



AVOIN TIEDE JA TUTKIMUS

AVOIMEN TIETEEN JA TUTKIMUKSEN KÄSIKIRJA OSA 2: TUTKIJALLE JA ORGANISAATIOILLE

(työversio 0.7))



Avoimen tiedon käsikirja v1.5, jonka tekijä on ATT-hanke, on lisensoitu Creative Commons Nimeä 4.0 Kansainvälinen -käyttöluvalla.

Tämä käsikirja on tarkoitettu tutkijoille, tutkimusryhmille ja tutkimusorganisaatioille auttamaan avoimen tieteen ja tutkimuksen hyödyntämisessä ja edistämässä. Käsikirjaa työstetään hankkeessa kaikkien kiinnostuneiden kesken. Hankkeen strategiaryhmän hyväksymät versiot julkaistaan avointiede.fi-sivustolle

Tutkijoiden ja tutkimusorganisaatioiden on pyrittävä asettamaan tutkimusaineistot, tutkimusmenetelmät, ja tutkimusjulkaisut tulevat avoimesti saataville huomioiden toimintaympäristön ja tutkimusetiikan. Tutkijoita ja organisaatioita tuetaan tässä muutoksessa ohjeistuksella, palveluilla sekä kannustimilla. 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 koneluetettavaa lisenssiä CC BY 4.0. Tutkimusaineistoja kuvailevat metatiedot suositellaan lisensoimaan käyttöluvalla CC0. Tutkijoita, tutkimusryhmiä ja tutkimusorganisaatioita kehoitetaan hyödyntämään olemassa olevia palveluita kuten esimerkiksi Euroopan tasolla suositellut PubMed central, ArXiv, Openaire+, Zenodo sekä Suomessa IDA, Etsin ja AVAA sekä korkeakoulujen ja tutkimuslaitosten omat julkaisuarkistot ja tutkimustietojärjestelmät .

Tähän on koottu toimenpiteitä avoimen tieteen ja tutkimuksen edistämiseksi toimijaryhmittäin.

TUTKIJAN TOIMENPITEET

- Huolellinen aineistonhallinnan suunnittelu ja toteuttaminen
- Tutkimuksen toistettavuudesta huolehtiminen riittävällä dokumentaatiolla
- Pysyvien tunnisteiden hankkiminen sekä viittauksista huolehtiminen
- Palvelujen hyödyntäminen (mm. tietoaarkistot, julkaisuarkistot, tutkimustietojärjestelmät)
- Tutkimuksesta kertominen julkisuuteen ja avoimen arvioinnin hyödyntäminen mahdollisuuksien mukaan
- Avoin julkaiseminen (artikkelit, aineistot menetelmät) ja tulosten lisensoiminen
- Rikkaiden metatietojen tuottaminen

TUTKIMUSRYHMIEN TOIMENPITEET

- Aineistonhallinnan periaatteista ja toimenpiteistä sopiminen
- Tekijyydestä, omistajuudesta ja jatkokäytöstä sopiminen
- Tutkimuksen toistettavuudesta huolehtiminen
- Palveluinfrastruktuurin hyödyntämiseen ja avoimeen tutkimusprosessiin kannustaminen
- Tutkimuksen tulosten lisensoiminen
- Avoimen lähdekoodin, standardien ja rajapintojen suosiminen ja hyödyntäminen
- Rikkaiden metatietojen tuottaminen tietoaineistoille, tutkimusmenetelmille ja julkaisuille

TUTKIMUSORGANISAATIOIDEN TOIMENPITEET

- Selkeän julkaisupolitiikan ja datapolitiikan laatiminen sekä kuvaus tutkijan oikeuksista avoimuuteen
- Avoimen tieteen ja tiedonhallinnan osaamisen tukeminen organisaatiossa ja osana opetusta
- Tutkimusaineistojen hallintaa edistävien ohjeiden ja toimien kehittäminen sekä näistä palkitseminen;
- Yhteisen palveluinfrastruktuurin hyödyntäminen sekä tarvittavan paikallisen infrastruktuurin ja tukipalvelujen tarjoaminen
- Tarvittavien indikaattoritietojen keruu
- Omien tietojärjestelmien avaaminen ja rajapintojen rakentaminen noudattaen standardeja
- Avoin yhteistyö muiden tutkimusorganisaatioiden kanssa

Avoimen tieteen ja tutkimuksen käsikirja (versio 2: tutkijalle ja organisaatioille)	0
AVOIMEN TIETEEN JA TUTKIMUKSEN KÄSIKIRJA	1
TUTKIJAN TOIMENPITEET	1
TUTKIMUSRYHMIEN TOIMENPITEET	1
TUTKIMUSORGANISAATIOIDEN TOIMENPITEET.....	1
AVOIMEN TIETEEN JA TUTKIMUKSEN KESKEISET LÄHTÖKOHDAT	3
KANSAINVÄLISET KEHITYSSUUNNAT	3
AVOIMEN TIETEEN JA TUTKIMUKSEN EDISTÄMINEN.....	4
AVOIMEN TIETEEN JA TUTKIMUKSEN HYVÄT KÄYTÄNNÖT	4
AVOIMUUDEN EDISTÄMINEN TUTKIMUSPROSESSISSA	5
TUTKIMUSTIEDONHALLINTA.....	6
TUTKIJAN muistilista.....	7
TUTKIMUKSEN TietoAINEISTOT	8
TUTKIJAN muistilista.....	11
TUTKIMUSAINEISTON METATIEDOT.....	11
TUTKIJAN muistilista.....	13
TUNNISTEET JA VIITTAUKSET	14
TUTKIJAN muistilista.....	15
AVOIN TUTKIMUSPROSESSI	15
TUTKIJAN muistilista.....	16
TULOSTEN JULKAISEMINEN	17
TUTKIJAN NÄKÖKULMA.....	17
JULKAISUARKISTOJEN NÄKÖKULMA	18
TUTKIJAN muistilista.....	19
AVOIMUUDET POLITIIKAT	20
TUTKIMUSORGANISAATION MUISTILISTA.....	20
PALVELUIDEN RAHOITUS JA HINNOITTELU	20
Käsitteet	22

AVOIMEN TIETEEN JA TUTKIMUKSEN KESKEISET LÄHTÖKOHDAT

Avoimuus on tieteen ja tutkimuksen keskeinen periaate, jonka 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. Tieteellisen tiedon luonteeseen kuuluu toistettavuus ja vastuullisuus. 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.

KANSAINVÄLISET KEHITYSSUUNNAT

Suuri joukko kansallisia ja kansainvälisiä tutkimusorganisaatioita ja tutkimuksen rahoittajia on asettanut tutkimuksen tulosten avoimuuden lisäämisen tavoitteekseen. Euroopan komission suositus ”Recommendation on Access to and Preservation of Scientific Information” asettaa linjoja ja tavoitteita eurooppalaisen tutkimuksen avoimuudelle. OECD:n 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. Avoimen tieteen sidosryhmillä - kuten rahoittajilla, korkeakouluilla ja niiden kirjastoilla, datakeskuksilla, tieteellisillä seuroilla ja tiedekustantajilla - on isoja haasteita ja mahdollisuuksia edessään tietoaaineistojen uudelleen hyödyntämisessä. Myös suomalaisten tutkijoiden ja tutkimusorganisaatioiden on syytä huolehtia siitä, että pysyvät kehityksessä mukana.

Yhä suurempi osa tieteellisistä julkaisuista ilmestyy open access -lehdissä, ja open access -periaatteella toimiva PLoS One on kasvanut maailman ylivoimaisesti suurimmaksi tieteelliseksi lehdeksi. Muutkin merkittävät tieteelliset kausijulkaisut kuten Nature ja Science julkaisevat yhä useammin avointa tiedettä käsitteleviä artikkeleita. Kustantajat kehittävät strategioita tutkimusdatan jakamiseen julkaisujen rinnalla. Monet erilaiset sidosryhmät julkisella ja yksityisellä sektorilla ottavat osaa tähän toimintaan. Erilaiset hankkeet kuten DataCite, preprintpalvelu arXiv, 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 vielä normi kaikilla tutkimusaloilla. Vaikka tutkimusdatan julkaisemiseen on kiinnostusta eri puolilla maailmaa, menestyksen avain on asiaan liittyvien, monia eri sidosryhmiä koskevien haasteiden ymmärtäminen.

Myös tutkijoiden puolella on herätty tutkimusaineistojen käytettävyyden ja aineistonhallinnan haasteisiin. 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 Foundation on muodostanut omia kansainvälisiä verkostojaan avoimen tieteen ympärille. Facebookissa Open Science Finland-ryhmällä on yli 300 jäsentä. Open Knowledge Foundationin toiminta keskittyy usein nimenomaan osaamisen lisäämiseen. Vuoden 2015 kansainvälisenä teemana on tiedepuolella mm. avoimuus korkeakoulutuksessa sekä avoimen tieteen koulutus ja opetus.¹

¹ ”Open Training for Open Science”, OKF Open Science Working Group, 21.12.2014 <http://science.okfn.org/2014/12/21/open-training-for-open-science/>

Euroopan tutkimusneuvosto (European Research Council) vaatii avointa julkaisua rahoittamaltaan tutkimukselta artikkeleiden ja monografioiden kohdalla ja suosittelee sitä tutkimusaineistoille.² Horizon 2020-ohjelman pilottiprojektissa vaaditaan myös tutkimusaineistojen avaamista.³ 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.⁴ 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.⁵ FOSTER, joka keskittyy avoimuuden koulutukseen, on kerännyt verkkosivuilleen paljon aineistoa avoimeen tietoon ja tutkimuksen tiedonhallintaan liittyen.⁶ Open Access-julkaisemista edistää myös SPARC Europe, joka on tutkimusorganisaatioiden verkosto⁷ 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.⁸

Maa- ja maailmanlaajuisesti julkaisuarkistot leviävät ja kasvavat,⁹ ja kansainväistä infrastruktuuria syntyy jatkuvasti lisää. Euroopan tutkimusneuvosto on sitoutunut ArXivin tukemiseen¹⁰. OpenAIREplus-projektissa¹¹ ratkotaan aineistojen, menetelmien ja aineistojen kesken viittaamisen sekä löytyvyyden haasteita. Myös tieteelliset kirjastot ovat aktiivisesti mukana eurooppalaisissa hankkeissa LIBER-järjestönsä kautta.¹²

AVOIMEN TIETEEN JA TUTKIMUKSEN EDISTÄMINEN

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 tieteentutkimuksen levittäminen yhteiskunnassa.

AVOIMEN TIETEEN JA TUTKIMUKSEN HYVÄT KÄYTÄNNÖT

² ERC Open Access Guidelines 2013, http://erc.europa.eu/sites/default/files/document/file/ERC_Open_Access_Guidelines-revised_2013.pdf

³ Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

⁴ Sowing the seed: Incentives and Motivations for Sharing Research Data, a researcher's perspective, s 10–11.

<http://www.knowledge-exchange.info/Default.aspx?ID=733>

⁵ ReCODE (Policy Recommendations for Open Access to Research Data in Europe) Research & Results

<http://recodeproject.eu/research/>

⁶ FOSTER (Facilitate Open Science Training for European Research) <https://www.fosteropenscience.eu/>

⁷ SPARC Europe <http://sparceurope.org/>

⁸ LERU Advice papers <http://www.leru.org/index.php/public/publications/category/advice-papers/>

⁹ OpenDOAR <http://www.opendoar.org/>

¹⁰ "European Research Council funds arXiv — a taste of changes to come", European Research Council 16.9.2013, <http://erc.europa.eu/european-research-council-funds-arxiv-%E2%80%94-taste-changes-come>

¹¹ OpenAIRE <https://www.openaire.eu/>

¹² LIBER <http://libereurope.eu/>

Tieteessä avoimuus on aina ollut tavoitteena hyvän laadun ja hyvien käytäntöjen edistämiseksi sekä väärennysten ja huonolaatuisen tutkimuksen estämiseksi. Tieteen itsekorjautuva luonne vaatii tutkimuksen läpinäkyvyyttä ja toistettavuutta. Avoimuus parantaa tutkimuksen laatua, koska avoimesti saatavilla olevat tulokset ja tietoaineistot 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 täsmentävistä suosituksista. Opetus- ja kulttuuriministeriön asettama tutkimuseettinen neuvottelukunta (TENK) on yhteistyössä suomalaisen tiedeyhteisön kanssa uudistanut vuonna 2012 tutkimuseettisen ohjeen hyvästä tieteellisestä käytännöstä ja sen loukkausepäilyjen käsittelemisestä¹³.

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.¹⁴ 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 hallintatavan¹⁵ tuloksena tietoja ei asiattomasti muuteta, tiedot eivät vahingoitu tai häviä käsittelyprosessin aikana ja tietojen virheettömyyteen ja laatuun voidaan luottaa. Tutkimuksen tietojenhallinnan laatu ja virheettömyys ovat itseisarvoja, joista ei tule aiheuttomasti 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 sen hallintaa, siksi huolellinen dokumentointi ja kuvailu on tärkeää.

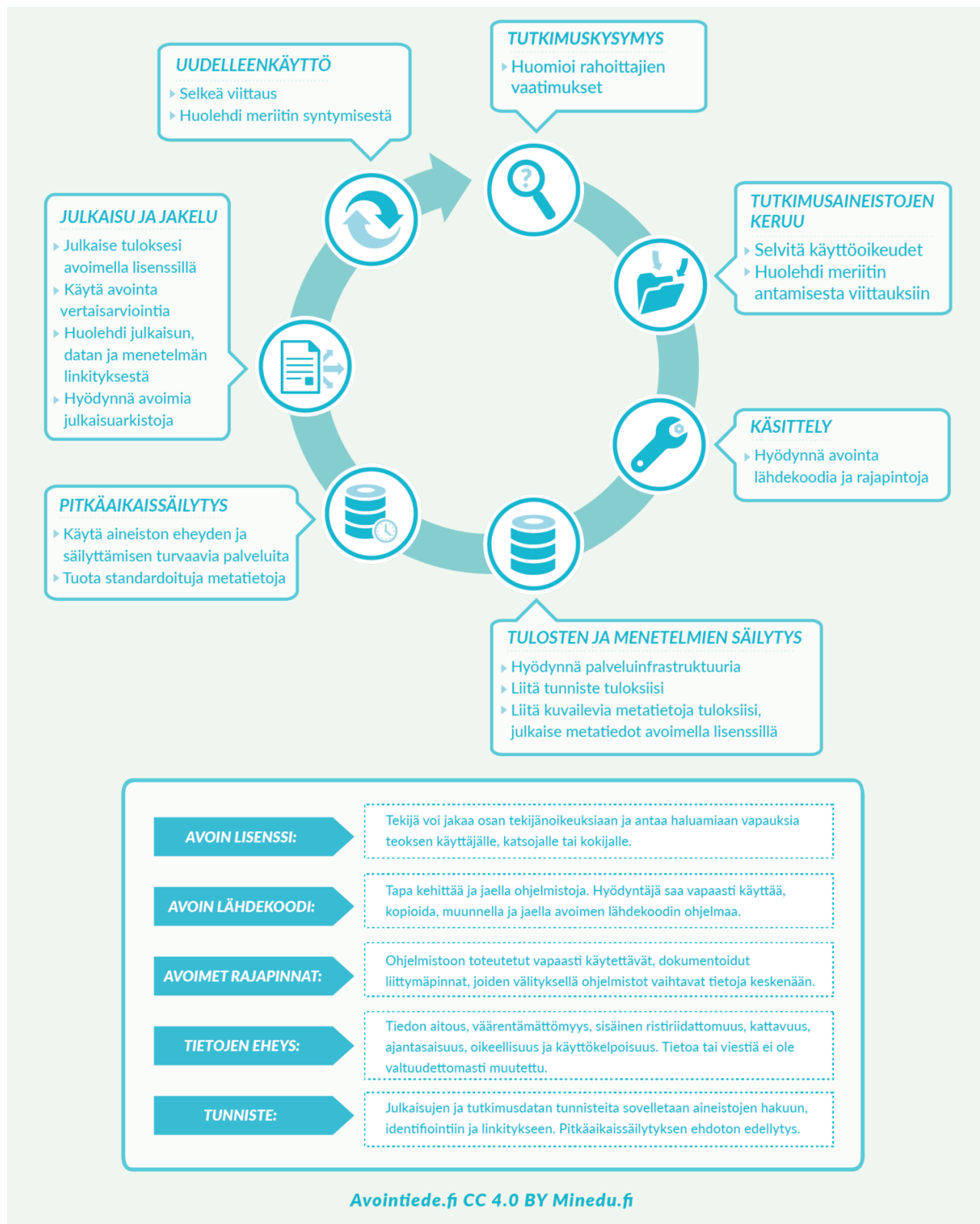
AVOIMUUDEN EDISTÄMINEN TUTKIMUSPROSESSISSA

Tutkimusprosessin kaikissa vaiheissa voidaan toimia avoimuutta edistävasti (Kuva 1).

¹³ Tutkimuseettinen neuvottelukunta. Hyvä tieteellinen käytäntö. <http://www.tenk.fi/fi/htk-ohje>

¹⁴ Hyvän laadun ja arvioinnin ohjeita mm. Helsingin Yliopisto. Laatu ja arviointi. Tutkimus. http://www.helsinki.fi/laatu_ja_arviointi/tutkimus.html

¹⁵ Hyvän tiedonhallintatavan ohjeita: Hyvän tiedonhallintatavan määrittely. Valtiovarainministeriön työryhmämuistioita 11/2000. http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/4127/4128_fi.pdf



Kuva 1. Avoinuuden edistäminen tutkimusprosessin eri vaiheissa.

TUTKIMUSTIEDONHALLINTA

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. 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. Silti on syytä huolehtia siitä, että suunnitelma on 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 korjattava.

Keskeinen osa tiedonhallinnan suunnittelua on tutkimusaineistojen hallinta. Jos se 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 Tutkimuksen tietoaineistot -hankkeessa toteutettu Datanhallinnan opas¹⁶ tai Yhteiskuntatieteellisen tietoarkiston opas¹⁷. 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.

Avointa tiedettä ja tutkimusta tukeva toimintakulttuuri tarvitsee tuekseen rahoittajien ja tutkimusjärjestelmän organisaatioiden selkeitä linjauksia ja hyviä käytäntöjä. Tämä tarkoittaa esimerkiksi sopimuksissa, rahoituksessa ja meritoitumisessa avoimuuden tukemista ja huomioimista. Tutkimusprosessissa vuorottelevat suunnittelu, toiminta ja toiminnan arviointi. Toimintatapojen muutos liittyy paitsi tutkimuksen rakenteisiin ja prosesseihin, myös arvoihin, asenteisiin ja käyttäytymiseen.

Kansallisen tietoinfrastruktuurin palveluvalikoiman tulee rakentua kestäväälle pohjalle ja sen tulee tarjota palveluita datan, 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 ja AVAA sekä korkeakoulujen ja tutkimuslaitosten omat julkaisuarkistot ja tutkimustietojärjestelmät sekä tuleva pitkäaikais säilytysratkaisu. Infrastruktuurien yhteentoimivuudesta on huolehdittava silloinkin kun aineistoja, menetelmiä tai julkaisuja ei voida avata vapaasti käytettäväksi. Löydettävyys ja käytettävyys ovat myös avoimen tieteen tärkeitä piirteitä. Palveluita suunnitellaan kokonaisvaltaisesti yhteistyössä.

TUTKIJAN MUISTILISTA

- Suunnittele tutkimuksen tiedonhallinta huolellisesti jo rahoitushakuvaiheessa ja/tai tutkimussuunnitelmaa tehdessä
 - Mitä aineistoja tutkimukseen liittyy? Muista sekä jo olemassa olevat että syntyvät aineistot
 - Selvitä niihin liittyvät omistus- ja oikeuskysymykset
 - Mitä sopimuksia tarvitaan?

¹⁶ <https://www.tdata.fi/tutkimusdatan-hallinta>

¹⁷ <http://www.fsd.uta.fi/tiedonhallinta/>

- Mitä dokumentaatiota on syytä tehdä aineistosta, sen käsittelystä ja tutkimusmenetelmistä?
- Selvitä rahoittajien, tutkimusorganisaatioiden ja aineistojen omistajien ja/tai informanttien oikeudet, vaatimukset ja suositukset. Etsi sopimus pohjia ja kysy neuvoja tutkimuksen tukipalveluista, tietoarkistoista, kirjastoista yms. Tietoa löytyy myös Avointiede.fi-sivustolta.
- 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 eri tuloksille
- Suunnittele ja kuvaa miten aineiston laatu varmistetaan. Kuvaile prosessit, tehdäänkö esim. vertaisarviointi sille erikseen? Julkaistaanko datapaperi? 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



Kuva 2. Tutkimusaineistojen hallinnan edistäminen

TUTKIMUKSEN TIETOAINIESTOT

Tutkimusaineistoilla tarkoitetaan niitä resursseja, joita tutkija tuottaa tai joita hän käyttää tutkimusprosessin aikana. Tutkimusaineistot voivat olla aineellisia tai aineettomia. Jotta aineisto olisi tutkimukseen kelpavaa, 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 on aina syytä tallentaa esimerkiksi metatietoihin, koodikirjoihin ja/tai muuhun dokumentaatioon. Yhdessä datan kanssa tästä kokonaisuudesta muodostuu tutkimusaineisto. Kun tutkimuksen tuloksia julkaistaan, on syytä aina huolehtia siitä, että myös tutkimusaineistot ovat kunnossa. Tutkimuspolitiikan laatijan,

tuutusorganisaation ja tutkusrahoittajien on huolehdittava siitä, että tarvittavia tukipalveluita ja resursseja on saatavilla luotettavasti ja pysyvästi, mutta tutkijoiden vastuulla on huolellinen suunnittelu, dokumentaatio sekä aineiston saattaminen käytettäväksi parhaaksi katsomallaan tavalla sopimuksia, lakeja ja hyvää etiikkaa noudattaen.

Aineistojen julkaiseminen on hyvän tieteellisen käytännön mukaista, sillä siten tutkimuksen tulokset ovat laajasti verifioitavissa. Aineistoja tai niiden osia saattaa jäädä käyttämättä, ja myös niiden julkaiseminen on suositeltavaa. Rahoittajien olisi hyvä huomioida tämä seikka aineistohallinnan suunnitelmia arvioidessa. Myös aineistojen tuottamisesta ja jakamisesta kertyy meriittä, mikä kannattaa mainita ansioluettelossa. Esimerkiksi TENK:n ansioluettelopohjassa on erillinen kohta ”ansiot tutkimus- ja tietoaineistojen tuottamisessa ja jakamisessa”.

Tietoarkistojen ja tieteellisten kirjastojen palveluun on tärkeä panostaa, jotta tutkimusjärjestelmässämme tuotetut aineistot olisivat mahdollisimman hyvälaatuisia. Tutkimusaineistojen hallinnassa vaaditaan yhteistyötä eri alojen asiantuntijoiden kesken. Tutkijoilla on ratkaisevan tärkeää tietoa omista aineistoistaan, mutta tiedonhallinnan yms. asiantuntijoita tarvitaan myös säilyvyyden, käytettävyyden ja löydettävyyden takaamiseksi. Tutkimuksen tietoinfraan 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 korkeakoulutukseen¹⁸, sillä jo opintojen yhteydessä syntyy aineistoja.

Avoimuuden edistäminen edellyttää tutkimusaineistoihin liittyvien oikeuksien oikeanlaista hallinnointia. Aineistoihin tarvitaan 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.

Eri toimijoiden on otettava huomioon erilaisia asioita eri näkökulmasta. Taulukossa 1 on kerrottu tutkijan ja tutkimusorganisaation näkökulmasta tutkimusaineistojen hallinnasta.

Taulukko 1. Tutkijan ja tutkimusorganisaation näkökulmat tutkimusaineistojen hallintaan

Toimijan näkökulma	Tutkija	Tutkimusorganisaatio
Millä rahoituksella tutkimusta tehdään	<p>Mikäli tutkimuksella ei ole ulkopuolista rahoitusta, tutkija tekee työtä ns. avoimessa tutkimuksessa, eikä aineistojen hallintaa ohjaamassa ole ulkopuolista tahoja. Tällöin riippuu tutkijan oman organisaation tavoista, miten hän saa määrätä aineistoistaan.</p> <p>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.</p>

¹⁸ Sowing the seed, s. 26, 30, Tutkijan kohtaamispaikat. Knowledge Exchange.

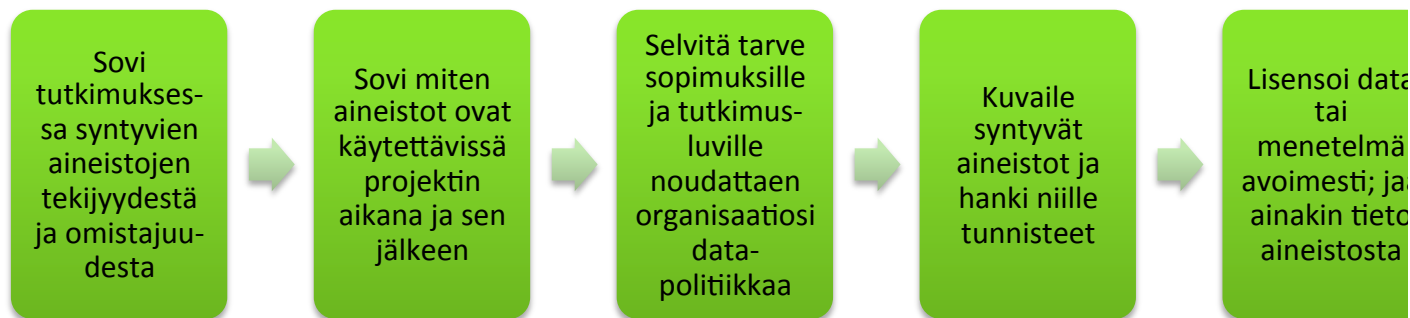
		Mikäli rahoittaja ei aseta mitään tutkimusaineistoja ja tuloksia koskevia ehtoja, organisaation omat ohjeet ja toimintatavat ratkaisevat, miten tutkimusaineistojen avointa saatavuutta edistetään.
Kuka tekee tutkimusta	Jos tutkimusta tehdään yhdessä muiden tutkijoiden kanssa, tulee tekijöiden sopia keskenään, miten kukin tekijä saa käyttää yhdessä aikaansaatuja aineistoja hankkeen aikana ja sen päätyttyä. Tässä yhteydessä on hyvä käydä lävitse kaikkien osapuolten oikeudet, tekijyyttä koskevat periaatteet, vastuut ja velvollisuudet sekä käyttöoikeuksia koskevat kysymykset.	
Miten tutkimuksen aineisto saadaan käyttöön	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. 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.	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.
Aineistojen säilytys ja jako	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).	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).
Kolmansien osapuolien aineistojen käyttö	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 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).	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 siitä, että aineistoja käytetään sopimusehtojen mukaista. 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).

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 erottelemaan 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 velvoitteet 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 tähän.

TUTKIJAN MUISTILISTA

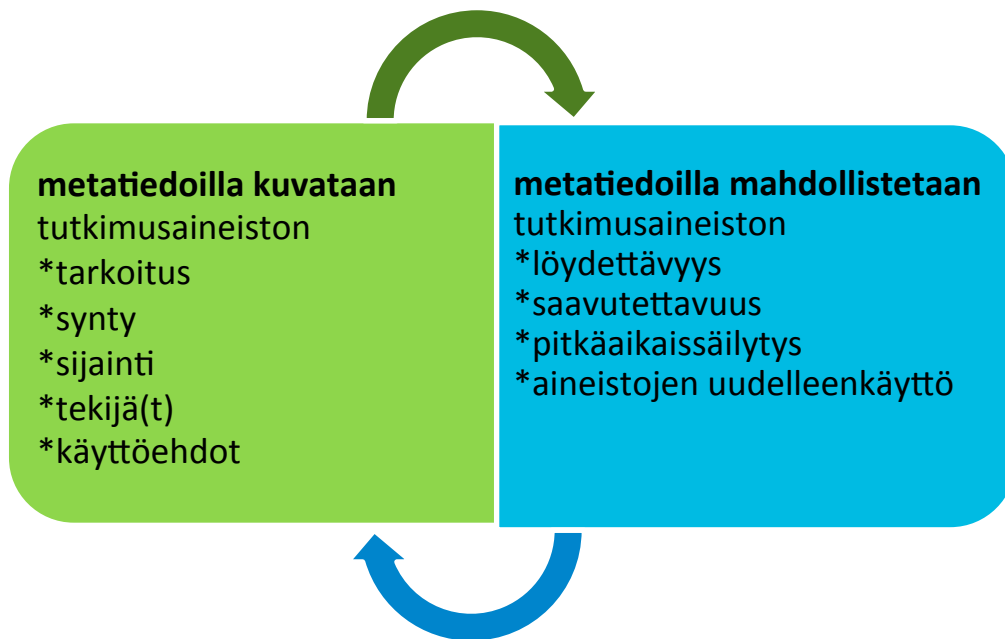
- Tee tarvittavat sopimukset tutkimusaineistoihin liittyen
 - Aineiston tuottajien kanssa
 - Aineiston omistajien kanssa
 - Aineiston säilyttäjien kanssa
 - Tutkijoiden välillä
 - Kustantajien tms. julkaisijoiden kanssa
 - Muut mahdolliset osapuolet
- Huomioi sekä tutkimusaika että aika tutkimuksen jälkeen
- Selvitä miten työsuhteesi ja rahoituksesi vaikuttaa asioihin
- Mitä avoimempiin ratkaisuihin voit päästä, sitä yksinkertaisempaa julkaiseminen yleensä on
- Joskus avaaminen voi silti vaatia paljonkin suunnittelua, sopimista ja esim. anonymisointia. Se voi kuitenkin parantaa tutkimuksen vaikuttavuutta ja laatua huomattavasti
- Huolehdi käyttöoikeuksista, lisenseistä ja metatiedoista
- Huolehdi siitä, että aineistosi on kuvailtu alasi tietoaarkistoon tai tietoaaineistojen hakupalvelu Etsimeen
- Myös hyödyntämättä jääneet tai nollatulokseen johtaneet aineistot kannattaa julkaista



Kuva 3. Tutkimusaineistojen avoimuuden edistäminen

TUTKIMUSAINIESTON METATIEDOT

Aineiston hyödyntämisen lähtökohtana on ymmärtää sopiiko se kyseessä olevaan tutkimuskysymykseen. Tutkimusaineistojen metatiedoilla tarkoitetaan yleisesti tutkimusaineistoa kuvailevaa rakenteistettua tietoa. Metatiedoilla kuvataan yksiselitteisesti tutkimusaineiston tarkoitus, synty, sijainti, tekijät ja käyttöehdot. Metatietojen käyttötarkoitus on mahdollistaa tutkimusaineistojen löydettävyyden, saavutettavuuden, pitkäaikais säilytys ja aineistojen uudelleen käyttö. Siten metatiedoilla on kaksi tarvisijaa: tutkimusaineiston tuottaja, joka antaa tarvittavat tiedot tuottamastaan aineistosta eli siitä tavasta ja kontekstista, jossa se on syntynyt, ja tutkimusaineiston hyödyntäjä, joka arvioi näiden tietojen perusteella mitä aineistoja on saatavilla, tunnistaa etsittyjä aineistoja, osaa valita tarpeitaan vastaavan aineiston ja hankkii tarvitsemansa aineiston eli saa tiedon sen sijainnista.



Kuva 4. Metatietojen tarkoitus.

Tutkimusala ja sen vallitsevat käytännöt 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. Tutkimusaineiston metatieto muodostuu metatiedon rakenteesta ja rakenteen mukaisesti määritellystä sisällöstä. Metatiedot voidaan jaotella kolmeen päätyyppiin käyttötarkoituksen mukaan. Kuvaileva metatieto kuvaa tutkimusaineiston 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. Hallinnollinen metatieto määrittelee tutkimusaineiston hallintaan tarvittavat hallinnolliset ja tekniset tiedot, esimerkiksi tutkimusaineiston käyttöehdot ja aineiston ja tulosten pitkäaikaissäilytykseen tarvittavat tekniset tiedot. Rakenteellinen metatieto kuvaa tutkimusaineiston rakennetta ja sen järjestystä, esimerkiksi aineiston eri osien suhdetta toisiinsa. Tutkimusaineiston hallinnolliset ja rakenteelliset metatiedot tuotetaan pääosin ohjelmallisesti.

Tutkimusaineistojen metatietojen perimmäinen käyttötarkoitus on siis vastata tutkimusaineistoihin liittyviin kysymyksiin. Taulukossa 2 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.

Taulukko 2. Tutkimusaineistoihin liittyviä kysymyksiä ja niihin vastauksen antavia metatietoja.

- **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 käytetty?	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)

Sovittujen käytänteiden noudattaminen, esimerkiksi kontrolloitujen asiansastojen käyttö kuvailussa ja 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 uudelleenkäytön kannalta. Laadukas metatieto on tutkimusaineiston käyntikortti, jonka tutkija voi antaa tutkimusaineistostaan siitä kiinnostuneille.

TUTKIJAN MUISTILISTA

- Huolehdi kaikkien aineistojesi ja julkaisujesi mahdollisimman laajasta kuvailusta
- Aikaan ja paikkaan liittyvät tiedot ovat myös tärkeitä, ilmoita ne ohjeiden mukaan
- Käytä oman alasi terminologiaa ja asiansanoja
- Muista myös vapaamuotoinen ja ainakin osittain yleistajuinen kuvaus: mistä on kysymys? (description)
- Suosi sanastoja, ontologioita ja yleisessä käytössä olevia termejä löytyvyyden takaamiseksi
- Lisenssi/käyttöehdot ovat tärkeä metatieto, koneluetettava formaatti kuten CC-lisenssi on suositeltava

- Muuttujakohtaiset metatiedot ja niiden kuvailu on hyvä tarkistaa yhdessä alan tietoarkiston kanssa
- Huomio aineistojen historia- ja versiotiedot yksiselitteen viittaamisen mahdollistamiseksi
- Esimerkiksi kirjastoista ja tietoarkistoista löytää kuvailun ammattilaisia, joilta kannattaa pyytää apua ja vinkkejä

TUNNISTEET JA VIITTAUKSET

Jotta tutkimustieto olisi mahdollisimman käytettävää ja löydettävää on syytä kiinnittää erityistä huomiota tietorakenteisiin ja semantiikkaan. Erilaisten aineistojen ja julkaisujen kohdalla on olemassa erilaisia käytäntöjä, jotka ovat peräisin eri tutkimus- ja julkaisuperinteistä. Avoimen tieteen järjestelmien ja käytäntöjen tulee huomioida eri yhteisöjen tarpeet, mutta myös tavoitella mahdollisimman hyvää yhteentoimivuutta ja löytyvyyttä. Kestävä tutkimus vaatii esimerkiksi pysyviä yksilöiviä tunnisteita. Vain näin voidaan turvata aineistojen ja viittausten säilyvyys pitemmällä aikavälillä. Sähköisessä ympäristössä myös jatkuvat päivitykset ja uusien versioiden syntyminen vaatii kaikilta toimijoilta suunnitelmallisuutta ja kestävyuden huomioimista tutkimuksen uskottavuuden vuoksi.

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. Lisäksi on huolehdittava siitä että tieto julkaisuista ja tietoaineistoista kulkeutuu myös omaan tutkimustietojärjestelmään sekä Etsin-palveluun riittävien kuvailu- ja tunnistetietojen kanssa.

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. URI-nimet (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.

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. Tutkimusdatalle 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 näin tiedonhakuja. Tunnisteet ovat osa metatietoa ja osa metatietoa hallinnoivia välineitä. Ontologioiden, rekistereiden ja auktoriteettipalveluiden avulla voidaan luoda tunnisteita kaikille keskeisille metatiedoille

Tietolähteitä ovat tyypillisesti olleet tutkimusjulkaisut, mutta myös tietoarkistoihin tallennettu tutkimusdata on sähköinen lähde ja menetelmänä käytetty sovellus kaipaavat viittaamista uskottavuuden ja toistettavuuden varmistamiseksi. Arkistoihin tallennetaan yleensä varsinainen tietoaineisto (esimerkiksi datatiedosto(t), koodikirja, kyselylomake jne.) sekä tietoaineistoa kuvailevaa metadataa. Näihin on mahdollista viitata kun niille on annettu pysyvä tunniste kuten URN- tai DOI-tunniste. Tutkimusdataan viitattaessa tärkeitä yksilöiviä tietoelementtejä ovat tekijä, tietoaineiston nimi, versio, tietoarkiston antama aineiston tunniste (useimmiten aineistonumero), keruuajankohta, kerääjä-, tuottaja- ja jakajatiedot sekä mahdollinen pysyvä tunniste. Lähde- ja tekstiviitteiden tekoa ohjaa kansallinen standardi SFS 5989 Lähde- ja tekstiviitteitä koskevat ohjeet. Monilla korkeakouluilla on em. standardiin pohjautuvia viittausohjeistoja, ja lisäksi tutkimusdataan viittaamista¹⁹ ohjeistaa Yhteiskuntatieteellinen tietoarkisto.

TUTKIJAN MUISTILISTA

- Hanki itsellesi tutkijan ORCID-tunnus ellei sinulla ole sellaista ja käytä sitä joka yhteydessä
- Huolehdi tunnisteiden käytöstä aina kun viitaa
- 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!

AVOIN TUTKIMUSPROSESSI

Sähköisessä maailmassa myös tutkimusprosessin tehostamiseen voi käyttää avoimuutta. Tieteelliseen tutkimusprosessiin liittyvää vuorovaikutus ennen tulosten julkaisemista on aiemmin useimmiten ollut rajattuna seminaareihin ja konferensseihin. Kirjapainoon nojaava kulttuuri on ohjannut tutkimusprosessin muotoutumiseen valmiiden painettujen tulosten julkaisemiseen tieteellisen tiedon manifestaationa. Ennen julkaisua vuorovaikutus on ollut rajoitettua tiedeyhteisön sisäiseksi asiaksi. Tänään kuitenkin 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 levittämiseksi 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, blogeja, 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.

¹⁹ Yhteiskuntatieteellinen tietoarkisto. Aineistoon viittaminen.

<http://www.fsd.uta.fi/fi/aineistot/jatkokaytto/viittaaminen.html>

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 julkaisu poissulje jatkojulkaisua vertaisarvioidussa lehdessä jos sellainen on suunnitelmassa. Niiden 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. 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.

TUTKIJAN MUISTILISTA

- 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 hallinnoit 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. huolelta moneen kertaan eri kanavissa
- Muista myös kuvien, videoiden, blogitekstien yms. lisenssit. Mitä enemmän CC-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?
- Muista, että vaikka kaikkea tietoa ei voi jakaa tai julkaista, sen olemassaolosta ja ominaisuuksista on silti tärkeä kertoa!

TULOSTEN 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 seurut ja kansainvälisellä tasolla myös kaupalliset kustantajat.

Vuosituhanen vaihteen jälkeen näkyvään rooliin nousut Open Access –liike (OA) on pyrkinyt löytämään vaihtoehtoja aiemmalle tilausmaksuihin ja julkaisujen rajoitettuun saatavuuteen perustuvalla toimintamallille. 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 ArXiv²⁰. 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 rektorit 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.²¹

TUTKIJAN NÄKÖKULMA

Periaatteessa kustantajien ylläpitämä kultainen open access, johon liittyvät kirjoittajamaksut, 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

²⁰ ArXiv-arkiston verkkosivu: <http://arxiv.org/>

²¹ Jyrki Ilva & Johanna Lilja: Kotimaiset tieteelliset lehdet ja avoin julkaiseminen. Selvitys mahdollisista rahoitusmalleista. 2014. <http://urn.fi/URN:ISBN:978-952-5995-09-1>

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 suosittelvat 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.

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. Julkaisun 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 omissa organisaatioissa tarjolla neuvontaa ja tukea julkaisemiseen liittyen, ja näitä tukipalveluita kannattaa hyödyntää.

JULKAISUARKISTOJEN NÄKÖKULMA

Julkaisuarkiston sisältämissä tietueissa pitäisi pystyä ilmaisemaan yksiselitteisellä ja yhdenmukaisella tavalla, liittykö tietueeseen kuvailun 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 määritellyyn sanastoon/ontologiaan perustuva tieto kuvailun 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.

Avoin tiede ja tutkimus-hankkeen julkaisuryhmän alustavia suosituksia tutkimusorganisaatioille:

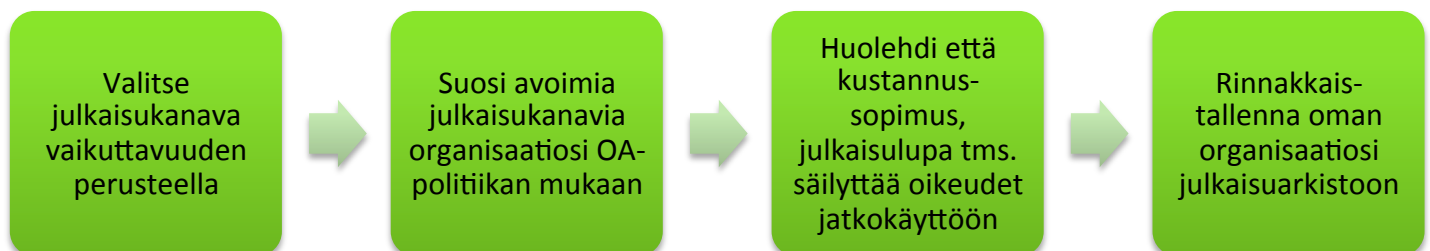
Tutkimusorganisaatiot

- 1 Vastaavat rinnakkaistallentamisen prosessien ja käytäntöjen tuomisesta osaksi tutkijoiden ja organisaation perustoimintaa.
- 2 Seuraavat rinnakkaistallentamisen kehittämisen vaikutusta julkaisujen avoimeen saatavuuteen organisaatiossa.
- 3 Kehittävät keskitettyjä palveluja rinnakkaistallentamisen käytännön toteuttamiseen.
- 4 Käyttävät avoimien aineistojen yhteydessä koneluettavia lisenssejä, kuten CC-lisenssi.

- 5 Kehittävät keskeytetyt palvelut ja prosessit tutkimusaineistojen hallinnan ja avoimen julkaisemisen mahdollistamiseksi.
- 6 Kannustavat tutkijoitaan julkaisemaan tutkimustuloksia myös laadukkaissa Open Access -julkaisufoorumeissa.
- 7 Kehittävät Open Access -lehtien kirjoittaja- ja tukimaksujen hallintaan ja käsittelyyn keskitettyjä palveluja ja rahastoja.
- 8 Osoittavat julkaisuarkistojensa kehittämiseen riittävät resurssit.
- 9 Huolehtivat tutkimustietojärjestelmien, julkaisuarkistojen ja data-arkistojen OpenAIRE-yhteentoimivuudesta (<https://guidelines.openaire.eu>).
- 10 Tutkivat mahdollisuudet linkittää rinnakkaistallentaminen julkaisutietojen keruu- ja tallennusprosessiin.
- 11 Kehittävät menetelmiä ja käytäntöjä avoimesta julkaisemisesta palkitsemiseen.

TUTKIJAN MUISTILISTA

- Huolehdi tutkimusjulkaisujesi mahdollisimman suuresta avoimuudesta.
- Jos haluat patentoida työsi tulokset, tee se ennen julkaisemista.
- Rinnakkaistallentaminen ei poista tekijänoikeutta
- Julkaisemalla avoimen jatkokäytön sallivalla lisenssillä (kuten CC BY 4.0) edistät mahdollisuuksia vapaasti jakaa ja hyödyntää julkaisuasi
- Muista että avoin julkaiseminen ei sinänsä poista tekijänoikeutta, mutta lisää näkyvyyttä.
- 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.
- Kysy neuvoja esimerkiksi kirjastosta kun teet kustannussopimuksia.
- Muista tunnisteet ja linkitykset, käytä aina niitä ahkerasti.
- Huolehdi hyvästä kuvailusta, pyydä neuvoja kirjastosta.
- 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.
- Huolehdi rinnakkaistallentamisesta sopivaan julkaisuarkistoon. Tarvittaessa voi käyttää useaa, se lisää näkyvyyttä (esim. kansainvälinen alakohtainen ja organisaation oma julkaisuarkisto).
- Myös harmaan aineiston julkaiseminen julkaisuarkistossa on hyvä vaihtoehto, silloin saa pysyvän tunnisteiden, linkin ja säilytyksen helposti järjestettyä. Esimerkiksi konferenssien verkkosivut eivät valitettavasti aina ole ikuisia. Harmaiden julkaisujen kohdalla tulee aina varmistaa etteivät ne poissulje jatkajulkaisua vertaisarvioidussa lehdessä jos sellainen on suunnitelmassa.
- Valitse mahdollisimman avoin lisenssi (miehellään CC-lisenssi).
- 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.



AVOIMUUDET POLITIIKAT

Tämä luku on osoitettu lähinnä tutkimusorganisaatiolle. 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.

TUTKIMUSORGANISAATION MUISTILISTA

- Laaditaan selvät suunnitelmat ja tavoitteet toimintakulttuurin kehittämiseksi
- Immateriaalioikeuksiin liittyvät omistajuudet sovitaan selkeästi. Lisensoimista suositetaan mahdollisuuksien mukaan. Suomen julkisen avoimen datan virallinen lisenssisuitus 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

PALVELUIDEN RAHOITUS JA HINNOITTELU

Suomalaisen tutkimusjärjestelmän toimijat käyttävät julkisesti rahoitettujen infrastruktuurien peruspalveluja maksutta. Lisäpalvelut voidaan hinnoitella räätälöidysti, esim. määritellen 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. Kaupallisten

toimijoiden mukanaolo sekä rahoittajina että käyttäjinä lisää infrastruktuurien käyttöastetta, mikä osaltaan kannustaa niiden ylläpitämiseen ja kehittämiseen.

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 (engl. *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 (engl. *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. <http://www.fsd.uta.fi/tiedonhallinta/osa7.html>

Altimetriikka

Huomioimalla muutakin kuin akatemian sisäistä tieteellistä toimintaa, voidaan tutkimuksen yhteiskunnallisen vaikuttavuuden arviointia kehittää. Altimetriikassa yritetään mitata näitä asioita, kuten julkaisemista suurelle yleisölle tai bloggaamista tai muuta aktiviteettia sosiaalisessa mediassa. Sähköinen ympäristö antaa myös mahdollisuuden tilastoida aineistojen latauskertoja, jakamisia tai suosituksia. Tämän tyyppisellä kommunikaatiolla voi olla vaikuttavuuden kannalta suurtakin merkitystä, ehkä tieteenalastakin hieman riippuen.

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 (engl. *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 julkaisuutuotannosta. Kansallisten sääntöjen mukaiset hakumuodot tallennetaan auktoriteettitietueisiin yhdessä kohdetta koskevien ID- tunnisteen (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 (engl. *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. Voit kokeilla eri ehtojen yhdistelmiä lisenssivalitsimella²². ATT-hanke suosittelee CC0:n, CC 4.0 BY:n tai tarvittaessa muiden yleisesti tunnettujen sopimuslisenssien käyttöä.

Avoimet rajapinnat (engl. *Open interface*)

Avoimilla rajapinnoilla tarkoitetaan sellaisia ohjelmistoon toteutettuja vapaasti käytettäviä, huolellisesti dokumentoituja liittymäpintoja, joiden välityksellä ohjelmistot vaihtavat tietoja keskenään.

Avoimet standardit (engl. *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 (engl. *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 (engl. *Open data*)

Avoimella datalla tarkoitetaan tutkimukselle, julkishallinnolle, organisaatioille, yrityksille tai yksityishenkilöille kertynyttä jalostamatonta informaatiota, joka on avattu ulkopuolisillekin vapaasti ja maksutta hyödynnettäväksi.

²² <http://creativecommons.fi/lisenssit/>

Avoim julkaiseminen (engl. *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.

Avoim lähdekoodi (engl. *Open source*)

Avoim 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.

Avoim tiede (engl. *Open Science*)

Avoim 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. Avoim tiede sisältää käytäntöjä, kuten tutkimusjulkaisujen avoimen saatavuuden (open access) edistäminen, tutkimusaineistojen avoim julkaiseminen, avoimen lähdekoodin ja avoimien standardien hyödyntäminen, sekä tutkimusprosessin julkinen dokumentointi niin kutsutun avoimen muistikirjan menetelmällä.

Avoim tieto (engl. *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, yhteistyöverkoston 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 *Altimetriikka*. 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 aineistonhallinnan ohjeistus.

Digitaalinen informaatio

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

Digitointi

Analogisessa muodossa olevan aineiston (esim. paperisen dokumentin) muuntaminen digitaaliseen muotoon.

DOAJ Directory of Open Access Journals

Kansainvälinen rekisteri, josta löytyy tietoja avoimista tieteellisistä lehdistä. <http://doaj.org/>

E-infrastruktuuri

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

Eheys (engl. 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

(Tietoaaineiston) Elinkaari

Tietoaaineiston olemassaolo kokonaisuutena alusta loppuun. Tietoaaineiston 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 (ESFRI)

EU-jäsenmaiden ja assosiaatiomaiden yhteistyöelin tutkimusinfrastruktuureja koskevan politiikan valmistelussa ja määrittelyssä. ESFRI koostuu tutkimusministeriöiden ja rahoittajien edustajista.

Embargo (engl. Holding period)

Karenssiaika. Rajoitus, jonka mukaan aineiston vapaa käyttö alkaa vasta säädetyn odotusajan jälkeen esimerkiksi julkaisun ilmestyttyä.

Ensikäyttö/primäärikäyttö (engl. Primary use)

Ensikäytöllä tarkoitetaan alkuperäisessä tutkimussuunnitelmassa ilmaistua tai muuta ensisijaista tutkimusaineiston käyttötarkoitusta.

Esiystieto (engl. *Representation Information*)

Osa säilytettävään informaatioon liitettävää metatietoa. Esiystiedon 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 tietosuovaatimuksen.

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.

Indikaattori

Osoitin, ilmaisin 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.

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 (engl. *Harvesting*)

Metatiedon automaattinen poiminta sovelluksesta toiseen, esimerkiksi jonkun tietoarkiston luettelosta Etsin-palveluun. Haravoinnissa sovelletaan yleensä jotakin rajapintastandardia kuten OAI-PMH protokollaa.

Harmaat julkaisut (engl. *Grey publications*)

Tieteelliset julkaisut, joita ei varsinaisesti ole vertaisarvioitu ja jotka voivat olla enemmän tai vähemmän valmiita. Tällaisia ovat esimerkiksi artikkeleiden 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 merkinnöt, 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.

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

Julkaiseminen

Julkaisuarkisto

Tieteellisten julkaisujen tallentamiseen ja avoimeen verkkojulkaisemiseen soveltuva tekninen järjestelmä ja sen ympärille rakennettut 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. Julkaisukanavien arviointi tapahtuu 23 tieteenalakohtaisessa asiantuntijapaneelissa. Julkaisufoorumi toimii Tieteellisten seurain valtuuskunnan yhteydessä. Hanke 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.

Julkisen 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 (engl. *Citizen Science*)

Laajan yleisön tekemää joukoistettua 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 asiansharrastajien 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*, 17.2.2013.
<http://filosofianpuutarha.blogspot.fi/2013/02/elakoon-kansalaistiede.html>

Keskustelupaperi (engl. *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. Parhimmillaan 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 otsikkoa 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 (engl. *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 (engl. *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 (engl. *Reference Information*)

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, kuvaluohjeita ja metatietomalleja) jotta löydettävyys olisi hyvä. Kannattaa aina käyttää sanastoja ja ontologioita ja noudattaa ohjeistusta tarkasti esimerkiksi nimimuotojen tai päivämäärien ilmaisussa.

Käsikirja

Tämä 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 auttamaan avoimen tieteen ja tutkimuksen hyödyntämisessä ja edistämässä. Käsikirja käsittää ATT:n keskeisiä lähtökohtia, avoimen tieteen ja tutkimuksen edistämisen ja toteuttamisen, kansainvälisen tilanteen ja keskeisiä käsitteitä.

Käytettävyys (engl. *Usability*)

Tietojärjestelmien käytettävyys määritellään seuraavasti: Valtiohallinnon tietoturvallisuuskäsitteistön (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 määrittyy aina suhteessa tutkimustehtävään, mutta edellyttää vähintään tietoja aineiston alkuperästä.

Käyttöoikeuksien hallinta

Joskus aineistoihin pääsyä on rajoitettava esimerkiksi tietosuojasyistä. Jonkun on silloin vastattava siitä, kenelle erilaisia oikeuksia aineistoihin annetaan (lukeminen, muokkaminen, 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 käyttäjätunnus ja salasana. 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 (engl. *Quality system*)

Laatujärjestelmä koostuu koulutuksen, tutkimuksen, yhteiskunnallisen vuorovaikutuksen ja vaikuttavuuden, henkilöstövoimavarojen, palvelujen ja johtamisen laatua varmistavista menettelyistä ja prosesseista.

Linkitetty data (engl. *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 käyttä samankaltaista tekniikkaa nimellä Knowledge Graph. Tunnisteita sisältävien linkkien käyttö parantaa löydettävyyttä ja näkyvyyttä.

Linkitetty data voi olla avointa dataa, jolloin puhutaan avoimesta linkitetystä datasta (engl. open linked data).

Lokitieto (engl. *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 selvitä esimerkiksi järjestelmän virhetilanteet, yhteydenotot tietokoneelta Internetiin sekä tietokoneelle Internetistä tulleet yhteydenottopyynnöt.

Löydettävyys (engl. *Findability*)

Löydettävyydellä tarkoitetaan usein sitä, miten hyvin tietty aineisto löytyy hakukoneen kautta. Usein halutaan edistää oman sisällön tai aineiston näkyvyyttä käyttämällä hakukoneoptimointia (engl. *Search Engine Optimization*). Tätä voi tehdä esimerkiksi huolehtimalla hyvästä linkittämisestä ja oikeiden avainsanojen käyttämisestä osoitteissa, otsikoissa tai kuvateksteissä. Löydettävyyden 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) (engl. *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 (engl. *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.

Paikkatieto (engl. *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.

OpenDOAR

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

Paikkatietoinfrastruktuuuri

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 (engl. *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 (engl. *Preprint*)

Preprintin tavoite on tuoda tutkimuksen alustavat tulokset tutkimusyhteisön arvioitavaksi. Tulokset julkaistaan yleensä muussa julkaisukanavassa myöhemmin. Ei vertaisarvoitu 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 (PID, engl. *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. Resolverpalvelun avulla tunnisteet voivat toimia toiminallisina 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 käyttäjäryhmille. Pääsynhallintaan kuuluu myös aineiston käyttötapahtumien rekisteröinti.

Raakadata

Raakadata on jalostamaton 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 digitoitun 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 määrittellä 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 tietoaaineistoja, joita kerätään rekistereihin ja tietojärjestelmiin. Lisäksi tilastotoimessa syntyy tilastollisia rekistereitä.

Rekisteritiedot

Rekisteritiedot ovat rekisteriviranomaisten ylläpitämää tietoaaineistoa tietystä, laissa määritellystä 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 veloitteita 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 kliinisiä ja näyteaineistoja) täydentävinä tietoina.

Saatavuus (engl. *Availability*)

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.

(Avoin) 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 huipputason 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 (engl. *Crowdsourcing*)

Käyttäjien tai muun laajemman joukon tai yleisön ottaminen mukaan aineiston jalostamiseen tai tuottamiseen. Tämä voi tapahtua verkon kautta tai muuten.

Tarkistussumma (engl. *Check sum*)

Informaatiosta (merkkijonosta tai tiedostosta) soveltuvaa matemaattista algoritmia käyttäen laskettu lukuarvo tai merkkijono, jonka avulla voidaan tunnistaa informaatiosta 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 keskittyminen

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 (engl. *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 (engl. *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 edistämässä 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.

Tietoaineistojen uudelleen hyödyntäminen

Uudelleen hyödyntämisellä tarkoitetaan sitä, että yksityiset ihmiset ja oikeushenkilöt voivat kopioida, muokata, julkaista ja levittää tietoaineistoja eri tarkoituksiin. Suurta osaa julkisin varoin kerätystä raakadatatista ja tutkimusaineistoista voidaan hyödyntää ja hyödynnetään niin tutkimuksessa kuin erilaisten arkipäivän tuotteiden ja palvelujen kehittämisessä.

Tietoaineiston hallintasuunnitelma

Hallintasuunnitelma ja –mekanismi, joka ohjaa 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 (engl. *Data Archive, Data Center*)

Organisaatio tai toiminta jossa ammattimaisesti kerätään, säilytetään, hoivataan ja tehdään saatavaksi tutkimuksen tietoaineistoja.

Tietoinfrastrukturi

Tietoinfrastrukturi on e-infrastruktuurin 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 (engl. *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, tietosuojakäytäntöihin, hinnoitteluun kuin saatavuuden ja käyttöehtojenkin määrittelyyn.

Tietosuoja

Ihmisten yksityisyyden suojeleminen. Sillä tarkoitetaan yksilöä koskevien tietojen suojaamista oikeudettomalta käytöltä henkilötietoja käsiteltäessä.. Aineistohallinnassa otetaan huomioon henkilön yksilöivän tiedon käsittely oikeudellisia säännöksiä noudattavin periaattein ja toimintakäytännöin. Lisää tietoa Tietosuojavaltuutetun toimiston ”Rekisteritutkimuksen tietosuojaoapas tutkijoille ja tietopyyntöjä käsitteleville viranomaisille

http://www.tietosuoja.fi/material/attachments/tietosuojavaltuutettu/tietosuojavaltuutetuntoimisto/oppaat/6Jf qBT5i3/Rekisteritutkimuksen_tietosuojaoapas_tutkijoille_ ja_tietopyyntoja_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, Terveyden 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 (*engl. Persistent Identifiers*)

Julkaisujen ja tutkimusdatan tai muiden entieettien tunnisteita sovelletaan muun muassa näiden aineistojen hakuun, identifiointiin ja linkitykseen. Aineistojen tunnisteet ovat lisäksi pitkäaikaissäilytyksen ehdoton edellytys. Tutkimusjulkaisujen tunnisteena voidaan käyttää julkaisutyyppistä riippuen esimerkiksi ISBN:ää (monografiat) tai erilaisia pysyviä tunnisteita (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 tunnisteita 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. URI-nimet (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 (linkki), OID (linkki), ontologiat (linkki), auktoriteettitiedot (linkki).

Turmeltuminen (*engl. 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 seurantaan, 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. Pörssiyritykset käyttivät liikevaihdostaan noin kaksi prosenttia tutkimus- ja tuotekehitykseen vuonna 2010.

Kansainvälinen yhteistyö. Suomi on mukana keskeisissä kansainvälisissä tutkimusyhteistyöelimissä 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 kuvaliemiä 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 tietoturvasäilytyskäsitteistö 1/2000).

Työpaperi (engl. *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 (engl. *Workflow*)

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

Uutuusvahti (engl. *Alert*)

Ohjelmiston komponentti, joka toistaa ennalta määritellyn haun halutuista väliajoin halutuista tietokannoista, ja ilmoittaa uutuuksista haluttaessa sähköpostilla tai jollain muulla sopivalla teknisellä ratkaisulla.

Verkkopalvelu

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

Vertaisarviointi (engl. *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.

Viitetieto (engl. *Reference Information*)

Osa säilytettävään informaatioon liitettävää metatietoa. Se on aineiston pysyvä tunniste, bibliografinen tieto tai muu vastaava tieto.

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

Viittaaminen (engl. *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.

Visualisointi

Tietoaineiston havainnoillistaminen graafisesti.

XML (*Extensible Markup Language*)

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 (engl. *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.

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.

Viides vapaus

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

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.