

Avoimen tieteen käsikirja tutkijoille ja tutkimusorganisaatioille

Avoimen tieteen käsikirja on tarkoitettu tutkijoille, tutkimusryhmille ja tutkimusorganisaatioille auttamaan avoimen tieteen ja tutkimuksen hyödyntämisessä ja edistämässä. Käsikirjaa on työstetty Avoin tiede ja tutkimus -hankkeessa vuosina 2014–2017. Käsikirjaa ei enää päivitetä. Käsikirjasta on olemassa myös lyhennetty, tutkijoille suunnattu versio [Forskarens handbok till öppen forskning](#).



Avoimen tieteen käsikirja, jonka tekijä on [Opetus- ja kulttuuriministeriö](#), on lisensoitu [Creative Commons Nimeä 4.0 Kansainvälinen -lisenssillä](#).

Avoimen tieteen ja tutkimuksen käsikirja

Tieteen kehityksessä mukana pysymiseksi myös Suomessa tulisi varmistaa riittävä avoimen tieteen osaaminen. Tutkijoiden ja tutkimusorganisaatioiden tulisi varmistaa tiedonhallinnan riittävä osaaminen, jotta tutkimusaineistot, -menetelmät, ja -tutkimusjulkaisut voidaan saattaa mahdollisimman avoimesti kaikkien saataville. Avoin tiede ja tutkimus -hanke tukee tutkijoita ja organisaatioita näissä pyrkimyksissä luomalla ohjeistuksia ja palveluita sekä kehittämällä kannusteita. Tieteen avoimuutta tukevana yleisenä periaatteena on, että tutkimusaineistojen ja -julkaisujen jatkokäyttöä ei rajoiteta tarpeettomasti ja aineistojen käyttöehdot määritellään selkeästi.

Tutkijoille avoimuus voi näyttäytyä sekä uhkana että mahdollisuutena. Toimintaympäristön muutos tarjoaa kuitenkin tutkijalle monia mahdollisuuksia edistää työtään ja sen vaikuttavuutta, mutta muutos vaatii usein paneutumista uusiin asioihin, kuten tiedonhallintaan tai tekijänoikeuksiin. Tämän käsikirjan on tarkoituksena olla tässä tutkijan ja tutkimushallinnon apuna.

Opetus- ja kulttuuriministeriön käynnistämä [Avoin tiede- ja tutkimus -hanke](#) suosittelee tutkimuksen tulosten julkaisemista avoimella lisenssillä tai avoimena lähdekoodina. Yleisenä periaatteena on, että tutkimusaineistojen ja -julkaisujen jatkokäyttöä ei rajoiteta tarpeettomasti ja käyttöehdot tuodaan selkeästi esille. Avoimille aineistoille suositellaan, standardimuotoista koneluettavaa lisenssiä [CC BY 4.0](#). Tutkimusaineistoja kuvailevat metatiedot suositellaan lisensoimaan käyttöluvalla [CC0](#). Tietokoneohjelmille hyvä lisenssi on [MIT](#)-lisenssi, sillä se sallii kaiken käytön, myös kaupallisen, mitä korostetaan sekä [EU:n](#) että [kansallisella tasolla](#). Tutkijoita, tutkimusryhmiä ja tutkimusorganisaatioita kehoitetaan hyödyntämään olemassa olevia palveluita kuten esimerkiksi Euroopan tasolla suositellut [PubMed central](#), [arXiv](#), [OpenAIRE](#), [Zenodo](#) sekä Suomessa [IDA](#), [Etsin](#), [Aila](#) ja [Kielipankki](#) sekä korkeakoulujen ja tutkimuslaitosten omat julkaisuarkistot ja tutkimustietojärjestelmät.

Table of Contents

Mitä avoimuus on?.....	3
Miksi avoimuus on hyvä ajatus?.....	4
Tutkimuksen tiedonhallinnan suunnittelu	6
Tutkijan tiedonhallinnan muistilista	7
Tutkimuksen tietoaaineistot	8
Tutkijan muistilista tietoaaineistoista	9
Aineistonhallinta: Tutkimuksen tietoaaineistot	10
Datan avaaminen.....	13
Datan jakamisen hyödyt	15
Tutkimusaineiston julkaisuprosessi.....	17
Korkeakoulujen linjaukset	18
Tietoturva	19
Riski.....	20
Toimenpide.....	20
Tutkimusaineistojen metatiedot	22
Tutkijan muistilista metatiedoista	24
Tunnisteet ja viittaukset	24
Viittaaminen tutkimusaineistoihin	26
Tutkijan muistilista tunnisteista ja viittauksista	27
Organisaatioille: Tunnisteet	27
Tutkimusprosessin avaaminen	29
Tutkijan muistilista tutkimusprosessin avaamisesta	31
Tutkimustulosten julkaiseminen	31
Tutkijan muistilista julkaisemisesta	33
Yleiset	33
Sopimukset, oikeudet ja lisenssit	33
Rinnakkaistallennus	34
Organisaation muistilista julkaisemisesta	34
Sosiaalinen media ja julkaiseminen.....	34
Tutkijan muistilista sosiaalisen median käyttöön.....	35
Julkaisuarkistot	36
Politiikat.....	37
Datapolitiikan tekeminen	38
Tiedonhallinnan palveluiden suunnittelusta	40
Avoimen tieteen käsikirjan sanasto.....	41

Mitä avoimuus on?

Avoimuus on tieteen ja tutkimuksen keskeinen periaate. Sen kautta voimme yhdessä luoda uusia mahdollisuuksia osallistua tutkimuksen tekemiseen laajemmin niin tiedeyhteisölle kuin päätöksentekijöille ja kansalaisille. Avoin tiede on noussut kansainvälisesti merkittäväksi tavaksi edistää tiedettä ja tieteen vaikuttavuutta yhteiskunnassa. Nousu johtuu avoimuuden tuomasta mahdollisuudesta tutkimustulosten laajempaan todentamiseen, läpinäkyvyyteen ja toistettavuuteen. Edellytyksenä on kuitenkin tutkimuksen tuottamien ja tarvitsemien julkaisujen, tietoaaineistojen, menetelmien, osaamisen ja tukipalveluiden laajamittainen saatavuus. Avoimuuden tulee heijastua niin tiedon hankintaan kuin tutkimus- ja arviointimenetelmiin. Muiden kiinnostus tutkimuksen hyödyntämiseen kertoo sen merkittävydestä. Kun tulokset ovat laajasti hyödynnettävissä tieteen vaikuttavuus kasvaa.

Avoin tiede ja tutkimus eivät ole vain kokoelma keinoja ja suosituksia, vaan tutkimuskulttuurin muutosprosessi

Tiede kuuluu kaikille. Avoimen tieteen ja tutkimuksen uusien toimintamallien avulla demokratisoimme tiedettä ja lisäämme sen vaikuttavuutta. Tutkimusprosessin digitalisoituminen ja avautuminen luo tutkijoille ja kaikille tutkimuksesta kiinnostuneille uusia mahdollisuuksia yhteistyöhön ja viestintään. Avoimuus antaa mahdollisuuksia uudistumiseen. Nuorten, aloittelevien ja tutkimusinfrastruktuurien ulkopuolella toimivien tutkijoiden tasavertaisten mahdollisuuksien takaaminen tietoon pääsyyn on myös tärkeää. Erityisesti kehittyvien maiden mahdollisuudet tutkittuun tietoon kasvavat avoimen tieteen käytäntöjen leviämisen avulla. Avoimuuden kautta syntyy kaikille periaatteellinen mahdollisuus kykyjensä mukaan tutkia, kritisoida, vakuuttua tai edistää tutkimusta. Tämä kasvattaa luottamusta tieteeseen sekä vauhdittaa yritystoimintaa, mutta vähintään yhtä tärkeää on kriittisen ajattelun ja tieteenlukutaidon levittäminen yhteiskunnassa.

Tieteessä avoimuus on aina ollut tavoitteena hyvän laadun ja hyvien käytäntöjen edistämiseksi sekä väärennösten ja huonolaatuisen tutkimuksen estämiseksi. Tieteen itsekorjautuva luonne vaatii tutkimuksen läpinäkyvyyttä ja toistettavuutta. Avoimuus parantaa tutkimuksen laatua, koska avoimesti saatavilla olevat tulokset ja tietoaaineistot mahdollistavat tieteellisten havaintojen kyseenalaistamisen ja verifioimisen. Näin tiede kehittyy ja korjaa itseään entistä nopeammin ja päällekkäistyöltä vältytään. Avoimen tieteen ja tutkimuksen avulla pyrimme edistämään tutkimuksen kestävyyttä, käytettävyyttä, saatavuutta sekä luottamusta. Tutkimuksen etiikkaan avoimuus kuuluu luonnollisena oletusarvona. Tutkimustulosten ja -aineistojen on oltava löydettävissä ja käytettävissä myös pidemmällä aikavälillä.

Hyvät käytännöt avoimen tieteen ja tutkimuksen edistämiseksi koostuvat yleisistä, velvoittavista tutkimuseettisistä ohjeista sekä hyvästä tutkimustiedon hallintotavasta, sekä avoimuuden peruslinjauksesta ja täsmäntävistä suosituksista. Opetus- ja kulttuuriministeriön asettama tutkimuseettinen neuvottelukunta (TENK) on yhteistyössä suomalaisen tiedeyhteisön kanssa uudistanut vuonna 2012 [tutkimuseettisen ohjeen hyvästä tieteellisestä käytännöstä](#) ja sen loukkausepäilyjen käsittelemisestä.

Avoin tiede ja tutkimus on jatkumo, joka kattaa koko tutkimusprosessin ja toimintakulttuurin. Tutkimustieto, siihen liittyvä osaaminen ja ymmärrys ovat hajaantuneet tutkimusjärjestelmän eri toimijoille, verkostoille ja yhteisöille. Tällaisessa kokonaisuudessa voimme kasvattaa avoimuutta

sekä sisäisesti avaamalla ja yhdessä kehittämällä tutkimuksen tiedonhallinnan prosesseja ja toimintamalleja että ulkoisesti avaamalla koko prosessia ja sen osia muille tutkijoille ja yhteiskunnalle ja antamalla muille mahdollisuuksia hyödyntää aineistoja, menetelmiä ja tuloksia.

Avoimeen ja hyvään tutkimustulosten hallintatapaan kuuluu toiminnan korkea taso ja hyvä laatu: rehellisyys, yleinen huolellisuus ja tarkkuus. Sama pätee tietenkin tutkimusaineistoihin näiden osana, mutta niiden laadun ominaisuuksia ovat lisäksi käytettävyys ja saatavuus, eheys ja virheettömyys sekä avoimuus ja luottamuksellisuus. Tietojen laadun vaaliminen on keskeinen hyvän tutkimustiedon hallintatavan vaatimuksista. Hyvän tutkimustiedon hallintatavan tuloksena tietoja ei asiattomasti muuteta, tiedot eivät vahingoitu tai häviä käsittelyprosessin aikana ja tietojen virheettömyyteen ja laatuun voidaan luottaa. Nämä ovat itseisarvoja, joista ei tule aiheettomasti tinkiä. Tietojen laatua ja virheettömyyttä varmistavat toimenpiteet tulee kuitenkin aina suhteuttaa tiedon merkitykseen ja tärkeyteen. Mitä tärkeämpi tieto, sitä enemmän kannattaa panostaa sen laatuun ja virheettömyyteen. Tietojen käyttötarkoitus ohjaa niiden hallintaa ja huolellinen dokumentointi ja kuvailu on tärkeää.

Miksi avoimuus on hyvä ajatus?

Tiedon käsittely sähköisessä muodossa luo uusia mahdollisuuksia tieteelliselle tutkimukselle. Sähköisessä muodossa olevaa tietoa on mahdollista jakaa ja kommentoida helposti ja nopeasti eri toimijoiden kesken. Tutkimusaineistojen ja tulosten avoin jakaminen on yksi tapa edistää tieteen avointa toimintamallia. Avoin toimintamalli käsittää myös tutkimusprosessin eri vaiheiden avaamisen. [Esimerkkinä](#) tästä on Human Genome Project, jossa uudenlainen avoin tutkimusprosessinaikainen kommunikaatiotapa mahdollisti ennennäkemättömän laajan ja tehokkaan tavan kartoittaa ihmisen geeniperimää. Myös [paleogenetiikasta](#) ja [tähtitieteestä](#) löytyy esimerkkejä laajalle viedystä avoimesta tieteestä.

Tieteen avoin toimintamalli on osoittautunut hyväksi tavaksi edistää tieteen läpinäkyvyyttä ja toistettavuutta, mikä kokonaisuutena lisää tieteen luotettavuutta. Avoimuus antaa tutkijoille ja tutkimusorganisaatiolle mahdollisuuden edistää myös oman tieteellisen toiminnan näkyvyyttä, mutta avoimuuden hyödyntäminen vaatii uudenlaista tutkimustiedon ja -prosessin hallintaan liittyvää osaamista.

Avoimuus edistää tutkimuksen löydettävyyttä sekä tieteellisten toimijoiden että suuren yleisön keskuudessa. Sen käytännöt avaavat huomattavia [mahdollisuuksia](#), mutta vaativat myös tutkijoilta hyvää suunnittelua ja harkintaa. Avoimuutta voivat rajoittaa esimerkiksi erilaiset eettiset ja juridiset esteet. Suuri osa suomalaisesta tutkimuksesta on tänään vaikeasti löydettävissä ja heikosti suuren yleisön saatavilla. Siksi pyrimme tukemaan parempaa tiedonhallintaa tämän käsikirjan avulla. Se on kirjoitettu asiantuntijoiden avoimena yhteistyönä Avoin tiede ja tutkimus -hankkeen puitteissa.

Suuri joukko kansallisia ja kansainvälisiä tutkimusorganisaatioita ja tutkimuksen rahoittajia on asettanut tavoitteekseen edistää avointen toimintamallien toteutumista tieteellisessä tutkimuksessa. Euroopan komission suositus "[Recommendation on Access to and Preservation of Scientific Information](#)" asettaa linjoja ja tavoitteita eurooppalaisen tutkimuksen avoimuudelle. OECD:n vuonna 2007 julkaiseman "[Principles and Guidelines for Access to Research Data from Public Funding](#)"-raportin ansiosta tutkimusdatan avoimuus on parantunut. [OECD Councilin](#) vuonna 2008 antama suositus Julkisen sektorin tuottaman tiedon saatavuudesta vaikutti suuresti [PSI-direktiiviin](#). Edellä mainituista suosituksista johtuen avoimen tieteen eri sidosryhmillä kuten rahoittajilla,

korkeakouluilla, kirjastoilla, datakeskuksilla, tieteellisillä seuroilla ja tiedekustantajilla on edessään sekä isoja mahdollisuuksia että haasteita avoimen tieteen saralla.

Avoimen tieteen edistyessä kaikkien saatavilla olevien tutkimusjulkaisujen ja aineistojen määrä tulee koko ajan kasvamaan. Suuntaus johtuu rahoittajien vaatimuksista, open access lehtien kuten [PLoS One](#) (maailman ylivoimaisesti suurin tieteellinen lehti) kasvavasta määrästä ja muiden avointa saatavuutta tukevien toimintatapojen kuten julkaisuarkistojen yleistymisestä. [Euroopan tutkimusneuvosto](#) (European Research Council) vaatii avointa julkaisua rahoittamaltaan tutkimukselta artikkeleiden ja monografioiden kohdalla ja suosittelee sitä tutkimusaineistoille. [Horizon 2020-ohjelman pilottiprojektissa vaaditaan myös tutkimusaineistojen avaamista](#). [Knowledge Exchange](#)-verkoston tekemän selvityksen mukaan avoimuuteen rohkaistaan yhä laajemmin ja sitä jopa vaaditaan monen suuren eurooppalaisen rahoittajan puolelta. Myös tutkimusorganisaatiot ovat itse lähteneet luomaan uusia käytäntöjä esimerkiksi Saksassa.

Euroopan komission rahoittamia projekteja ovat myös [RECODE](#) ja [FOSTER](#). RECODE on nostanut esiin tutkimusaineistoihin liittyviä kysymyksiä ja on viimeistelemässä suosituksia aineistojen arkistoinnin ja levittämisen suhteen. FOSTER, joka keskittyy avoimuuden koulutukseen, on kerännyt verkkosivuilleen paljon aineistoa avoimeen tietoon ja tutkimuksen tiedonhallintaan liittyen. Open Access-julkaisemista edistää myös [SPARC Europe](#), joka on tutkimusorganisaatioiden verkosto yhteistyössä pohjois-amerikkalaisen sisarverkostonsa [Scholarly Publishing and Academic Resources Coalitionin](#) kanssa. Myös eurooppalainen tutkimusyliopistojen verkosto [League of European Research Universities](#) (LERU) on luonut ohjeistusta [tutkimusaineistojen hallintaan](#) ja [avoimeen julkaisemiseen](#).

Yleinen kiinnostus avoimeen tieteseen on lisääntynyt ja monet merkittävät tieteelliset kausijulkaisut kuten Nature ja Science julkaisevat yhä useammin avointa tiedettä käsitteleviä artikkeleita. Tiedejulkaisuissa pyritään kehittämään myös käytäntöjä, jotka tukevat tutkimusaineiston ja tutkimusmenetelmiin liittyvän tiedon julkaisemista tutkimustulosten ohessa. Erilaiset hankkeet kuten [DataCite](#), preprintpalvelut [arXiv](#), [bioarxiv](#) ja [socarxiv](#), EU-projektien julkaisuja kokoava palvelu [OpenAIRE](#) sekä dataa kokoava [Zenodo](#), sekä tieteenalakohtaiset tietoarkistot kuten [PANGAEA](#) ja [Dryad](#), kertovat avoimen tieteen yleistymisestä ja tärkeydestä maailmalla. Myös monet suuret tutkimusinfrastruktuurit (kuten [ESFRI](#)-hankkeet) jakavat tuloksiaan avoimissa palveluissa. Tutkimusdatan jakaminen ei ole vallitseva käytäntö kaikilla tutkimusaloilla erityyppisistä haasteista riippuen, mutta siihen liittyvä kiinnostus on herännyt.

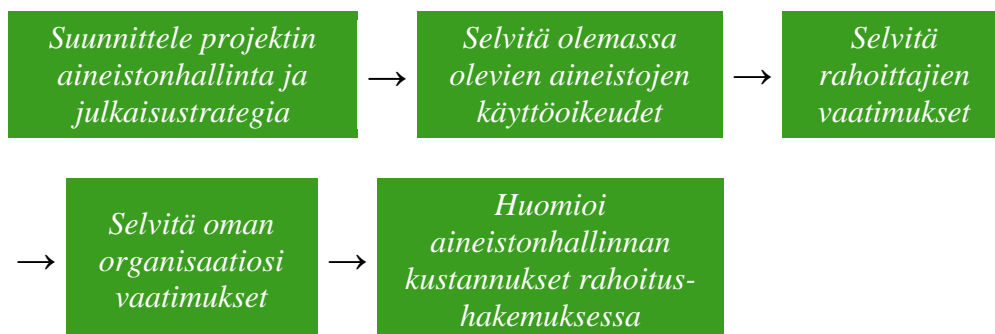
Avoimen toimintamallin soveltamiseen tieteellisessä tutkimuksessa liittyy erityyppisiä haasteita, joita pyritään ratkaisemaan kansainvälisissä ja kansallisissa verkostoissa. Kansainvälisen [Research Data Alliance](#)-verkoston (RDA) mottona onkin "Research Data Sharing without Barriers" ja verkoston työryhmät pohtivat muun muassa aineistojen uudelleenkäyttöön ja kuvailemiseen liittyviä kysymyksiä. Lisäksi Suomessakin aktiivinen [Open Knowledge Network](#) on muodostanut omia kansainvälisiä verkostojaan avoimen tieteen ympärille. Facebookissa [Open Science Finland](#)-ryhmällä on yli 700 jäsentä. Open Knowledge Networkin toiminta keskittyy usein nimenomaan osaamisen lisäämiseen.

Tutkimuksen tiedonhallinnan suunnittelu

Tutkimukseen liittyy paljon tiedonhallintaa. Tiedonhallintaan on perinteisesti liitetty tiedonhaku ja viitteiden hallinta, jotka edelleen ovat niitä osia joita yhdistetään lähinnä kirjasto-osaamiseen. Tämän päivän tutkija painiskelee kuitenkin monen muunlaisen, yhä useammin sähköisen, tiedon kanssa: muistiinpanojen, kirjeenvaihdon, tutkimusdatan, kuvien, äänitteiden, sovellusten, artikkeliversioiden, sopimusten, tutkimustietojärjestelmien, julkaisuarkistojen, koodikirjojen, lokien, suunnitelmien, hakemusten, budjettien ja raporttien, joitakin mainitaksemme. Usein lisäksi tutkimusprojekteissa tietoa jaetaan kollegoiden kesken, jolloin versioitakin saattaa syntyä useita. Tutkimuksen toistettavuus ja todennettavuus ovat tavoitteita, jotka on hyvä pitää mielessä.

Kaiken tämän tiedon hallinnan suunnitteluun on syytä panostaa. Kirjastoalan ammattilaisten olisi hyvä olla tässä osallisena ja tutkijan ja tutkimusorganisaation tukena. Kannattaa esimerkiksi selvittää mahdollisia yhteistyöalustoja, sovelluksia, tietoarkistoja yms. jo suunnitteluvaiheessa, jotta kustannukset voi ottaa huomioon. Myös lakiin, sopimukseen ja etiikkaan liittyvät asiat on syytä ottaa haltuun jo tutkimuksen suunnitteluvaiheessa. Suunnitelma on kuitenkin elävä dokumentti, johon kirjataan myös muutokset. Sitä on syytä pitää ajan tasalla osana projektin dokumentaatiota. Olosuhteet tai tarpeet saattavat muuttua, tai valittu ratkaisu saattaa osoittautua huonoksi. Siinä tapauksessa on tietenkin suunnitelmia päivitettävä.

Keskeinen osa tiedonhallinnan suunnittelua on tutkimusaineistojen hallinta. Aineistoihin kuuluu yhä useammin myös dataa. Jos sen hallinta epäonnistuu, koko tutkimus saattaa menettää uskottavuutensa. Hyvä suunnittelu on usein myös (avoimen) julkaisemisen edellytys. Yhä useampi tutkimuksen rahoittaja vaatiikin aineistonhallintasuunnitelmaa osana projektisuunnitelmaa. Tarkoituksena on edistää hyvää suunnittelua ja rohkaista tutkijoita jo ennen tutkimuksen aloittamista miettimään tutkimusaineistojen laatua ja käytettävyyttä. Aineistonhallinnan suunnitteluun on olemassa myös suomalaisia ohjeita, esimerkiksi [Yhteiskuntatieteellisen tietoarkiston opas](#). Nämä oppaat voivat toimia tukena kaikkien alojen tutkijoille. Lisäksi on hyvä suunnitella julkaisukanavia kaikille tutkimuksessa syntyville tuloksille, kuten menetelmille, artikkeleille jne. On syytä jo alusta pitäen selvittää mihin eri sopimukset, lait ja ohjeistukset velvoittavat tutkijaa ja miten tutkimus on mahdollista tehdä siten, että mahdollisimman paljon voidaan julkaista mahdollisimman avoimesti. Myös julkaisuajankohtia kannattaa miettiä avoimin mielin: kaikkea ei ehkä tarvitse julkaista vasta tutkimusprojektin päätteeksi. Varhainen avaaminen ja julkaiseminen voi tuoda yllättäviäkin hyötyjä omalle tutkimukselle.



Tutkimustiedonhallintaa tukemaan on kehitetty palveluita ja järjestelmiä (tutkimusinfrastruktuuria) muun muassa [FIRI:n](#) toimesta sekä Avoin tiede ja tutkimus -hankkeen yhteydessä. Tavoitteena on

palveluvalikoima, joka rakentuu kestäväälle pohjalle ja joka tarjoaa palveluita aineistojen, menetelmien ja julkaisujen tallentamiseen, hakemiseen ja säilyttämiseen. Näiden palveluiden tuloksena tietojen saatavuuteen, säilyvyyteen, virheettömyyteen ja laatuun voidaan luottaa. Palveluita ovat esimerkiksi Euroopan tasolla suositellut [PubMed Central \(PMC\)](#), [arXiv](#), [OpenAire+](#), [Zenodo](#) sekä Suomessa [IDA](#), [Etsin](#), [Kielipankki](#) ja [Aila](#) sekä korkeakoulujen ja tutkimuslaitosten omat [julkaisuarkistot](#) ja tutkimustietojärjestelmät sekä tuleva pitkäaikaissaatavuusratkaisu.

Tutkimusjärjestelmän on huolehdittava näiden infrastruktuurien yhteentoimivuudesta silloinkin kun aineistoja, menetelmiä tai julkaisuja ei voida avata vapaasti käytettäviksi. Löydettävyyden ja käytettävyyden ovat avoimen tieteen tärkeitä piirteitä. Palveluita suunnitellaan siksi kokonaisvaltaisesti ja yhteistyössä.

Tutkijan tiedonhallinnan muistilista

- Suunnittele tutkimuksen tiedonhallinta huolellisesti jo rahoitushakuvaiheessa ja tutkimussuunnitelmaa tehdessä. Aineistonhallinnan suunnittelutyökalu [DMPTuuli](#) tarjoaa tähän tukea.
 - Mitä aineistoja tutkimukseen liittyy?
 - Muista sekä jo olemassa olevat että syntyvät aineistot.
 - Selvitä niihin liittyvät omistus- ja oikeuskysymykset. Mitä sopimuksia tarvitaan?
 - Mitä dokumentaatiota on syytä tehdä aineistosta, sen käsittelystä ja tutkimusmenetelmistä?
 - Miten tutkimuksen toistettavuus voidaan varmistaa?
- Selvitä rahoittajien, tutkimusorganisaatioiden ja aineistojen omistajien ja mahdollisten tutkimuskohteiden tai informanttien oikeudet, vaatimukset ja suositukset. Etsi sopimus pohjia ja kysy neuvoja tutkimuksen tukipalveluista, tietoarkistoista, kirjastoista yms. Tietoa löytyy myös Avointiede.fi-sivustolta.
- Valitse mahdollisimman avoin lisenssi, näin turvaat itsellesikin pääsyn aineistoihin ja tuloksiin tulevaisuudessa.
- Tietokoneohjelmille hyvä ja avoin lisenssi on [MIT-lisenssi](#), tietokannoille [CC0](#) ja aineistoille [CC-BY 4.0](#) (CC-lisenssin valintaopas, [pdf](#))
- Selvitä oman alasi aineistojen säilytyspaikat ja tekniset edellytykset siirtää, jakaa ja kopioida aineistoja tarpeen mukaan
 - Selvitä mitä ohjeita ja standardeja liittyy samankaltaisiin tutkimusaineistoihin.
 - Missä muodossa ne ovat ja missä niitä säilytetään ja julkaistaan?
 - Miten niitä kuvaillaan ja miten ne kannattaa dokumentoida, jotta kollegat ja muut osaavat käyttää niitä oikein?
 - Entä menetelmät?
- Selvitä ja määrittele ainakin alustavasti julkaisukanavat tutkimuksen tuotoksille ja tuloksille.
- Suunnittele ja kuvaile miten aineiston laatu varmistetaan. Kuvaile prosessit, tehdäänkö esim. vertaisarviointi sille erikseen? Julkaistaanko erillinen artikkeli datasta? Selvitä alasi parhaat käytännöt. Aineiston laatuun vaikuttavat:
 - virheettömyys
 - määräystenmukaisuus
 - johdonmukaisuus
 - ymmärrettävyys
 - yksiselitteisyys
 - ajantasaisuus
- Arvioi ja huomioi myös avaamiseen liittyvät kustannukset

- Mieti mitä aineistoille tapahtuu tutkimuksen loputtua. Mitä mahdollisesti säilyttämisen arvoisia aineistoja tai menetelmiä syntyy? Miten voit huolehtia siitä, että aineistoista voisi tehdä mahdollisimman avoimia ja tutkimuksestasi toistettavaa? Säilyttämisen tärkeyteen vaikuttavat muun muassa:
 - Aineiston ainutlaatuisuus (sitä ei voi tuottaa uudestaan).
 - Historiallinen tai tieteellinen merkittävyys (esim. täydentää tärkeää tietovarantoa).
 - Suuri uudelleenkäytön mahdollisuus.
 - Sen taloudellinen arvo

Tutkimuksen tietoaineistot

Kun tutkimuksen tuloksia julkaistaan, myös tutkimusaineistojen on oltava kunnossa. Tutkimusaineistoilla tarkoitetaan niitä resursseja, joita tutkija tuottaa tai käyttää tutkimusprosessin aikana. Niitä tarvitaan tutkimuksen validointiin ja ne voivat myös antaa lisää meriittiä, jos ne on julkaistu. Tutkimuspolitiikan laatijan, tutkimusorganisaation ja tutkimusrahoittajien on huolehdittava riittävästä, luotettavista palveluista ja resursseista. Tutkijan vastuulla on huolellinen suunnittelu, dokumentaatio sekä aineiston jakaminen parhaaksi katsomallaan tavalla sopimuksia, lakeja ja hyvää etiikkaa noudattaen.

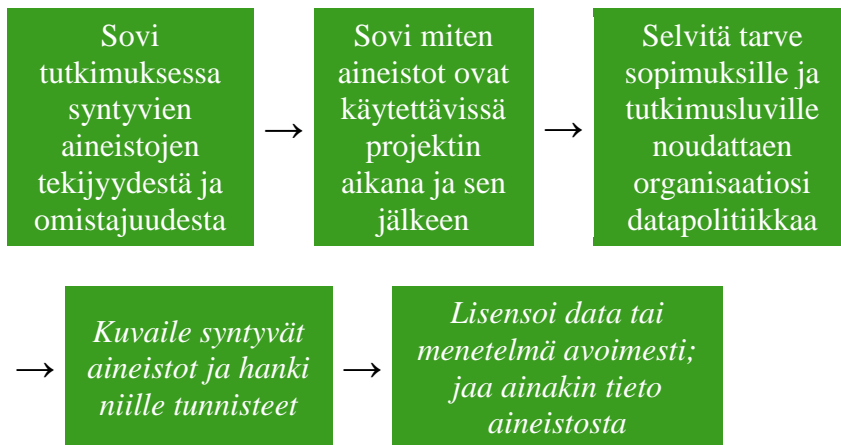
Jotta aineisto olisi tutkimukseen kelpaavaa, siihen täytyy liittyä tietoja ainakin sen alkuperästä. Esimerkiksi datan liitteeksi tarvitaan kuvailevaa ja teknistä tietoa siitä mitä informaatiota se sisältää. Siksi tutkimusdataan liittyy paljon tietoa siitä, miten se on rakenteistettu ja koodattu, miten se on syntynyt ja miten sitä on käsitelty. Nämä tiedot voi tallentaa esimerkiksi tietoarkiston luettelon metatietoihin, koodikirjoihin tai muuhun dokumentaatioon. Yhdessä datan kanssa tästä kokonaisuudesta muodostuu tutkimusaineisto.

Aineistojen julkaiseminen on hyvän tieteellisen käytännön mukaista, koska siten tutkimuksen tulokset ovat laajasti verifioitavissa. Myös käyttämättä jääneiden aineistojen julkaisua suositellaan. Rahoittajien olisi hyvä huomioida tämä aineistohallinnan suunnitelmia arvioitaessa. Tutkimusaineistojen tuottamisesta ja jakamisesta kertyy meriittiä. Siksi ne kannattaa mainita ansioluettelossa. Esimerkiksi TENK:n [ansioluettelopohjassa](#) on erillinen kohta "ansiot tutkimus- ja tietoaineistojen tuottamisessa ja jakamisessa".

Tutkimusorganisaatioiden on tärkeä panostaa tietoarkistojen ja tieteellisten kirjastojen palveluun, jotta tuotetut aineistot olisivat mahdollisimman hyvälaatuisia. Tutkimusaineistojen hallinnassa vaaditaan yhteistyötä eri alojen asiantuntijoiden kesken. Tutkijoilla on ratkaisevan tärkeää tietoa omista aineistoistaan, mutta tiedonhallinnan ja muita asiantuntijoita tarvitaan säilyvyyden, käytettävyyden ja löydettävyyden varmistamiseksi. Tutkimuksen tietoinfrastruktuuriin liittyy siksi aina palveluita, yhteistyötä ja vuorovaikutusta, jota on vaalittava ja johon on annettava kannusteita ja resursseja. Äärimmäisen tärkeää on myös tiedonhallinnan suunnittelun opetuksen integrointi korkeakoulutukseen, sillä jo opintojen yhteydessä syntyy aineistoja.

Avoimuuden edistäminen edellyttää tutkimusaineistoihin liittyvien oikeuksien oikeanlaista hallinnointia. Selkeät datapolitiikat ja sopimukset selkeyttävät tilannetta. Aineistoihin tarvitaan esimerkiksi pysyvä oikeus valmistaa maksutta kappaleita, muokata aineistoja sekä siirtää oikeuksia edelleen kolmansille osapuolille, ellei siirto-oikeutta haluta jostain syystä rajoittaa esimerkiksi vain tutkimustarkoituksiin ja tutkijoille. Oikeuksien hallinnoinnissa toimijoita ovat tutkija(t), organisaatiot joissa tutkijat työskentelevät, aineistojen säilyttäjät ja edelleenjakelijat. Seuraavassa luvussa olevassa [taulukossa](#) on kerrottu tutkijan ja tutkimusorganisaation näkökulmasta

tutkimusaineistojen hallinnasta tekijänoikeuksien kannalta. [Tietoarkiston sivuilta](#) löytyy tietoa muista eettisistä ja juridista kysymyksistä.



Myös aineistojen säilyttäjän näkökulma tulee huomioida. Aineistot tulevat säilytykseen ja levitykseen erilaisilla ehdoilla, joten aineiston säilyttäjän tulee kyetä hallinnoimaan aineistojen erilaisia käyttöoikeuksia: mitä, miten ja kenelle aineistoja voi luovuttaa. Aineiston säilyttäjän pitää tarvittaessa kyetä erottelamaan erilaiset käyttäjät, tutkijat, opiskelijat ja muut henkilöt. Aineistoihin liittyvät henkilötietolain vaatimukset tulee ottaa huomioon samoin kuin mahdolliset salassapitovelvoitteet. On mietittävä, miten tällaisten tietojen kanssa toimitaan ja kuinka kauan veloitteet ovat voimassa. Aineiston säilyttäjä ei voi vastata aineistoihin liittyvistä kolmansien osapuolien oikeuksien loukkauksista, mutta sillä pitää olla valmius reagoida tällaisiin väitteisiin. Mikäli sopimuslissenssiin perustuvaa oikeutta luovuttaa tekijänoikeudella suojattujen aineistojen digitaalisia kopioita ei ole, tulee aineiston säilyttäjän saada oikeudenhaltijoilta lupa.

Tutkijan muistilista tietoaineistoista

- Kirjaa jo tutkimussuunnitelmaan, miten tulet säilyttämään ja käsittelemään tutkimusaineistoa hankkeen aikana ja sen jälkeen
- Selvitä eettiset, juridiset ja sopimukselliset kysymykset ennen tutkimuksen aloittamista. Katso [Tietoarkiston aineistonhallinnan käsikirja](#) ja [Avoimen tieteen käsikirjan luku aineistonhallinnasta](#). Kielentutkijalle myös [Kielipankin ohje](#).
- Tee tarvittavat sopimukset tutkimusaineistoihin liittyen
 - Aineiston tuottajien kanssa
 - Tutkimuskohteiden kanssa
 - Aineiston omistajien kanssa
 - Aineiston säilyttäjien kanssa
 - Tutkijoiden välillä
 - Kustantajien tms. julkaisijoiden kanssa
 - Muut mahdolliset osapuolet
- Huomioi sekä tutkimusaika että aika tutkimuksen jälkeen
- Selvitä miten työsuhteesi ja rahoituksesi vaikuttaa asioihin
- Mitä avoimpiin ratkaisuihin voit päästä, sitä yksinkertaisempaa julkaiseminen yleensä on
- Avaaminen saattaa vaatia paljonkin suunnittelua, sopimista ja esim. anonymisointia, mutta se voi kuitenkin parantaa tutkimuksen vaikuttavuutta ja laatua
- Huolehdi käyttöoikeuksista, lissensseistä ja riittävästä kuvailusta ja dokumentaatiosta

- Huolehdi siitä, että aineistosi on kuvailtu alasi tietoaarkistoon ja tietoaaineistojen hakupalvelu [Etsimeen](#)
- Myös hyödyntämättä jääneet tai nollatulokseen johtaneet aineistot kannattaa julkaista

Aineistohallinta: Tutkimuksen tietoaaineistot

Aineistohallinnan suunnittelu on tärkeä osa tutkimuksen suunnittelua. Sen avulla voidaan varmistaa tutkimusaineistojen järkevä jatkokäyttö. Aineistojen keruun, käytön, säilytyksen ja jatkokäytön suunnittelu selkeyttää myös tutkijan ja tutkimusryhmän omaa toimintaa. Suunnitelma on yhä useammin myös [rahoituksen ehto](#). [Tuuli-hankkeessa](#) on kerätty [tietoa eri tutkimusorganisaatioiden ohjeistuksesta](#) aineistohallinnan suhteen. Ohjeita on myös viety aineistohallinnan suunnittelutyökaluun [DMPTuuliin](#), mikä helpottaa suunnittelua. Tilanne kehittyy kuitenkin nopeasti ja ajankohtaista tietoa kannattaakin aina hakea myös oman organisaation verkkosivuilta tai tutkijapalveluista.

Aineistojen hallintaa helpottamaan kannattaa jo tutkimuksen alkuvaiheissa laatia erillinen suunnitelma, jota tarkennetaan ja päivitetään tutkimushankkeen edetessä. Tarpeelliset kysymykset, joihin suunnitelmassa tulisi vastata, ovat hyvin koottuna esimerkiksi [Yhteiskuntatieteellisen tietoaarkiston \(FSD\) Tutkimusaineistojen hallinnan käsikirjassa](#) seuraavasti:

- Mitä aineistoa kerätään?
- Miten aineisto kerätään?
- Mitkä ovat tutkimusaineiston omistus-, hallinta- ja tekijänoikeudet?
- Kuka antaa käyttöoikeudet aineistoon?
- Miten tutkittavia informoidaan?
- Mitä ohjelmistoja käytetään aineiston tallentamiseen ja prosessointiin?
- Miten varmistetaan aineiston (tekninen) laatu?
- Mitä tiedostoformaatteja ja tallennusmedioita käytetään?
- Minkälaiset oikeudet eri käyttäjäryhmillä on tiedostojen lukemiseen ja käsittelyyn?
- Miten hoidetaan varmuuskopiointi?
- Miten aineiston käsittely dokumentoidaan?
- Miten tallennetaan aineiston keruuta ja sisältöä kuvaava tieto (metadata)?
- Miten varmistetaan tietosuojaa?
- Miten huolehditaan tietoturvasta?
- Mitä aineistolle tapahtuu projektin päättyttyä? ([FSD 2013](#))

Tietoaaineistoja koskee sekä tekijänoikeuteen että henkilötietoihin liittyviä lakeja ja säädöksiä, jotka on tärkeä huomoida jo tutkimusta suunniteltaessa. Usein tarvitaan erillisiä sopimuksia, joihin vaikuttavat myös rahoittajien ja tutkimusorganisaatioiden ehdot. Jos aineistossa on ihmisiä koskevia tietoja on syytä tarkistaa Yhteiskuntatieteellisen tietoaarkiston [Aineistohallinnan käsikirjan](#) ohjeet ja suositukset. Sieltä löytyy myös [sopimukseen liittyvää tärkeää tietoa](#). Kieliaineistojen kohdalla kannattaa tutustua [Kielipankin ohjeisiin](#). Alla olevasta taulukosta taas löytyy tietoa tekijänoikeuksiin liittyvistä asioista.

Toimijanäkökulma	Tutkija	Tutkimusorganisaatio
<p>Millä rahoituksella tutkimusta tehdään?</p>	<p>Mikäli tutkimuksella ei ole ulkopuolista rahoitusta, tutkija tekee työtä ns. avoimessa tutkimuksessa, eikä aineistojen hallintaa ohjaamassa ole ulkopuolista tahoa. Tällöin riippuu tutkijan oman organisaation tavoista, miten hän saa määrätä aineistoistaan. Mikäli taas tutkimusta tehdään ulkopuolisella rahoituksella, ns. sopimustutkimus tai täydentävän rahoituksen tutkimus, täytyy tarkistaa rahoittajan ehdot:</p> <ul style="list-style-type: none"> • edellyttääkö rahoittaja aineistojen säilyttämistä ja avointa saatavuutta, vai • onko tulosten ja aineistojen osalta joitain rajoituksia, esimerkiksi aineisto on luovutettava mukana olevien yritysten käyttöön, vai • annetaanko rahoituksen saajan valita toimintatapansa. 	<p>Mikäli tutkimusta ei tehdä ulkopuolisella rahoituksella, organisaation omat ohjeet ja toimintatavat ratkaisevat, miten tutkimusaineistojen avointa saatavuutta edistetään:</p> <ul style="list-style-type: none"> • siirtääkö organisaatio itselleen oikeuksia ja määrääkö organisaatio aineistojen käytöstä, vai • säilyykö määräysvalta tutkijalla, jolloin on mietittävä, miten tutkijaa tuetaan, sekä • miten tutkijat sopivat keskenään aineistojen käytöstä. <p>Kun tutkimusta tehdään ulkopuolisella rahoituksella, organisaatio on sopimusosapuoli ja vastaa siitä, että rahoittajan ehtoja noudatetaan. Organisaation on tarvittaessa siirrettävä itselleen riittävät käyttö- ja omistusoikeudet aineistoihin rahoitusehtojen noudattamiseksi. Tässä tapauksessa organisaatio huolehtii siitä, miten ja millä ehdoin aineistoihin myönnetään oikeuksia. Mikäli rahoittaja ei aseta mitään tutkimusaineistoja ja tuloksia koskevia ehtoja, organisaation omat ohjeet ja toimintatavat ratkaisevat, miten tutkimusaineistojen avointa saatavuutta edistetään.</p>
<p>Kuka tekee tutkimusta?</p>	<p>Jos tutkimusta tehdään yhdessä muiden tutkijoiden kanssa, tulee tekijöiden sopia keskenään, miten kukin tekijä saa käyttää yhdessä aikaansaatuja aineistoja hankkeen aikana ja sen päätyttyä. Tässä yhteydessä on hyvä käydä lävitse kaikkien osapuolten oikeudet,</p>	

	<p>tekijyyttä koskevat periaatteet, vastuut ja velvollisuudet sekä käyttöoikeuksia koskevat kysymykset.</p>	
<p>Miten tutkimuksen aineisto saadaan käyttöön?</p>	<p>Jos tutkijoilla on oikeus määrätä aineistoistaan, eikä oikeuksia ole siirretty organisaatiolle, tutkimuksen tekijöiden tulisi päättää, millä ehdoin he haluavat antaa tekemiään aineistoja muiden käytettäväksi eli päättää lisensointitapa.</p> <p>Jos tutkimusaineistona syntyy esimerkiksi haastattelumateriaalia tms. aineistoa, joka sisältää henkilötietoja tai joka voi olla tekijänoikeudella suojattu, tarvitaan tällaisen aineiston käyttämiseen henkilöiden suostumus. Suostumusasiakirja on laadittava henkilötietolain puitteissa siten, että avoin saatavuus on mahdollista.</p>	<p>Organisaation varoilla ostetut materiaalit kuuluvat organisaatiolle eli aineistoihin voi liittyä kahdenlaisia oikeuksia: tutkijan immateriaalioikeuksia ja määräysvaltaa sekä organisaation omistusoikeus materiaaliin. Tutkijoiden ja organisaation on sovittava käyttöehdoista.</p>
<p>Aineistojen säilytys ja jako</p>	<p>Tutkijan pitää hyvissä ajoin, mielellään jo tutkimusta aloittaessaan miettiä, mitä, miten ja missä aineistot säilytetään, jotta ne ovat mahdollisimman hyvin ja avoimesti muiden käytettävissä. Tähän vaiheeseen liittyy myös säilytettävän aineiston metatietojen lisääminen ja muun muassa tiedot aineistoon kohdistuvista oikeuksista (esimerkiksi tekijänoikeuden haltija) ja käytön rajoituksista (esimerkiksi salassapito määräaikoineen).</p>	<p>Mahdollisimman aikaisessa vaiheessa tulisi miettiä yhteistyössä tutkijan kanssa, mitä, miten ja missä aineistot säilytetään, jotta ne ovat mahdollisimman hyvin ja avoimesti muiden käytettävissä. Tähän vaiheeseen liittyy myös säilytettävän aineiston metatietojen lisääminen ja muun muassa tiedot aineistoon kohdistuvista oikeuksista (esimerkiksi tekijänoikeuden haltija) ja käytön rajoituksista (esimerkiksi salassapito määräaikoineen).</p>
<p>Kolmansien osapuolien aineistojen käyttö</p>	<p>Jos käytetään muiden tekemiä aineistoja, tulisi aineistojen käytöstä sopia oikeudenhaltijoiden kanssa siten, että avoin saatavuus on mahdollista, ellei käyttö ole sallittu lain nojalla. Tutkijoiden pitää myös muistaa tarkistaa, että he käyttävät</p>	<p>Jos käytetään muiden tekemiä aineistoja, tulisi aineistojen käytöstä sopia oikeudenhaltijoiden kanssa siten, että avoin saatavuus on mahdollista, ellei käyttö ole sallittu lain nojalla. Lisäksi on aina tarkistettava ja huolehdittava</p>

	<p>aineistoja sopimusehtojen mukaista. Jos aineistoihin ei ole mahdollista saada riittäviä oikeuksia, tällaiset aineistot pitää merkitä ja huolehtia siitä, että näitä aineistoja käytetään vain rajoitetusti. Tähän vaiheeseen liittyy myös säilytettävän aineiston metatietojen lisääminen ja muun muassa tiedot aineistoon kohdistuvista oikeuksista (esimerkiksi tekijänoikeuden haltija) ja käytön rajoituksista (esimerkiksi salassapito määräaikoineen).</p>	<p>siitä, että aineistoja käytetään sopimusehtojen mukaista.</p> <p>Jos aineistoihin ei ole mahdollista saada riittäviä oikeuksia, tällaiset aineistot tulee merkitä ja on huolehdittava siitä, että tällaisia aineistoja käytetään vain rajoitetusti. Tähän vaiheeseen liittyy myös säilytettävän aineiston metatietojen lisääminen ja muun muassa tiedot aineistoon kohdistuvista oikeuksista (esimerkiksi tekijänoikeuden haltija) ja käytön rajoituksista (esimerkiksi salassapito määräaikoineen).</p>
--	---	---

Datan avaaminen

Tutkimuksen julkaisijat ja rahoittajat edellyttävät yhä useammin, että tutkimuksen tausta-aineistot tuodaan mahdollisimman avoimesti saataville viimeistään tutkimuksen päätyttyä.

Tutkimusaineistojen avoimella saatavuudella ja jaettavuudella edistetään tutkimusaineistojen jatkokäyttöä, josta hyötyvät aineiston tuottajan itsensä lisäksi muut tutkijat ja parhaimmillaan yhteiskunnat maailmanlaajuisesti. Aineistojen avaamisessa suositellaan kuitenkin lisensoijia, joiden avulla tutkija voi itse määrittellä datansa julkisuuden astetta ja käyttöoikeuksia. On myös tärkeä huomioida rahoittajien ja tutkimusorganisaation ohjeistukset ja vaatimukset, lain ja tutkimusetiikan lisäksi.

Avoimuuden asteista

Toisin kuin tieteellisten artikkeleiden kohdalla, jotka ovat joko julkaistuja tai eivät, on totuttu ajattelemaan, tutkimusaineistolla voi olla erilaisia julkisuuden asteita. Aineiston julkisuuden aste voi määräytyä joko lainsäädännöllisten tai tallennusteknisten perusteiden mukaan.

Teknisestä avoimuudesta: Mitä tarkoittavat avoin ja koneluettava formaatti?

Avoin formaatti tarkoittaa yleisesti ottaen jotain ei-kaupallista formaattia, jonka käyttämiseen ei tarvita mitään kaupallista ohjelmaa. Esimerkiksi ASCII-muotoiset txt-tiedostot aukeavat millä vaan notepadin tapaisella ohjelmalla millä tahansa käyttöjärjestelmällä, mutta Microsoft Wordilla tehtyjä doc-dokumentteja ei saa välttämättä auki oikeannäköisenä ilman maksullista Wordia. Vastaavasti taulukkomuotoinen tieto kannattaa tallentaa ja jakaa pilkkueroteltuna, tekstimuoisena csv-tiedostona enemmän kuin Excel-taulukkona. Jaettava aineisto pitäisi mahdollisuuksien mukaan tallentaa sellaisessa formaatissa, jossa sitä pystytään käyttämään mahdollisimman monilla käyttöjärjestelmillä ilman maksullisia ohjelmia.

Koneluetteavuus tarkoittaa, että tieto on sellaisessa muodossa että tietokone pystyy helposti käsittelemään tietoja. Yleinen PDF-tiedostomuoto ei esimerkiksi ole teknisesti katsottuna erityisen

joustava eikä siinä julkaistu tieto ole helposti uudelleen käytettävää. Koneluettavat formaatit ovat usein XML-pohjaisia formaatteja. Monilla aloilla on standardeja XML-pohjaisia formaatteja ko. alan aineistojen siirtoon. Esim. paikkatietoalalla tällainen formaatti on GML. Näiden erityisalojen ohjelmistoihin, esim. paikkatieto-ohjelmistoihin, sisältyy yleensä mahdollisuus tallentaa aineistot kyseisessä formaatissa tai muunnostoiminto kyseiseen formaattiin.

Tallennusteknisistä avoimuuden asteista konkreettisen kuvan antaa Tim Berners-Leen esittämän ns. [viiden tähden malli](#) täydennettynä [Linked Data Finlandin](#) kahdella tähdellä:

1* Julkaise tiedostosi missä tahansa formaatissa avoimella lisenssillä, esim. Creative Commons-lisenssillä

2* Julkaise tiedostosi rakenteisessa muodossa, esim. XML tai CSV-taulukko

3* Käytä avoimia, ei-kaupallisia tiedostomuotoja, esim. CSV, ei Excel

4* Käytä URI-tunnisteita

5* Linkitä tiedostosi muihin kontekstin luomiseksi

6* Datan ja sen skeeman dokumentointi (automaattinen)

7* Datan laadun automaattinen validointi (ja korjailu)

Kyse on siis siitä, miten dynaamista ja käyttökelpoista aineisto on. Linkitetyssä avoimessa datassa rakenne on koneluettavaa ja yksiselitteistä, eli yleensä helppokäyttöistä moneen eri tarkoitukseen. Mitä enemmän tunnisteita datassa on, sen parempaa se usein on. Näissä teknisissä asioissa tutkijan kannattaa kääntyä tiedonhallinnan asiantuntijoiden puoleen, esimerkiksi oman alan tietoarkistoon.

Koneluettavuudesta saat lisätietoja mm. seuraavista linkeistä: [Helsinki Region Infoshare](#) ja [Open Data Handbook](#).

Avoimuudesta käyttöoikeuksien suhteen

Täysin avoimesti saatavilla saattaa olla vain tiedot aineistosta, ja aineiston varsinainen käyttöön saaminen edellyttää käyttöoikeussopimuksen eli lisenssin hyväksymistä tai yksilöllisesti laadittua sopimusta. Tutkimusaineiston julkaiseminen verkossa ei vielä tarkoita, että aineistolla saisi tehdä mitä tahansa. Rajoituksia voi olla sekä teknisiä että juridisia. Julkaistua aineistoa saa ilman lupaakin katsoa, muttei välttämättä koskea. Muutenkin kannattaa miettiä voiko aineistosta julkaista osia, vaikkei kaikkea ehkä voi julkaista avoimesti. Aineiston käyttöehdoista määrää aina tutkimusaineiston tekijä tai se, kenelle tekijä on siirtänyt oikeudet määrätä aineistosta ([tekijänoikeuslaki 404/1961](#)). Aineistot voivat olla joko kaikille täysin avoimia tai käyttö sallitaan vain tiettyyn tarkoitukseen (yleensä tieteelliseen tutkimukseen). Aineistoissa saattaa olla joko arkaluontoisia ([henkilötietolaki 523/1999](#) 11 §) tai salassa pidettäviä ([laki viranomaisten toiminnan julkisuudesta 621/1999](#) 24 §) tietoja tai liike- tai ammattisalaisuuksia, jotka tutkimustuloksia julkaistaessa on jätettävä pois. Aineiston käyttö voi olla myös rajoitettua. Tällaisen aineiston käyttöön saaminen (esim. lataaminen) voi edellyttää tapauskohtaisen sopimuksen tekemistä aineiston tekijän kanssa. [Avoin tiede ja tutkimus -hanke](#) suosittelee [Creative Commons 4.0 -lisenssien](#) (CC BY) käyttöä, ellei aineiston sisällöstä muuta johdu. Creative Commons Suomen

toimintaa vetävät Helsinki Institute of Information Technology (HIIT) ja Aalto-yliopiston Taiteiden ja muotoilun koulun Medialaboratorio.

[Etsin - tutkimuksen tietoaaineistojen hakupalvelussa](#); on aineiston tietojen yhteydessä maininta, millaisin ehdoin aineistoa saa käyttää. Aineiston tekijä voi antaa käyttäjälle luvan tehdä aineistolla mitä vaan eli luopua kaikista oikeuksistaan, mutta silloinkin on huomattava, että sekä tekijänoikeuslaki että hyvä tieteellinen käytäntö edellyttävät, että tekijä mainitaan hyvän tavan mukaisesti. Näin on, vaikka aineisto olisi julkaistu CC0-lisenssillä. Tekijä saa ja voi asettaa aineiston käytölle tiettyjä ehtoja. Näillä ehdoilla hän antaa käyttöluvan aineistoonsa eli lisensoi aineistonsa. Jos aineisto julkaistaan jollain avoimella lisenssillä (esim. CC0 tai CC BY 4.0), niin aineiston julkaisija antaa kaikille aineiston lataajille oikeuden käyttää ja muokata aineistoa lisenssissä määritellyillä ehdoilla. Silloin käyttäjien ei tarvitse erikseen tehdä tai allekirjoittaa sopimusta eikä aineiston tuottajan tarvitse tietää, ketkä aineistoa käyttävät. Creative Commons 4.0 Nimeä -lisenssi on ollut myös pohjana työryhmällä, jonka tehtävänä oli laatia suositus Suomen julkishallinnon avoimiin tietoaaineistoihin sovellettavasta, kansainvälisesti yhteentoimivasta avoimen datan käyttöluvasta. [JHS 189 Avointen tietoaaineistojen käyttö lupa](#) on tänään Suomessa virallinen suositus. Lisenssin valintaa auttamaan on olemassa oma [sivustonsa](#). Myös [Avoindata.fi](#) tukee Creative Commons-käyttö lupaa. Lisää tietoa löydät käsikirjan [Aineistonhallinta](#)-osasta.

Embargo eli julkaisuviive

Globaalien ilmiöiden tutkiminen edellyttää usein työlästä ja kallista aineiston keruuta esimerkiksi kokeilla tai mittauksilla – ja toisaalta aineiston hyvää saatavuutta kansainvälisesti. Tällaisissa tapauksissa aineiston julkaisemiselle sovitaan usein suoja-/karanteeniaika (proprietary period, embargo), jonka aikana aineiston tuottajat voivat valmistella oman tutkimuksensa ja artikkelinsa ennen aineiston julkaisua muun tiedeyhteisön ja jatkotutkimuksen hyödynnettäväksi. Suoja-ajan pituus vaihtelee esimerkiksi kustantajan, rahoittajan tai tutkimusorganisaation mukaan. Tyypillisesti se on noin 1-3 vuotta datan keruuhetkestä alkaen.

Tieteellisten julkaisujen kuten artikkelien tapauksessa embargo tarkoittaa usein kustantajan määrittelemää viiveaikaa, jonka aikana artikkeleita ei saa julkaista vapaaseen verkkokäyttöön. Embargo-aika lasketaan alkavaksi lehden ilmestymispäivästä, joko elektronisen version tai painetun lehden ilmestymispäivästä riippuen siitä, kumpi tapahtuu aiemmin. Embargoaika vaihtelee kustantajien välillä, mutta on normaalisti noin 6-12 kk. Kaikilla kustantajilla viivevatimusta ei ole.

Datan jakamisen hyödyt

Tutkimusaineistojen avoimella saatavuudella ja jaettavuudella edistetään tutkimusaineistojen jatkokäyttöä, josta on hyötyä paitsi aineiston tuottajalle, myös rahoittajille, tiedeyhteisölle ja yhteiskunnalle niin kansallisesti kuin kansainvälisestikin.

Mitä avoin saatavuus käytännössä tarkoittaa?

Tutkimusaineistoilla voi olla useita julkisuusasteita täysin avoimesta julkisuudesta [lisensoilla](#) tai kertaluontoisilla sopimuksilla rajattuihin käyttöoikeuksiin. Julkisuuden aste on tutkimusaineiston omistajan määritettävissä, mutta hänen on otettava huomioon [aineiston käyttöä rajoittavat lait](#), kuten henkilötietolaki, julkisuuslaki ja tekijänoikeuslaki. Lisenssien hyödyntäminen on suositeltavaa.

Hyödyt aineiston tuottajalle

Aineiston julkaiseminen tuo tutkijalle tai tutkijaryhmälle julkisuutta kansainvälisesti. Avoimesti saatavilla oleviin tieteellisiin julkaisuihin viitataan useammin kuin pelkästään maksullisissa lehdissä julkaistuihin artikkeleihin. Lisäksi julkaisut, joiden tausta-aineisto on avointa, keräävät enemmän viittauksia kuin julkaisut, joiden tausta-aineisto ei ole saatavilla (ks. esim. Artikkelit [Data reuse and the open data citation advantages](#) ja [Scientists Who Share Data Publicly Receive More Citations](#)). Tunnettavuus puolestaan voi auttaa lisärahoituksen saamisessa. Avoimuus lisää tutkijoiden välistä kommunikointia madaltamalla sen kynnyksiä. Avoimuuden ansiosta tutkija voi saada tuntemattomiltakin aineistostaan palautetta, jonka pohjalta aineistoa voidaan parantaa edelleen. Palaute voi edesauttaa uusien tutkimuskysymysten asettelua ja uusien vastausten löytämistä analysoitaviin ongelmiin. Aineiston julkaiseminen tiedeyhteisön saataville osana avoimen tieteen prosessia voi kannustaa myös muita tutkijoita jakamaan aineistojaan. Datan jakamisessa ja avaamisessa on erittäin suositeltavaa hyödyntää lisenssejä, jotka varmistavat tutkijalle tietyt oikeudet.

Tutkija voi jo nyt lisätä julkaisemansa aineistot ansioluetteloonsa. Suomen Akatemia suosittelee käyttämään rahoitushakemuksissa [TENK Tutkijan ansioluettelomallia](#). Toivottavaa onkin, että jatkossa tätä kertyvää ansiotietoa hyödynnetään rahoituspäätöksissä.

Hyödyt tiedeyhteisölle

Tiedeyhteisö hyötyy tutkimusaineistojen jakamisesta esimerkiksi seuraavin tavoin:

- Tieteenteko tehostuu ja nopeutuu: Aineistojen jatkokäyttö säästää merkittävästi resursseja, sillä aineistojen kerääminen on usein kallista ja työlästä. Jotkut aineistot ovat niin ainutlaatuisia, että niiden uudelleen kerääminen on joko mahdotonta, hyvin kallista tai eettisesti arveluttavaa. (FSD 2013)
- Tiedeyhteisö voi käyttää saatavilla olevaa tutkimusaineistoa päätelmien ja tulosten luotettavuuden arviointiin, uudelleentulkintaan tai uusiin tarkoituksiin - uusia tutkimusongelmia tarkasteltaessa ja uusilla tieteenaloilla. Aineistojen avoin saatavuus voi edistää myös uusien tutkimusmenetelmien kehittämistä. (FSD 2013)
- Laajat, saatavilla olevat aineistot mahdollistavat ajalliset ja paikalliset vertailut (FSD 2013).
- Datan jakaminen tuo takeita tiedon säilyvyydestä yksittäisten tutkimushankkeiden päättyessä.
- Aineistojen avoin saatavuus voi tehostaa opetusta ja uusien tutkijasukupolvien koulutusta.
- Tutkimusaineistojen avoin ja helppo saatavuus parantaa tutkimuksen innovatiivisuutta sekä tieteenalojen sisäistä ja välistä verkostoitumista ja tutkijoiden tunnettuutta. Ohessa voi syntyä keskustelua tai uusia tutkimuspolkuja. (FSD 2013)
- Internetissä avoin tieto löytyy helpommin eri hakukoneilla.
- [OECD:n suosituksen](#) mukaan julkisrahoitteisen tutkimusdatan avoin saatavuus on olennainen edellytys myös tulokselliselle kansainväliselle tiedeyhteistyölle.

Hyödyt rahoittajille

Tieteen avoimuus ja läpinäkyvyys, johon kuuluu aineistojen avoin saatavuus, helpottaa myös rahoittajien työtä:

- Resurssit on helpompi kohdentaa oikeisiin kohteisiin, kun rahoittaja tietää, mitä aineistoja on olemassa. Väärennöstä tai plagiointia epäiltäessä data on tarjolla arviointia varten.
- Rahoittaja voi huomioida tutkijan ansiot aineistojen tuottamisessa tiedejulkaisujen lisäksi.

Tutkimusdatan avoimuuden hallinta esimerkiksi lisensoinnin avulla on kuitenkin suositeltavaa, jotta tutkija voi etukäteen määritellä omat oikeutensa dataan.

Hyödyt yhteiskunnalle

Julkisilla varoilla rahoitetun tutkimuksen tulosten avoin saatavuus lisää tutkimuksen läpinäkyvyyttä, uskottavuutta ja hyödyllisyyttä veronmaksajien näkökulmasta. Avoimuus edistää demokratiaa ja parantaa poliittisen päätöksenteon läpinäkyvyyttä ja laatua, kun sekä kansalaisilla että päätöksentekijöillä on päätösten kannalta oleellinen tutkimustieto paremmin käytettävissään. Avoimista aineistoista on hyötyä niin julkishallinnolle kuin yrityksillekin: ne voivat luoda uusia mahdollisuuksia innovaatiotoiminnalle, erityisesti PK-sektorille. Oleellisen tärkeää on että aineistot ovat yksiselitteisesti tulkittavissa ja riittävästi kuvailtuja jatkokäyttöä ajatellen.

Kansainvälisesti aineistojen ja julkaisujen avoimuus edistää tieteen ja tutkimuksen tasa-arvoisuutta ja kehitystä ja demokratisoi tiedon saatavuutta sekä kansallisesti että globaalisti tarjoten tasa-arvoiset mahdollisuudet aineistojen käyttöön saamiselle myös niille tutkijoille (esimerkiksi kehitysmaissa), joilla ei ole varaa kalliiden aineistojen ostamiseen. Tämä voi parhaimmillaan vaikuttaa myönteisesti köyhien maiden kehitykseen. (Ristikartano 2010)

Tutkimusaineiston julkaisuprosessi

Valitse ja valmistele aineisto

Tutkimusaineistojen julkaisua säätelevät lait, tutkimusetiikka ja rahoittajien ja tutkimusorganisaation ohjeet. Jotta tutkimusdatasta tulisi valmis aineisto, se on kuvailtava ja dokumentoitava tarpeeksi hyvin. Kun valitset julkaistavia aineistoja, kannattaa aloittaa esimerkiksi sellaisista, joka on joko valmiiksi helposti julkaistavassa muodossa, tai sellaisesta, jolla tiedät jo olevan kysyntää. Voit myös halutessasi olla etukäteen yhteydessä potentiaalisiin uudelleenkäyttäjiin. Se voi auttaa valitsemaan julkaistavia aineistoja. Voit tutustua myös [Julkinen data-sivustoon](#) tai [Open Knowledge Open Data Handbookiin](#). Datan avaamisen [MOOC-kurssin aineistot](#) löytyvät myös verkosta.

- Kaikkea dataa ei tarvitse julkaista kerralla
- Tarkasta datan laatu ja oikeellisuus
- Suorita tarvittaessa tarpeellinen anonymisointi, katso Tietoarkiston [ohjeet](#)
- Määrittele ja merkitse datan kuvailutiedot mahdollisimman varhaisessa vaiheessa, katso [lisäohjeita](#) kuvailusta ja metadatatista
- Huolehdi tunnisteista, viittauksista ja linkityksistä, katso tunnisteista

- Muunna data avoimeen ja, jos mahdollista, koneluettavaan formaattiin. Koneluettavuus parantaa aineiston löytämis-, hyödyntämis- ja yhdistelymahdollisuuksia ohjelmallisesti, esim. webissä. Mahdollisuus koneluettavuuteen on kuitenkin alakohtaista, eivätkä kaikki aineistot ole muunnettavissa. Lue lisää teknisestä avoimuudesta käsikirjan [edelliseltä sivulta](#).

Liitä aineistoon mahdollisimman avoin lisenssi

Lisensointi kertoo aineiston tuottajan ja käyttäjän oikeudet, ja näin ollen parantaa aineiston uudelleenkäyttämismahdollisuuksia huomattavasti. Epäselvät tai kokonaan puuttuvat käyttöehdot saattavat rajoittaa sen käyttöä paljon enemmän kuin sen haltija on tarkoittanut. Lisenssiin valintaan voivat vaikuttaa myös organisaation politiikat, rahoittajan vaatimukset tai muut sopimukset. Katso myös [Tutkijan tiedonhallinnan muistilista](#) sekä luku [Tutkimuksen tietoaineistot](#).

Tallenna aineisto luotettavaan tallennuspaikkaan

Kannattaa valita tallennuspaikka, joka on säilytyksen suhteen luotettava ja stabiili, sekä avoimen saatavuuden mahdollistava: aineisto tai ainakin sen metatiedot ovat avoimesti nähtävillä, maksuttomasti saatavilla, koneluettavassa ja helposti ladattavassa muodossa sekä kenen tahansa löydettävissä ja hyödynnettävissä. [Palvelut-sivulla](#) esitellään olemassa olevia kotimaisia tallennuspalveluja. Tutustu myös alakohtaisiin kansainvälisiin palveluihin esimerkiksi [re3data](#)-rekisterin avulla. Oman organisaation tutkimuspalveluista tai kirjastosta saa myös apua. Mikäli valitset ulkomaisen säilytys- tai julkaisupaikan muista varsinkin silloin erikseen huolehtia siitä, että aineisto tulee näkyviin [Etsimessä!](#)

Kuvalle aineisto ja anna sille pysyväistunniste viimeistään nyt

Aineiston kuvailun voi myös itse tehdä suoraan [Etsin](#)-palveluun. Osa tietoarkeistojen luetteloista kopioidaan sinne automaattisesti, älä siinä tapauksessa luo uusia tietueita tai tunnisteita. [Lisätietoja löytyy Avointiede-sivuston Palvelut-osioista](#). Metatiedosta ja kuvailusta lisätietoja käsikirjan [Metatiedot](#)-osasta ja tunnisteista voi lukea kohdasta [Tunnisteet ja viittaukset](#). Pysyväistunniste, kuten [URN](#) tai [DOI](#), on kansainvälisesti uniikki ja säilyy samana, vaikka aineiston sijainti vaihtuisi. Pysyväistunniste mahdollistaa viittaukset julkaisujen ja aineistojen välillä, mikä edistää tutkijan ansioitumista. Jotkut palvelut, kuten [Etsin](#) ja [IDA](#), generoivat automaattisesti URN-tunnisteen. Käyttäjä voi myös antaa muualta hankitun tunnisteen lisätessään aineiston palveluun.

Sovelluksen tekeminen

Pohdi riittääkö aineiston julkaiseminen pelkkänä datana vai kannattaisiko sitä tehdä käytettävämmäksi tarjomalla myös joitakin sovelluksia? Hyödyntämällä työkaluja aineistojen avaamiseen helpotat niiden käyttöä. Katso tarkemmin [AVAA-julkaisualusta](#).

Korkeakoulujen linjaukset

Ensisijainen vastuu hyvien tieteellisten käytäntöjen noudattamisesta tutkimusdatan suhteen on tutkijoilla. [Tutkimuseettisen neuvottelukunnan](#) mukaan "vastuu kuuluu kuitenkin myös koko tiedeyhteisölle: tutkimusryhmille ja niiden vastuullisille tutkijoille, tutkimusyksiköiden johtajille sekä tutkimusta harjoittavien organisaatioiden johdolle" (TENK 2012). Hyvän tieteellisen käytännön edistäminen samoin kuin tieteelliseen tutkimukseen kohdistuvien hyvän tieteellisen käytännön loukkausepäilyjen käsitteleminen kuuluvat ennen muuta tieteellistä tutkimusta harjoittaville organisaatioille. Joka tapauksessa tarvitaan lisäksi toimenpiteitä tukemaan

tutkimusdatan pitkää elinkaarta ja käytettävyyttä. Useat korkeakoulut ja rahoittajat ovatkin luoneet ohjeistuksia ja velvoitteita datanhallinnan käytäntöjen yhtenäistämiseksi ja tutkimusdatan jatkokäytön mahdollistamiseksi.

Korkeakoulut

Suomen korkeakoulujen rehtorien neuvosto on allekirjoittanut [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#)-julistuksen 2005. Käytännössä korkeakoulujen periaatelinjaukset, toimintatavat ja käytännöt datanhallinnan suhteen vaihtelevat jonkin verran korkeakouluittain.

Korkeakoulujen sitoutuminen datanhallinnan käytäntöjen yhtenäistämiseen ja aineistojen keruun, käytön, säilytyksen ja jatkokäytön suunnittelun tukemiseen on hyvin tärkeää. Datanhallinnan käytäntöjen pitäisi olla selvillä jo tutkimuksen alkuvaiheessa, jolloin tutkimusdatan avoimuus ja jatkokäytön mahdollisuus varmistettaisiin. Korkeakoulujen tulisi myös pyrkiä tarjoamaan tarvittavat tukipalvelut järkevän datapolitiikan toteuttamisen mahdollistamiseksi.

Korkeakoulujen tieto- ja tutkimushallinnosta sekä korkeakoulujen verkkosivuilta saa lisää tietoa korkeakoulukohtaisen datapolitiikan käytännöistä. Linkkejä tutkimusorganisaatioiden linjauksiin ja ohjeistuksiin on koottu [toisaalle](#) Avoin tiede ja tutkimus -sivuille. Tutkijan on syytä selvittää oman organisaationsa linjaukset jo tutkimuksen suunnitteluvaiheessa.

Tietoturva

Tietoturvallisuus tarkoittaa yleisesti ottaen sitä, että suojattava kohde, tässä tapauksessa tutkimusdata, suojataan riittävästi riskien varalta luottamuksellisuuden, eheyden ja saatavuuden osalta. Tietyissä tapauksissa tulee myös kiinnittää huomiota datan kiistämättömyyteen sekä tietosuojariskeihin. Tutkijan osalta kyse on lähinnä siitä, että tietoaineistoja käsitellään huolellisesti ja esimerkiksi riittävästä käytön- ja pääsynhallinnasta huolehditaan, kun sellaista tarvitaan. Neuvoja ja apua tähän saa esimerkiksi oman organisaation tukipalveluista tai [Yhteiskuntatieteellisestä tietoarkistosta](#). Tutkimusaineistojen säilytyspaikkoja ja –palveluita on aina syytä pohtia kriittisesti: ovatko ne riittävän luotettavia ja turvallisia? Tieteelliseen käyttöön tarkoitetut palvelut ovat usein luotettavampia ja esimerkiksi lainsäädännön kannalta turvallisempi valinta kuin puhtaan kaupalliset yleiset palvelut. Jos aineistossa on sensitiivistä dataa, on sen kanssa oltava erityisen huolellinen.

Joskus aineistoihin pääsyä on rajoitettava esimerkiksi tietosuojasyistä. Jonkun on silloin vastattava siitä, kenelle erilaisia oikeuksia aineistoihin annetaan (lukeminen, muokkaaminen, jakaminen). Tästä vastaa yleensä tietojen ja järjestelmien omistaja, joskus tutkija. Käytönhallinta on tekninen menettely, jolla pyritään rajaamaan pääsyä tietojärjestelmään tai tietoihin. Käytönhallinnassa voidaan hyödyntää esimerkiksi tunnistamista (autentikointia) ja valtuuksia (auktorisointia) siten, että tietojärjestelmään pyrkivältä käyttäjältä vaaditaan käyttäjätunnusta ja salasanaa. Suomen tutkimusjärjestelmässä yleinen käytetty menetelmä on [Haka-järjestelmä](#), johon liittyy myös aineistojen käyttöluupien hallinnointiin käytettävä [REMS-palvelu](#). Pääsynhallintaan liittyy myös fyysinen pääsy laitteisiin. Tietoturvan ja aineiston laadun kannalta on tärkeää, että järjestelmä kerää käytöstä mahdollisimman tarkkoja lokitietoja.

Koska tietoturvallisuus ja muu turvallisuus, kuten fyysinen turvallisuus ja henkilöturvallisuus, liittyvät kiinteästi yhteen, datan kohdalla turvallisuuden osa-alueita on mielekästä käsitellä yhtenä kokonaisuutena.

Riskillä tarkoitetaan haittaa tai vahinkoa joka voi johtua tapahtumasta, teosta tai passiivisuudesta. Riski voidaan ajatella haitallisen tapahtuman todennäköisyytenä kerrottuna sen vaikutuksen haitalla. Kun riskejä arvioidaan, tulee päätellä esimerkiksi aiempaan kokemukseen perustuen, kuinka todennäköinen tietty riski on. Riskin haitallisuus riippuu suojattavan kohteen luonteesta. Riskit voivat johtua tuottamuksellisista teoista, mutta myös vahingoista ja vioista, jälkimmäiset ovat tyypillisesti yleisimpiä datan käsittelyssä.

Tietomurron yritys verkkopalveluissa on hyvin tavallinen tapahtuma, kun taas datakeskuksen täydellinen tuhoutuminen on hyvin epätavallista. Mikäli tutkimusdata on helposti generoitavissa uudelleen, ei sen tuhoutumisen vaikutus ole merkittävä.

Datan luottamuksellisuus tarkoittaa, että vain valtuutetut ja oikeutetut tahot voivat käsitellä dataa. Mikäli data on julkista, pääsyä dataan ei tarvitse rajata, muussa tapauksessa data tulee luokitella pääsyn mukaan. Valtionhallinto käyttää erityisiä asetukseen ([VnA 682/2010](#)) perustuvaa luokittelua, valtionhallinnon ulkopuolella voidaan data luokitella vapaammin, esim. ryhmiin julkinen, rajattu käyttö, sisäinen, luottamuksellinen tai salainen. Datan omistajan tulee päättää miten sen data luokitellaan ja kenellä tulee olla oikeus käsitellä dataa.

Datan eheys tarkoittaa, että dataa voivat muuttaa vain ne tahot, joilla on siihen oikeus. Eheys tarkoittaa myös, että data on sisäisesti eheä siten, että se ei ole esimerkiksi korruptoitunut virheiden vuoksi.

Datan saatavuus tarkoittaa, että data on oikeutettujen käyttäjien saatavissa ja käytettävissä, kuten on sovittu. Liian suppeat pääsyoikeudet voivat esimerkiksi johtaa siihen, että data ei ole saatavissa.

Joissain tapauksissa myös datan kiistämättömyys on tärkeää. Silloin pitää pystyä osoittamaan, että data on alkuperäisessä tilassa, ja että sitä ei ole muutettu tarkoituksella tai vahingossa.

Tietosuojan tarkoitus on taata yksityisyyden suoja, jottei kenenkään henkilötietoja käsitellä tai paljasteta ilman lupaa. Henkilötietojen käyttöä säädellään mm. [henkilötietolaissa \(523/1999\)](#). Tietoturvatavoitteita (luottamuksellisuus, eheys, saatavuus) varmistetaan suojaamalla dataa riskien varalta erityisten turvallisuuskontrollien avulla. Turvallisuuskontrollit voivat olla mm. teknisiä kontrolleja, tai sopimuksiin liittyviä ja oikeudellisia kontrolleja. Usein kontrollit ovat kuitenkin kalliita toteuttaa tai ne voivat haitata datan saatavuutta, eivätkä ne myöskään välttämättä poista riskejä kokonaan, vaan vähentävät tai lieventävät riskejä kohtuulliselle tasolle. Kaikkien riskien varalta ei ole järkevää tai aina mahdollistakaan suojautua.

Taulukossa on lueteltu yleisiä datan hallintaan liittyviä riskejä sekä niiden varalta toteutettavia turvallisuuskontrolleja

Riski	Toimenpide
Datan menetys vian tai virheen vuoksi	Varmuuskopiointi, versionhallinta
Luottamuksellisen datan vuotaminen internetiin	Pääsynhallinta, käyttöoikeudet, datan salaus, salassapitosopimukset

Datan korruptoituminen	Tarkistussummat, varmuuskopiointi, versionhallinta, eheystarkistukset
Data ei ole saatavissa	Palvelutasosopimukset, varmuuskopiot, luotettavat palvelutoimittajat
Tutkija joutuu oikeudelliseen edesvastuuseen henkilötietojen väärinkäytöstä	Sopimukset, pääsynhallinta, datan salaus
Järjestelmäylläpitäjät käyttävät dataa väärin	Sopimukset, tietoturvasertifikaatit
Tietomurto	Työaseman, verkon ja palvelimien suojaus, hyvät salasanat ja muu tunnistus
Haittaohjelmat	Järjestelmien suojaus, luotettavat järjestelmät
Urkinta ja vakoilu	Järjestelmien suojaus, luotettavat järjestelmät
Oikeudellinen vastuu sopimusten tai lakien rikkomisesta	Sopimusten ja lainsäädännön noudattamisen varmistaminen, vastuuvakuutus
Datan menetys	Pitkäaikaisen tallennuksen turvaaminen
Data ei ole löydettävissä	Metatietojen käyttö, luotettavat palvelutoimittajat

Kun tutkimusryhmä tai yksittäinen tutkija miettii datan suojaamista, kannattaa lähteä liikkeelle systemaattisesti ja järkevästi. Tempoilu huhujen perusteella tuottaa harvoin hyvää turvallisuutta, kohtuullinen perusasioista huolehtiminen on usein paljon tärkeämpää.

Turvallisuuden varmistaminen kannattaa aloittaa datan turvallisuustekijöiden tunnistamisella sekä turvallisuusluokittelulla. Onko data luonteeltaan julkista vai tuleeko käyttöä rajata? Yleensä ainakin datan muuttamiseen kannattaa vaatia erityiset valtuudet. Data, joka on muutettavissa anonyymisti, on harvoin luotettavaa, ainakaan jos muutamista ei valvota. Myös datan eheys- ja saatavuusvaatimuksia kannattaa miettiä.

Riskit ja turvallisuuskontrollit kannattaa kirjata ja miettiä, miten suojaavat toimenpiteet toteutetaan ja kenen toimesta. Usein kannattaa ulkoistaa tekniset rutiinit palveluntoimittajille, mikäli mahdollista. Näin datan omistaja voi siirtää riskit pois itseltään.

Lähteitä ja lisätietoa:

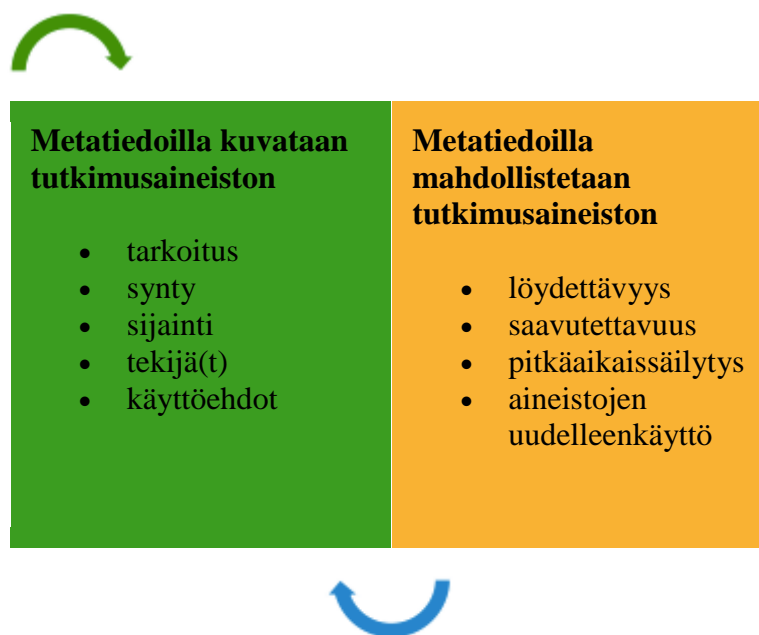
[Viestintäviraston uutisia tietoturvariskeistä](#)

[Valtionhallinnon tietoturvaohjeet](#)

Tutkimusaineistojen metatiedot

Aineiston hyödyntäminen edellyttää käyttäjän arviota sen soveltuvuudesta kyseiseen tutkimukseen. Tietoa saatetaan tarvita aineiston synnystä, alkuperäisestä käyttötarkoituksesta, saatavuudesta, sen tekijöistä ja käyttöehdoista. Esimerkiksi näiden asioiden kuvausta rakenteistettuna tietona kutsutaan metatiedoiksi.

Metatietoja voi olla tiedoston sisällä tai erillisenä luettelona esimerkiksi tietokannassa. Ne voidaan jaotella kolmeen päätyyppiin käyttötarkoituksen mukaan. 1. Kuvaileva metatieto kuvaa aineiston sisältöä ja luonnetta. Kuvailevasta metatiedosta vastaa joko tutkija tai jokin kolmas osapuoli, kuten informaattikko. Kuvailevan metatiedon sisällön tuottamisessa on suositeltavaa käyttää valmiita sanastoja, luokituksia ja ontologioita, jotka usein onkin integroitu aineistojen säilytys- tai hakupalvelujärjestelmiin. 2. Hallinnollinen metatieto määrittelee tutkimusaineiston hallintaan tarvittavat hallinnolliset ja tekniset tiedot, esimerkiksi aineiston käyttöehdot ja aineiston ja muiden tuotosten pitkäaikaissäilytykseen tarvittavat tekniset tiedot. 3. Rakenteellinen metatieto kuvaa tutkimusaineiston rakennetta ja sen järjestystä, esimerkiksi aineiston eri osien suhdetta toisiinsa. Aineiston hallinnolliset ja rakenteelliset metatiedot voidaan yleensä tuottaa pääosin automaattisesti.



Metatiedot mahdollistavat aineistojen löytymisen, saatavuuden ja uudelleenkäytön. Kertoessaan tarvittavat tiedot tuottamastaan aineistosta sen tuottaja saa näkyvyyttä ja mahdollistaa sen avulla meritoitumisen. Aineiston hyödyntäjä voi näiden tietojen perusteella arvioida mitä aineistoja on saatavilla, tunnistaa aineistojen kontekstin, valita tarpeitaan vastaavan aineiston ja hankkia tarvitsemansa aineiston eli saada tiedon sen sijainnista.

Tutkimusala ja sen vallitsevat käytänteet vaikuttavat tuotettavan metatiedon määrään. Useilla aloilla on kehitetty juuri kyseisen tutkimusalan aineistojen erityispiirteet huomioonottavia

metatietomalleja. Nämä metatietomallit näkyvät tutkijalle usein aineistojen säilytys- tai hakupalveluiden kautta, jolloin käytettävä palvelu ohjaa tutkijaa metatietojen tuottamisessa.

Tutkimusaineistojen metatietojen perimmäinen käyttötarkoitus on siis vastata tutkimusaineistoihin liittyviin kysymyksiin. Taulukossa alla on esitetty esimerkinomaisesti joitakin tutkimusaineistoihin liittyviä kysymyksiä ja niihin vastauksen antavia metatietoja. Metatiedon rakenteita muodostettaessa tulee huomioida olennaiset kysymykset eli ne asiat mitä tutkimusaineistojen sisällöstä ja olemassaolosta halutaan jatkossa tietää. Tämän tyyppinen kysymyksiin perustuva lähestymistapa voi auttaa purkamaan metatietomalleja selkokielelle ja ymmärtämään aineistojen säilytys- ja hakupalveluiden taustalla vaikuttavia malleja.

Sovittujen käytänteiden noudattaminen, tieteenalakohtaisten vaatimusten huomioiminen, aineiston riittävän kattava ja täsmällinen kuvailu sekä tehdyn kuvailun ymmärrettävyys, yksiselitteisyys ja -käsitteisyys, esimerkiksi käyttöehtoja ja aineiston omistajuutta koskevien kilpailevien tulkintojen esiintymättömyys, ovat merkittäviä asioita metatiedon laadun ja sitä kautta aineiston uudelleen käytön kannalta. Laadukas metatieto on tutkimusaineiston käyntikortti, jonka tutkija voi antaa tutkimusaineistostaan siitä kiinnostuneille.

Mitä?	
Mikä on tutkimusaineiston nimi?	Aineiston nimi (title)
Mitä tutkimusaineisto sisältää?	Kuvaus (description)
Mihin tutkimusalaan aineisto liittyy?	Tieteenalaluokitus (classification)
Mihin asioihin tutkimusaineisto liittyy?	Aihe (subject)
Minkä kielinen aineisto on?	Aineiston kieli (language)
Mitä muuttujatietoja aineistossa on? (esimerkki)	Muuttujatieto (dimension)
Mihin julkaisuun aineisto liittyy?	Julkaisu (publication)
Missä?	
Missä organisaatiossa aineisto on tehty?	Organisaatio (organization)
Missä projektissa aineisto on tehty?	Projekti (project)
Missä maassa aineisto on tehty?	Maa (country)
Missä aineisto on löydettävissä?	Katalogi (catalog)
Missä muodossa aineisto on saatavilla?	Tiedostotyyppi (format)
Milloin?	
Miltä ajalta tutkimusaineisto on olemassa?	Ajallinen kattavuus (temporal)
Milloin tutkimusaineisto on avattu?	Avaushetki (issued)
Milloin tutkimusaineistoa on muokattu?	Muokkaushetki (modified)
Kuka?	
Kuka on tehnyt tutkimusaineiston?	Tekijä (author)
Kuka omistaa tutkimusaineiston?	Omistaja (owner)

Kuka antaa tutkimusaineiston käyttöön?	Jakelija (distributor)
Miten?	
Miten tutkimusaineisto on kerätty?	Menetelmän kuvaus (method)
Miten menetelmästä saa lisätietoa?	Menetelmään liittyvät julkaisut (link)
Miten tutkimusaineistoa saa käyttää?	Käyttöehdot (rights)
Miten aineistosta saa lisätietoa?	Yhteystiedot (distributor contact)
Miten aineistoon linkitetään?	Aineiston tunniste (identifier)
Miten tutkimusaineistoon viitataan julkaisussa?	Lähdeviitteen muoto (citation)
Miksi?	
Miksi tutkimusaineisto on tuotettu ja avattu?	Tutkimusaineiston hallintasuunnitelma (DMP)
Mihin tutkimusaineisto on tarkoitettu?	Aihe (subject) / Kuvaus (description)

Tutkijan muistilista metatiedoista

- Huolehdi kaikkien aineistojesi ja julkaisujesi mahdollisimman laajasta kuvailusta ja metatietojen avoimuudesta
- Aikaan ja paikkaan liittyvät tiedot ovat myös tärkeitä, ilmoita ne ohjeiden mukaan. Aineistojen säilytys- tai hakupalvelut usein ohjaavat metatietojen täyttämässä.
- Käytä oman alasi terminologiaa ja asiasanoja.
- Muista myös vapaamuotoinen ja ainakin osittain yleistajuinen kuvaus: mistä on kysymys? (description).
- Suosi valmiita sanastoja, ontologioita ja yleisessä käytössä olevia termejä löytyvyyden takaamiseksi .
- Lisenssi/käyttöehdot ovat tärkeä metatieto, koneluettava formaatti kuten [CC-lisenssi](#) on suositeltava. Käytä mahdollisimman sallivaa lisenssiä.
- Huomioi aineistojen historia- ja versiotiedot yksiselitteisen viittaamisen mahdollistamiseksi.
- Esimerkiksi kirjastoista ja tietokannoista löytää kuvailun ammattilaisia, joilta kannattaa pyytää apua ja vinkkejä.
- Metatietomallesta kiinnostuneet voivat tutustua esim. [Ailaan](#), [Kansallisen digitaalisen kirjaston metatietosuositukseen](#), Digital Curation Centren [metatietorekisteriin](#).

Tunnisteet ja viittaukset

Kestävä tutkimus vaatii myös pysyviä, yksilöiviä tunnisteita. Tunniste on esimerkiksi jonkin julkaisun, organisaation, aineiston tai henkilön pysyvä, uniikki tunnus, jolla voidaan varmistaa, että kyseessä on juuri se yksilö tai yksikkö, johon halutaan viitata. Esimerkiksi tutkijan identifioivasta [ORCID-tunnisteesta](#) on hyötyä varsinkin, jos tutkija vaihtaa nimeä tai nimienkirjoitusasuja on useita tai samannimisä tutkijoita on useita. Tunnisteiden käytön avulla voidaan myös turvata viittausten kestävyys pitemmällä aikavälillä ja linkkien rikkoutumiselta

välttyä. Sähköisessä ympäristössä järjestelmisen vaihtumiset, jatkuvat päivitykset ja uusien versioiden syntyminen vaatii kaikilta toimijoilta suunnitelmallisuutta ja kestävyuden huomioimista tutkimuksen uskottavuuden varmistamiseksi. Myös esimerkiksi asiasanoilla voi olla tunnisteet, jolloin ne voivat olla monikielisiä tai niiden keskinäisiä suhteita voidaan kuvata koneluettavassa muodossa. Tällaista palvelua tarjoaa esimerkiksi [Finto-ontologiapalvelu](#).

Avoimen tieteen järjestelmien ja käytäntöjen tulee huomioida eri yhteisöjen tarpeet, mutta myös tavoitella mahdollisimman hyvää yhteentoimivuutta ja löytyvyyttä. Jotta tutkimustieto olisi mahdollisimman käytettävää ja löydettävää on syytä kiinnittää erityistä huomiota tietorakenteisiin ja semantiikkaan eli merkitysten ilmaisemisen muotoihin, kuten sanavalintoihin ja tunnisteisiin, joiden olisi hyvä olla mahdollisimman yhtensopivia.

Erilaisten aineistojen ja julkaisujen kohdalla on olemassa erilaisia käytäntöjä, jotka ovat peräisin eri tutkimus- ja julkaisuperinteistä. Yleisesti tutkimushankkeissa (mm. [ERC](#), [Horizon2020](#)) suositellaan tieteenalakohtaisten julkaisukanavien ja -arkistojen käyttöä, silloin kun niitä on tarjolla. Niissä aineistot saavat oikean kontekstin ja sopivan rakenteen. Arkistoa valittaessa on kuitenkin syytä kriittisesti tarkastella palveleeko arkisto riittävän hyvin avoimuutta ja uudelleenkäyttöä. Tietoarkisto on antanut [ohjeistusta](#) sähköisiin aineistoihin viittaamisesta. Lisäksi on huolehdittava siitä että tieto julkaisuista ja tietoaineistoista kulkeutuu myös omaan tutkimustietojärjestelmään sekä [Etsin](#)-palveluun riittävien kuvailu- ja tunnistetietojen kanssa.

Yhtenäiset yksilöntikäytännöt ovat erityisen tarpeellisia, kun tietojärjestelmät vaihtavat informaatiota avointen rajapintojen kautta, useiden eri järjestelmien välillä. Tällöin käytettävät tietosisällöt mm. koodistot ja niiden versiot on tunnistettava eli yksilöitävä. Kansainvälisesti on yksilöity mm. puhelinnumerot, sähköpostiosoitteet, kirjat ([ISBN-numero](#)), tuotteet ([EAN-koodi](#)). Muita yksilöntitunnuksia ovat mm. URI-nimet (Uniform Resource Identifiers).

Tutkimusjulkaisujen tunnisteena voidaan käyttää julkaisutyyppistä riippuen esimerkiksi [ISBN](#):ää (monografiat) tai erilaisia pysyviä tunnisteita (persistent identifiers, PIDs), joista [Handle](#)-tunnistetta käytetään esimerkiksi julkaisuarkistoissa, [DOI](#):ta kaupallisten kustantajien järjestelmissä ja [URN-tunnusta](#) kansalliskirjastojen digitaalisissa kokoelmissa. Tutkimusaineistoille sovelletaan kansainvälisissä ja kansallisissa hankkeissa lähes yksinomaan PID-tunnisteita, suomalaisessa tutkimusjärjestelmässä yleisimmin (esim. Opetus- ja kulttuuriministeriön hankkeet) URN:ää. Useat suomalaiset palvelut jakavat julkaisuille ja aineistoille Kansalliskirjaston myöntämän URN-tunnisteen. Aineistojen lisäksi tarvitaan tunnisteita tutkijoille sekä tutkimusprosessiin osallistuville juridisille henkilöille (yliopistot ja muut korkeakoulut sekä tutkimuslaitokset; tiedekunnat ja niiden laitokset, tutkimusryhmät). Näiden tunnisteiden soveltamista Suomessa selvitetään erikseen.

[Hyvä tieteellinen käytäntö](#)-ohjeen mukaan tutkijoiden tulee ottaa muiden tutkijoiden työ ja saavutukset asianmukaisella tavalla huomioon kunnioittamalla heidän tekemäänsä työtä ja viittaamalla heidän julkaisuihinsa asianmukaisella tavalla. Tunnisteiden ja muiden standardien avulla organisoidut tietosisällöt voidaan paremmin saada kansalliseen ja kansainväliseen käyttöön, sillä tunnisteista voidaan muodostaa pysyviä verkko-osoitteita. Tunnisteet tukevat näin tiedon globaalia löytyvyyttä ja täsmentävät tiedonhakuja. Ontologioiden, rekistereiden ja auktoriteettipalveluiden avulla voidaan luoda tunnisteita kaikille keskeisille metatiedoille.

Viittaaminen tutkimusaineistoihin

Tutkimusaineistoihin viittaamisessa pätevät yleensä samat perussäännöt kuin kaikessa viittaamisessa: tavoitteena on varmistaa, että muut tutkijat voivat löytää saman aineiston ja aineiston tekijät saavat heille kuuluvan meriitin. Usein viittausohjeita löytyy arkistojen verkkosivuilta, mutta myös esimerkiksi kustantajien määräykset vaikuttavat viittaustapaan. Aina kannattaa pyrkiä antamaan mahdollisimman täydelliset tiedot viitteessä ja/tai lähdeluettelossa.

Sähköisessä ympäristössä aineiston saatavilla ja käytettävänä pitäminen voi pitemmällä aikavälillä olla haastavaa. Siksi on tärkeää käyttää tutkimusaineistojen julkaisemiseen ja arkistointiin asianmukaisia palveluita, jotka tarjoavat niille pysyvän tunnisteen (Persistent Identifier), kuten URN tai DOI. Vain näin voidaan varmistaa, että aineisto todella on identtinen aiempaan verrattuna, myös pitemmän ajan kuluttua. Hyvät pysyvät tunnisteet ovat myös hyperlinkkejä, joiden avulla pääsee aineiston luokse verkossa. Pysyvän tunnisteen voi ilmoittaa vain lyhyenä tunnisteena, mutta ilmoittaminen täydellisenä http-alkuisena (esim. URN) tai https-alkuisena (esim. DOI) verkko-osoitteena helpottaa aineiston luokse pääsyä.

Tunnisteiden käyttö ja huolellinen kuvailu parantavat aineiston näkyvyyttä. Huolellinen viittaaminen mahdollistaa puolestaan myös yhä tärkeemmän koneluettavuuden, jonka avulla kasvavia tietomääriä on mahdollista hallita ja hyödyntää. Linkittämällä tutkimustuotoksia toisiinsa ne linkittyvät myös tekijöihin, jotka näin saavat näkyvyyttä ja tiedon laatu paranee.

Viitteessä on tietenkin tärkeää kertoa aineistosta tavanomaisen lähteen tavoin sen alkuperästä vaikka se olisikin täysin avointa dataa. On tärkeää dokumentoida mahdollisimman tarkkaan sellainen tiedon poiminta ja käsittely, joka vaatii kuvailua tutkimuksen todentamiseksi. Eri alojen käytännöt vaihtelevat siinä, missä ja miten tällaiset asiat tehdään, mutta on tärkeää, että etsivä löytää tarvittavat tiedot.

Tutkijan on syytä muistaa myös omien tutkimusaineistojensa arkistointi, riittävä kuvailu ja käytön maininta julkaisussa. Kaikkien tuotosten pitäisi löytyä esimerkiksi artikkelin lähteistä ja lähdeluettelosta, jotta niistä kertyy meriittiä. Jotkut tutkimusorganisaatiot edellyttävät myös organisaation mainintaa viittauksessa.

Käytettäviä tietoelementtejä voivat olla: Tekijä, Tekijän organisaatio, Organisaatio, Julkaisupäivä, Nimi, Versio, Aineiston tyyppi, Julkaisija, Jakaja, Arkisto, verkko-osoite, Pysyvä tunniste.

Esimerkkejä viittauksista

Tekijä(t), Vuosi, Aineiston nimi, Arkisto/Julkaisija/Jakaja, versio, pysyvä tunniste, esim. <http://urn.fi/urn:nbn:fi:lb-2016050903>

Tekijä(t), Vuosi, Aineiston nimi, Arkisto/Julkaisija/Jakaja, pysyvä tunniste esim. <https://doi.org/10.5284/1000389>

Tekijä(t), (Tekijän organisaatio), Vuosi, Aineiston nimi, Arkisto/Julkaisija/Jakaja, versio, pysyvä tunniste.

Tekijä(t), (vuosi): Aineiston nimi, Versio, Julkaisija, [aineiston tyyppi (esim sähköinen aineisto, korpus tms)], pysyvä tunniste.

Kannattaa tutustua myös [Tietoarkiston viittausohjeisiin](#).

Viittausten tekemiselle löytyy myös kansallinen standardi: Lähde- ja tekstiviitteitä koskevat ohjeet, SFS 5989, joka perustuu ISO 690:2009 standardiin. Tätä standardia ollaan parhaillaan uusimassa.

Tutkijan muistilista tunnisteista ja viittauksista

- Hanki itsellesi oma tutkijan tunniste [ORCID-tunnus](#), ellei sinulla ole sellaista ja käytä sitä joka yhteydessä
- Huolehdi tunnisteiden käytöstä aina kun viittaat
- Muista viitata myös omaan aineistoosi ja mainita se lähdeluettelossa
- Vinkkejä aineistoihin viittamiseen löytyy esim. [Tietoarkiston sivuilta](#), [Portti-verkkolehdestä](#), [Data Pubista](#) tai [DataCitestä](#)
- Menetelmiin (lähinnä sovelluksiin) viittamisesta voi lukea [Portti-verkkolehdestä](#) tai [Force11-ohjeista](#)
- Huolehdi omien julkaisujesi ja aineistojesi tunnisteista
- Hanki aina uudelle versiolle uusi tunniste, jotta viittaukset ovat luotettavia
- Vältä useamman tunnisteiden hankkimista samalle kohteelle
- Käytä tunnisteita ja linkityksiä ahkerasti!

Organisaatioille: Tunnisteet

Hyvien tutkijapalveluiden toteuttaminen ja tehokas palveluarkkitehtuuri vaativat organisaatioilta tunnisteiden hallintaa. Myös kansallisten arkkitehtuurien toimeenpano edellyttää organisaatioiden kiinnostavan huomiota semanttiseen yhteentoimivuuteen ja tiedon linkittämisen mahdollistamiseen. Tämä vaatii myös yksilöivien tunnisteiden ja niiden pysyvyyden pohdintaa. Ideaalitulenteessa verkossa näkyvää tunnistetta klikkaamalla pääsee aina alkuperäisen, yksilöidyn tiedon äärelle ja myös kone osaa tulkita tämän linkityksen sekä ymmärtää minkä tyyppisestä sisällöstä tai asiasta on kyse.

Tarjolla on sekä organisaatioiden sisäiseen käyttöön toimivia ratkaisuja että ulkopuolisia PID-palveluita. Esimerkiksi [paikatiedolle](#), [julkaisuille ja digitaalisille resursseille](#), [tutkijoille](#) sekä [opetus- ja koulutusosalalle](#) on olemassa eritasoisia ohjeistusta ja palveluita yksilöivien tunnisteiden ja niiden käytön osalta. Asiasanoille löytyy puolestaan tunnisteita [Finto](#)-palvelusta, jossa on myös hallinnonalan sanastoja. Finto tarjoaa myös [tieteenalaluokistusta](#). Kehitteillä on myös kansallinen Nimitietopalvelu. Yhteisten tunnisteiden, kuten [ORCID:in](#) käyttö helpottaa yleensä toimintaa huomattavasti ja on siksi suositeltavaa.

Mikäli käytössä on omia järjestelmiä joissa on verkkoon näkyviä tunnisteita, on syytä huolehtia siitä, että ne ovat vähintään [EU-kommission ohjeiden](#) sekä [W3C:n suositusten](#) mukaisia. Jos kyseessä on tavallinen URI sen vakaudesta on pidettävä huolta. Ulkoista palvelua valittaessa on puolestaan syytä tarkistaa, että järjestelmä on teknisesti luotettava, auktoritatiivinen, joustava metadatan esittämisen suhteen sekä yhteentoimiva omien ja kansallisten järjestelmien kanssa. Myös resolvoinnin tarvetta on syytä pohtia.

Erilaisia PID-järjestelmiä käytetään yhä useammin, sillä yksinkertaiset URI-tunnisteet eivät välttämättä ole riittävän pysyviä verkko-osoitteiden, -sivustojen tai organisaatioiden rakenteiden muuttuessa. Domainia valittaessa kannattaakin mieluummin valita domain-nimi, joka kuvaa

kyseistä tietovarantoa mahdollisimman hyvin, eikä organisaation nimeä sisältävää domain-nimeä. URI-tunnisteet voivat kuitenkin olla pysyviä niin kauan kuin niitä hallinnoiva organisaatio omistaa oikeuden kyseiseen verkko-osoitteeseen. PID-järjestelmän tunnisteet ovat puolestaan pysyviä niin kauan kuin palvelu on olemassa, eikä esimerkiksi verkkosivun osoitteen muutos vaikuta niihin. PID-järjestelmän, kuten esim. Handlen, ylläpitäminen vaatii kuitenkin jatkuvaa teknistä ylläpitoa ja osaamista. Content negotiation -mekanismin avulla järjestelmä voi myös sopeuttaa vastauksensa riippuen kyselyn agentista siten, että esimerkiksi selaimelle vastataan html-dokumentilla ja RDF-lukijan saadessa RDF-tiedoston. Tämä mahdollistaa uusien formaattien lisäämisen järjestelmään myöhemmin.

[Ruotsin yhteentoimivuussihteeristö](#) on kehittänyt organisaatioita varten seuraavan tarkistuslistan, jonka avulla voi lähestyä tunnisteiden hallintaa:

- onko organisaatiollasi datapolitiikka?
- onko organisaatiollasi tunnistepolitiikka?
- ovatko organisaatiosi tunnisteet sisäisesti yksilöllisiä?
- ovatko organisaatiosi tunnisteet globaalisti yksilöllisiä?
- ovatko organisaatiosi tunnisteet sisäisesti pysyviä?
- ovatko organisaatiosi tunnisteet globaalisti pysyviä?
- Mikäli käytössänne on vain organisaation sisäisiä tunnisteita, ovatko ne sellaisia, että niitä voidaan käyttää suositusten mukaisesti osana ulkoista pysyvää yksilöivää tunnistetta?
- onko käytettyjen tunnisteiden luonti ja käyttö organisaatiossa hallittua?
- toimiiko linkkien päivitys organisaatiossanne?
- onko tarvetta ulkoisen resolverin käytölle?
- jos pysyvät tunnisteet ovat URI-tunnisteita, ovatko ne EU-komission suosituksen mukaisia? ["http://{domain}/{tyyppi}/{käsite}/{viite}"]

Mikäli käytössänne on PID-järjestelmä:

- käytetäänkö sitä vain sisäisesti vai myös ulkoisesti?
- onko järjestelmää integroitu muihin järjestelmiin?
- onko järjestelmälle olemassa hallintasuunnitelmaa?
- koulutetaanko järjestelmän käyttäjiä?
- onko järjestelmä luotettava koskien esim. tietoturvaa ja kopiointia?
- mikäli käytätte ulkopuolisen osapuolen tarjoamia tunnisteita tai resolvointia, onko niitä ylläpitävä taho auktoratiivinen ja luotettava?
- onko järjestelmä joustava ja tarjoaako se tietoa sen laatua heikentämättä?
- onko järjestelmä yhteentoimiva (kansallisten ja kansainvälisten tai avointen standardien mukainen)?

Tutkimusorganisaatioilla on suuri vastuu hallinnoida ylläpitämiensä järjestelmien tunnisteita ja niiden pysyvyyttä. Jotta tutkijat voivat turvallisesti viitata julkaisuihin ja muihin tuotoksiin sekä tuoda sekä itselleen, että organisaatiolleen näkyvyyttä ja vaikuttavuutta, organisaatioiden on tuettava ja ohjattava tutkijoita tunnisteiden käytössä sekä hallinnoitava omia tunnisteitaan suunnitelmallisesti.

Lisää lukemista

[Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. \(ed.\) San Diego CA: FORCE11; 2014.](#)

Digisam: Checklista: [Beständiga identifierare](#)

[EU-komission ohje](#)

[Uniform Resource Identifier](#), Wikipedia

Kansalliskirjasto:[URN:NBN \(National Bibliography Numbers\)](#)

[Uniform Resource Name \(URN\)](#)

[Persistent URL \(PURL\)](#)

[Digital object identifier \(DOI\)](#)

[ORCID](#)

[Suomen ORCID-sivu](#)

[EPIC](#)

[Archival Resource Key \(ARK\)](#)

[Open URL](#)

[Handle-järjestelmästä](#), Wikipedia

[Best Current Practice RFC 7320 - URI Design and Ownership](#)

["Implementing Persistent Identifiers: Overview of concepts, Guidelines and Recommendations"](#), Consortium of European Research Libraries European Commission on Preservation and Access, 2006

["Persistent and unique identifiers"](#), CLARIN, 2008.

[Australian National Data Service ANDS Guides: Persistent Identifiers Working Level](#)

Aineistoa W3C:ltä

[Data Identification](#)

[URI Design Principles](#)

[Cool URISs don't change](#)

[Cool URIs for the Semantic Web](#)

[Research Data Alliancen video PID-järjestelmistä](#)

Tutkimusprosessin avaaminen

Sähköisessä maailmassa avoimuutta voi käyttää myös tutkimusprosessin tehostamiseen ja vaikuttavuuden lisäämiseen. Kirjapainoon nojaava kulttuuri on ohjannut tutkimusprosessin muotoutumiseen valmiiden painettujen tulosten julkaisemiseen tieteellisen tiedon manifestaationa. Vuorovaikutus ja tiedon jakaminen ennen julkaisua on myös käytännön syistä ollut rajoitettua tiedeyhteisön sisäiseksi asiaksi, mutta tänään tutkimusta voi jakaa ja tehdä yhdessä laajemmassa piirissä, jopa globaalisti.

Tekstin ja tiedon versioiden ja kopioiden lukematon määrä ja laajan yleisön pääsy lähes rajattomien aineistojen äärelle luo paineita myös tutkitun, tieteellisen tiedon aktiiviselle avaamiselle ja levittämislle mahdollisimman laajalla rintamalla. Myös siitä kertominen miten tutkimusta tehdään, mihin tieteellinen tieto perustuu ja miten se syntyy on yhä tärkeämpää. Tutkimustahan usein tehdään myös koska se on jännittävää ja mielenkiintoista ja tästä voisikin kertoa avoimemmin

kaikille kiinnostuneille. Tänä päivänä on myös uskomuksiin perustuvaa väärää ja jopa vaarallista tietoa runsaasti saatavilla. Siksi tieteellisen prosessin avoimuutta on syytä lisätä osallistamalla mahdollisimman laaja yleisö aina kun se on eettisesti ja juridisesti mahdollista. Avoimen prosessin aikana mahdollisesti löytyvät virheet, heikkoudet tai uusi tieto ovat, samoin kuin seminaareissa tai konferensseissa, omiaan tehostamaan työtä ja parantamaan tutkimuksen laatua ja uskottavuutta.

Tieteellistä prosessia voi avata esimerkiksi hyödyntämällä verkossa olevia ilmaisia palveluita kuten Flickr, SlideShare, YouTube, Vimeo, blogalustat, Twitter, Facebook tai muita vastaavia palveluita. Nämä voivat hyvin toimia myös oman projektin tai tiedeyhteisön apuvälineinä keskustelussa tai tiedonvaihdossa. Niissä on kuitenkin se etu, että ne voivat samalla toimia avoimena kanavana laajemmalle yleisölle. Aktiivisesti sosiaalista mediaa käyttävä tutkija saa myös ajan mittaan hyvän alustan levittää tieteellistä tietoa ja korjata verkossa tai perinteisessä mediassa esiintyviä virheellisiä käsityksiä. Tiedeyhteisön aktiivinen suora läsnäolo verkossa on erittäin toivottavaa. Näiden aineistojen hallintaa ja arkistointia on myös hyvä suunnitella alusta pitäen. Palaamme myöhemässä [luvussa](#) sosiaalisen median käyttöön.

"Julkaistua tietoa ei voi varastaa"
Tutkija Samuli Ollila

Niin sanotut harmaat julkaisut ovat perinteisesti olleet juuri "puolivalmiita" tuloksia ja eri ideoiden esittelyä ja testaamista esimerkiksi konferenssipapereiden tai luentoesitysten muodossa. Harmaiden julkaisujen kohdalla on hyvä varmistaa, ettei niiden julkaiseminen poissulje jatkojulkaisua vertaisarvioidussa lehdessä jos sellainen on suunnitelmassa. Tulosten laajemmasta nopeasta julkaisemisesta on usein pitäydytty perinteisistä syistä kuten esimerkiksi siksi, ettei niitä pidetä vielä ihan varmoina (virheiden riski) tai siksi että omaa mahdollista etumatkaa ei haluta riskeerata (kilpailu) minkä takia ei kaikkea kerrota avoimesti. Tulosten tai ideoiden "varastamista" tai oman uskottavuuden menettämistä saatetaan pelätä. Aikainen julkaiseminen verkossa (esimerkiksi esityskalvojen tallentaminen julkaisuarkistoon) on kuitenkin samalla tapa julkisesti liittää tutkimus omaan työhön ja altistaa se mahdolliselle kritiikille *ennen* lopullisen tutkimusjulkaisun tuottamista, samoin kuin esimerkiksi seminaarikäsittely. Sitä voidaankin pitää eräänlaisena avoimena vertaisarviointina, joka voi vain nostaa tutkimuksen laatua. Lisäksi tällaiset julkaisut antavat muille tutkijoille mahdollisuuden viitata myös keskeneräiseen tutkimukseen, mikä voi olla arvokasta, jos kyseessä on pitkä tutkimusprojekti jonka tuloksia joudutaan odottamaan pitkään. Yleiseen informaationlukutaitoon ja lähdekritiikkiin on kiinnitettävä huomiota, mutta julkiseen keskusteluun ja vapaaseen sanaan on myös voitava luottaa, sillä tieteellinen tieto on aina perusteltavissa ja mahdollista näyttää toteen. Tutkijoiden on siis syytä olla tarkkoja oikean kontekstin antamisesta tuotoksilleen ja kirjata tarpeelliset varaukset selkeästi, jokaiseen yhteyteen. Koska tieto leviää verkossa nopeasti ja kontrolloimatta, on hyvä esimerkiksi esityskalvoja julkaistaessa huolehtia siitä, että niistä ilmenee päivämäärä, esitysyhteys jne., jotta kuka tahansa pystyy tekemään oikean lähdekriittisen arvioinnin siitä, miten siinä esitettyjä tietoja voi käyttää.

Tutkimusprosessia voidaan myös avata kutsumalla mukaan kiinnostuneita osaavia harrastajia tai laajaa yleisöä tuottamaan tai käsittelemään aineistoja. Tällä tavoin ihmisiä osallistamalla voidaan mahdollisesti myös lisätä tutkimuksen yhteiskunnallista vaikuttavuutta. Laadunvarmistusta voidaan tehdä tarpeen mukaan eri tavoin, esimerkiksi toistamalla samoja tehtäviä eri tekijöillä, tekemällä pistokokeita ja analysoimalla aineistoja tilastollisesti. Tätä on syytä suunnitella tarkoin jo ennen tutkimuksen aloittamista tutkimustehtävän tarpeet huomioiden. Kaikki prosessit ja muutokset on tietenkin syytä dokumentoida huolellisesti. On myös hyvä antaa osallistuneille tietoa tutkimuksen etenemisestä ja tuloksista.

Tutkijan muistilista tutkimusprosessin avaamisesta

- Suunnittele tutkimusprosessin aikana tapahtuvaa viestintää jo etukäteen
- Viestintä voi olla sekä projektinsisäistä että ulkoista, huomioi kummatkin
- Mieti miten avointa projektinsisäinen viestintä voi olla, avoimuus saattaa avata uusia mahdollisuuksia!
- Ulkoinen viestintä on sekä tiedeyhteisöön että laajemmalle yleisölle suunnattua
- Muista blogit ja sosiaalinen media, ne ovat tämän päivän "avoin seminaari" ja tehokas verkostoitumisväline
- Lehdistötiedotteita ja mediakontakteja kannattaa hyödyntää, pyydä apua organisaatiosi ammattilaisilta
- Mieti missä vaiheessa mitään voi julkaista (dataa, aineistoja, menetelmiä, työpapereita, esityskalvoja, videoita)
- Miten hallinnoitte projektissa aineistojen jakamisen ja versioinnin?
- Minkälaisia yhdessä kirjoittamisen tai muun yhdessä tuottamisen muotoja ja alustoja voitte käyttää?
- Miten viestinnälliset materiaalit voi linkittää? Haluatteko arkistoida niitä?
- Käytä paljon ristiinlinkityksiä ja käytä omia tekstejä ja kuvia yms. huoletta moneen kertaan eri kanavissa
- Muista myös kuvien, videoiden, blogitekstien yms. lisenssit. Mitä enemmän [Creative Commons-lisenssejä](#), todennäköisesti sen parempi näkyvyys
- Jaa ja linkitä ahkerasti myös kollegoiden hyviä aineistoja. Verkossa on hyvä olla aulis ja tärkeää olla korrekti
- Osallistu keskusteluun blogeissa, twitterissä yms. Ole positiivinen ja rakentava. Silloin voit luontevasti myös mainostaa omia aineistoja
- Voiko täysin raakaa dataa julkaista heti esim. [Open notebook](#)-tyyppisesti?
- Koodia voin jakaa ja työstää avoimesti esimerkiksi [GitHub](#)-alustalla
- Avointa kirjoittamista voi tehdä wiki-alustoilla, blogeissa tai vaikka Googlen Drive-palvelussa. Muista [tietoturva](#)!
- Erilaiset alustat kuten [Eduuni](#), [BaseCamp](#), [f1000](#) tai [Slack](#), tai sosiaaliset mediat tarjoavat mahdollisuuksia keskusteluun myös tutkijoiden kesken. Muista arkistointi!
- Muista, että vaikka kaikkea tietoa ei voi jakaa tai julkaista, sen olemassaolosta ja ominaisuuksista on silti tärkeä kertoa!

Mallidokumentti:

- [Lupa puhujilta esitysten taltioimiseen ja julkaisemiseen](#)

Tutkimustulosten julkaiseminen

Tieteellinen julkaiseminen poikkeaa monella tavalla muusta julkaisutoiminnasta. Tieteellisillä julkaisukanavilla katsotaan yleensä olevan neljä perustehtävää:

- julkaisujen vertaisarviointi (laadunvarmistus)
- julkaisujen rekisteröiminen
- tiedonvälitys
- julkaisujen arkistointi

Kirjoittajille ei yleensä tieteellisissä lehdissä makseta heidän tekemästään työstä, vaan he saattavat päinvastoin joutua maksamaan lehdelle käsittely- tai kirjoittajamaksuja saadakseen artikkelinsa julkaistua. Tekijän julkaisusta saamat hyödyt eivät siis ilmene suoraan rahallisena korvauksena, vaan ne tulevat kiertotietä – toisaalta julkaisut tuovat esiin tutkimushankkeiden ja -organisaatioiden toiminnan tuloksia, toisaalta ne edistävät tutkijan omaa meritoitumista.

Sen lisäksi, että tieteellinen julkaiseminen palvelee tieteellistä keskustelua ja tiedonvälitystä, tieteellisillä julkaisukanavilla ja niihin kytketyllä vertaisarvioinnilla on siis merkittävä rooli akateemisen tutkimuksen laadunvalvonnan ja tutkijoiden meritoitumisen kannalta. Tieteellinen julkaiseminen tapahtuu käytännössä suurimmaksi osaksi yliopistojen ja tutkimuslaitosten ulkopuolella. Vaikka näiden organisaatioiden henkilökunnan kirjoittajina, vertaisarvioijina ja lehtien toimittajina tekemällä työllä on suuri merkitys, päävastuussa julkaisukanavien toiminnasta ja sen kehittämisestä ovat yleensä tieteelliset seurat ja kansainvälisellä tasolla myös kaupalliset kustantajat.

Vuosituhaten vaihteen jälkeen näkyvään rooliin noussut Open Access –liike (OA) on pyrkinyt löytämään vaihtoehtoja aiemmalle tilausmaksuihin ja julkaisujen rajoitettuun saatavuuteen perustuvalla toimintamallilla. Liikkeen taustalla on pidempi historia, jonka varhaisia virstanpylväitä ovat 1990-luvun kuluessa vähitellen yleistyneet verkossa ilmestyvät open access-lehdet ja fysiikan alan vuonna 1991 perustettu avoin pre print-arkisto [arXiv](#). ArXiv arkistoon tallennettiin lopulta lähes kaikki tieteenalalla julkaistavat uudet artikkelit jo ennen kuin ne ilmestyivät perinteisissä lehdissä.

Poliittisella tasolla open access -liike nousi yleiseen tietoisuuteen 2000-luvun alkuvuosina, kun sen periaatteet kiteytettiin [Budapestin \(2001\)](#) ja [Berliinin \(2003\)](#) julistuksissa, joista jälkimmäisen myös suomalaisyliopistojen rehtorit allekirjoittivat vuonna 2006. Käytännössä liike pyrki edistämään kahta rinnakkaista toimintamallia: avoimiin tieteellisiin lehtiin perustuvaa kultaista tietä (Gold OA) ja avoimiin julkaisuarkistoihin perustuvaa vihreää tietä (Green OA). Vihreän ja kultaisen tien eli rinnakkaistallennuksen ja avointen lehtien keskinäisestä priorisoinnista on käyty paljon keskustelua. Vaikka molemmat edistävät avoimuutta, niiden tuottamat hyödyt ovat erilaisia. Silti kultainen ja vihreä tie pitäisi nähdä toisiaan täydentävinä ratkaisuin, ei kilpailevina vaihtoehtoina.

Julkaisutiedoista on tullut Suomessakin keskeinen tieteellisen toiminnan laadun ja tuloksellisuuden mittari viimeistään yliopistojen vuonna 2013 voimaan tulleen uuden rahoitusmallin myötä. Sekä tutkimustietojärjestelmät että julkaisuarkistot ovat tutkimusorganisaatioiden kannalta keskeisiä tietojärjestelmiä tässä suhteessa.

Tutkijan näkökulma

Periaatteessa kustantajien ylläpitämä kultainen open access, johon liittyvät kirjoittajamaksut (niin kutsutut APC, eli Article Processing Charge) ei sulje pois julkaisuarkistojen käyttöä, joka aina on suositeltavaa. Tallentamalla julkaisunsa julkaisuarkistoon tutkija voi varmistaa julkaisun saatavuuden pitkällä aikavälillä, sekä sen löytyvyyden ja koneluettavuuden. On hyvä muistaa, ettei julkaisuarkiston käyttö millään tavalla poista tekijänoikeudellista suojaa, mutta se mahdollistaa esimerkiksi selkeän viittaamisen, löytyvyyden ja tiedonlouhinnan. Myös tutkijan organisaatio ja muut toimijat hyötyvät niiden piirissä tuotettujen tutkimustulosten löytyvyydestä ja näkyvyydestä. Siksi jotkut organisaatiot ja rahoittajat suosittelevat tai jopa vaativat open access -julkaisemista, mikäli se on mahdollista. Tekijä voi kuitenkin antaa julkaisulleen esimerkiksi haluamansa [Creative Commons](#)-lisenssin, jos haluaa edistää julkaisun tai sen osien leviämistä vielä

voimakkaammin. Julkaisukanavaa valittaessa on syytä varoa ns. saalistavia huonolaatuisia lehtiä. Julkaisufoorumi on myös pyrkinyt [huomioimaan](#) nämä.

Julkaisusopimuksia tehdessä tutkijan kannattaa siis varmistaa, että pitää itsellään riittävät oikeudet rinnakkaistallentamiseen. Kustantajan linjauksista riippuen julkaisuarkistoon voi muutenkin useimmiten tallentaa ainakin julkaisun jonkun version, esimerkiksi lopullisen julkaistun version tai viimeisen tarkastetun version. Tietoa näistä saa [Romeo/Sherpa](#)-palvelusta mutta asia kannattaa aina varmistaa kustantajalta. Julkaisuarkistossa kuvailutiedoissa tulee ilmaista mistä versiosta on kysymys. Useimmat kustantajat suostuvat rinnakkaistallentamiseen ainakin tietyn viiveen eli ns. embargo-ajan kuluttua, esimerkiksi kuuden kuukauden tai vuoden jälkeen. Myös monografioiden kohdalla voi hyvin huomioida rinnakkaistallentamisen mahdollisuuden. Usein kirjoittajalla on kahden vuoden jälkeen oikeus julkaista aineisto uudestaan, ellei uusia painoksia ole tehty. Erityistä huomiota on kuitenkin syytä kiinnittää kuvitukseen, sillä kuviin liittyvät tekijänoikeudet voivat vaatia niiden poistamista rinnakkaistallennettavasta versiosta, ellei asiasta ole sovittu erikseen tai kuvien tekijänoikeuden voida katsoa umpeutuneen. Tutkijoille on yleensä omassa organisaatiossa tarjolla neuvontaa ja tukea julkaisemiseen liittyen, ja näitä tukipalveluita kannattaa hyödyntää.

Tutkijan muistilista julkaisemisesta

Yleiset

- Suosi avoimia julkaisukanavia aina kun se on mahdollista.
- Pyri mahdollisimman nopeaan avoimuuteen ja julkaisemiseen. Saat todennäköisesti enemmän näkyvyyttä ja viittauksia. Vältä pitkiä embargoaikoja ja kustantajia, jotka vaikeuttavat esimerkiksi rinnakkaistallentamista.
- Huolehdi hyvästä kuvailusta, pyydä neuvoja kirjastosta.
- Muista tunnisteet ja linkitykset, käytä aina niitä ahkerasti. Noudata kustantajan kanssa sovittuja tapoja viitata ja linkittää alkuperäiseen dokumenttiin.
- Julkaise ja linkitä mahdollisimman laajasti myös menetelmiä ja aineistoja, se parantaa sekä näkyvyyttä että uskottavuutta.
- Tutkimusaineistojen julkaisemista voi myös jalostaa esimerkiksi visualisointisovellusten avulla (ks [Avaa-palvelu](#)). Tämä lisää edelleen niiden kiinnostavuutta ja arvoa.
- Mieti muita julkaisumuotoja ja kanavia kuin tieteellinen artikkeli, voiko jotakin kertoa visualisoinnilla, videolla tms. tavalla?
- Ole tarkka etiikasta, huolehdi tietosuojasta. Hyvä dokumentointi ja avoimuus ovat myös osa tutkijan etiikkaa.

Sopimukset, oikeudet ja lisenssit

- Jos haluat patentoida työsi tulokset, tee se ennen julkaisemista.
- Kysy neuvoja esimerkiksi kirjastosta kun teet kustannussopimuksia.
- Rinnakkaistallentaminen tai avoin julkaiseminen eivät poista tekijänoikeutta, mutta lisäävät näkyvyyttä.
- Julkaisemalla avoimen jatkokäytön sallivalla lisenssillä (kuten [CC BY 4.0](#)) edistät mahdollisuuksia vapaasti jakaa ja hyödyntää julkaisuasi.

Rinnakkaistallennus

- Huolehdi rinnakkaistallentamisesta sopivaan julkaisuarkistoon. Tarvittaessa voi käyttää useaa, se lisää näkyvyyttä (esim. kansainvälinen alakohtainen ja organisaation oma julkaisuarkisto).
- Selvitä mitä julkaisuarkistoa rahoittajasi ja organisaatiosi edellyttävät tai suosittelevat.
- Myös [harmaan aineiston](#) julkaiseminen julkaisuarkistossa on hyvä vaihtoehto, silloin saa pysyvän tunnusteen, linkin ja säilytyksen helposti järjestettyä. Esimerkiksi konferenssien verkkosivut eivät valitettavasti aina ole ikuisia.

Esimerkkidokumentit:

- [Lupa rinnakkaisjulkaisemiseen kustantajalta \(rtf\)](#)
- Muutos kustannussopimukseen, joka mahdollistaa Horizon2020:n edellyttämän avoimuuden ([Word](#)) (Sherpa, [html](#))

Organisaation muistilista julkaisemisesta

- Laaditaan selvät suunnitelmat ja tavoitteet toimintakulttuurin kehittämiseksi.
- Tutkijoille tarjotaan riittävä tuki ja osaaminen.
- Immateriaalioikeuksiin liittyvät omistajuudet sovitaan selkeästi. Lisensioimista suositetaan mahdollisuuksien mukaan. Suomen julkisen avoimen datan virallinen lisenssisuositus on [CC 4.0 BY](#).
- Syntyviin tuloksiin niin datan, julkaisujen kuin menetelmienkin osalta sovelletaan kansallisia avoimen tieteen linjauksia, ellei asiasta ole tehty erillissopimusta (esim. kaupalliset toimijat).
- Tutkimusinfrastruktuurien oleellisten prosessien tulee olla helposti liitettävissä muihin kansallisiin tai kansainvälisiin prosesseihin avoimien rajapintojen kautta.
- Tutkimusinfrastruktuurit rakennetaan avoimia rajapintoja ja standardeja hyödyntäen.
- Vältetään yhden toimittajan ratkaisuja, aina kun se on mahdollista. Tutkimusinfrastruktuurin tuottamisessa ja kehittämisessä käytetään avointa lähdekoodia.
- Noudatetaan hyviä tiedonhallintatapoja.
- Noudatetaan yleisiä tutkimuseettisiä periaatteita ja hyvää tieteellistä käytäntöä.
- Palvelut ovat helposti löydettävissä ja hyvin kuvailtuja.
- Laadunvarmistusjärjestelmä on avoimesti kuvailtu.
- Käytettävyys- ja käyttötilastot ovat avoimesti saatavilla.

Sosiaalinen media ja julkaiseminen

Sosiaalinen media ja vastaavat julkaisu- ja kommunikaationalustat ovat myös tutkijoille yhä tärkeämpiä työvälineitä. Niihin liittyy kuitenkin asioita, joita on hyvä huomioida kun niitä käytetään tutkimuksessa tai siitä viestimessä. Sosiaalisen median palveluissa voi tuottaa yksin tai yhdessä muiden kanssa dokumentteja, esityksiä ym. sekä tallentaa ja jakaa niitä tai käydä muuten vain keskustelua omilla kirjoituksilla. Mikäli työstät kirjoituksia tai dokumentteja yhdessä muiden tutkijoiden kanssa, on hyvä sopia etukäteen kirjallisesti yhdessä tehdyn dokumentin käyttöehdoista kaikkien tekijöiden kanssa. Mikäli haluat liittää dokumenttiin tai kirjoitukseesi jonkun muun tekemiä aineistoja, kuvia, videoita tms. on tarkistettava, että niihin on riittävät käyttöoikeudet. Tekijöihin on aina viitattava. Mikäli tekijä on nimimerkki, sekin kelpaa.

Kuvien muokkaukseen ja jakoon on verkossa tarjolla myös valmiita kuvapankkeja ja kuvankäsittelyohjelmia. Muistathan kuitenkin aina tarkistaa, millaisia käyttöehtoja verkossa oleviin aineistoihin liittyy. [Creative Commonsilla](#) on oma hakupalvelunsa lisensoituja aineistoja varten. Myös esimerkiksi [Wikipedia Commonsista](#) löytyy paljon selkeästi lisensoituja aineistoja.

Jos aiot julkaista sellaisia kuvia tai nauhoitteita, joista henkilöt on tunnistettavissa, tulee ko. henkilöiltä aina pyytää lupa julkaisemiseen. Jos esim. haluat julkaista konferenssimateriaalia sosiaalisessa mediassa, tulee kaikilta konferenssin puhujilta ja esittäjiltä pyytää lupa materiaalien jakoon.

Avoimen julkaisemisen ja rinnakkaistallentamisen ratkaisuja ei tule perustua oletukseen, että sosiaalisen median palvelu toimii varmasti. Tutkimusjulkaisujen tai tutkimusdatan tallentaminen sosiaalisen median palveluihin ei myöskään yleensä täytä rahoittajien open access ehtoja. Esimerkiksi [Horizon2020](#)-tutkimusrahoitusta saavien hankkeiden osalta tulee avoimen julkaisemisen kanavat valita siten, ettei sosiaaliseen mediaan, kuten esim. ResearchGate palveluun, tukeuduta rahoittajan rinnakkaistallentamisen vaatimuksen täyttäjänä. Rinnakkaistallentamisessa tulee huomioida myös oman organisaation ohjeistus ja avoimen tallentamisen vaatimukset oman organisaation julkaisuarkistoon. Verkossa olevia verkostopalveluita voi sen sijaan käyttää julkaisuarkistoitujen julkaisujen disseminaatioon.

[Tutkijan muistilista sosiaalisen median käyttöön](#)

Kun käytät sosiaalista mediaa tutkimuksen tekemisessä tai julkaisemisessa otathan huomioon lisäksi seuraavat asiat:

- Jos tavoitteena on tutkimustulosten suojaaminen patentilla tai muulla rekisteröitävällä, immateriaalioikeudellisella suojalla, joka vaatii ehdotonta uutuutta, älä käsittele aihetta tai jaa mitään aiheeseen liittyvää tietoa sosiaalisen median palveluissa. Tällainen tutkimus tulee pitää salassa ainakin patenttihakemuksen tai muun rekisteröintihakemuksen jättämiseen saakka.
- Et voi julkaista sosiaalisessa mediassa mitään sellaista tietoa, joka tutkimusta koskevissa sopimuksissa on määritelty luottamukselliseksi.
- Jos tutkimuksessa käsitellään haastatteluja tai muita aineistoja, joihin liittyy henkilötietoja, huolehdi henkilöiden yksityisyyden suojasta ja tietosuojasta. Älä mainitse esimerkiksi haastateltavien nimiä tai muita tietoja, joista heidät voi tunnistaa.
- Monet kustantajat edellyttävät että heille julkaistaviksi luovutettua materiaalia ei ole aikaisemmin julkaistu. Älä laita kirjoituksiasi sosiaaliseen mediaan, ennen kuin olet varmistanut, millä ehdoin voit julkaista ne muualla.
- Ota huomioon lakien ja oman organisaatiosi määräykset siitä, miten tietoturvasta on huolehdittava. Sosiaalisen median palveluntarjoaja ei yleensä ota mitään vastuuta datan tietoturvasta.

Muista vastuusi työnantajan edustajana:

- Keskustele tutkimusryhmässä siitä, mitä asioita tutkimusryhmän jäsen voi kertoa tutkimuksesta esimerkiksi blogissaan tai muutoin sosiaalisessa mediassa. Noudata tutkimushankkeen sopimuksissa olevia rajoituksia siitä, millainen materiaali on luottamuksellista.

- Muistathan, että opiskelijoita, jotka eivät ole työsuhteessa yliopistoon, tulee erityisesti informoida tutkimushankkeeseen osallistuvien velvollisuuksista ja heidän osaltaan on tarvittaessa sovittava kirjallisesti toimintatavoista, joihin muut työsuhteen perusteella ovat velvoitettuja.

Julkaisuarkistot

Tämä luku on kirjoitettu julkaisuarkistoja ylläpitäville organisaatioille.

Julkaisuarkiston sisältämissä tietueissa pitäisi pystyä ilmaisemaan yksiselitteisellä ja yhdenmukaisella tavalla, liittykö tietueeseen kuvaillun kohteena olevan aineiston sisältävä kokotekstitiedosto tai -tiedostoja. Tiedon pitää olla haravoitavissa aineiston metadatan mukana muihin järjestelmiin. Tietueen pitäisi sisältää selkeästi ilmaistu tieto siitä, onko sen yhteyteen tallennettu kokotekstisisältö julkaisuarkiston kautta vapaasti luettavissa ja millä ehdoilla - esim. jonkin CC-lisenssin puitteissa - sitä voidaan muuten hyödyntää. Tiedon pitää olla haravoitavissa aineiston metadatan mukana muihin järjestelmiin.

Tietueiden metadatan pitää sisältää riittävän tarkat bibliografiset tiedot kuvailun kohteena olevasta aineistosta tai julkaisusta. Tietueisiin pitää sisältyä yhteisesti määriteltyyn sanastoon/ontologiaan perustuva tieto kuvaillun aineiston julkaisu- tai aineistotyyppistä (esim. vertaisarvioitu tieteellinen artikkeli). Etenkin rinnakkaistallennettujen artikkelien kohdalla pitää ilmaista yhdenmukaisella tavalla mikä versio arkistoon on tallennettu.

Julkaisuarkistoon tallennettujen aineistojen yhteydessä on suositeltavaa hyödyntää pysyviä tunnisteita (URN, Handle, DOI), joihin pohjautuvia pysyviä osoitteita voidaan käyttää julkaisuihin viitattaessa. Julkaisuarkistoissa on suositeltavaa käyttää julkaisuihin liittyville henkilöille ORCID- tai ISNI-tunnisteita sekä organisaatioille ISNI-tunnisteita niiltä osin kuin se on prosessien ja muiden järjestelmien – mm. kirjastotietokannat ja tutkimusjärjestelmät – kanssa sovitun työnjaon puitteissa mielekästä.

Aineistojen käyttöä on suositeltavaa tilastoida vähintään kokotekstitiedostojen latauksien tasolla esim. COUNTER suosituksen mukaisia teknisiä käytäntöjä noudattaen.

Alla Avoin tiede ja tutkimus-hankkeen julkaisuryhmän alustavia suosituksia tutkimusorganisaatioille. Niitä noudattamalla on mahdollista edistää avointa tiedettä ja oman tutkimuksen näkyvyyttä.

Tutkimusorganisaatiot:

- Vastaavat rinnakkaistallentamisen prosessien ja käytäntöjen tuomisesta osaksi tutkijoiden ja organisaation perustoimintaa.
- Seuraavat rinnakkaistallentamisen kehittämisen vaikutusta julkaisujen avoimeen saatavuuteen organisaatiossa.
- Kehittävät keskitettyjä palveluja rinnakkaistallentamisen käytännön toteuttamiseen.
- Käyttävät avoimien aineistojen yhteydessä koneluettavia lisenssejä, kuten CC-lisenssi.
- Kehittävät keskeytetyt palvelut ja prosessit tutkimusaineistojen hallinnan ja avoimen julkaisemisen mahdollistamiseksi.
- Kannustavat tutkijoitaan julkaisemaan tutkimustuloksia myös laadukkaissa Open Access - julkaisufoorumeissa.

- Kehittävät Open Access -lehtien kirjoittaja- ja tukimaksujen hallintaan ja käsittelyyn keskitettyjä palveluja ja rahastoja.
- Osoittavat julkaisuarkistojensa kehittämiseen riittävät resurssit.
- Huolehtivat tutkimustietojärjestelmien, julkaisuarkistojen ja data-arkistojen [OpenAIRE-yhteentoimivuudesta](#)
- Tutkivat mahdollisuudet linkittää rinnakkaistallentaminen julkaisutietojen keruu- ja tallennusprosessiin.
- Kehittävät menetelmiä ja käytäntöjä avoimesta julkaisemisesta palkitsemiseen.

Politiikat

Tämä luku on suunnattu ensisijaisesti tutkimusorganisaatioille. Avoimen tieteen ja tutkimuksen toteuttaminen vaativat niiltä ennen kaikkea yhteistyötä, koordinaatiota ja asiakaslähtöisyyttä järjestelmien ja palveluiden suunnittelussa. Toimintakulttuurin tulisi manifestoitua strategioissa, arvoissa ja toimintasuunnitelmissa, mutta toteutuakseen se vaatii myös toimenpiteitä palveluiden ja järjestelmien kehittämisen tasolla sekä selkeän kommunikaation kautta, esimerkiksi organisaation ohjeistuksessa tutkijoille, tai siinä miten tutkimuksen tuloksia ja tutkimusaineistoja esitellään tai voi hakea verkossa. Avoimuus edellyttää myös sitä, että organisaatio noudattaa yhteisiä yleisiä standardeja ja tekee yhteistyötä muiden toimijoiden kanssa oman tutkimuksen ja omien aineistojen näkyvyyden ja saavutettavuuden takaamiseksi. Myös omia erillISRatkaisuja tehtäessä on syytä ottaa huomioon suurempi konteksti sekä loppukäyttäjän tarpeet, jotta tutkimusympäristöstä tulisi tehokas ja avoin.

- Laaditaan selvät suunnitelmat ja tavoitteet toimintakulttuurin kehittämiseksi.
- Immateriaalioikeuksiin liittyvät omistajuudet sovitaan selkeästi. Lisensioimista suositetaan mahdollisuuksien mukaan. Suomen julkisen avoimen datan virallinen lisenssisuositus on CC 4.0 BY.
- Syntyviin tuloksiin niin datan, julkaisujen kuin menetelmienkin osalta sovelletaan kansallisia avoimen tieteen linjauksia, ellei asiasta ole tehty erillissopimusta (esim. kaupalliset toimijat).
- Tutkimusinfrastruktuurien oleellisten prosessien tulee olla helposti liitettävissä muihin kansallisiin tai kansainvälisiin prosesseihin avoimien rajapintojen kautta.
- Tutkimusinfrastruktuurit rakennetaan avoimia rajapintoja ja standardeja hyödyntäen.
- Vältetään yhden toimittajan ratkaisuja, aina kun se on mahdollista. Tutkimusinfrastruktuurin tuottamisessa ja kehittämisessä käytetään avointa lähdekoodia.
- Noudatetaan hyviä tiedonhallintatapoja.
- Noudatetaan yleisiä tutkimuseettisiä periaatteita ja hyvää tieteellistä käytäntöä.
- Palvelut ovat helposti löydettävissä ja hyvin kuvailtuja.
- Laadunvarmistusjärjestelmä on avoimesti kuvailtu.
- Käytettävyys- ja käyttötilastot ovat avoimesti saatavilla.

Suomalaisen tutkimusjärjestelmän toimijat käyttävät julkisesti rahoitettujen infrastruktuurien peruspalveluja maksutta. Lisäpalvelut voidaan hinnoitella räätälöidysti, esim. määrittellen kaupallisille toimijoille korkeamman tariffin infrastruktuurin ylläpitämisen ja kehittämisen

rahoittamiseksi. Oikein tehdyllä hinnoittelupolitiikalla parannetaan mahdollisuuksia ylläpitää ja kehittää tutkimusinfrastruktuureja. Tutkimusinfrastruktuurille on tärkeää laatia infrastruktuurin koko elinkaaren kattava pitkän aikavälin rahoitusstrategia, jossa ennakoidaan myös mahdollisia toiminnan luonteessa tapahtuvia muutoksia. Rahoitus pohja on sitä vakaampi, mitä enemmän erilaisia toimijoita on mukana. Mahdollisia konsortioiden osapuolia ovat mm. ministeriöt, korkeakoulut, tutkimuslaitokset, sairaalat, yritykset ja säätiöt. Hyvin suurissa hankkeissa tulee harkita Pohjoismaisia tai muita kansainvälisiä yhteisjärjestelyjä. Käytännön hinnoittelutyössä voidaan ottaa huomioon esimerkiksi markkinalähtöisiltä toimijoilta perittävä korkeampi käyttömaksu. Maksu voi mahdollistaa käyttökulujen kattamisen lisäksi myös investointien rahoittamisen. Kaupallistentoimijoiden mukanaolo sekä rahoittajina että käyttäjinä lisää infrastruktuurien käyttöastetta, mikä osaltaan kannustaa niiden ylläpitämiseen ja kehittämiseen.

Datapolitiikan tekeminen

Datapolitiikan kirjoittamisella on tärkeä olla organisaation ylimmän johdon tuki ja mandaatti. Läheinen yhteistyö on eduksi ja helpottaa prosessia. Muita osapuolia, joita on hyvä osallistaa ja sitouttaa politiikan kirjoittamiseen, ovat kirjasto/tietopalvelu, tiedonhallinta sekä tietotekniikka ja muut organisaation omat ja tutkijoille suunnatut oleelliset tukipalvelut, kuten juristit. Myös tutkijoita on tärkeä kuulla ja osallistaa prosessin aikana.

Politiikan sisällön on oltava riittävän laaja ja selkeä. Suuressa, monialaisessa organisaatiossa voi olla tarkoituksenmukaista aloittaa työ yleiseltä tasolta, kuten esimerkiksi Helsingin yliopisto on tehnyt. Askel kerrallaan edetessä vältetään siltä, että tärkeiden yleisten linjausten tekeminen viivästyy kohtuuttomasti. Erillisiä politiikoita tai täsmennyksiä kannattaa joskus tehdä organisaation eri yksiköille tai toimintoille. Yleistä datapolitiikkaa on syytä tukea erillisillä käytännön ohjeilla ja tiekartoilla. Kaikkia asioita ei tarvitse käsitellä kaikissa dokumenteissa tai kaikilla tasoilla, mutta dokumenttien on syytä muodostaa koherentti ja ristiriidaton kokonaisuus.

On hyvä aloittaa käymällä läpi lait, sopimukset ja muut datapolitiikkaan mahdollisesti vaikuttavat dokumentit ja ohjeistukset. Seuraavaksi kannattaa hahmotella sisällysluetteloa ja aikataulua. Sisällysluettelo voi näyttää vaikka tältä ([DCC:n esimerkki](#)):

I. Header-osio, logo yms

II. Otsikko

III. Tavoitteet

IV. Periaatteet

V. Kattavuus (mitä toimijoita, mitä tutkimusta (kenen rahoittamaa jne), mitä aineistoja politiikka koskee)

VI. Poliitiikka

VII. Roolit ja vastuut

a. Organisaatio

- b. Vastaavat tutkijat (PI)
- c. Muut tutkijat ja tutkimusavustajat
- d. Muut tutkimusta avustavat tai tukevat henkilöt

VIII. Poliitiikan hyväksyminen ja päivittäminen (kuvataan prosessit, aikataulut ja muutokset)

IX. Footer-osio (sivunumero, versionumero, tila jne)

X. Liitteet (määritelmät, viittaukset muihin oleellisiin dokumentteihin tai otteet niistä yms)

Poliitiikan ja muiden tarvittavien ohjeiden teko onkin laaja hanke, johon vaaditaan monenlaista osaamista ja asiantuntijuutta sekä ennen kaikkea yhteistyötä. Aluksi on kuitenkin hyvä käydä läpi alla olevat asiat ja määritellä millä tavalla ja millä tasolla on järkevä edetä. Mitkä alla olevista kysymyksistä ratkaistaan missäkin vaiheessa ja missä dokumenteissa? Kysymykset perustuvat osin [Iso-Britanniassa tuotettuihin tutkimusdatapolitiikka-ohjeisiin](#). Osa asioista on hyvinkin yksityiskohtaisia, eikä niihin tarvitse ottaa kantaa heti alkuvaiheessa, vaan ne saattavat vaatia selvittelyä, suunnittelua ja uusien palveluiden kehittämistä. Organisaatiossa on hyvä miettiä mikä on oman toiminnan kannalta järkevin ratkaisu.

1. Poliitiikan kattavuus: Ketä politiikkaa koskee? Minkä alojen dataa politiikka koskee? Millä kielillä toimitaan (tarvitaanko esimerkiksi dokumenteille ja metadatalle käännöksiä ja kieliversioita)? Minkä tyyppistä dataa politiikka koskee? Mitä muita aineistoja paitsi dataa on huomioitava (menetelmät, koodistot yms)? Mitä versioita datasta on olemassa ja mikä on versionhallinnan tarve? Mikä on kohderyhmä? Mikä on käyttäjähallinnon ja turvatasojen tarve? Huomioitavat lait, säädökset ja sopimukset? Mikä on organisaation vastuu tutkimusdatasta tutkimuksen aikana? Mitä palveluita organisaatio tarjoaa tutkijoille ja kenen toimesta?

2. Aineistojen metatietojen hallinta: Missä ja miten metatietoja luodaan ja säilytetään? Miten metatietoja hallinnoidaan? Mikä on metatietojen avoimuuden aste? Miten avoimuus toteutetaan teknisesti (sallitaanko haravointi, miten ja millainen)? Mitä metatietoja tarvitaan? Mitä standardeja käytetään? Miten metatiedot ja rajapinnat dokumentoidaan? Miten metatietojen laatu varmistetaan? Miten metatietoja voidaan tulevaisuudessa kehittää (rakenteet, sanastot, ontologiat, pitkäaikaissaatavuus)?

3. Omistajuus ja käytöhallinta. Aineistojen luovutus organisaatiolle: Ketkä saavat luovuttaa aineistoja organisaatiolle ja mitä aineistoja saa luovuttaa? Mikä on datan luovuttamisen prosessi? Onko olemassa rajoituksia koon tai formaattien suhteen? Mitä muita vaatimuksia aineistolle on olemassa? Kenen vastuulla on aineistojen luovuttajien tukeminen? Tehdäänkö normalisointia ja kenen toimesta? Mitä formaatteja on käytössä? Kuka vastaa mahdollisista konversioista ja riittävästä dokumentaatiosta ja kuvailusta? Millaista dokumentaatioita luovutuksen yhteydessä syntyy? Miten aineiston laatu varmistetaan? Miten tietoturva varmistetaan? Kuka vastaa riittävästä anonymisoinnista ja tekijänoikeuksista? Mitä organisaatio voi tehdä datalle ja missä vaiheessa ja miten oikeudet syntyvät (miten menetellään jos esimerkiksi data halutaan siirtää muualle arkistoitavaksi tai konvertoida?) Mitä oikeuksia aineistoihin liittyy ja millä periaatteilla käyttö lisensoidaan?

4. Aineistojen jakaminen: Millaista käytöhallintaa tarvitaan ja kuka vastaa siitä? Miten ja millaisilla periaatteilla aineistoja jaetaan?

5. Aineistojen säilyttäminen: Miten pitkäaikainen säilytys ja saatavuus turvataan? Kuka vastaa säilytysuunnitelmista ja priorisoinneista? Tähän on suunnitteilla [kansallinen ratkaisu](#), mutta kehitystä on hyvä seurata. Miten suhtaudutaan muualle kotimaassa ja ulkomailta arkistoituihin aineistoihin, miten niihin linkitetään?

6. Aineistojen poistaminen ja exit plan: Millä perusteilla ja miten aineistoja voidaan poistaa ja/tai tuhota? Miten menetellään esimerkiksi laittomien tai tekaistujen aineistojen paljastuessa? Mitä tapahtuu jos organisaatio lakkaa olemasta?

Suomalaisten tutkimusorganisaatioiden linjauksiin voit tutustua [Tutkimusorganisaatioiden linjaukset ja ohjeistukset](#) -sivulla.

[Australian National Data Services](#)

[DDC Policy resources](#)

[Akers et al: Building Support for Research Data Management: Biographies of Eight Research Universities. International Journal of Digital Curation.2014, Vol. 9, No. 2, pp. 171-191. doi:10.2218/ijdc.v9i2.327](#)

[Dillo & Doorn: The Front Office–Back Office Model: Supporting Research Data Management in the Netherlands. International Journal of Digital Curation. 2014, Vol. 9, No. 2, pp. 39-46. doi:10.2218/ijdc.v9i2.333](#)

Tiedonhallinnan palveluiden suunnittelusta

Avoin tiede ja tutkimus -hankkeessa on tehty tiettyjä [arkkitehtuurilinjauksia](#), joita organisaatioiden on hyvä huomoida, jotta palveluita voidaan tuottaa yhteentoimivasti ja tehokkaasti. Arkkitehtuuriperiaatteet ovat seuraavat:

- Mahdollistetaan tiedon maksimaalinen avoimuus ja sen jatkokäyttö primääriläheistä
- Huomioidaan tiedon koko elinkaari ja sen hallinta
- Helpotetaan tutkimustyötä ja tuetaan tieteen perusolemusta (itsensäkorjaavuus, toistettavuus)
- [Viitearkkitehtuurissa](#) luodaan määrittymät, joiden avulla tunnustetaan avoimen tieteen toteutumisen ehdot
- Tuetaan kerroksellista, modulaarista ja kustannustehokasta ratkaisumallia, jossa hyödynnetään olemassa olevia osia

Tietoturvasta huolehtiminen ja tutkijoiden tukeminen siinä sekä oikeudellisissa asioissa ovat tärkeitä ja runsasta vuorovaikutusta ja asiantuntemusta vaativia tehtäviä laadukkaiden palveluiden varmistamisessa.

Tiedonhallinnan palveluihin kuuluvat seuraavanlaiset asiat:

- Tuki aineistonhallinnan suunnitelman teossa. Tähän kuuluvat mm työkalujen ylläpito ja käyttötuki, tuki lakiasioissa sekä eettisissä menettelyissä ja kysymyksissä.

- Koko tutkimusprosessin aikana tarvittavan säilytystilan järjestäminen sekä aineistonhallinta ja tuki datankäsittelyssä siten, että se tapahtuu sääntöjen mukaisesti ja datan laatu on toivottua
- Tuki hyvän metadatan ja dokumentaation tuottamisessa ja datan oikeanlaisesta näkyvyydestä huolehtimien
- Hankkeen päättymisen jälkeen tutkijoille tukea aineistojen arkistointiin liittyvissä päätöksissä sekä arkistointi
- Lopullinen luettelointi ja löydettävyydestä huolehtiminen

(Mukaillen: Graham Pryor: "Who's doing data? A spectrum of roles, responsibilities and competencies" teoksessa *Delivering research data management services. Fundamentals of good practice*, 2014. s 44-45)

Kaikkia palveluita ei ole yleensä tarkoituksenmukaista järjestää jokaisessa organisaatiossa, mutta saadakseen tutkimusaineistot hyvin haltuunsa organisaation tulisi huolehtia siitä, että tutkijoilla todella on riittävästi tietoa mistä tukea saavat. Tärkeitä yleisiä ja tieteenalariippumattomia kansallisia palveluita ovat [Tuuli](#) ja [Etsin](#).

Brittiläinen tietoarkisto DDC on tehnyt organisaatioille itsearviointityökalun [CARDIO](#).

Avoimen tieteen käsikirjan sanasto

Tämä ei ole virallinen sanasto, vaan ohjeellinen luettelo tärkeistä käsitteistä selityksineen. Varsinaista sanastotyötä tehdään erikseen tietomallityön yhteydessä. Tässä oleva sanasto voi toimia lähinnä yleisenä keskustelunavauksena.

Aineiston avoimuus Aineisto on kenen tahansa käytettävissä mihin tahansa tarkoitukseen, myös kansainvälisesti, kunhan lähteeseen viitataan hyvän tieteellisen käytännön mukaisesti. Se ei välttämättä tarkoita maksuttomuutta; esimerkiksi aineiston tulkinnasta tai irrottamisesta voidaan periä palveluhinnaston mukainen korvaus mikäli tulkinta tai irrottaminen edellyttää työpanosta.

Aineistotyyppi (*Resource Type*) Aineistotyyppillä kuvaillaan millaista aineisto on, esimerkiksi onko kyseessä fyysinen objekti, dataa, kuvia, äänitiedostoja vai sovellus. Periaatteessa aineistotyyppi käsittää aineistot, jotka syntyvät yhden prosessin tuloksena ja joille luodaan omantyyppinen, kuvaileva metatieto. Aineistotyyppikäsitteen avulla voidaan tarkentaa periaatteita: aineistotyyppi käsittää aineistot, joita koskevat samat periaatteet esimerkiksi pitkäaikaissäilytyksen teknisessä suunnittelussa.

Aitous (*Authenticity*) Datan oikeellisuus ja väärentämättömyys, vertaa termiin *eheys*.

Anonymisointi Tunnistetietojen poistaminen tai muuntaminen siten että tunnistaminen ei enää onnistu eikä henkilöitä voi identifoida. Tunnistetietoja ovat esimerkiksi nimet, osoitteet, puhelinnumerot tai henkilötunnukset. Tunnistaminen voi tapahtua myös eri taustamuuttujia yhdistelemällä, joten on syytä olla huolellinen myös niiden suhteen. Taustamuuttujia ovat esimerkiksi ikä, sukupuoli, koulutus, ammattiasema, kieli tai kansallisuus. Anonymisointia voi tehdä monella eri tavalla. Ks. esimerkiksi [Yhteiskuntatieteellisen tietoarkiston Datanhallinnan käsikirja](#).

Altmetriikka Tutkimuksen yhteiskunnallisen vaikuttavuuden arviointia voidaan kehittää huomioimalla muutakin kuin viittausmääriä ja tieteensaisäistä toimintaa. Altmetriikassa mitataan julkaisemista suurelle yleisölle, bloggaamista tai muuta aktiviteettia sosiaalisessa mediassa. Altmetriikan avulla yritetään löytää vastauksia siihen, mitä tieteessä tapahtuu juuri nyt. Mistä keskustellaan? Mikä kiinnostaa toisia tutkijoita, mikä kansalaisia? Sähköinen ympäristö antaa myös mahdollisuuden tilastoida aineistojen latauskertoja, jakamisia tai suosituksia. Tämän tyyppisellä kommunikaatiolla voi olla vaikuttavuuden kannalta suurtakin merkitystä.

Asiakirjajulkisuus Julkisuuslain mukaan yleiset asiakirjat ovat julkisia. Yleisiä asiakirjoja ovat viranomaisen laatimat ja antamat sekä sille lähetetyt tai annetut asiakirjat, jotka ovat sen hallussa. Asiakirjan käsite laissa on varsin laaja. Kirjallisten dokumenttien ohella asiakirjoina pidetään myös teknisiä tallenteita ja kuvallisia esityksiä. Asiakirja, joka on valmisteltavana viranomaisella, ei ole vielä julkinen, ei myöskään viranomaisen omassa keskuudessa syntynyt ehdotus, luonnos, mietintö, lausunto, muistio tai muu selvitys. Salassa on pidettävä sellainen asia tai asiakirja, josta niin on lailla säädetty. Julkisuuden kannalta asiakirjat voidaan jakaa julkisiin asiakirjoihin, (ei-julkisiin) valmisteluasiakirjoihin ja salaisiin asiakirjoihin.

Auktorisointi Valtuutus. Todennetulle käyttäjälle annettu lupa tietyn tiedon, suojattavan kohteen tai muun palveluelementin käyttöön.

Auktoriteettipalvelut (*Authority services*) Auktoriteettitiedosto tukee henkilön tai yhteisön identifiointia erottamalla toisistaan samannimiset henkilöt mm. aikamääreen, ammatin jne. avulla. Toisaalta auktoriteettitiedosto kokoaa yhteen tekijänä tai kohteena olevan henkilön/yhteisön eri nimimuodot tai nimen variaatiot takaamaan saman tekijän teosten löytyvyyden nimenmuodosta riippumatta. Auktoriteettitiedosto organisoii tietokantaa myös linkittämällä toisiinsa saman henkilön eri julkiset identiteetit, kuten todellinen nimi ja salanimi. Samalla tavoin linkitetään toisiinsa henkilöiden ja yhteisöjen nimenmuutoksista syntyneet vanhat ja uudet nimet. Myös niiden keskinäisiä suhteita ja historiaa voidaan kuvata koneluettavassa muodossa. Kansalliskirjastot luovat nimiauktoriteettitiedostoja oman maansa julkaisutuotannosta. Kansallisten sääntöjen mukaiset hakumuodot tallennetaan auktoriteettitietueisiin yhdessä kohdetta koskevien ID- tunnisteiden (ISNI, ORCID) ja nimen varianttimuotojen kanssa. Tiedot yhdistetään kansainvälisissä tietokannoissa (esim. ISNI- ja ORCID –kannat, Virtual International Authority File VIAF), jotka ovat vapaasti selailtavissa verkossa. Nykyään julkishallinnon suosituksena on julkaista nämä tiedot myös avoimena linkitettyinä datana.

Autentikointi Todentaminen/tunnistautuminen. Käyttäjän henkilöllisyyden varmistaminen.

Autenttisuus Aitous, väärentämättömyys, luotettavuus.

Avoimet lisenssit (*Open License*) Tutkimusdata ja -julkaisut ovat yleensä tekijänoikeuksien suojaamia. Näiden aineistojen avoin käyttö voidaan kuitenkin mahdollistaa sopimuksin. Creative Commons -lisensseillä jaat osan tekijänoikeuksiasi ja annat haluamasi vapaudet teoksen käyttäjälle, katsojalle tai kokijalle. Eri ehtoja yhdistelemällä voit jakaa oikeuksiasi itsellesi ja tilanteeseen sopivalla tavalla. ATT-hanke suosittelee CC0:n, CC 4.0 BY:n tai tarvittaessa muiden yleisesti tunnettujen sopimuslisenssien käyttöä. <http://creativecommons.fi/lisenssit/>

Avoimet rajapinnat (*Open interface*) Avoimilla rajapinnoilla tarkoitetaan sellaisia ohjelmistoon toteutettuja vapaasti käytettäviä, huolellisesti dokumentoituja liittymäpintoja, joiden välityksellä ohjelmistot vaihtavat tietoja keskenään.

Avoimet standardit (*Open standard*) Avoin standardi tarkoittaa vakiointia, eli yhteistä sopimusta, jolloin standardia noudattavat pystyvät toteuttamaan saman, tai keskenään yhtenevän sovelluksen. Avoin standardi on kaikkien saatavilla, joten kuka tahansa saa halutessaan tiedon siitä ja voi noudattaa sitä.

Avoimet tutkimusympäristöt (*Open research environments, Open research infrastructures*) Tutkimusympäristö on tutkimusvälineiden, -laitteistojen, -aineistojen ja -palveluiden kokonaisuus, joka mahdollistaa tutkimuksen tekemisen. Tutkimusympäristöillä vahvistetaan tutkimusyhteistyötä ja tutkimuskapasiteettia. Tutkimusympäristöt voivat sijaita yhdessä paikassa tai olla hajautettuja kokonaisuuksia tai virtuaalisia infrastruktuureja. Avoin tutkimusympäristö tarjoaa pääsyn kokonaisuuteen, tutkimusprosessiin, joilla tutkimustuloksia tuotetaan. Avoin tutkimusympäristö edellyttää, että tutkimustulokset, julkaisut ja tausta- aineistot – ovat tiedeyhteisön vapaasti saatavissa.

Avoin data (*Open data*) Avoimella datalla tarkoitetaan tutkimukselle, julkishallinnolle, organisaatioille, yrityksille tai yksityishenkilöille kertynyttä jalostamatonta informaatiota, joka on avattu ulkopuolisillekin vapaasti ja maksutta hyödynnettäväksi.

Avoin julkaiseminen (*Open publishing*) Tutkimusjulkaisujen (artikkelien, raporttien, monografioiden) open access- julkaisemisella tarkoitetaan yksikertaisimmillaan julkaisun saattamista tietoverkkoon ja oikeutta lukea, kopioida, tulostaa ja linkittää kokonaisia tieteellisiä julkaisuja. Open access -julkaisemisella tarkoitetaan tieteellisen tiedon vapaata levitystä. Tieteellinen julkaisu on avoimesti saatavissa, kun se on sekä tiedeyhteisön että suuren yleisön luettavissa Internetistä ilmaiseksi ja esteettömästi.

Avoin lähdekoodi (*Open source*) Avoin lähdekoodi on tapa kehittää ja jakaa tietokoneohjelmistoja. Hyödyntäjä saa vapaasti käyttää, kopioida, muunnella ja jaella avoimen lähdekoodin ohjelmaa. Avoimessa ohjelmistojen kehitysmallissa sekä ideat että toteutukset ovat kaikkien nähtävissä ja hyödynnettävissä. Kehitystä ei hallinnoi vain yksi yritys, vaan mukana on maailmanlaajuinen yhteisö, joka koostuu sekä yksityishenkilöistä että yrityksistä. Kaikki voivat osallistua kehitystyöhön ja ohjelmistovirheet on mahdollista löytää ja korjata nopeasti. Tämä johtaa usein korkeaan laatuun, hyvään tietoturvaan ja yhteentoimiviin ohjelmistoihin.

Avoin tiede (*Open Science*) Avoin tiede tarkoittaa pyrkimyksiä edistää avoimia toimintamalleja tieteellisessä tutkimuksessa. Keskeinen tavoite on tutkimustulosten, tutkimusdatan ja tutkimuksessa käytettyjen menetelmien julkaiseminen siten, että ne ovat kaikkien halukkaiden tarkasteltavissa sekä käytettävissä. Avoin tiede sisältää käytäntöjä, kuten tutkimusjulkaisujen avoimen saatavuuden (open access) edistäminen, tutkimusaineistojen avoin julkaiseminen, avoimen lähdekoodin ja avoimien standardien hyödyntäminen, sekä tutkimusprosessin julkinen dokumentointi niin kutsutun avoimen muistikirjan menetelmällä.

Avoin tieto (*open knowledge*) Avoimella tiedolla viitataan digitaalisessa muodossa oleviin sisältöihin ja dataan, joita kuka tahansa voi vapaasti ja maksutta käyttää, muokata ja uudelleen jakaa. Avoimen tiedon määritelmän mukaisesti aineiston tulee olla kokonaisuudessaan saatavilla käyttökelpoisessa ja muokattavassa muodossa Internetin kautta ja sen tulee olla lisensoitu niin, että sen käyttöä, muokkausta ja uudelleenjakelua ei rajoiteta.

Bibliometriikka Bibliometriikka on tutkimusalue, joka tutkii tieteellistä julkaisutoimintaa kvantitatiivisin menetelmin. Aikaisemmin bibliometrinen tutkimus suuntautui enemmän tieteen rakenteen tutkimiseen, kuten esimerkiksi nousevien alojen tunnistamiseen, yhteistyöverkostojen

kuvaamiseen tai yksilötason julkaisutuottavuuden tekijöiden tarkastelemiseen. Bibliometrisia menetelmiä hyödynsivät enimmäkseen alalle erikoistuneet asiantuntijat. Viimeaikainen kehitys on kuitenkin luonut haasteita bibliometriikan tutkimusalalle, ja nykyisin yhä suurempi osa bibliometrisesta tutkimuksesta liittyy arviointiin soveltuvien indikaattoreiden kehittämiseen. Ks myös **altmetriikka**. Jos julkaisukäytäntöjen moninaisuuden vuoksi on vaikeaa antaa yleispätevää määritelmää tieteelliselle julkaisulle, vielä haastavampaa on kaikenkattavasti määritellä laadukas julkaisu. Julkaisun laadun ominaisuuksia voivat olla esimerkiksi sen vaikuttavuus ja yhteiskunnallinen relevanssi. Bibliometriikassa tieteellistä vaikuttavuutta mitataan usein julkaisun saamien viittausten määrällä. Myös vaihtoehtoisia julkaisun laatua kuvaavia indikaattoreita on kehitetty esimerkiksi luokittelemalla lehtiä ja kustantajia niiden laadun perusteella. Viittausindikaattoreihin liittyy useita tunnettuja ongelmia, kuten negatiiviset viittaukset tai että ne suosivat valtavirtatutkimusta. Siitä huolimatta osa luonnon- ja lääketieteen tutkijoista kannattaa viittausindikaattoreiden käyttöä tutkimuksen arvioinnissa. Näillä aloilla on korkea yksimielisyys siitä, mikä on hyvää tutkimusta. Yhteiskunta- ja humanistisissa tieteissä näkemykset tutkimuksen laadun kriteereistä vaihtelevat enemmän

Data Digitaalinen aineisto, johon ei liity tulkintaa (vrt. käsite *tieto*, jolla viitataan tulkittuun informaatioon)

Datapolitiikka Organisaation aineistohallinnan ohjeistus.

Digitaalinen informaatio Digitaaliseen muotoon tallennettu tai muunnettu (esim. skannattu dokumentti) aineisto sekä siihen liittyvä metatieto.

Digitointi Analogisessa muodossa olevan aineiston (esim. paperidokumentin) muuntaminen digitaaliseen muotoon.

DOAJ Directory of Open Access Journals Kansainvälinen rekisteri, josta löytyy tietoja avoimista tieteellisistä lehdistä. <http://doaj.org/>

E-infrastrukturi Termi e-infrastrukturi viittaa tutkimusta tukevaan sähköisten palveluiden ympäristöön, jossa kaikki tutkijat — työskennellessään sekä koti-instituutioissaan että kansallisissa tai kansainvälisissä tieteellisissä hankkeissa — saavat pääsyn palveluihin ja resursseihin. E-infrastrukturi liittyy yhteen tutkimuslaitteistot, tietoaineistot, laskentaresurssit, ohjelmistot, tietoverkot sekä palvelut.

Eheys (Integrity) Eheyden tarkistamisella tarkoitetaan sitä, että säilytettävän datan (tiedoston, bittijonon) sisällöstä laskettua tarkistussummaa verrataan samasta datasta aikaisemmin laskettuun tarkistussummaan. Vertailun perusteella voidaan luotettavasti havaita mahdolliset muutokset datassa. Vertaa aitous ja tarkistussumma

(Tietoaineiston) Elinkaari Tietoaineiston olemassaolo kokonaisuutena alusta loppuun. Tietoaineiston vaiheet raakadatan tuottamisesta tuotteeseen ja loppukäsittelyyn.

Emulaatio Vanhentuneen käyttöjärjestelmä- ja sovellusohjelmistoympäristön käytön mahdollistaminen uudistuneessa laiteympäristössä, siten että alkuperäinen käyttökokemus säilyy. Emulaatio voidaan toteuttaa joko ohjelmistotasolla (esimerkiksi DOS-käyttöjärjestelmän emulointi) tai laitteistotasolla (esimerkiksi Intel 80386- prosessorin emulointi).

ESFRI (*European Strategy Forum on Research Infrastructures*) EU-jäsenmaiden ja assosiaatiomaiden yhteistyöelin tutkimusinfrastruktuureja koskevan politiikan valmistelussa ja määrittelyssä. ESFRI koostuu tutkimusministeriöiden ja rahoittajien edustajista.

Embargo (*Holding period*) Karenssiaika. Rajoitus, jonka mukaan aineiston vapaa käyttö alkaa vasta säädetyin odotusajan jälkeen esimerkiksi julkaisun ilmestyttyä.

Ensikäyttö/primäärikäyttö (*Primary use*) Ensikäytöllä tarkoitetaan alkuperäisessä tutkimussuunnitelmassa ilmaistua tai muuta ensisijaista tutkimusaineiston käyttötarkoitusta.

Esitystieto (*Representation Information*) Osa säilytettävään informaatioon liitettävää metatietoa. Esitystiedon muodostaa aineiston luettavuuden mahdollistava tieto, tiedostomuodon kuvaus, semantiikka ja muu tieto, jolla tiedoston bittijonot voidaan muokata ihmisen tai koneen luettavaan ja ymmärrettävään muotoon.

Etäkäyttöjärjestelmä Etäkäyttöjärjestelmän avulla tutkijat pääsevät käsiksi aineistoihin (yleensä internetin kautta) ilman aineistojen siirtämistä tutkijan omalle koneelle. Tutkija voi tarkastella, yhdistellä ja analysoida aineistoja etäkäyttötyöpöydän avulla. Alkuperäisaineistot säilyvät palomuurien takana turvaten aineistojen korkean tietosuojavaatimuksen.

Federoitu järjestelmä Laaja järjestelmä koostuu useista eri osajärjestelmistä. Federointi voidaan toteuttaa esim. tietokanta- tai sovellustasolla, mutta se voi olla myös käyttäjän tunnistamista ja henkilötietojen ylläpitoa koskeva järjestelmä.

Haka-tunnistautuminen Haka on Suomen käytetyin korkeakoulujen ja tutkimuslaitosten käyttäjätunnistusjärjestelmä. Haka perustuu luottamusverkostoon, jonka jäsenet - tutkijat, opiskelijat ja muu henkilöstö - voivat käyttää kotiorganisaationsa käyttäjätunnuksia kirjautuessaan moniin eri palveluihin. Kotiorganisaation tietohallinto vastaa käyttäjiensä käyttäjätiedoista ja henkilöllisyyden todentamisesta. Hakassa olevien palvelujen käyttäjätiedot saadaan suoraan käyttäjän kotiorganisaatiosta. Haka-tunnistautumista käytetään myös esimerkiksi Etsin-palvelussa.

Hallinnollinen metatieto Metatieto, jonka avulla digitaalinen aineisto on hallittavissa ja säilytettävissä. Sen osa-alueet ovat tekninen, käyttöoikeuksiin liittyvä sekä pitkäaikaissäilytyksen metatieto. Perinteinen ns. kuvaileva metatieto sisältää hallinnollisen metatiedon elementtejä, mutta ei yhtä kattavasti kuin varsinaiset hallinnollisen metatiedon formaatit kuten still-kuvien teknisen metatiedon MIX.

Haravointi (*Harvesting*) Metatiedon automaattinen poiminta sovelluksesta toiseen, esimerkiksi jonkun tietoarhivon luettelosta Etsin- palveluun. Haravoinnissa sovelletaan yleensä jotakin rajapintastandardia kuten OAI-PMH protokollaa.

Harmaat julkaisut (*Grey publications*) Tieteelliset julkaisut, joita ei varsinaisesti ole vertaisarvioitu ja jotka voivat olla enemmän tai vähemmän valmiita. Tällaisia ovat esimerkiksi artikkeleiden tarkastamattomat preprint-versiot tai erilaiset työ- ja keskustelupaperit. Niiden julkaiseminen edistää tieteen avoimuutta ja nopean palautteen saamista tiedeyhteisöltä. Julkaisu voi tapahtua julkaisuarkistossa.

Harmonisointi Yhteensopivaksi tekeminen.

Henkilötieto Kaikenlaiset luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavat merkinnät, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi. Henkilötietojen käsittelyllä tarkoitetaan henkilötietojen keräämistä, tallettamista, järjestämistä, käyttöä, siirtämistä, luovuttamista, säilyttämistä, muuttamista, yhdistämistä, suojaamista, poistamista, tuhoamista sekä muita henkilötietoihin kohdistuvia toimenpiteitä. Henkilörekisterillä tarkoitetaan käyttötarkoituksensa vuoksi yhteenkuuluvista merkinnöistä muodostuvaa henkilötietoja sisältävää tietojoukkoa, jota käsitellään osin tai kokonaan automaattisen tietojenkäsittelyn avulla taikka joka on järjestetty kortistiksi, luetteloksi tai muulla näihin verrattavalla tavalla siten, että tiettyä henkilöä koskevat tiedot voidaan löytää helposti ja kohtuuttomitta kustannuksitta. (Henkilötietolaki 22.4.1999/523)

Hosting-palvelu Käyttöympäristön tarjoaminen palveluna, esimerkiksi web-palvelimen toiminnan ja ylläpidon tuottaminen ulkopuolisille.

Hybridi-OA Tilausmaksullisen ja maksuttoman käytön yhdistelmä. Kirjoittaja voi maksamalla kustantajan määrittelemän kirjoittajamaksun avata artikkelinsa avoimesti saataville. Muilta osin lehti on käytettävissä vain tilausmaksun maksaneille lukijoille. Kustantaja kerää tuloja tällöin sekä kirjoittajilta että lukijoilta.

Indikaattori Osoitin, ilmainen tai tilastollinen luku, jonka avulla voidaan moninaista tietoa tiivistää helpommin hallittavaan ja ymmärrettävään muotoon. Indikaattori on siis laadullinen mittari, joka parhaimmillaan antaa laajoja ja monimutkaisiakin tietoja yksinkertaisemmassa muodossa. Indikaattori voi toimia myös apuvälineenä niin tavoitteiden asettamisessa ja seurannassa kuin suunnittelussa ja päätöksenteossakin.

INSPIRE-direktiivi INSPIRE-direktiivi 2007/2/EY luo yleiset puitteet tiettyjen viranomaisten hallinnassa olevien paikkatietoaineistojen saatavuudelle ja käytölle Euroopassa. Direktiivin toimeenpano perustuu ns. kansallisten paikkatietoinfrastruktuurien yhteentoimivuuden vaiheittaiseen kehittämiseen. Direktiivin toimeenpanoa varten on Suomessa säädetty laki paikkatietoinfrastruktuurista ja lakia täydentävä asetus.

Integraatioympäristö/arkkitehtuuri Integraatioarkkitehtuuri kuvaa miten eri sovellusten joukko jakaa ja välittää tietoa toisilleen. Se kuvaa ne periaatteet, joilla sovelluksen liittymät muihin järjestelmiin toteutetaan. Integraatioarkkitehtuuria määriteltäessä otetaan kantaa miten kokonaisuus integroidaan yhtenäisesti, ei vain yksittäisen liittymän integrointiin tapauskohtaisesti.

Irrottamiskustannus Käytetään myös muotoa irrotuskustannukset. Tässä raportissa kulut, jotka aiheutuvat aineistojen muokkauksesta tai kopioinnista luovutusta varten, aineiston toimittamisesta tai muista aineistojen luovuttamiseen liittyvistä toimista.

Jatkokäyttö/sekundäärikäyttö (*Secondary Research, re-use*) Jatkokäytöllä viitataan käyttöön, jossa joko ensi- tai jatkokäyttäjät hyödyntävät tutkimusaineistoa muuhun kuin alkuperäiseen esim. tutkimussuunnitelman sisältämään tutkimustarkoitukseen.

Jakaminen

Joukkoistaminen (Crowdsourcing) ks. Talkoistaminen

Julkaiseminen

Julkaisuarkisto Tieteellisten julkaisujen tallentamiseen ja avoimeen verkkojulkaisemiseen soveltuva tekninen järjestelmä ja sen ympärille rakennetut palvelut. Teknisen kehitystyön ja ylläpitotyön lisäksi julkaisuarkiston olemassaolo edellyttää mm. aineistojen tallentamiseen, kokoelmien hallinnointiin ja asiakaspalveluun liittyvien prosessien ja palvelujen järjestämistä. Julkaisuarkistot voivat olla joko organisaatio- tai tieteenalakohtaisia.

Julkaisufoorumi Julkaisufoorumi ylläpitää ja kehittää tieteellisten julkaisukanavien tasoluokitusta. Tavoitteena on ollut luoda järjestelmä, jossa tieteellistä julkaisutoimintaa voidaan arvioida määrän lisäksi myös laadullisesti. Järjestelmän perustana on kaikki tieteenalat kattava tasoluokitus erityyppisille julkaisukanaville: tieteellisille sarjoille, konferensseille ja kirjakustantajille. Julkaisukanavat arvioidaan asiantuntijapaneelissa. Julkaisufoorumi toimii Tieteellisten seurain valtuuskunnan yhteydessä. Julkaisufoorumi perustettiin Suomen yliopistot UNIFI ry:n aloitteesta vuonna 2010. Ensimmäinen tasoluokitus valmistui vuonna 2011. Tutkimusorganisaatiot ovat autonomisia soveltaessaan erilaisia arviointimenetelmiä oman toimintansa kehittämisessä, mutta tulisi huomioida, että Julkaisufoorumi-luokitus soveltuu tutkimuksen arvioinnin työvälineeksi ainoastaan suurten julkaisumäärien tasolla. Luokitus ei sovellu tieteenalojen välisten vertailujen tekemiseen, eikä se voi korvata julkaisujen sisältöön perustuvaa vertaisarviointia yksittäisten tutkijoiden tai tutkimusryhmien arvioinnissa.

Julkinen sektori Kansantalouden osa, joka on valtion tai kuntien omistuksessa. Julkiseen sektoriin lasketaan valtion ja kuntien varsinaisen toiminnan lisäksi laajimmillaan muun muassa julkiset sosiaaliturvarahastot (esimerkiksi Kansaneläkelaitos), kuntayhtymät, valtion tutkimuslaitokset, liikelaitokset ja valtionyhtiöt.

Julkinen sektorin tiedot Käytetään myös termiä PSI-tieto (Public sector information). Julkisorganisaatioiden keräämää ja tuottamaa tietoa (esim. digitaaliset kartat, säätiedot, liikennetiedot, liike- ja kansantaloudelliset tiedot, lakitiedot).

Julkinen tieto Julkista tietoa kuvaa kaksi piirrettä: ensinnäkin, julkisen tiedon muodostaa se, mikä jää tekijänoikeuslainsäädännön tai muun lainsäädännön ulkopuolelle. Toisaalta julkinen tieto voi tarkoittaa julkisen hallinnon tuottamaa tietäntyyppistä tietoa, jota tuotetaan viranomaistoimintana. Julkishallinnon tuottama tieto on osa laajempaa kategoriaa julkisen sektorin tiedot.

Julkishyödyke Taloustieteellinen käsite hyödykkeestä, jota lukuisat henkilöt voivat käyttää useita kertoja sen kulumatta tai jonka kulutus ei estä muita kuluttamasta samaa hyödykettä.

Julkisrahoitteinen tutkimus Tutkimus, jonka rahoitus pääosin tulee verovaroista suoraan tai epäsuorasti (esimerkiksi Helsingin yliopiston Luomukselle myöntämä budjettirahoitus, ministeriöiden, Tekesin, Suomen Akatemian ja EUn hankemäärärahat)

Kansalaistiede (*Citizen Science*) Laajan yleisön tekemää joukkoistettua tutkimusta, jota on tehty esimerkiksi tähtitieteen tai biologian alalla tiedon keräämisessä. "Kuka tahansa voi koska tahansa havaita jotakin tieteellisesti kiinnostavaa. Esimerkiksi Taivaanvahtiin kirjatut tulipallohavainnot voivat auttaa asiantuntijoita laskemaan tulipallon etenemisreitit ja mahdollisen putoamispaikan, vaikka havaitsijat olisivatkin enimmäkseen satunnaisia kulkijoita. Tietysti tietotaidon kasvaessa myös harrastajan mahdollisuudet lisääntyvät. Tiede on kyllä osannut jo pitkään hyödyntää puoliammattilaisten asianharrastajien panosta monilla aloilla, mutta kansalaistiede tuo tieteen tekemisen myös aidosti tavallisten, vailla kummempia erikoistaitoja olevien ihmisten ulottuville. Ympäristöetikon sydäntäni lämmittää tietysti kovasti se että kansalaistieteelliseen tutkimukseen soveltuvat hyvin monet sellaiset alat, joiden tuloksista on hyötyä ympäristönsuojelun, ja sitä kautta

myös kansalaisten oman hyvinvoinnin kannalta. Hyödyllisiä eivät ole ainoastaan kansalaisvoimin kerätyistä aineistoista saatavat tieteen tuotteet, vaan ennen kaikkea se, että kansalaistieteeseen osallistuminen nostaa tavallisten ihmisten tietoisuutta ympäristönsä tapahtumista ja auttaa hahmottamaan arkisten pikku havaintojen suhdetta suureen kuvaan ja ympäristössä tapahtuviin muutoksiin." Saara Reiman: "Eläköön kansalaistiede", blogissa [Filosofian puutarhassa](#) 17.2.2103.

Keskustelupaperi (*Discussion paper*) Keskustelupaperin tavoite on tuoda tutkimuksen alustavat tulokset tutkimusyhteisön arvioitavaksi. Tulokset julkaistaan yleensä muussa julkaisukanavassa myöhemmin.

Koneluettavuus Koneluettavuus tarkoittaa, että tieto on rakenteistettu siten, että kone pystyy käsittelemään tietoja. Pilkku- tai sarkaineroteltu data sopii parhaiten taulukkomuotoisen tiedon tai nimi-arvoparien julkaisemiseen. Etuna on, että lähes jokaiselta tietokoneelta löytyy taulukkolaskentaohjelma, jolla tällaiset tiedostot voidaan avata tutustuttaviksi ja muokattaviksi. Monimutkaisempien tietorakenteiden esittämiseen esimerkiksi XML, JSON ja RDF ovat yleisen tason standardeja. Parhaimmillaan koneluettava tieto on yksilöity pysyvillä tunnisteilla, kuten ureilla, joita voi linkittää muihin lähteisiin esimerkiksi ontologioiden avulla. Esimerkiksi verkkosivuilla merkintäkielenä käytettävä HTML soveltuu hyvin asiakirjojen rakenteen kuvailuun (mikä on otsikko ja mikä leipätekstiä), mutta itse tietosisältöä se ei tarjoa koneluettavassa muodossa. Esimerkiksi jos kunnan verkkosivuilla luetellaan kaikkien kirjastojen osoitteet ja aukioloajat on ihmisen helppo löytää haluamansa tieto sivuilta, mutta koneellisesti on vaikea erottaa osoitteita, aukioloaikoja ja muuta sivuilta löytyvää informaatiota toisistaan. Koneluettavuuden näkökulmasta on pystyttävä yhdistämään eri aineistojen metatiedot, jotta koneellisesti voidaan todeta, kirjastojen «osoitteiden» olevan verrannollisia «katuosoitteeseen» uimahalleista puhuttaessa. Aina tietueiden yhdistäminen (harmonisoiminen) ei kuitenkaan ole näin suoraviivaista. Hyvän esimerkin tarjoavat työttömyysluvut, jotka Tilastokeskus ja työ- ja elinkeinoministeriö laskevat eri tavalla. Näitä lukuja ei voida helposti yhdistää.

Kokonaisarkkitehtuuri (*Enterprise Architecture*) Kuvaa kuinka eri elementit – organisaatioyksiköt, ihmiset, toimintaprosessit ja tietojärjestelmät – liittyvät toisiinsa ja toimivat kokonaisuutena.

(Tietoaineistojen) Kuratointi Käytetään myös termiä hoivaaminen. Aineistojen sisällön ylläpitoon ja päivityksiin liittyvät toimenpiteet, joilla turvataan aineistojen eheys ja käytettävyys.

Kultainen OA (*Gold OA*) Koko julkaisu (lehti tai kirja) on maksuttomasti käytettävissä kustantajan tai muun välittäjän verkkopalvelussa. Julkaisun avoin saatavuus perustuu usein kustantajan keräämiin kirjoittajamaksuihin.

Kuvaileva metatieto, kuvailutieto (*Descriptive Metadata*) Kohdetta kuvaileva metatieto, joka on liitetty luovuttajan toimesta kohteeseen siirrettäessä aineisto säilytettäväksi. Kuvaileva metatieto mahdollistaa aineiston löytämisen, tunnistamisen, valinnan sekä käyttöön saamisen ja ymmärrettävyyden (KDK).

Kuvailu Kuvailu on kuvailevan metatiedon tuottamista. Kuvaileva metadata kertoo dokumentista tai aineistosta ja sen sisällöstä (esimerkiksi tekijä, nimeke ja aihe). Ilman kuvailevaa metatietoa data ei kelpaa tutkimusaineistoksi, eikä se ole löydettävissä tai hyödynnettävissä. Kuvailu tulisi tehdä mahdollisimman rikkaasti mutta myös noudattaen sääntöjä (standardeja, kuvailuohjeita ja metatietomalleja), jotta löydettävyys olisi hyvä. Kannattaa aina käyttää sanastoja ja ontologioita ja noudattaa ohjeistusta tarkasti esimerkiksi nimimuotojen tai päivämäärien ilmaisussa.

Käsikirja Käsikirja perustuu opetus- ja kulttuuriministeriön asettaman poikkihallinnollisen Avoin tiede ja tutkimus -hankkeen (ATT) työhön tiedon saatavuuden ja avoimen tieteen edistämiseksi. Käsikirja on tarkoitettu tutkijoille, tutkimusryhmille, tutkimusorganisaatioille, päättäjille ja rahoittajille sekä kansalaisille avuksi avoimen tieteen ja tutkimuksen hyödyntämisessä ja edistämisessä.

Käytettävyys (*Usability*) Tietojärjestelmien käytettävyys määritellään seuraavasti: Valtiohallinnon tietoturvasuuskäsitteistön (1/2000) mukaan käytettävyys tarkoittaa 1) ominaisuutta, että tieto, tietojärjestelmä tai palvelu on siihen oikeutetuille saatavilla ja hyödynnettävissä haluttuna aikana ja vaaditulla tavalla ja 2) helppokäyttöisyyttä. Jotta tieto olisi käytettävää, sen on oltava myös ymmärrettävää. Tässä dokumentissa kyse on usein *tietoaineistojen* käytettävyydestä tai käyttökelpoisuudesta tutkimuskäyttöön. Tämä määrittäytyy aina suhteessa tutkimustehtävään, mutta edellyttää vähintään tietoja aineiston alkuperästä.

Käyttöoikeus ja käyttörajoitus Käyttörajoitus on ehto, joka liittyy aineiston käyttöön ja joka perustuu sopimukseen, lakiin tai etiikkaan, ja joka toteutuessaan voi estää pääsyn aineistoon tai määrittelee miten aineistoa saa käyttää. Joskus aineistojen saatavuutta tai käyttöä on rajoitettava esimerkiksi tietosuojasyistä. Jonkun on silloin vastattava siitä, kenelle erilaisia oikeuksia aineistoihin annetaan (lukeminen, muokkaaminen, jakaminen). Saatavuudesta vastaa yleensä tietojen ja järjestelmien omistaja, joskus tutkija. Käyttöoikeus on puolestaan valtuus käyttää jotakin aineistoa. Käytöllä voidaan tarkoittaa esimerkiksi aineiston näyttämistä, kopioimista tai muokkaamista. Käytönhallinta on tekninen menettely, jolla pyritään rajaamaan pääsy tietojärjestelmään tai tietoihin. Käytönhallinnassa voidaan hyödyntää esimerkiksi tunnistamista (autentikointia) ja valtuuksia (auktorisointia) siten, että tietojärjestelmään pyrkivältä käyttäjältä vaaditaan käyttäjätunnusta ja salasanaa. Suomen tutkimusjärjestelmässä yleinen käytetty menetelmä on Haka-järjestelmä. Pääsynhallintaan liittyy myös fyysinen pääsy laitteisiin. Tietoturvan ja aineiston laadun kannalta on tärkeää, että järjestelmä kerää käytöstä mahdollisimman tarkkoja lokitietoja.

Laatujärjestelmä (*Quality system*) Laatujärjestelmä koostuu koulutuksen, tutkimuksen, yhteiskunnallisen vuorovaikutuksen ja vaikuttavuuden, henkilöstövoimavarojen, palvelujen ja johtamisen laatua varmistavista menettelyistä ja prosesseista.

Linkitetty data (*Linked data*) Eri lähteissä sijaitseva rakenteinen data, jossa on lähteen sisäisiä ja lähteiden välisiä linkkejä. Linkitetty data perustuu standardoituihin verkkotekniikoihin kuten HTTP:hen, RDF:ään tai URI:in. Linkitettyä dataa on hyödynnetty esimerkiksi Kulttuurisampo-portaalissa (<http://www.kulttuurisampo.fi/>), joka sisältää tietoja usean suomalaisen museon, kirjaston, arkiston ja muun organisaation kokoelmista sekä erityyppisiä verkkoaineistoja. Portaalin kautta voi etsiä tietoja kaikkien osallistujaorganisaatioiden aineistoista. Linkittäminen on osa ns. semanttista webiä, ja sitä käytetään laajasti, esim. Google käyttää samankaltaista tekniikkaa nimellä Knowledge Graph. Tunnisteita sisältävien linkkien käyttö parantaa löydettävyyttä ja näkyvyyttä. Linkitetty data voi olla avointa dataa, jolloin puhutaan avoimesta linkitetystä datasta (open linked data).

Lokitieto (*log data*) Tietojärjestelmän muistiin automaattisesti kirjautuva tapahtumatieto. Lokitieto voi sisältää erilaisia tunnistautumistietoja. Lokitieto voi koskea muun muassa sitä, kuka järjestelmää on käyttänyt tai miten ja milloin järjestelmää on käytetty. Lokitiedoista voivat selvittää esimerkiksi järjestelmän virhetilanteet, yhteydenotot tietokoneelta Internetiin sekä tietokoneelle Internetistä tulleet yhteydenottoopyynnöt.

Löydettävyys (*Findability*) Löydettävyydellä tarkoitetaan usein sitä, miten hyvin tietty aineisto löytyy hakukoneen kautta. Usein halutaan edistää oman sisällön tai aineiston näkyvyyttä käyttämällä hakukoneoptimointia (*Search Engine Optimization*). Tätä voi tehdä esimerkiksi huolehtimalla hyvästä linkittämisestä ja oikeiden avainsanojen käyttämisestä osoitteissa, otsikoissa tai kuvateksteissä. Löydettävyuden voi kuitenkin myös käsittää laajemmin. Esimerkiksi, ellei julkaisulle anneta oleellisia kuvailutietoja kuten sopivia asiasanoja vaikkapa julkaisuarkistossa, sitä ei hakija löydä kun sitä tarvitsee. Löydettävyys eroaa löytyvyydestä, jolla mitataan hakujärjestelmän kykyä löytää relevantteja hakuosumia. Löydettävyys kuvaa yksittäisen aineiston ominaisuuksia, joihin sen tuottaja usein voi itse vaikuttaa huolehtimalla hyvien kuvailutietojen syntymisestä. Yhä useammin kuvailutiedot leviävät avointen rajapintojen ja haravointien kautta myös muihin palveluihin ja hakukoneisiin, ja voivat näin edistää tutkimustiedon leviämistä.

Metatieto/metadata 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 kirjallisuudessa käyttää myös termiä metadata. Metatiedot sisältävät sekä aineiston kuvailutietoja että teknisiä, järjestelmän metatietoja.

Meriitti / Ansio Tieteellinen tai virka-ansio.

Mikrodata Mikrodatalla tarkoitetaan aineistoja, jotka sisältävät tietoa yksittäisistä henkilöistä tai talousyksiköistä. Mikrodata on kerätty yleensä kyselytutkimuksin väestöotokseen poimituilta henkilöiltä tai kokonaisväestöltä väestönlaskennoissa. Tietoja voidaan poimia myös rekistereistä.

Migraatio (säilyttävä kopiointi) (*Migration*) Digitaalisen informaation muuntaminen tai kopioiminen siten, että alkuperäisen formaatti tai säilytysmedia muuttuu. Esimerkiksi tiedosto muunnetaan eri ohjelmalla luettavaan muotoon alkuperäisen tiedostomuodon vanhentuuessa; tai tiedoston nimi muuttuu, jolloin viittaukset sisältötietoon täytyy päivittää uutta tiedostonimeä vastaaviksi. Muunnoksella pyritään säilyttämään digitaalisen aineiston luettavuus ja ymmärrettävyys.

My Data Omadata tai My Data on alun perin Britanniasta lähtöisin oleva termi. Sillä viitataan käytäntöihin, joilla ihmisille annetaan oikeus ja pääsy heistä kerättyyn dataan kuten ostotietoihin, liikennetietoihin, teletietoihin, energiatietoihin, eri verkkopalveluihin kertyvään dataan jne. Keskeistä My Data -ajattelussa on, että data on teknisesti helposti käytettävissä ja yksilö voi hallita, kuinka sitä hyödynnetään ja jaetaan edelleen (okf.fi)

Nanojulkaisu (*Nanopublication*) Julkaisutapa, jossa yksittäiseen tietoon, väittämään tai oletukseen (*assertion*) liittyy metadataa sen luojasta ja alkuperästä. Tämän avulla tietovarantoja on helpompi yhdistellä jalostaa ja yksittäiseen tietoon voi myös viitata. Käytössä lähinnä lääketieteen ja biologian alalla tällä hetkellä. Perustuu datan linkittämiseen.

Normalisointi 1. Metatietoelementin (esimerkiksi aika) esitystavan muuntaminen yhtenäiseen (standardi)muotoon. 2. Metatietoformaattien tietoelementtien semanttisten vastaavuuksien määrittely.

Omistajuus Ontologia Formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä.

OpenDOAR Kansainvälinen rekisteri, josta löytää tietoa ja tilastoja julkaisurekistereistä. Ks. <http://www.opendoar.org/>

Paikkatieto (*geographic information, spatial data*) Paikkatiedolla tarkoitetaan kaikkea tietoa, joka sisältää välittömän tai välillisen viittauksen tiettyyn paikkaan tai maantieteelliseen alueeseen. Paikkatietoaineistot kuvaavat tiettyä teemaa tai ilmiötä kattaen rajatun maantieteellisen alueen. Paikkatietoteemoja ovat mm. maa- ja kallioperä, vesistöt ja ilmasto, kasvillisuus ja eläimistö, maankäyttö, kiinteistöt, rakennukset ja väestö sekä toimipaikat, liikenne- ja tietoliikenneverkot sekä johtoverkot. Digitaaliset ilma- ja satelliittikuvat ovat keskeistä paikkatietoaineistoa.

Paikkatietoinfrastrukturi Paikkatietoinfrastruktuurilla, osana yhteiskunnan tietoinfrastruktuuria, tarkoitetaan yhteisesti sovitulla tavalla tuotettuja paikkatietoaineistoja ja paikkatietopalveluja, näiden kuvailuja ja teknisiä toteutuksia sekä tietojen saatavuutta ja käyttöä koskevia periaatteita ja prosesseja.

Perusrekisterit Viranomaisen ylläpitämä koko maan kattava hallinnollinen rekisteri. Perusrekistereihin kuuluvat: Väestötietojärjestelmä (henkilötiedot, rakennus- ja huoneistotiedot), Kiinteistötietojärjestelmä (kiinteistörekisteri, lainhuuto- ja kiinnitysrekisteri), Yritys- ja yhteisötiedot (kaupparekisteri, yritys- ja yhteisötietojärjestelmän tiedot, yhdistysrekisteri sekä yritys- ja toimipaikkarekisteri).

Perustietovarannot Perustietovarannot ovat julkisen hallinnon keskeisiä tietovarastoja, joita hyödynnetään laajalti yli viranomaisrajojen. Perusrekisterit ovat yksi keskeinen tietovaranto. Lisäksi perustietovarantoihin kuuluvat myös perusrekisterien kaltaiset rekisterit, tiettyjä käyttötarkoituksia palvelevat rekisterit sekä merkittävät paikkatietoaineistot.

Pitkäaikaissäilytys (Digital preservation) Sähköisessä muodossa olevan aineiston rakenteen, eheyden ja sisällön pitkäaikaista huolenpitoa. Säilytysaika on vähintään 10 vuotta ja enimmillään rajaton eli aineisto säilytetään pysyvästi, mikä vaatii huomattavaan paljon suunnittelua ja metatietoja.

Primääriaineisto Tai primäärikappale on aineiston päivitettävä versio. Tällä versiolla on metatiedot ja omistaja.

Preprint-versio (*Preprint*) Preprintin tavoite on tuoda tutkimuksen alustavat tulokset tutkimusyhteisön arvioitavaksi. Tulokset julkaistaan yleensä muussa julkaisukanavassa myöhemmin. Ei vertaisarvioitu ennen julkaisua.

PSI-direktiivi Euroopan Unionin direktiivit EU PSI 2003/98/EC ja 2013/37/EU, jotka ohjaavat julkisen datan avaamiseen.

Pysyvä linkki Toiminnalliseen tunnuksen perustuva URI (verkko-osoite), joka viittaa pysyvästi samaan objektiin riippumatta kyseisen objektin web-osoitteen mahdollisista muutoksista. Pysyvä linkki perustuu yleensä URN-tunnukseen, jonka resoluutiopalvelu muuntaa ajantasaiseksi web-osoitteeksi tai -osoitteiksi. Ks myös *Pysyvä tunniste*.

Pysyvä tunniste, pysyväistunniste (PID, *Persistent identifier*) Tunniste on kohteen yksilöivä ainutkertainen merkkijono, jonka avulla kohteeseen voidaan yksiselitteisesti viitata. Pysyvä tunniste on koko kohteen elinajaksi tallennettu tunniste. Digitaalisille materiaaleille on useita eri tunnistejärjestelmiä, esim. URN tai DOI-tunnisteet. Resolver-palvelun avulla tunnisteet voivat toimia toiminnallisina tunnuksina eli pysyvinä linkkeinä verkossa.

Pääsynhallinta Mekanismit, joilla hallitaan käyttäjien tunnistamista (*autentikointi*) ja sitä, mihin informaatioon kukin käyttäjä pääsee käsiksi (*auktorisointi*). Saatavuustieto on välttämätön metatieto, jos pääsyä halutaan hallita tietojärjestelmän avulla esimerkiksi tietoarkistossa tai säilytyspalvelussa. Suomessa käytämme yleisesti REMS- palvelua. Tekniset ratkaisut ja palvelut, joilla voidaan antaa tietoturvan ja -suojan vaatimukset huomioon ottava pääsy aineistoon halutuille käyttäjärhyhmille. Pääsynhallintaan kuuluu myös aineiston käyttötapahatunien rekisteröinti.

Raakadata Raakadata on jalostamatonta dataa jota syntyy moninaisissa muodoissa, ja joka voi elinkaarensa aikana kokea monia transformaatioita ja eri käsittelyvaiheita. Luonnontieteissä raakadataa tuottavat usein tutkimusinstrumentit kuten teleskoopit, satelliitit, synkrotronit ja kasvavassa määrin tietokonesimulaatiot. Humanistisissa ja yhteiskunnallisissa tieteissä raakadataa muodostuu esimerkiksi haastatteluista ja kyselyistä.

Rakenteellinen metatieto Metatiedon tyyppi, joka kuvaa miten kohteen komponentit järjestetään. Rakenteellisella metatiedolla voidaan esimerkiksi kertoa digitoidun kirjan, jossa jokainen sivu on erillinen tiedosto, sivujärjestys (KDK).

Rajapinta (avoin/suljettu) Ohjelmisto tai ohjelmistokomponentti, jolla eri ohjelmistot, tietokannat tai tietojärjestelmät voivat vaihtaa tietoja keskenään. Englanniksi Application Programming Interface (API). Ohjelmiston, järjestelmien, tai teknisten laitteiden rajapinnat voidaan määritellä joko suljetuiksi tai avoimiksi. Rajapintaa, jonka kaikki ominaisuudet ovat julkisia ja jota voi käyttää ilman rajoittavia ehtoja (esimerkiksi laatia rajapintaa hyödyntävän ohjelman ilman rajapinnan valmistajan erillistä hyväksyntää tai pakollisia lisenssimaksuja) kutsutaan avoimeksi. Suljetun rajapinnan ominaisuudet eivät ole julkisia tai niitä ei saa käyttää ilman rajoittavia ehtoja.

Rekisterit Julkisella sektorilla syntyy viranomaistoiminnan yhteydessä kattavia tietoaineistoja, joita kerätään rekistereihin ja tietojärjestelmiin. Lisäksi tilastotoimissa syntyy tilastollisia rekistereitä.

Rekisteritiedot Rekisteritiedot ovat rekisteriviranomaisten ylläpitämää tietoaineistoa tietyistä, laissa määritellyistä kohdejoukosta. Rekisteritiedot ovat yksikötason tietoja tarkasteltavasta ilmiöstä (henkilö, yritys, ympäristö jne.), jotka tallennetaan viranomaisten keräämiin hallinnollisiin tai tilastollisiin rekistereihin.

Rekisteriviranomaiset Rekisteriviranomaiset ylläpitävät yhteiskunnan perusrekistereitä tai toimialuekohtaisena viranomaisena hallinnollisia seuranta- ja valvontatietoa sisältäviä rekistereitä ja tietojärjestelmiä. Näitä tehtäviä hoitavilla virastoilla ja laitoksilla on usein velvoitteita raportoida asiantiloista.

Rekisteritutkimus Rekisteritutkimus on tutkimusta, jossa käytetään hyväksi rekisteritietoja. Tutkimus voi perustua kokonaan rekisteritietoihin tai rekisteritietoja voidaan käyttää muita aineistoja (esimerkiksi haastattelu- tai kyselyaineistoja tai klinisiä ja näyteaineistoja) täydentävinä tietoina.

Saatavuus (Access) Saatavuus määrittelee, onko tieto saatavilla käyttötarkoituksen mukaisesti; periaatteellisesti sekä teknisesti ja muiden toimintatarpeiden vaatimusten mukaisesti.

Sanasto Luettelo jossain kielessä tai ympäristössä sallituista sanoista tai termeistä luokitteluineen, määritelmineen, kuvauksineen ja esimerkkeineen. Sanastoja voidaan laatia monin eri tavoin ja eri käyttötarkoituksiin. Sanastotyyppejä ovat muun muassa terminologiset sanastot, asiasanastot, ontologiat sekä tietojärjestelmien ja sovellusten integrointia tukevat sanastot. Tietoteknisessä

ympäristössä sanastoilla kuvataan käsitteiden merkityksiä siten, että eri tietojärjestelmät voivat ymmärtää käsittelemäänsä tietoa. (JHS 179).

Siirtokelpoinen tiedostomuoto Tiedostomuoto, jota voi käyttää kun tietoa siirretään pitkäaikaissäilytykseen. Katso tarkemmin Kansallisen digitaalisen kirjaston verkkosivut kdk.fi

Sekundääriaineisto Primääriaineistosta otettu kokonainen tai osittainen kopio. Sekundääriaineistoon viitattaessa on aina mainittava alkuperä ja aikatieto (viittaus JHS179 Master Data Management -liite).

Semantiikka Semantiikka eli merkitysoppi tutkii esimerkiksi sanojen ja ajatusten merkitystä. Kielen merkitystä tutkiessaan se on kielitieteen alalaji, joka tutkii esimerkiksi sanan, virkkeen tai kokonaisten tekstien merkitystä ja viittauksia. Semantiikalle oleellista ovat ilmaisujen väliset suhteet, kuten antonymia, homonymia, hyponymia, meronymia, polysemia ja synonymia.

(Avoim) Standardi Valtuutetun tahon tietyllä alalla käytettäväksi hyväksymä tai alalla laajasti omaksuttu tuotetta tai toimintaa koskeva malli, mitta, ominaisuus tai nimitys (jota voi hyödyntää ilman rajoittavia ehtoja).

Sui generis –suoja Luo ylimääräisen suojakerroksen mm. aineistoille, jotka muuten olisivat vapaasti käytettävissä.

SHERPA/RoMEO Palvelu, jota ylläpitää Nottinghamin yliopisto, josta voi tutkia tieteellisten lehtien suhtautumista rinnakkaistallennukseen ja julkaisuarkistojen käyttöön.
<http://www.sherpa.ac.uk/romeo/>

SHOK – Strategisen huippuosaamisen keskittymä Strategisen huippuosaamisen keskittymät (SHOK) ovat yritysten, yliopistojen ja tutkimuslaitosten muodostamia yhteistyöliittymiä osaamisen jalostamiseen. Keskittymien käynnistäminen perustuu tiede- ja teknologianeuvoston (nyk. innovaatio- ja tutkimusneuvosto) vuonna 2006 tekemään linjaukseen. SHOK:ien tavoitteena on tarjota huipputasoinen tutkimusyksiköille ja tutkimustuloksia hyödyntäville yrityksille uuden tavan tehdä tiivistä ja pitkäjänteistä yhteistyötä keskenään. Keskittymät ovat sovelluslähtöisiä ja ne tukevat monitieteisyyttä. Teknologian edistämiskeskus Tekes vastaa pääosin SHOK:ien julkisesta rahoituksesta.

Skeema Rakennemalli, jolla kuvataan esimerkiksi dokumentin sisältöä.

Standardisalkku Yhteisesti sovittu valikoima standardeja, esimerkiksi kulttuuriorganisaatioiden järjestelmien välisen yhteistoiminnan edellyttämä, KDK:n kokonaisarkkitehtuuriin kuuluva standardi-kokonaisuus. Se määrittelee erityisesti sovellusten tietosisältöjä sekä sovellusten välisiä rajapintoja. Standardisalkkua ylläpidetään osana KDK-kokonaisarkkitehtuuria.

Sähköinen allekirjoitus Digitaalisessa muodossa oleva tieto, joka on liitetty tai joka loogisesti liittyy muuhun digitaaliseen tietoon ja jota käytetään allekirjoittajan henkilöllisyyden todentamisen välineenä (Laki sähköisistä allekirjoituksista 14/2003).

Talkoistaminen (*Crowdsourcing*) Käyttäjien tai muun laajemman joukon tai yleisön ottaminen mukaan aineiston jalostamiseen tai tuottamiseen. Tämä voi tapahtua verkon kautta tai muuten. Myös muodossa joukkoistaminen. Kohdennetulle asiantuntijaryhmälle suunnattu "niche sourcing" on samankaltainen, usein vapaaehtoiselle työlle perustuva toimintamalli.

Tarkistussumma (*Check sum*) Informaatiosta (merkkijonosta tai tiedostosta) soveltuva matemaattista algoritmia käyttäen laskettu lukuarvo tai merkkijono, jonka avulla voidaan tunnistaa informaatioissa syystä tai toisesta tapahtuneet muutokset. Jos yksikin bitti muuttuu, informaatiosta uudelleen laskettu lukuarvo ei täsmää aiemmin lasketun ja tallennetun lukuarvon kanssa. Tarkistussumman laskemiseen voidaan käyttää esimerkiksi MD5-algoritmia.

Tekijänoikeus Tutkimusjulkaisut ovat yleensä tekijänoikeuksien suojaamia, kun ne ylittävät teoskynnyksen. Aineistojen avoin käyttö voidaan kuitenkin mahdollistaa sopimuksin. ATT-hanke suosittelee CC0:n, CC 4.0 BY:n tai tarvittaessa muiden yleisesti tunnettujen sopimuslisenssien käyttöä.

Temaattinen keskittymä Tiettyyn tutkimukselliseen asiakokonaisuuteen keskittyvä toiminta.

Tiedonhallinta Tiedon keruu, organisointi ja tallentaminen siten, että se on helposti löydettävissä ja käytettävissä.

Tiedonkeruu Tässä yhteydessä tarkoitetaan useimmiten opetus- ja kulttuuriministeriön vuosittaista tiedonkeruuta yliopistojen toiminnasta (KOTA). Tuloksia voi tarkastella esim <http://vipunen.fi/fi-fi> tai julkaisutietoportaalista Juuli <http://www.juuli.fi/> . Tietoja käytetään yliopistojen rahoitusneuvotteluissa.

Tiedonlouhinta (*Data mining*) Tiedonlouhinta tarkoittaa joukkoa menetelmiä, joilla pyritään oleellisen informaation löytämiseen suurista tietomassoista. Tiedonlouhinta voidaan soveltaa hyvin laaja-alaisesti, sillä lähtökohdaksi tarvitaan ainoastaan raakadataa. Tyypillisesti tiedonlouhinnassa käytetty tietoaaineisto on esimerkiksi mittauksia teollisuusprosessista, otteita asiakastietokannasta tai vaikkapa web-palvelimen loki-tiedostoja. Määritelmänä tiedonlouhinta ei rajaa käytettäviä menetelmiä.

Tiedonpoiminta Tiedonpoiminnalla tarkoitetaan prosessia, jolla kerättyä tietoa muutetaan koneluettavaan muotoon.

Tiedostomuoto (*File format*) Sovittu tapa tallentaa informaatiota digitaaliseen muotoon. Esimerkiksi *Portable Document Format* (pdf).

Tiekartta Tiekartta perustuu opetus- ja kulttuuriministeriön asettaman poikkihallinnollisen Avoin tiede ja tutkimus – hankkeen (ATT) työhön tiedon saatavuuden ja avoimen tieteen edistämiseksi. Tiekartta on tarkoitettu suomalaisille tutkijoille, tutkimusryhmille, tutkimusorganisaatioille, päättäjille, rahoittajille sekä kansalaisille etenemissuunnitelmaksi avoimen tieteen ja tutkimuksen edellytyksien, sovellusten ja tulosten hyödyntämisen edistämiseksi vuosina 2014-2017. Tiekartan painotukset selkeytyvät ja toimenpiteet täsmentyvät hankkeen kuluessa kansallisen ja kansainvälisen tilannekuvan tarkentuessa erilaisten selvitysten kautta. Tiekartan taustatietona ovat aiemmat selvitykset ja julkaisut.

Tieteen neljäs paradigma Dataintensiivinen tutkimus, jonka nykyaikaiset tietokoneet, ohjelmistot ja tutkimusmenetelmät mahdollistavat. Se antaa tutkijoille työkalut tehdä kokeita ja laskuja eri suuruusluokissa, skaaloissa ja määrissä kuin koskaan ennen.

Tietoaaineistojen uudelleen hyödyntäminen Uudelleen hyödyntämisellä tarkoitetaan sitä, että yksityiset ihmiset ja oikeushenkilöt voivat kopioida, muokata, julkaista ja levittää tietoaaineistoja eri tarkoituksiin. Suurta osaa julkisin varoin kerätystä raakadatasta ja tutkimusaineistoista voidaan

hyödyntää ja hyödynnetään niin tutkimuksessa kuin erilaisten arkipäivän tuotteiden ja palvelujen kehittämisessä.

Tietoaineiston hallintasuunnitelma Hallintasuunnitelma ja -mekanismi ohjaavat digitaalisen informaation keruuta, laadunvalvontaa, analysointia, julkaisemista ja varastointia. Suomen akatemian edellyttämässä hallintasuunnitelmassa kerrotaan, miten tutkimuksessa käytettävät aineistot hankitaan, miten aineistoja käytetään ja säilytetään ja miten mahdollistetaan aineistojen käyttö tutkimuksen päätyttyä.

Tietoarkisto (*Data Archive, Data Center*) Organisaatio tai toiminta jossa ammattimaisesti kerätään, säilytetään, hoivataan ja tehdään saatavaksi tutkimuksen tietoaineistoja.

Tietoinfrastruktuuuri Tietoinfrastruktuurin osa, joka sisältää kaikki tietoaineistojen tuottamiseen, tallentamiseen, jakamiseen ja hyödyntämiseen tarvittavat yhteentoimivat peruspalvelut ja työkalut.

Tietojärjestelmäarkkitehtuuri Tietojärjestelmän arkkitehtuuri kuvaa kohdealueensa rakenneosat, niiden ulospäin näkyvät ominaisuudet ja niiden väliset yhteydet ja riippuvuudet. Arkkitehtuuri muodostaa rungon järjestelmän suunnittelulle ja toteutukselle sekä ohjaa järjestelmän rakenteen kehittämistä järjestelmän elinkaaren ajan. Se toimii myös keskusteluvälineenä järjestelmän kehittämisen ja ylläpitämisen sidosryhmien (johto, käyttäjät, suunnittelijat) välillä.

Tietojen eheys (*Data Integrity*) Tietojen eheys tarkoittaa "1. (Tietojen tai tietojärjestelmän) aitous, väärentämättömyys, sisäinen ristiriidattomuus, kattavuus, ajantasaisuus, oikeellisuus ja käyttökelpoisuus. 2. Ominaisuus, että tietoa tai viestiä ei ole valtuudettomasti muutettu, ja että mahdolliset muutokset voidaan todentaa kirjausketjusta." (Valtionhallinnon tietoturvallisuuskäsitteistö 1/2000).

Tietokanta Tietokanta eli tietovarasto on kokoelma tietoja, joilla on yhteys toisiinsa. Tietokantaan voidaan tallentaa eri formaateissa olevaa tietoa, esimerkiksi tekstiä, ääntä ja videokuvaa. Tietokannan rakentaminen perustuu siihen tallennettavan tiedon mallintamiseen ja esittämiseen yleensä jonkin kuvauskielen tai tekniikan avulla. Tietokantoja luokitellaan usein niiden rakentamiseen käytetyn ohjelmointimallin mukaan.

Tietokantaohjelmisto/ympäristö Tietokantaohjelmisto/hallintajärjestelmä on kokoelma varusohjelmistoja, joiden tehtävänä on tukea varsinaista tietokantamoottoria (database engine) mm. optimoimalla kyselyjä käyttäen hyväksi tallennettua hakutilastoa (statistics) ja indeksejä. Tietokannan hallintajärjestelmät sisältävät myös ylläpitoa helpottavia apuohjelmia joiden avulla tietokantamoottorin säilyttämä tieto voidaan varmuuskopioida ja palauttaa sekä viedä ja tuoda muihin järjestelmiin, sekä käyttöliittymän.

Tietopolitiikka Tietopolitiikalla tarkoitetaan tässä yhteydessä kansallisen tason koordinoitua politiikkasuunnitelmaa ja -ohjelmaa, joka ohjaa toiminnan ja käytäntöjen kehittymistä eri hallinnonaloilla, ja joka sisältää eri toimijoiden roolien ja työnjaon määrittelyn sekä suunnitelmat niin immateriaalioikeuksien hallintaan, tietosuojakysymyksiin, hinnoitteluun kuin saatavuuden ja käyttöehtojenkin määrittelyyn.

Tietosuoja Ihmisten yksityisyyden suojeleminen. Sillä tarkoitetaan yksilöä koskevien tietojen suojaamista oikeudettomalta käytöltä henkilötietoja käsiteltäessä. Aineistonhallinnassa otetaan huomioon henkilön yksilöivän tiedon käsittely oikeudellisia säännöksiä noudattavin periaattein ja

toimintakäytännöin. Lisää tietoa Tietosuojavaltuutetun toimiston "Rekisteritutkimuksen tietosuojapöytäkirjoille ja tietopyyntöjä käsitteleville viranomaisille http://www.tietosuoja.fi/material/attachments/tietosuojavaltuutettu/tietosuojavaltuutetuntoimisto/opaat/6Jf_qBT5i3/Rekisteritutkimuksen_tietosuojapöytäkirjoille_ja_tietopyyntöjä_kasitteleville_viranomaisille.pdf

Tietotilinpäätös Organisaation raportti siitä, millaista tietoa/tietoaineistoja heillä hallinnoidaan, miten niitä jaetaan ja hyödynnetään.

Tietoturva Tietoturvalla tarkoitetaan aineistojen suojaamista tuhoutumiselta, muuttumiselta tai varkaudelta aineiston käsittelyn, säilytyksen ja tiedonsiirron aikana. Tähän liittyy hallinnollisia ja teknisiä toimia, joilla varmistetaan se, että tiedot ovat vain niiden käyttöön oikeutettujen saatavilla, ettei tietoja voida muuttaa muiden kuin siihen oikeutettujen toimesta mutta että tiedot ja tietojärjestelmät ovat niiden käyttöön oikeutettujen hyödynnettävissä. Tietoturvaan liittyviä käsitteitä ovat pääsynvalvonta, salassapito, yksityisyys, todennus, eheys ja turvallisuus.

Tilastolliset tietosuojamenetelmät Yleisnimitys menetelmille, joiden tavoitteena on julkaistaviin tai tutkimuskäyttöön luovutettaviin taulukko- ja yksikköaineistoihin vaikuttavien yksiköiden yksityisyyden suojaaminen.

Tilastotieto/ tilastomuotoinen tieto Tilastointia varten hankittu yksikötason tieto (esim. henkilötason tai yritystason tieto) tai tehtyyn tilastoon perustuva aggregoitu tieto. Termiä käytettäessä on syytä tarkentaa onko kyse yksikötason tiedosta vai aggregoidusta tiedosta.

Tilastoviranomaiset Tilastoviranomaisilla on oikeus kerätä tietoja tilastotarkoituksiin tilastolaissa säädetyn tiedonantovelvollisuuden perusteella. Tilastoviranomaisia ovat Maa- ja metsätalousministeriön tietopalvelukeskus Tike, Terveystieteiden ja hyvinvoinnin laitos THL, Tilastokeskus ja Tullihallitus. Tilastoviranomaisten lisäksi valtion tilastotoimeen kuuluu muita valtion virastoja tai laitoksia, tilastoja laativia muita viranomaisia kuten Ilmailuhallinto, Ilmatieteen laitos, MTT, Maanmittauslaitos, Metla, RKTL, SYKE ja TEM.

Tilastoyksikkö Tilastoyksiköllä tarkoitetaan tarkastelun perusyksikköä, eli luonnollista henkilöä, kotitaloutta, talouden toimijaa ja muita yrityksiä, johon tiedot viittaavat.

Tulkintatieto (*Representation Information*) Osa säilytettävään informaatioon liitettävää metatietoa. Tulkintatiedon muodostaa aineiston luettavuuden mahdollistava tieto, tiedostomuodon kuvaus, semantiikka ja muu tieto, jolla tiedoston bittijonot voidaan muokata ihmisen tai koneen luettavaan ja ymmärrettävään muotoon.

Tunnisteet (*Persistent Identifiers*) Julkaisujen ja tutkimusdatan tai muiden entiteettien tunnisteita sovelletaan muun muassa näiden aineistojen hakuun, identifiointiin ja linkitykseen. Aineistojen tunnisteteet ovat lisäksi pitkäaikaissäilytyksen ehdoton edellytys. Tutkimusjulkaisujen tunnisteenä voidaan käyttää julkaisu-tyypistä riippuen esimerkiksi ISBN:ää (monografiat) tai erilaisia pysyviä tunnisteteita (persistent identifiers, PIDs), joista Handle-tunnistetta käytetään julkaisuarkistoissa, DOI:ta kaupallisten kustantajien järjestelmissä ja URN-tunnusta kansalliskirjastojen digitaalisissa kokoelmissa. Tutkimusdatalle sovelletaan kansainvälisissä ja kansallisissa hankkeissa lähes yksinomaan PID-tunnisteita, TTA- ja ATT-hankkeissa URN:ää. Aineistojen lisäksi tarvitaan tunnisteteita tutkijoille sekä tutkimusprosessiin osallistuville juridisille henkilöille (yliopistot ja muut

korkeakoulut sekä tutkimuslaitokset; tiedekunnat ja niiden laitokset, tutkimusryhmät). Näiden tunnisteiden soveltamista Suomessa selvitetään erikseen.

Yhtenäiset yksilöintikäytännöt ovat erityisen tarpeellisia, kun tietojärjestelmät vaihtavat informaatiota avointen rajapintojen kautta, useiden eri järjestelmien kanssa. Tällöin käytettävät tietosisällöt mm. koodistot ja niiden versiot on tunnistettava eli yksilöitävä. Kansainvälisesti on yksilöity mm. puhelinnumerot, sähköpostiosoitteet, kirjat (ISBN-numero), tuotteet (EAN-koodi). Yksilöinnissä voidaan käyttää ISO OID-yksilöintitunnusta. ISO OID- yksilöintitunnus on kansainvälisesti vain yhteen objektiin liitettävä numeroarvo, joka yksilöi kyseisen objektin yksiselitteisesti ISO:lle varatussa yksilöintijärjestelmässä. Muita yksilöintitunnuksia ovat mm. URInimet (Uniform Resource Identifiers) ja UUID (Universal Unique Identifier).

Yksilöintitunnuksella (object identifier) tarkoitetaan ilmiön, kuten tutkimusjulkaisun, tunnistamista yksikäsitteisellä tunnuksella, jolla se voidaan erottaa muista vastaavista. Yksikäsitteisyys voi olla paikallista, alueellista, kansallista tai kansainvälistä. HTK-ohjeen mukaan: *"Tutkijat ottavat muiden tutkijoiden työn ja saavutukset asianmukaisella tavalla huomioon niin, että he kunnioittavat muiden tutkijoiden tekemää työtä ja viittaavat heidän julkaisuihinsa asianmukaisella tavalla ja antavat heidän saavutuksilleen niille kuuluvan arvon ja merkityksen omassa tutkimuksessaan ja sen tuloksia julkaistessaan."*

Tunnisteiden ja muiden standardien avulla organisoidaan tietosisällöt kansalliseen ja kansainväliseen käyttöön. Tunnisteet tukevat tiedon globaalia löytyvyyttä ja täsmentävät näin tiedonhakuja. Tunnisteet ovat osa metatietoa ja osa metatietoa hallinnoivia välineitä. Suositeltavia tunnisteita ovat mm. URN, OID, ontologiat, auktoriteettitiedot.

Turmeltuminen (*Corruption*) Informaation sisällön, eheyden tai käytettävyyden menetys tai muuttuminen. Turmeltumisen voi aiheuttaa esimerkiksi laitevika, ohjelmavirhe, tietojen epäonnistunut migraatio, huolimaton ylläpito tai järjestelmään kohdistunut hyökkäys.

Tutkimusaineisto, tutkimuksen tietoaaineisto Tutkimusaineistoilla tarkoitetaan niitä resursseja, joita tutkija tuottaa tai joita hän käyttää tutkimusprosessin aikana. Tutkimuksen tietoaaineistolla tarkoitetaan digitaalisessa, analogisessa ja fyysisessä muodossa olevia tieteellisen tutkimuksen perusaineistoja ja niistä jalostettuja aineistoja, joihin tutkimuksen löydökset ja tulokset perustuvat. Tutkimusaineistot voivat olla aineellisia tai aineettomia. Jotta aineisto olisi tutkimukseen kelpaavaa, siihen täytyy liittyä tietoja ainakin sen alkuperästä. Tietoaaineistoihin yleensä kuuluvat sekä tutkimuksen tuottamat että tutkimuksen hyödynnettävissä olevat aineistot. Esimerkiksi datan liitteeksi tarvitaan kuvailevaa ja teknistä tietoa siitä mitä informaatiota se sisältää. Siksi tutkimusdataan liittyy paljon tietoa siitä, miten se on rakenteistettu ja koodattu, miten se on syntynyt ja miten sitä on käsitelty. Nämä tiedot on aina syytä tallentaa esimerkiksi metatietoihin, koodikirjoihin ja/tai muuhun dokumentaatioon. Yhdessä datan kanssa tästä kokonaisuudesta muodostuu tutkimusaineisto. On siis huomattava, että kaikki tutkimukselle hyödylliset tietoaaineistot eivät välttämättä ole alun perin kerätty tutkimuskäyttöön vaan esimerkiksi hallinnollista seuranta, valvontaa, raportointia ja tilastointia varten.

Tutkimusinfrastrukturi Tutkimusinfrastruktuurit ovat tutkimusvälineiden, laitteistojen, aineistojen ja palveluiden varanto, joka mahdollistaa tutkimus- ja kehitystyön innovaatio toiminnan eri vaiheissa, tukee organisoitunutta tutkimustyötä sekä ylläpitää ja kehittää tutkimuskapasiteettia.

Tutkimusjärjestelmä Suomen tutkimus- ja innovaatiojärjestelmä on kokonaisuus, jonka muodostavat uuden tiedon ja osaamisen tuottajat, niiden hyödyntäjät sekä näiden väliset

vuorovaikutussuhteet. Innovaatiojärjestelmän keskeiset osat ovat koulutus, tutkimus ja tuotekehitys sekä näihin nojaava yritystoiminta. Tutkimus- ja innovaatiojärjestelmä nojaa yksityisen ja julkisen sektorin erilaisiin rooleihin ja yhteistyöhön. Yksityinen sektori rahoittaa järjestelmästä noin kaksi kolmasosaa ja julkinen sektori noin kolmanneksen. Uutta tietoa tuottavat erityisesti yliopistot, tutkimuslaitokset ja korkeakoulut mutta myös yritykset ja muut yhteiskunnan ja talouden uudistumiseen tähtäävät laitokset. Yritykset tuottavat pääosin tuotekehitykseen, uusien tuotteiden, palveluiden ja prosessien kehittämiseen liittyvää tietoa. Tietoa hyödyntävät useimmiten yritykset, yhdistykset, yksityiset kansalaiset sekä yhteiskunnan kehittämisestä vastaavat julkiset instituutiot kuten poliittiset päätöksentekijät ja hallinto. Tutkimustiedon merkitys yhteiskunnan ja talouden kehittämisessä on jatkuvasti kasvanut, mikä lisää hyvän yhteistyön ja verkottumisen merkitystä sektoreiden sisällä ja yksityisen sekä julkisen sektorin välillä.

Tutkimusjärjestelmän organisaatioita ja rakenteita ovat muun muassa:

- *Korkeakoulut.* Yliopistojen tehtäviin kuuluu tieteellinen- ja taiteellinen sivistys, tutkimus ja siihen perustuva ylin opetus mukaan lukien tutkijankoulutus. Ammattikorkeakoulujen tehtävänä on harjoittaa työelämää ja aluekehitystä tukevaa ja alueen elinkeinorakenteen huomioon ottavaa soveltavaa tutkimus- ja kehitystyötä.
- *Valtion tutkimuslaitokset ja muut tutkimuslaitokset.* Eri hallinnonalojen tutkimuslaitoksilla ja muilla tutkimuslaitoksilla on korkeakoulusektorin tutkimusyksiköiden ohella merkittävä asemansa innovaatiojärjestelmässä tieteelliseen tietoon perustuvan yhteiskunnan kehittäjinä.
- *Ohjelmatoiminta.* Suomen Akatemia ja Tekes pyrkivät tutkimusohjelmatoiminnallaan panostamaan tulevaisuuden kannalta keskeisiin aloihin.
- *Huippuyksiköt.* Oman tieteenalansa kansainvälisessä kärjessä olevat yksiköt uudistavat tutkimusta, kehittävät luovia tutkimusympäristöjä ja kouluttavat uusia tutkijoita.
- *Tutkimusyhteisöt.* Strategisen huippuosaamisen keskittymissä (SHOK) toteutetaan yritysten, yliopistojen ja tutkimuslaitosten yhdessä määrittelemää tutkimusstrategiaa.
- *Yritysmaailma.* Pörssiyhtiöt käyttivät liikevaihdostaan noin kaksi prosenttia tutkimus- ja tuotekehitykseen vuonna 2010.
- *Kansainvälinen yhteistyö.* Suomi on mukana keskeisissä kansainvälisissä tutkimusyhteistyöelimityksissä ja organisaatioissa sekä osallistuu aktiivisesti eurooppalaiseen tutkimusyhteistyöhön. Kansainvälisen yhteistyön avulla vahvistetaan suomalaisen tutkimuksen laatua ja vaikuttavuutta.

Tutkimustietojärjestelmä Tietojärjestelmä, johon organisaatiossa kerätään tietoja tutkijoiden toiminnasta. Joissain tapauksissa yhdistetty julkaisuarkistoon, jotta julkaisujen kuvallisuus ja tallentaminen on käyttäjälle helpompaa.

Tutkimustulosten avoimuus ja luottamuksellisuus Luottamuksellisuus on "1. Tietojen säilyminen luottamuksellisina ja tietoihin, tietojenkäsittelyyn ja tietoliikenteeseen kohdistuvien oikeuksien säilyminen vaarantumiselta ja loukkaukselta. 2. Se missä määrin luottamuksellisuutta pidetään tärkeänä." (Valtionhallinnon tietoturvaluokituskäsitteistö 1/2000).

Työpaperi (*Working paper*) Työpaperin tavoite on tuoda tutkimuksen alustavat tulokset tutkimusyhteisön arvioitavaksi. Tulokset julkaistaan yleensä muussa julkaisukanavassa myöhemmin. Julkaisukäytännöt vaihtelevat tieteenaloittain, ja joillakin aloilla työpapereita pidetään tärkeinä julkaisukanavina.

Työvuoro (*Workflow*) Työvuoro eli työnkulku on tutkimusprosessin automaatio, jonka aikana dokumentteja, tietoa ja tehtäviä siirretään suoritettaviksi proseduraalisten sääntöjen mukaisesti.

Uutuusvahti (*Alert*) Ohjelmiston komponentti, joka toistaa ennalta määritellyn haun halutuin väliajoin halutuista tietokannoista, ja ilmoittaa uutuuksista haluttaessa sähköpostilla tai jollain muulla sopivalla teknisellä ratkaisulla.

Valtion tilastotoimi Tilastoviranomaiset ja muut tilastoja tuottavat viranomaiset. Nämä vastaavat mm. Suomen virallisen tilaston ja Euroopan tilastojärjestelmään kuuluvien tilastojen (ESS) tuotannosta. Valtion tilastotoimen tehtävänä on palvella yhteiskunnan yleistä tietotarvetta tuottamalla yhteiskuntaoloja ja niiden kehitystä kuvaavia tilastotietoja yhteiseen käyttöön. Tilastotoimen yleisenä perustana on tilastolaki.

Valtion tutkimuslaitokset Valtion omistamat tuottavat tietoa pääasiassa oman aihealueensa tilasta ja kehityksestä. Ne myös ylläpitävät oman hallinnonalansa seurantatietoa sisältäviä tietojärjestelmiä. Voidaan käyttää myös termejä tutkimusviranomaiset, tutkimusta harjoittavat viranomaiset tai valtion tutkimustoimintaa harjoittavat organisaatiot.

Verkkopalvelu Internet-verkkoon liitetyn tietojärjestelmän antama palvelu, myös Internetissä oleva multimedia- tai sisältökokonaisuus.

Vertaisarviointi (*Peer Review*) Vertaisarviointi eli referee-käytäntö on alun perin tieteellisen julkaisemisen tapa, jossa lehteen tai julkaisuun lähetetyt artikkelit arvioi sekä julkaisusarjan toimitus että sen valitsevat ulkopuoliset asiantuntijat. Vertaisarvioinnissa tarkistetaan julkaistavaksi tarjotun artikkelin asiasisällön ja tieteellisen merkittävyyden lisäksi myös sen kieliasu sekä varmistetaan, että tekstin rakenne vastaa tieteellisen kirjoittamisen käytäntöjä (esimerkiksi tekstin tiiviys, kuvien ja taulukoiden selkeys sekä lähdeviitteiden käyttö) ja että se sopii kyseiseen julkaisukontekstiin.

Viitetietue (*Reference Information*) Osa säilytettävään informaatioon liitettävää metatietoa. Se on aineiston pysyvä tunniste, bibliografinen tieto tai muu vastaava tieto.

Vihreä OA (*Green OA*) Julkaisun tai sen osan rinnakkaiskopion tallentaminen tieteenala- tai organisaatiokohtaiseen julkaisuarkistoon, jossa se on vapaasti saatavilla joko heti tai ennalta määrätyn embargoajan jälkeen.

Viides vapaus Tutkijoiden, tiedon ja teknologian vapaa liikkuvuus, joka tarvitsee tuekseen avoimen tieteen ja sitä tukevan e- infrastruktuurin.

Viittaaminen (*Citation*) Viite on tietolähteeseen viittaava merkintä, joka voidaan sijoittaa juoksevaan tekstiin tekstiviitteeksi, sivun alareunaan alaviitteeksi, julkaisun tai sen osan loppuun loppuviitteeksi tai lähdeviitteeksi esimerkiksi lähdeluetteloon. Sähköisessä maailmassa sen tulisi myös sisältää pysyvä tunniste viittauksena lähteeseen.

Viranomaisaineisto Viranomaisaineisto voi olla paperinen tai sähköinen teksti, kuvat, ratkaisu, valmisteluasiakirja, viranomaiselle toimitetut tai toiminnan yhteydessä laaditut asiakirjat, rekisterit ja rekisteritiedot. Aineistot tulevat julkisiksi asian käsittelyn päätyttyä tai selvityksen valmistuttua käyttötarkoitukseensa muutoin keskeneräisessä asiassa.

Visualisointi Tietoaineiston havainnollistaminen graafisesti.

XML (*Extensible Markup Language*)n Laite- ja ohjelmistoriippumaton avoin standardi, joka tarjoaa syntaksin dokumenttien osien määrittelylle sekä standardoidun tavan kuvata osien välisiä suhteita. Sitä voidaan soveltaa sekä dokumentti- että tietoaineistomuotoisen tiedon tallentamisessa.

XML –skeema (*XML Schema*) W3C:n standardoima teknologia, jolla voidaan kuvata XML-dokumenttien rakenne. Skeemoilla voidaan kuvata esimerkiksi XML-muotoisen levytiedon tai vaikkapa myynti-ilmoituksen sallittu rakenne. Hyvä syy tällaisen kuvailun tekemiseen on yhteisen sanaston rakentaminen. XML-skeema on kuin standardoitu tapa puhua XML:stä.

Yhteentoimivuus Toiminnan eri tasojen ja toimijoiden sisäinen sekä niiden välinen saumaton toiminta. Toiminnan tasoja ovat muun muassa strategiat, palvelut, prosessit, organisaatiot, tiedot, määritelmät, termistö ja teknologia. Yhteentoimivuuden parantamiseksi julkisessa hallinnossa on ryhdytty hyödyntämään kokonaisarkkitehtuurisuunnittelua.