomurtoihin käytettävien resurssien kasvun myötä. Tietojen suojaus ja suojausmenetelmien murto ovat jatkuvaa kilpajuoksua. Myös huomattavat viat ja häiriöt tietojärjestelmissä ja toimintaprosesseissa muodostavat merkittäviä riskitekijöitä tiedon eheyteen ja saatavuuteen. Tietoturva-poikkeamat muodostavat myös riskejä tietosuoja vaatimusten täyttämiseksi. To-



dettakoon, että nykyisessä toimintaympäristössä tietomurroilla on jokin vaikeasti arvioitavissa oleva todennäköisyys kaikista hyvistä tietoturvakäytännöistä ja varautumisista riippumatta.

Samasta syystä ulkomailla sijaitsevia tai kansainvälisten yritysten ylläpitämiä pilvipalveluita ei turvallisuussyistä voi suositella käyttäväksi arkaluonteisen tiedon saatavuuspalveluina. Lisäksi testattavana oikeudessa on minkä valtion lakeja noudatetaan.¹³

¹³ [Pilvessä piilee riskejä, tietosuoja – Tietoturvan ja tietosuojan erikoislehti](#) 2, 2015
