

TTA-minimimetatietomalli

Työryhmän jäsenet

- Mari Kleemola, Yhteiskuntatieteellinen tietoarkisto, puheenjohtaja
- Panu Kalliokoski, CSC Tieteen tietotekniikan keskus, sihteeri 8/2013 asti
- Mika Wahlroos, CSC Tieteen tietotekniikan keskus, sihteeri 9/2013 alkaen
- Eero Mikkola, Metsäntutkimuslaitos
- Hanna Koivula, Helsingin Yliopisto
- Jaana Kilkki, Kansallisarkisto
- Juha Hakala, Kansalliskirjasto
- Kimmo Koskinen, Helsingin Yliopisto
- Osma Suominen, Aalto-yliopisto / Kansalliskirjasto
- Pekka Henttonen, Tampereen yliopisto
- Pekka Järveläinen, CSC Tieteen tietotekniikan keskus
- Saija Ylönen, Tilastokeskus

Versiohistoria

- versio 1.0, 4.12.2012: TTA-ohjausryhmän hyväksymä versio
- versio 1.1, 20.5.2013: Päivitykset sanastosuosituksen perusteella
- Versio 1.2, 9.12.2013: TTA-ohjausryhmän 9.12. hyväksymä versio. Tekijän taustaorganisaatio on pakollinen tieto, jos tekijä on henkilö. Käyttöoikeuksien muotoilutapa jätetty avoimeksi odottamaan mm. KDK:n vastaavia suosituksia.

Lisenssi

Creative Commons Suomi CC-BY-NC-SA 3.0

Metatietomallin käyttäjälle annetaan oikeus jakaa eli kopioida, levittää, näyttää ja esittää teosta sekä valmistaa muutettuja teoksia seuraavilla ehdoilla:

- Opetus- ja kulttuuriministeriö nimetään teoksen tekijäksi (ei kuitenkaan siten, että ilmoitus viittaisi lisenssinantajan tukevan lisenssinsaajaa tai teoksen käyttötapaa).
- Epäkaupallinen osapuoli eli käyttäjä ei saa käyttää teosta kaupallisesti.
- Mikäli käyttäjä tekee muutoksia tai käyttää teosta omien teostensa pohjana, tulee johdannaisteos jakaa samalla tai saman kaltaisella lisenssillä.

Sisällysluettelo

TTA-minimimetatietomalli.....	1
Työryhmän jäsenet.....	1
Versiohistoria.....	1
Lisenssi.....	1
Johdanto.....	2
Tutkimusaineiston metatieto.....	2
Tavoitteena yhteentoimivuus.....	2
Minimimetatieto on pienin yhteinen nimittäjä.....	3
Uudet ja vanhat aineistot.....	4
Yhteenveto minimimetatietomallista.....	4
Hajautettujen metatietojen käsittely.....	4
Historia- ja muutostiedot.....	5
Minimimetatietomallin elementtien tarkat kuvaukset.....	5
1. Metatiedon tunniste.....	5
2. Metatiedon muokkaamisaika.....	6
3. Aineiston tunniste.....	6
4. Muokkaamisaika.....	7
5. Jakelijan yhteystieto.....	7
6. Aineiston nimi.....	8
7. Käyttöehdot.....	8
8. Aihe.....	9
9. Toimija.....	9
10. Aineiston kieli.....	10
Liite 1. Metatietomallin vastaavuus yleisiin metatietostandardeihin.....	12
Liite 2. Metatietotyöryhmän kartoittamat metatietostandardit.....	12
Liite 3. Minimimetatietomallin mukainen esimerkkikuvailu.....	13

Johdanto

Tutkimusaineiston metatieto

Ilman kuvailevaa tietoa eli metatietoa tutkimusaineisto on vain merkityksetön kokoelma yksittäisiä tiedostoja, numeroita tai merkkejä. Laadukkaat ja ajantasaiset metatiedot ovat edellytys tutkimusaineiston elinkaaren hallinnalle, aineiston löytymiselle ja tehokkaalle (jatko)hyödyntämiselle. Metatietoa tarvitaan sekä ihmistä että tietojen koneellista käsittelyä varten.

Tutkimusaineistolla tarkoitetaan tässä ensisijaisesti alun perin tutkimuskäyttöä varten kerättyä sähköistä tietoa.

Tutkimusaineisto voi koostua esimerkiksi yksittäisestä tiedostosta, useammasta tiedostosta tai vaikkapa tietokannasta. Tutkimusaineiston kuvailija päättää, minkälainen kokonaisuus muodostaa kuvailtavan tutkimusaineiston. Yleisesti ottaen tutkimusaineiston tulisi olla sellaisenaan hyödynnettävissä oleva tietokokonaisuus. Tutkimusaineisto kannattaa kuvailla pienimpinä kokonaisuuksina, joina se on vielä itsenäisesti hyödynnettävissä.

Tavoitteena yhteentoimivuus

Työryhmän työn perustana ovat olleet kansainväliset ja kotimaiset suositukset ja standardit, ja työryhmä tekee yhteistyötä alan kotimaisten hankkeiden kanssa (esimerkiksi KDK, AHAA ja VM metatiedon esiselvityshanke). Tavoitteena on välttää päällekkäisyydet ja varmistaa eri järjestelmien yhteentoimivuus sekä tietojen yhteismitallisuus sekä kansallisesti että kansainvälisesti. Tiedot tulee kirjata vain yhteen paikkaan ja jo olemassa olevia rakenteita hyödyntäen. Esimerkiksi julkaisujen tietoihin viitataan linkittämällä aineiston tiedoista julkaisun

tietoihin julkaisutietokannassa, eikä julkaisun täysiä tietoja viedä TTA-metatietomalliin.

Tutkimusaineistojen metatiedolle on olemassa erilaisia malleja, suosituksia, formaatteja ja standardeja. Tähän mennessä TTA-metatietotyöryhmä on listannut yli 30 erilaista eri formaattia. Suurin osa niistä on kansainvälisiä ja kehitetty yksittäisen tieteenalan tarpeeseen tai tietyn aineistotyyppin kuvailuun. Ne eroavat toisistaan muun muassa siinä, miten yksityiskohtaisesti tutkimusaineiston voi niiden avulla kuvailla ja siinä, kuinka vakiintuneesti niitä ylläpidetään. Tietyt, tutkimusaineistoa hyvin yleisellä tasolla kuvailevat peruselementit toistuvat samankaltaisina eri formaateissa, ja myös tässä minimimetatietomallissa.

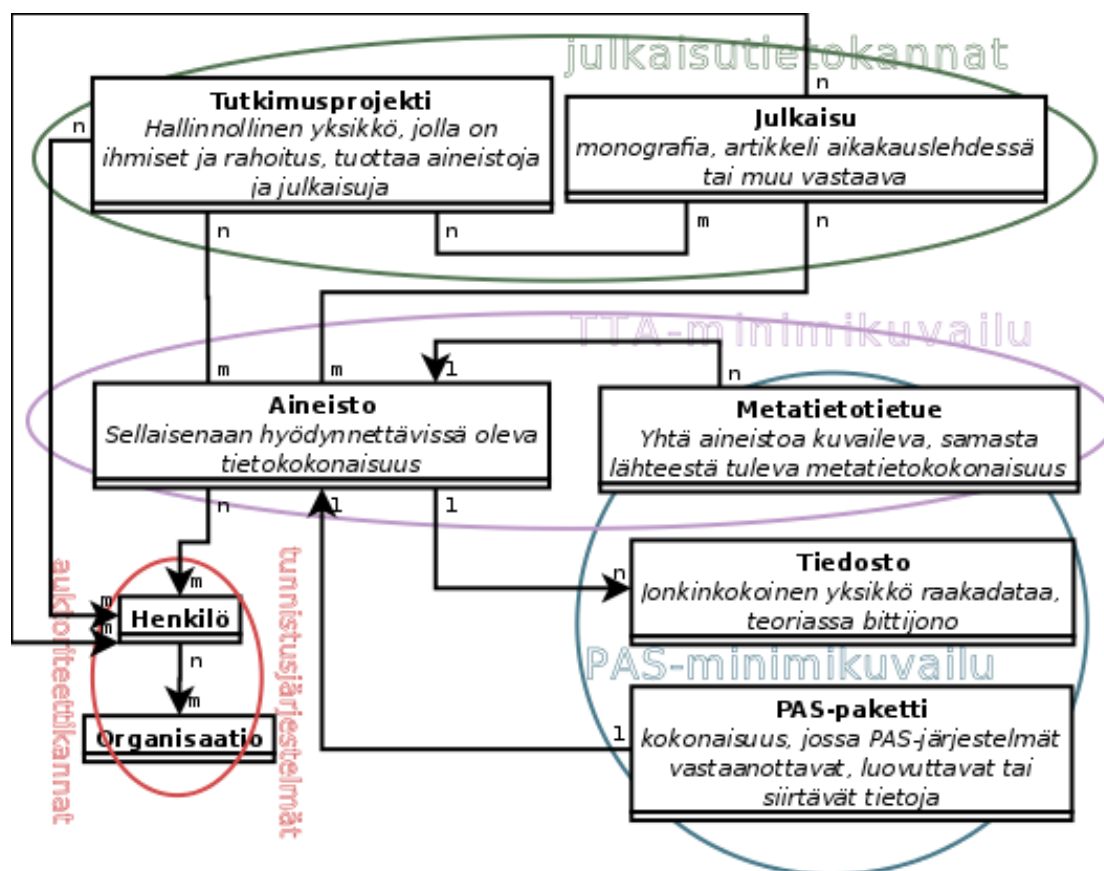
Eryteisesti tutkimusaineistojen kuvailuun tarkoitettujen formaattien lisäksi on olemassa sekä kotimaisia että kansainvälisiä sähköisten dokumenttien kuvailusuosituksia (esimerkiksi JHS143, INSPIRE).

TTA-minimimetatietomalli ei pyri tarkasti määrittämään, missä muodossa metatiedot talletetaan tai esitetään tekstinä tai käyttöliittymissä, vaan määrittää vain ne vähimmäistiedot, jotka jokaisesta tutkimusaineistosta on tiedettävä.

Minimimetatieto on pienin yhteinen nimittäjä

Tutkimusaineistoon liittyy monta eri asiaa, kuten esimerkiksi tutkimus, tutkimusprojekti, erilaisia toimijoita, aineistokokoelmia, yksittäisiä tiedostoja, julkaisuja ja palveluja.

Minimimetatietomalli keskittyy kuvailemaan tutkimusaineistoa ja siihen liittyvää metatietotietuetta. Se määrittelee pienimmän yhteisen nimittäjän eli ne pakolliset tieteenalasta riippumattomat tiedot, jotka jokaisesta tutkimusaineistosta on tallennettava. Vaikka minimimetatietomalli ei kuvaile esimerkiksi tutkimusaineistoon liittyviä julkaisuja tai aineiston tuottanutta projektia, ne pystyy linkittämään tutkimusaineistoon.



Minimimetatietomalli on luotu TTA-palveluita varten, mutta sitä voivat hyödyntää kaikki tutkimusaineistojen kanssa toimivat tahot. Esimerkiksi rahoittaja voi edellyttää syntyviltä tutkimusaineistoilta minimimallin mukaista kuvailua.

Minimimetatietomallin mukaista aineiston kuvailua voi ja useimmiten tuleekin rikastaa erilaisilla tieteenala-, aineistotyyppi- ja aineistokohtaisilla elementeillä tai tutkimuksen kontekstia kuvaavilla tiedoilla. TTA-hanke antaa myös suosituksia sekä lisäkuvailutiedoista että käytettävistä sanastoista, luokituksista ja ontologioista.

Uudet ja vanhat aineistot

Vanhoille, aikaisemmin tuotetuille aineistoille ei voi aina asettaa samanlaisia kuvailuvaatimuksia kuin uusille, koska usein tarvittavia tietoja ei yksinkertaisesti ole käytettävissä tai kohtuullisin kustannuksin selvitettävissä. TTA-palveluissa uusilta aineistoilta vaaditaan täydellisempiä toimijatietoja kuin vanhoilta aineistoilta.

Yhteenveto minimimetatietomallista

Minimimetatietomallin elementit sisältävät tietoa sellaisista aineiston tai metatietotietueen ominaisuuksista, jotka pitää olla tiedossa jokaiselle aineistolle.

Alla olevassa taulukossa on listattu minimimetatietomallin elementit.

Määrä-kenttä tarkoittaa, kuinka monta kertaa kenttä voi esiintyä kuvailtavalle aineistolle: 1 = vain kerran, 1* = useita kertoja, 0* = useita kertoja, voi puuttua. Se, että aineiston kieli -kenttä voi puuttua, ei tarkoita, että tieto on valinnainen, vaan ettei aineiston ymmärtämisessä auta minkään kielen tunteminen. Tuottaminen-kenttä sisältää huomioita, mikä taho tuottaa kyseisen metatiedon: t = tallennuspalvelu automaattisesti, k = aineistokatalogi automaattisesti, y = aineiston julkaisija käsin.

nro	elementin nimi	määrä	tuottaminen	mihin viittaa
1.	Metatiedon tunniste	1	t/k	linkki (PID)
2.	Metatiedon muokkaamisaika	1	t/k	ajankohta
3.	Aineiston tunniste	1*	t/k+y	linkki (PID)
4.	Muokkaamisaika	1	t/y	ajankohta
5.	Jakelijan yhteystieto	1*	t/k+y	FOAF-kuvaus
6.	Aineiston nimi	1	y(t)	merkkijono
7.	Käyttöehdot	1*	y	
8.	Aihe	1*	y(t)	käsitelinkki/merkkijono
9.	Toimija	1*	t/k+y	tunniste/merkkijono
10.	Aineiston kieli	0*	y	ISO 639 -kielikoodi

Tallennuspalvelulla tarkoitetaan tässä dokumentissa sellaista palvelua, johon voi tallettaa aineistoja ja josta käsin aineistoja voi julkaista. Aineistokatalogilla tarkoitetaan sellaista palvelua, johon säilötään metatietoja olemassa olevista aineistoista, mutta itse aineistot säilytetään muualla.

Hajautettujen metatietojen käsittely

Monet metatietoelementit voivat sisältää viittauksia (linkkejä) muualla ylläpidettyihin tietoihin. Tällaisia metatietoelementtejä ovat tunnisteet, käyttöehtotieto, aihe sanat ja toimijat. Linkit voivat olla eri palveluista saatavia

pysyväistunnisteita¹ tai sellaisenaan käytettävissä olevia linkkejä www-osoitteisiin, joista löytyy lisätietoja kyseisestä aineistosta, käyttöehdoista, käsitteestä tai toimijasta. Tällöin metatietojen talletus on siinä mielessä hajautettua, että metatieto sisältää tiedot aineistoon liittyvistä toimijoista tai aiheista, mutta tarkempaa tietoa esim. toimijoiden eri nimiversioista tai asiasanojen kieliversioista pidetään yllä toisaalla kuten auktoriteettitietokannoissa ja sanastopalveluissa.

Linkkejä sisältävää metatietoa käsitellessä pitää noudattaa tiettyjä yleisiä periaatteita:

1. Metatietojen syöttöä tukevissa käyttöliittymissä, kuten TTA-metatietokatalogin aineistonjulkaisulomakkeessa, on ohjelmiston autettava vapaiden tekstiviittausten (nimien, asiasanojen) muuntamisessa linkeiksi.
2. Linkkien ja viittausten tulee viedä osoitteisiin, joista on saatavissa sekä ihmisen hyödynnettävää tietoa (linkkiä seuraamalla) että konekäsiteltävää tietoa esim. RDF-muodossa.
3. Metatietoja käsittelevien palveluiden tulee hakea automaattisesti ja säännöllisesti linkkien takana olevat tiedot ja tallettaa ne tekstihakuja varten sekä säilyttää niistä täysi historia. Esimerkiksi aineistoja on voitava hakea tekijän nimen perusteella, vaikka tekijätieto annettaisiin linkkinä tekijätietokantaan; samoin hakujen pitää toimia, vaikka viitattu kohde olisikin siirtynyt muualle tai lakannut olemasta.

Minimimetatietomalliin ei sisälly suoraan elementtiä, joka kertoisi aineiston sijainnin tai osoitteen. Tämä tieto kuitenkin sisältyy epäsuorasti Tunniste-kenttään, sillä tunnisteet viittaavat ulkopuolisiin rekistereihin, joissa puolestaan pidetään yllä varsinaista sijaintitietoa. Sijaintitiedon talletustapa riippuu kuitenkin käytetystä tunnistetyypistä. PID-tunnisteisiin liittyy yleensä jonkinlainen resolver-palvelu, josta tunnistetta vastaavan osoitteen voi hakea. URL-tunnisteet perustuvat Internetin nimijärjestelmään ja niiden sijainti haetaan DNS-järjestelmästä. Voidaan myös kuvitella, että aineistoon viitattaisiin esim. ISBN-tunnisteella, jolloin sijaintitieto löytyy kirjastojen tietokannoista.

Historia- ja muutostiedot

Aineiston historia- ja muutostiedoille ei ole minimimetatietomallissa erillistä kenttää. Sen sijaan samalle aineistolle annetut metatietotietueet muodostavat yhdessä dokumentaation aineiston tapahtumahistoriasta. Metatiedon muokkaamisaikojen perusteella samantunnisteisista metatiedoista on mahdollista muodostaa kokonaisuus, josta käyvät ilmi toimijoiden, käyttöehtojen ynnä muiden vastaavien muutokset.

Minimimetatietomallin elementtien tarkat kuvaukset

1. Metatiedon tunniste

Metatiedon nimi: metatiedon tunniste (metadata identifier)

Kuvaus: Metatietojen tunniste on yksiselitteinen, pysyvä tunniste yhdelle metatietotietueelle, joka muodostuu yhden tahon yhdelle aineistolle antamista metatiedoista. Kun metatietoja päivitetään, pysyvä tunniste ei muutu.

Miten tuotetaan: metatiedot vastaanottava palvelu (tallennuspalvelu/aineistokatalogi) lisää tunnisteiden automaattisesti, mikäli sellaista ei vielä ole.

Elementtien määrä: ei voi puuttua, ei voi toistua.

¹Esimerkiksi joihinkin toimijoihin voi viitata ISNI- tai ORCID-tunnisteella, aineistoihin erilaisilla PID-tunnisteilla, kuten URN- tai DOI-tunnisteilla, ja projekteihin rahoituspäätösnumeroilla

Perustelu: Kun metatietoja kerätään eri lähteistä, tunnisteiden perusteella voidaan tunnistaa alun perin samasta lähteestä tulleet kuvailut ja yhdistää ne (tai valita ajantasaisin tietue metatietojen muutospäivän perusteella). Saman tunnisteiden metatiedot muodostavat myös muutoshistorian, jonka kautta voi tarkastella metatietoihin tulleita muutoksia.

Tarkenteet: Tunnistejärjestelmä, esim. Handle, URN

Muotoilu: globaalisti uniikki (ASCII-)merkkijono, mieluiten URI

Esimerkki: URN:NBN:fi-fe201204234822

Esimerkki: URN:OID:0.9.2342.19200300.100.4

2. Metatiedon muokkaamisaika

Metatiedon nimi: metatiedon muokkaamisaika (metadata date of modification)

Kuvaus: ajankohta, jolloin metatietotietue on viimeksi päivitetty; tarkemmin sanottuna viime ajankohta, jolloin on tuotettu metatietotietue, jolla on tämä (metatietotietueessa annettu) tunniste. Metatiedon muokkaamisaika ei ole suoraan riippuvainen aineiston päivityksestä: metatietoja voi täydentää tai korjata aineiston muuttumatta. Päivittäminen merkitsee metatietotietueen sisällöllistä päivittämistä, ei esimerkiksi migraation tai metatietojen siirron yhteydessä tehtyä automaattista muunnosta.

Miten tuotetaan: aineistokatalogi lisää automaattisesti ottaessaan vastaan päivityksen.

Elementtien määrä: ei voi puuttua, ei voi toistua.

Perustelu: muutospäivä tekee mahdolliseksi vertailla metatietojen eri versioita ja valita niistä ajantasaisimman. Myös metatietojen muutosten tarkkailu ja suhteuttaminen aineiston muutoksiin voi tarjota hyödyllistä tietoa aineiston muutoksista.

Muotoilu: päivämäärä ja kellonaika sekunnin tarkkuudella ISO 8601 -muodossa vvvv-kk-ppTt:mm:ss tai aikavyöhykkeen kanssa vvvv-kk-ppTt:mm:ss±tt:mm. Mikäli aikavyöhyke puuttuu, oletetaan aikaleiman viittaavan Suomen aikaan. Muita ISO 8601 -muotoja ei saa käyttää.

Esimerkki: 2012-04-30T12:30:05

Esimerkki: 2011-11-13T09:15:09+00:00

3. Aineiston tunniste

Metatiedon nimi: aineiston tunniste (identifier)

Kuvaus: yksiselitteinen tunniste aineistolle, jota metatietotietue kuvailee. Jokaisella aineistolla tulee olla pysyvä tunniste (PID), joka on ainoastaan tämän aineiston tunniste. Aineistolla voi olla useampia pysyväistunnisteita; ainakin yhden näistä tulee olla sellainen, että se muuttuu aina, kun aineiston sisältö muuttuu (eli saman aineiston eri versioilla pitää olla ainakin yksi eroava tunniste). TTA-palvelut voivat muodostaa tunnisteiden aineistolle automaattisesti, jos sellaista ei vielä ole. Tunniste toimii myös epäsuorasti aineiston sijaintitietona, sillä tunniste voi olla aineistoon osoittava URL (www-osoite) tai viittaus muualla olevaan sijaintirekisteriin (kuten PID-resolveriin). Tunniste-kenttään ei pidä merkitä aineistoon liittyvien julkaisujen tai henkilöiden tunnisteita. Niille on omat kenttensä metatietomallissa.

Miten tuotetaan: aineiston vastaanottava tallennuspalvelu hankkii automaattisesti aineistolle tunnisteiden, jos sillä ei vielä sellaista ole. Aineiston tekijä, tuottaja tai jakelija voi myös hankkia tunnisteiden itse haluamastaan PID-palvelusta ja merkitä sen aineiston metatietoihin. Mikäli metatiedoista puuttuu aineiston tunniste, metatietopalvelu hankkii sellaisen automaattisesti.

Elementtien määrä: ei voi puuttua, voi toistua; toistetut elementit ovat kaikki aineiston tunnisteita.

Perustelu: aineiston tunnisteiden perusteella pystytään tunnistamaan samaan aineistoon liittyvät metatiedot yhdistelyä / rinnakkain esittämistä varten. Tilanteissa, joissa metatiedot kulkeutuvat useampaa kautta samaan aineistokatalogiin, tunniste auttaa duplikaattien poistamisessa.

Tarkenteet: Tunnistejärjestelmä, esim. Handle, URN

Muotoilu: globaalisti uniikki (ASCII-)merkkijono, mieluiten http-URI, joka toimii myös linkkinä

Esimerkki: <http://urn.fi/URN:NBN:fi:metla-201205031004>

Esimerkki: <http://hdl.handle.net/1828/31698>

4. Muokkaamisaika

Metatiedon nimi: muokkaamisaika (date of modification)

Kuvaus: Kertoo, milloin aineiston sisältö on viimeksi muuttunut; tarkemmin sanottuna viime ajankohta, jolloin on tuotettu aineisto, jolla on nämä (metatietotietueessa luetellut) tunnisteet. Automaattista muunnosprosessia esim. tiedostomuotojen välillä ei katsota sisällön muutokseksi. Vastaa JHS143-suosituksen elementtiä aikamääre (tarkenne muokkaamisaika) ja DC-metatietoformaatin kenttää date.modified.

Miten tuotetaan: tallennuspalvelu lisää automaattisesti aineistoa muutettaessa; muille kuin tallennuspalvelun kautta tuleville aineistoille tekijä lisää käsin.

Elementtien määrä: ei voi puuttua, ei voi toistua.

Perustelu: Muokkausajankohdan perusteella voi tunnistaa ja vertailla aineiston eri versioita ja aineiston käyttäjä pystyy arvioimaan aineiston ajantasaisuutta. Aineistojen tuoreus on myös yleinen hakukriteeri.

Muotoilu: päivämäärä ja kellonaika sekunnin tarkkuudella ISO 8601 -muodossa vvvv-kk-ppTt:mm:ss tai aikavyöhykkeen kanssa vvvv-kk-ppTt:mm:ss±tt:mm. Mikäli aikavyöhyke puuttuu, oletetaan aikaleiman viittaavan Suomen aikaan. Muita ISO 8601 -muotoja ei saa käyttää.

Esimerkki: 2012-02-29T00:00:00+02:00

5. Jakelijan yhteystieto

Metatiedon nimi: jakelijan yhteystieto (distributor contact)

Kuvaus: Osoite tai yhteydenottopiste, jonka kautta voi tavoittaa aineistosta vastuussa olevia tai enemmän tietäviä tahoja. Aineistolla on oltava ainakin yksi osoite, johon kyselyt voi kohdentaa.

Miten tuotetaan: tallennuspalvelu tai aineistokatalogi tarjoaa sisäänkirjautumistiedoista saatuja aineiston julkaisijan tietoja oletukseksi, mutta julkaisija voi käsin korjata tiedot toisiksi.

Elementtien määrä: ei voi puuttua, voi toistua. Toistetut elementit ovat eri vastuutahojen yhteystietoja.

Perustelu: Yhteystietoja tarvitaan, jos aineiston uudelleenkäyttöön liittyy haasteita, joita käyttäjä ei pysty itse ratkaisemaan. Tilanteissa, joissa aineiston saatavuus, laatu tai ymmärrettävyys aiheuttavat käyttäjälle hankaluuksia, on yhteyshenkilön turvautuminen usein ainoa vaihtoehto. Myös palautetta, kontribuutiota ja korjauksia aineistoon pystyy yleensä toimittamaan, jos jokin yhteyshenkilö on tiedossa. Jakelijan rooli on luonteva ensimmäiseksi yhteydenottopisteeksi.

Muotoilu: FOAF-yhteystietomäärittely, jossa ovat mukana vähintään nimi, puhelinnumero, sähköpostiosoite ja kotisivu; muitakin yhteystietoja voi olla mukana. Yhteystieto kannattaa kohdistaa vastuuroolille organisaatiossa, mutta myös yksityishenkilö voi olla yhteystietona, jos vakiintuneempaa toimijaa ei ole.

Esimerkki:

```
<foaf:Organisation>
  <foaf:name xml:lang="fin">Yhteiskuntatieteellinen
tietoarkisto</foaf:name>
  <foaf:phone rdf:resource="tel:+358401901432"/>
  <foaf:mbox rdf:resource="mailto:fsd@uta.fi"/>
  <foaf:homepage rdf:resource="http://www.fsd.uta.fi/fi"/>
</foaf:Organisation>
```

Esimerkki:

```
<foaf:Person>
  <foaf:name xml:lang="fin">Panu Kalliokoski</foaf:name>
  <foaf:phone rdf:resource="tel:+358415323835"/>
  <foaf:mbox rdf:resource="mailto:panu.kalliokoski@csc.fi"/>
  <foaf:homepage rdf:resource="http://www.csc.fi/staff/pkalliok"/>
</foaf:Person>
```

6. Aineiston nimi

Metatiedon nimi: aineiston nimi (title)

Kuvaus: tekstimuotoinen otsake, joka kuvailee aineiston sisällön tiiviisti. Aineiston nimen perusteella ihmisen tulisi saada yleiskäsitys siitä, mitä aihepiiriä aineisto käsittelee. Aineiston nimen tulee myös olla niin yksilöivä, ettei muulla aineistolla ole todennäköisesti samaa nimeä.

Miten tuotetaan: aineiston tekijä tai jakelija määrittelee käsin. Oletuksena voi tallennuspalvelussa olla esim. kuvailtavan tiedoston nimi.

Elementtien määrä: ei voi puuttua, ei voi toistua (paitsi eri kieliversioina).

Perustelu: Aineiston nimellä voidaan viitata aineistoon eri yhteyksissä. Esimerkiksi aineistoon tai sen metatietoihin vievät linkit on luontevaa nimetä aineiston nimen mukaan.

Muotoilu: UTF-8-merkkijono ja kolmikirjaiminen ISO-639-kielitarkenne

Esimerkki: Väestönlaskennan pitkittäisaineisto 1999-2011 (fin-kielitarkenteen merkintä riippuu tiedostomuodosta)

Esimerkki: SMEAR-mittausaseman ilmakehäärvot 4/2010 (fin-kielitarkenteen merkintä riippuu tiedostomuodosta)

7. Käyttöehdot

Metatiedon nimi: käyttöehdot (access rights)

Kuvaus: Sisältää tiedot, millä tavalla ja ehdoilla aineistoa voi käyttää. Vastaa Dublin Core -elemettiä DC.Rights (Oikeudet). Rakenteistetusti kuvataan, saako aineiston käyttöön avoimesti vai pitääkö käyttölupaa hakea, ja onko hakemusprosessi automatisoitu vai manuaalinen.

Miten tuotetaan: aineiston omistaja tai jakelija määrittelee käsin.

Elementtien määrä: ei voi puuttua, voi toistua; toistetut elementit ovat vaihtoehtoisia käyttöehtoja aineistolle.

Perustelu: Aineiston käyttöehdot ovat aineistojen jatkokäyttäjälle tärkeä aineiston valintakriteeri. Käyttäjät saattavat haluta etsiä avoimia, maksuttomia tai välittömästi saatavilla olevia aineistoja tai käyttää vain aineistoja, jotka ovat yhdisteltävissä tai uudelleenjulkaistavissa muokkaamisen jälkeen.

Muotoilu: Tarkennetaan, kun KDK:n käyttöoikeustyöryhmän suositus valmistuu.

8. Aihe

Metatiedon nimi: aihe (subject)

Kuvaus: Käsitteitä ja sanoja, jotka liittyvät tavalla tai toisella aineiston sisältöön, luonnehtivat aineistoa tai kertovat sen asiayhteydestä tai merkityksestä. Aiheiden määrittelyssä on suositeltavaa käyttää asiasanoja, ontologioita ja/tai luokitusjärjestelmiä. Vastaa Dublin Core -elementtiä Subject ja JHS143-elementtiä Aihe. Asiasanoilla kannattaa kuvailla ainakin aineiston aiheet, asiayhteydet, tutkimuskohteet, tutkitut ominaisuudet, tieteenala(t), keruumenetelmä(t) ja analyysimenetelmä(t).

Miten tuotetaan: aineiston tekijä tai jakelija lisää käsin. Tallennuspalvelu voi ehdottaa asiasanoja esim. tiedoston hakemistopolusta tai käyttäjätunnuksen/työalueen asiakkuustiedoista (esim. projekti, tieteenala tai rahoittaja), mutta tietojen syöttäjän on tarkistettava nämä.

Elementtien määrä: ei voi puuttua, voi (ja yleensä täytyykin) toistua; toistetut elementit ovat kukin aineistoa eri tavoin kuvaavia asiasanoja, joiden keskinäistä suhdetta ei ole tarkemmin määritelty; voivat esimerkiksi kuvata aineiston eri osia tai koko aineistoa eri näkökulmista.

Perustelu: Elementti mahdollista aineiston kuvailun yleisellä tasolla sekä rakeenteellisesti että vapaan tekstin avulla. Sisällön kuvailu mahdollistaa aineiston linkittämisen samankaltaisiin tai samaan aiheeseen liittyviin aineistoihin sekä suuremmiksi kokonaisuuksiksi. Aiheen kuvailu tarjoaa tietoa aineiston sisällöstä helpottaen aineiston soveltuvuuden arvioimista.

Muotoilu: jokin seuraavista: (1) tarkenne, joka kertoo sanaston + termi UTF-8-merkkijonona; (2) URL-viittaus käsitteeseen; (3) vapaa UTF-8-merkkijono + merkkijonon kieli. Huom! URL-viittauksia käsitellään TTA-palveluissa tiettyjen sääntöjen mukaan, katso kohta "hajautetun metatiedon käsittely" yllä.

Esimerkki: <http://www.yso.fi/onto/yso/p19612>

Esimerkki: työolot (sanastotarkenne CESSDA ilmaistaan eri tavoin eri formaateissa)

Esimerkki: Baikal-järven ympäristö (kielitarkenne fin ilmaistaan eri tavoin eri formaateissa)

9. Toimija

Metatiedon nimi: toimija (agent)

Kuvaus: Taho (esimerkiksi henkilö, organisaatio tai projekti), joka tekee tai on tehnyt aineistolle jotain, tai jolta voi kysyä lisätietoja aineistosta. Usein toimija on myös vastuussa aineiston päivittämisestä tai on alun perin koonnut aineiston, jota ei päivitetä. Toimijalla voi olla seuraavat roolit: tekijä, muu tekijä, tuottajaprojekti, jakelija/julkaisija, omistaja ja rahoittaja. Nämä roolit vastaavat JHS143-suosituksen vastaavan nimisiä rooleja toimija- ja julkisuus-kentissä.

Pakollisia toimijatietoja kaikille aineistoille ovat tekijä ja jakelija/julkaisija. Uusille aineistoille pakollisia tietoja ovat lisäksi aineiston tuottajaprojekti, rahoittaja (mikäli nämä ovat olemassa) ja aineiston omistaja. Jos tekijä on henkilö, tiedoissa tulee olla mukana myös tekijän taustaorganisaatio. Projekti yksilöidään viittaamalla suoraan projekti- tai rahoituspäätöstietokantoihin tai, jos ei ole yksiselitteistä tapaa viitata näihin, rahoittajan nimellä ja rahoituspäätöksen numerolla.

Aineiston kuvailuun tulee tuottaa mahdollisimman täydelliset aineiston toimijoiden tiedot. Tämän tarkastaminen automatisoidusti metatietotietueen perusteella on kuitenkin mahdotonta. Esimerkiksi, vaikka aineistolle olisi merkitty rahoittaja, ei ole mahdollista koneellisesti tarkistaa, ettei jokin toinen rahoittaja ole jäänyt merkitsemättä.

Miten tuotetaan: henkilö, joka lisää aineiston tallennuspalveluun / aineistokatalogiin, merkitään automaattisesti aineiston jakelija/julkaisijaksi ja tallennuspalvelu lisää projekti- ja rahoittajatiedon käyttöoikeustietojen perusteella; muita toimijoita voi lisätä käsin.

Elementtien määrä: ei voi puuttua, voi toistua; toistetut elementit sisältävät aineiston eri toimijoiden tiedot.

Perustelu: Eri rooleissa olevat toimijat saavat näkyvyyttä ja krediittiä, kun tieto esiintyy mm. aineistoluetteloissa. Toimijatieto on myös yleinen tilastointi- ja hakukriteeri aineistoille, ja sen perusteella on mahdollista selvittää aineiston taustoja. Toimijoiden perusteella voi etsiä myös toisiinsa liittyviä aineistoja ja yhdistää joitain aineistoja julkaisuihin.

Tarkenteet (roolit): tekijä (creator/author) = henkilö tai organisaatio, joka on alun perin luonut aineiston; muu tekijä (contributor) = henkilö tai organisaatio, joka tekijä-kentässä tahojen lisäksi on osallistunut merkittävästi aineiston luomiseen; tuottajaprojekti (project) = projekti, joka on teettänyt tai muuten saanut aikaan aineiston tuottamisen; jakelija/julkaisija (distributor/publisher) = taho, jolla on oikeus jakaa aineistoa edelleen tai joka on asettanut aineiston julkisesti saataville; omistaja (owner/rights holder) = henkilö tai organisaatio, joka nykyhetkellä päättää aineiston käytöstä; rahoittaja (funder) = aineiston tuottamisen rahoittanut taho.

Muotoilu: viittaus, jonka perusteella pystyy saamaan lisätietoja toimijasta, esimerkiksi viittaus ISNI-tietokantaan, PID, henkilöä kuvailevan FOAF-tiedoston URL, projektin rahoituspäätöksen yksilöivä linkki tai projektin kotisivun URL; mikäli mitään tällaista ei ole käytettävissä, käytetään toimijan nimeä; lisäksi roolitarkenne

Esimerkki: <http://danbri.org/foaf.rdf#danbri> (roolina muu tekijä; tämä ilmaistaan eri tavoin eri formaateissa)

Esimerkki: <https://kata.csc.fi/user/pkallio> (roolina julkaisija; tämä ilmaistaan eri tavoin eri formaateissa)

Esimerkki: <urn:ISNI:1422-4586-3573-0476> (roolina tekijä; tämä ilmaistaan eri tavoin eri formaateissa)

10. Aineiston kieli

Metatiedon nimi: aineiston kieli (language)

Kuvaus: Kieli, jonka taitamista aineiston ymmärtäminen ja käyttäminen vaatii. Esimerkiksi kieli, jolla haastattelu on kirjoitettu, tai numeerisen aineiston kohdalla kieli, jolla on nimetty aineiston muuttujat on tai muuttujien tai arvojen selitteet. Vastaa Dublin Core -elementtiä Language ja JHS143-elementtiä kieli.

Miten tuotetaan: jakelija lisää käsin.

Elementtien määrä: voi puuttua, voi toistua; puuttuva elementti tarkoittaa, ettei aineistolla ole kieltä (esimerkiksi numeerinen aineisto ilman selitteitä); toistettu elementti tarkoittaa, että aineistossa on käytetty useita kieliä.

Perustelu: Voi olla hakukriteeri aineistolle. Vaikuttaa siihen, onko aineisto käyttäjän hyödynnettävissä.

Muotoilu: Kieli tallennetaan ISO 639-standardin mukaisena kolmikirjaimisena koodina

Esimerkki: eng

Esimerkki: fin

Liite 1. Metatietomallin vastaavuus yleisiin metatietostandardeihin

Tässä taulukossa esitellään, miten TTA-minimimetatietomallin kentät vastaavat olemassa olevia metatietostandardeja.

no	metatietomalli	DC(terms)	DDI Codebook
1	Metatiedon tunniste	-	docDscr / citation / titlStmt / IDNo
2	Metatiedon muokkaamisaika	X-MetadatalastModified	docDscr / citation / (prodStmt / prodDate, verStmt / version / date)
3	Aineiston tunniste Identifier (@scheme)		stdyDscr / citation / titlStmt / IDNo (/ agency)
4	Muokkaamisaika	Modified	stdyDscr / citation / (prodStmt / prodDate, verStmt / version / date)
5	Jakelijan yhteystieto	Publisher / (foaf:mbox, foaf:phone, foaf:homepage)	stdyDscr / citation / distStmt / contact
6	Aineiston nimi	Title	stdyDscr / citation / titlStmt / titl
7	Käyttöehdot	Rights, Available, License, AccessRights	stdyDscr / dataAccs / useStmt / restrctn
8	Aihe	Subject (@scheme)	stdyDscr / stdyInfo / subject / keyword (/ vocab)
9	Toimija	Creator, Contributor, Publisher	stdyDscr / citation / (rspStmt / AuthEnty, rspStmt / othld, prodStmt / producer, distStmt / distrbtr)
10	Aineiston kieli	Language	-

Liite 2. Metatietotyöryhmän kartoittamat metatietostandardit

Tässä liitteessä on lueteltu standardit, joita metatietotyöryhmä on käyttänyt lähtökohtana selvitystyössään.

Dublin Core ja sen seuraajat	JHS143, DataCite, DC for collections
RDF-metadatasuositukset	RDF, OWL, VoiD
PID-käytännöt	URN, DOI, Handle
Data Documentation Initiative	DDI Codebook (DDI2) ja DDI Lifecycle (DDI3)
Arkistojen suositukset	OAIS, EAD, KDK:n hallinnolliset ja rakenteelliset metatiedot, SÄHKE, PREMIS
Muut tieteenalakohtaiset suositukset	SDMX (ESMS), EML, CESSDA:n suositellut elementit, DataONE, RDF Data Cube, CoSSI, CMDI, Darwin Core, ESRI metadata, ISO 19115:2003, INSPIRE
Tutkimusten kuvailukäytännöt	CERIF/CRIS, RIF-CS, Dryad, CSMD

Liite 3. Minimimetatietomallin mukainen esimerkkikuvailu

Tämä liite antaa esimerkin, kuinka minimimetatietomallin mukaiset tiedot voidaan esittää RDF/XML-muodossa. Minimimetatietomalli ei ole sidottu tähän formaattiin, vaan samat tiedot pystyy esittämään monissa muissakin formaateissa.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <!-- Metatietotietueen tiedot, rdf:about on tämä dokumentti -->
  <rdf:Description rdf:about="">
    <!-- Metatiedon tunniste(et) -->
    <dct:identifier>urn:nbn:fi:csc-kata2012082200012</dct:identifier>
    <!-- Metatiedon muokkaamisaika -->
    <dct:modified
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-08-22T16:54:29</dct:modi
fied>
    <foaf:primaryTopic rdf:resource="urn:nbn:fi:csc-ida2012082200002"/>
  </rdf:Description>

  <!-- Aineiston tiedot, rdf:about on aineistoversion tunniste -->
  <rdf:Description rdf:about="urn:nbn:fi:csc-ida2012082200002">
    <!-- Eri aineistoversioille yhteinen tunniste -->
    <dct:identifier>urn:nbn:fi:csc-ida2012081100001</dct:identifier>
    <!-- Muokkaamisaika -->
    <dct:modified
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-08-22T10:37:08</dct:modi
fied>
    <!-- Jakelijan yhteystieto -->
    <dct:publisher>
      <foaf:Organization rdf:about="http://kata.csc.fi/user/CSC">
        <foaf:name xml:lang="fi">CSC - Tieteen tietotekniikan keskus oy,
aineistopalvelut</foaf:name>
        <foaf:mbox rdf:resource="mailto:esimerkkiaineistopalvelu@csc.fi"/>
        <foaf:phone rdf:resource="tel:+358912345789"/>
        <foaf:homepage rdf:resource="http://www.csc.fi/tutkijalle/aineistopalvelut"/>
      </foaf:Organization>
    </dct:publisher>
    <!-- Aineiston nimi -->
    <dct:title xml:lang="fi">Esimerkkiaineisto</dct:title>
    <!-- Käyttöehdot -->
    <!-- Kuvattu tässä METS Rightsin mukaisesti -->
    <dct:rights rdf:parseType="Literal" xmlns="http://www.loc.gov/METS/">
      <RightsDeclaration RIGHTSCATEGORY="PUBLIC DOMAIN"/>
    </dct:rights>
    <!-- Aihe -->
    <dct:subject rdf:resource="http://www.yso.fi/onto/yso/p2205"/>
    <dct:subject rdf:resource="http://www.yso.fi/onto/yso/p2880"/>
    <!-- vapaan asiananan voi merkitä näin tai viittaamalla omaan asianastoon -->
    <dct:subject xml:lang="fi">havainnollistaminen</dct:subject>
    <!-- Toimija -->
    <dct:creator rdf:resource="urn:isni:0000000121032683"/>
    <dct:creator rdf:resource="https://orcid.org/0000-0003-0296-7410"/>
    <dct:rightsHolder rdf:resource="http://kata.csc.fi/user/CSC"/>
    <dct:contributor>
      <foaf:Project rdf:about="http://www.aka.fi/rahoituskohteet/123456789">
        <foaf:name>Esimerkkiprojekti</foaf:name>
        <foaf:homepage rdf:resource="http://www.csc.fi/projektit/esimerkkiprojekti"/>
        <rdfs:comment>Suomen akatemian rahoituspätös 123456789</rdfs:comment>
      </foaf:Project>
    </dct:contributor>
    <!-- Aineiston kieli -->
    <dct:language rdf:resource="http://lexvo.org/id/iso639-3/fin"/>
  </rdf:Description>
</rdf:RDF>
```