

TerveSuomi-portaalin metatietomäärittely

Ver. 2.0 (18.6.2009)

Osma Suominen, Kim Viljanen, Eero Hyvönen
email: etunimi.sukunimi@tkk.fi



SEMANTIC COMPUTING RESEARCH GROUP
FinnONTO project, 2003-2010
<http://www.seco.tkk.fi>

Sisältö

1	Johdanto	1
1.1	Rajaukset	1
1.2	Tehdyt muutokset	1
2	Metatieto TerveSuomi-portaalissa	4
2.1	Metatiedon esittäminen	4
2.2	Metatietoarvojen tyypit	4
2.3	Metatietokenttien pakollisuus ja toistettavuus	4
2.4	Metatiedon laajennettavuus	5
2.5	Kieli- ja formaattiversiointi	5
3	Metatietokentät	6
3.1	Identifointitunnus (dc:identifier)	6
3.2	URL-osoite (ts:url)	7
3.3	Nimeke (dc:title) ja kuvaus (dc:description)	7
3.4	Kieli (dc:language)	7
3.5	Julkistamis-, hyväksymis- ja muokkausaika (dcterms:issued, dcterms:dateAccepted ja dcterms:modified)	8
3.6	Julkaisija ja tekijä (dc:publisher, dc:creator)	8
3.7	Kuuluu kokoelmaan (dcterms:isPartOf)	9
3.8	Oikeudet (dc:rights)	9
3.9	Tallennuskielto (ts:noindex)	9
3.10	Julkaisun laji (ts:genre)	9
3.11	Esitysmuoto (dc:type)	9
3.12	Formaatti (dc:format)	9
3.13	Aihe (dc:subject) ja Avainsana (ts:keyword)	10
3.14	Kohdeyleisö (dcterms:audience)	10
4	Metatiedon esittäminen RDF/XML-muodossa	11
4.1	RDF/XML-dokumentin rakenne	11
4.2	Dokumentin tunnistet ja URL-osoite	11
4.3	Vapaatekstikentät	12
4.4	Rakenteiset merkkijonot	12
4.5	Käsiteviittaukset	12
4.6	Yksilöt	13
4.7	Dokumentti- ja yksilöviittaukset	14

4.8	Esimerkki RDF/XML-muotoisesta metatiedosta	14
4.9	Organisaatiokohtaisten yksilöiden esittäminen	15
5	Metatiedon esittäminen osana HTML-sivua	16
5.1	Kenttätunnisteet	16
5.2	Profilimäärittely	16
5.3	Nimiavaruusmäärittelyt	17
5.4	Dokumentin tunniste ja URL-osoite	17
5.5	Vapaatekstikentät	17
5.6	Rakenteiset merkkijonot	17
5.7	Tallennuskielto	18
5.8	Käsite-, dokumentti- ja yksilöviittaukset URI-tunnisteella	18
5.9	Yksilöviittaukset vapaatekstimuodossa	18
5.10	Käsiteviittaukset termin avulla	18
5.11	Esimerkki metatiedosta HTML-dokumentissa	20
6	Sanastot ja ontologiat	21
7	Apuvälineet metatiedon tuottamiseen	22
7.1	ONKI-palvelimen toiminnallisuudet	22
7.1.1	ONKI-selain	22
7.1.2	ONKI-valitsin	22
7.1.3	ONKI Web Service -rajapinta	22
7.2	Saha-metatietoeditori	23
7.3	Palauteraportit	23
8	Kenttien vastaavuudet JHS 143- ja Sähke2-malleissa	24

1 Johdanto

TerveSuomi on kansallinen terveystietoportaali, joka toimii väylänä eri organisaatioiden tuottamaan terveysaiheiseen sisältöön siten, että tiedon löydettävyys kansallisella tasolla paranuu. Nykyisellään tietoa tuotetaan satojen eri organisaatioiden toimesta ja julkaistaan hajautetusti kunkin organisaation verkkosivuilla, jolloin kansalaisen on vaikeaa löytää kaikkea sitä terveyteen liittyvää aineistoa, joka olisi relevanttia hänen kulloisenkin tietotarpeen kannalta. TerveSuomi-portaali tarjoaa keskitetyn väylän tämän sisällön löytämiseen ja käyttämiseen.

FinnONTO-tutkimushankkeen¹ ja sitä seuranneen Semantic Web 2.0 (FinnONTO 2.0) -hankkeen² osana kehitettävässä TerveSuomi-järjestelmässä hyödynnetään uusimpia ns. semanttisia viestintäteknikoita (semanttinen web, ontologiat, metatieto) siten, että käyttäjälle tarjotaan mahdollisimman helppokäyttöinen, tehokas ja ”älykäs” käyttöliittymä tarjolla olevan tietosisällön käyttämiseen. Portaalin toiminnan edellytyksenä on kuitenkin, että sen sisältämistä dokumenteista on tuotettu laadukasta, standardimuotoista metatietoa. Metatiedoissa kerrotaan mm. dokumenttien julkaisija, tekijät, keskeiset päivämäärät, aihesisällöt ja näiden perusteella portaali osaa ohjata käyttäjän oikeaan tietoon.

Tässä dokumentissa kuvataan TerveSuomi-portaalin metatiedon luonne, määrä ja esitystavat siten, että tiedontuottajat pystyisivät tämän dokumentin perusteella tuottamaan oikeanlaista ja oikeanlaisessa muodossa tarvittavat metatiedot. Dokumentti perustuu kansainvälisiin Dublin Core -metatietosuositukseen [DCES, 2008, DCTERMS, 2008] sekä kansallisiin metatietostandardeihin, kuten JHS 143 -suositukseen [JHS 143, 2004], Sähke2-malliin [SÄHKE2, 2008] sekä Dublin Coren suomalaiseen versioon [DC-SFS, 2002, SFS 5895, 2001].

Työ on tehty osana FinnONTO- ja SW 2.0-tutkimushankkeita, joiden rahoittajina toimivat TEKESin lisäksi yli 30 eri organisaatiota. Tässä kuvattu työ on tehty Teknillisen korkeakoulun Mediatekniikan laitoksella Semanttisen laskennan tutkimusryhmässä (SeCo)³.

1.1 Rajaukset

Metatietokentille ei pääsääntöisesti ole määritelty oletusarvoja, koska oletusarvot ovat hyvin tiedontuottajakohtaisia (esim. julkaisija, käyttöoikeudet). Metatiedon tuottamista helpottavat oletusarvot tulisi määrittellä järjestelmäkohtaisesti.

Metatiedolle on määritelty HTML- ja RDF/XML-muotoiset esitysmuodot muttei erillistä, RDF:stä riippumatonta XML-esitysmuotoa. Toisaalta RDF/XML-esitysmuoto on määritelty niin, että sen mukaista metatietoa on varsin suoraviivaista tuottaa myös pelkin XML-työkaluin.

1.2 Tehdyt muutokset

Metatietomäärittelyn versio 1.0 julkaistiin tammikuussa 2007. Tämän jälkeen metatietomallia on yksinkertaistettu poistamalla määrittelystä sellaisia metatietokenttiä, joita ei käytännössä ole tarvittu TerveSuomi-portaalin prototyypissä tai sen tuotantoversiossa. Toisaalta skeemaa on myös muokattu erityisesti aiheiden ja kohderyhmien määrittelyn osalta, jotta portaalin käyttämä metatieto vastaisi paremmin loppukäyttäjien tarpeita. Aiheita on nyt mahdollista kuvata myös vapain asiansoin. Julkaisut on mahdollista liittää julkaisukokoelmiin, joille on määritelty oma kuvaustapansa. Muutokset metatietokenttiin on esitetty taulukossa 1.

Lisäksi metatietomäärittelyä on päivitetty vastaamaan uusimpia Dublin Core -suosituksia. Määrittelyä on myös pyritty saadun palautteen ja kertyneiden kokemusten perusteella kehittämään siten, että se vastaisi paremmin TerveSuomi-portaalin tiedontuottajaverkoston ja tietojärjestelmien toteuttajien tarpeita. Yhteydet JHS 143 -suositukseen [JHS 143, 2004] ja Sähke2-metatietomalliin [SÄHKE2, 2008] on myös pyritty tuomaan selkeämmin esiin.

¹<http://www.seco.tkk.fi/projects/finnonto/>

²<http://www.seco.tkk.fi/projects/sw20/>

³<http://www.seco.tkk.fi>

Taulukko 1: Metatietomäärittelyn ensimmäisen version (1.0, 26.1.2007) jälkeen tehdyt muutokset metatietokenttiin.

Nimi	Tunniste	Muutos
Tiivistelmä/Kuvaus	dcterms:abstract / dc:description	Tiivistelmä-kenttä korvattu yleiskäyttöisemmällä Kuvaus-kentällä.
Kieli	dc:language	Kielletty toisteiset arvot eli julkaisulla on täsmälleen yksi kieli. Kielikoodit ilmaistaan RFC4646-merkintäjärjestelmällä aieman RFC3066-järjestelmän sijaan. Käytännössä yleisimmät kielikoodit (fi, sv, en) pätevät kuitenkin edelleen.
Julkaisun laji	ts:genre	Kielletty toisteiset arvot eli julkaisulla on täsmälleen yksi laji/genre. Julkaisun laji -sanasto on jonkin verran muuttunut TerveSuomi-portaalin prototyypin käyttämästä versiosta.
Esitysmuoto	dc:type	Kielletty toisteiset arvot eli julkaisulla on täsmälleen yksi esitysmuoto.
Mediatyyppi	dcterms:medium	Poistettu kenttä toistaiseksi tarpeettomana.
Aihe	dc:subject	Mahdollisia sanastoja ovat YSA:n, MeSHin ja TESA:n lisäksi myös Terveystieteen edistämisen ontologia TERO sekä STAMETA-asiasanasto.
Avainsana	ts:keyword	Lisätty kenttä vapaille, asiasanastoihin ja ontologioihin kuulumattomille asiasanoille.
Kohdeyleisö	dcterms:audience	Kohdeyleisö-sanasto on korvattu TerveSuomi-portaalin osioluokituksella, joka on muodostettu siten, että se vastaa portaalin keskeisiä kohderyhmiä. Toisteiset arvot on kielletty, eli julkaisulla on täsmälleen yksi kohderyhmä.
Alueellinen kattavuus	dcterms:spatial	Poistettu kenttä toistaiseksi tarpeettomana.
Ajallinen kattavuus	dcterms:temporal	Poistettu kenttä toistaiseksi tarpeettomana.
Kuuluu	dcterms:isPartOf	Muutettu kentän määritelmää. Arvot ovat nyt PublicationCollection-tyyppisiä yksilöitä, joiden rakenne on myös määritelty.
Lähde	dc:source	Poistettu kenttä toistaiseksi tarpeettomana. Lähteet kuvataan tarvittaessa tekstisisällössä.
Viittaa	dcterms:references	Poistettu kenttä toistaiseksi tarpeettomana.
Käännöksen lähde	ts:isTranslationOf	Poistettu kenttä toistaiseksi tarpeettomana.
Aikaisempi formaatti	dcterms:isFormatOf	Poistettu kenttä toistaiseksi tarpeettomana.

HTML-esitysmuotoon on lisätty `profile`-attribuutti, joka mahdollistaa metatiedon koneellisen tulkinnan GRDDL-menetelmän [GRDDL, 2007] avulla, kuten Dublin Coren uusi suositus HTML-metatiedosta edellyttää [DC-HTML, 2008].

RDF/XML-esitysmuodossa on siirrytty käyttämään `rdf:datatype`-attribuuttia rakenteisten merkkijonojen koodausjärjestelmän ilmaisemiseen uusimman Dublin Core -suosituksen mukaisesti [DC-RDF, 2008]. RDF-esitysmuotoon on myös määriteltä tapa ilmaista aiheet ja muut käsiteviittaukset tekstimuotoisina asiasanoina, kun aiemmin ne oli pakko ilmaista URI-tunnisteella. Tämä helpottaa RDF-muotoisen metatiedon tuotantoa sellaisilla järjestelmillä, jotka eivät osaa käsitellä URI-tunnisteita.

Metatietomäärittelydokumenttia on kehitetty lisäämällä siihen luku, joka käsittelee metatiedon tuottamisen apuvälineitä. Johdantolukuja on lyhennetty, jotta määrittelystä ei tulisi kohtuuttoman pitkä ja raskaslukuinen.

2 Metatieto TerveSuomi-portaalissa

2.1 Metatiedon esittäminen

Metatieto koostuu TerveSuomessa paitsi tekstisisällöstä (esimerkiksi nimeke) myös yhteisistä, yksikäsitteisistä käsiteviittauksista (esimerkiksi dokumentin sisältöluokitus). Yksikäsitteisten käsiteviittausten ansiosta eri tiedontuottajien sisältö voidaan yhdistää toisiinsa. Dokumenttien, käsitteiden, henkilöiden ja organisaatioiden yksilöintiin käytetään pääasiasa URI-tunnuksia, joiden ansiosta kaikki viittaukset voidaan esittää täsmällisessä muodossa. Tavallisimmin URI-tunnukset⁴ ovat www-selaimesta tuttuja URL-osoitteita, mutta myös URN-tunnusten käyttö on mahdollista (ks. kohta 3.1).

TerveSuomi-metatietoskeema ("metatietokentät") on Dublin Coren metatietoformaatin mukainen. Mukaan on otettu Dublin Coren peruselementtijoukko [DCES, 2008] sekä Dublin Coren määrittelemiä tarkenteita [DCTERMS, 2008]. Kaksi lisätarkennetta on tässä projektissa määritelty: URL-osoite `dc:identifier`-kentän tarkenteena sekä julkaisutyyppi (genre) `dc:type`-kentän tarkenteena. Tarkenteiden käyttö mahdollistaa kentän täsmällisemmän käyttötarkoituksen ja semantiikan määrittelemisen menettämättä kuitenkaan kytkentää alkupe räiseen Dublin Core -elementtijoukkoon.

Kenttäjoukon valinnassa ja kenttien määritelmässä on soveltuvin osin huomioitu myös Dublin Core -metadatformaatin suomalainen versio [DC-SFS, 2002, SFS 5895, 2001] sekä JHS 143 -suositus [JHS 143, 2004].

2.2 Metatietoarvojen tyypit

Metatietokenttien arvot voivat olla joitakin seuraavista tyypeistä:

Vapaatekstin koneellinen tulkinta on pääsääntöisesti hankalaa tai ei onnistu. Tällaisia tietoja ovat mm. dokumentin nimeke ja kuvaus, henkilöiden ja organisaatioiden nimet (standardoinnista huolimatta kirjoitusasu voi vaihdella ja samannimisiä henkilöitä voi olla useita) sekä vapaat asiasanat joita ei ole sidottu tiettyyn sanastoon.

Rakenteiset merkkijonot, kuten päivämäärät, kielikoodit ja tiedostoformaatit esitetään standardinmukaisessa muodossa.

Käsiteviittaukset esitetään viitteinä jaetuissa ontologioissa määriteltyihin käsitteisiin. Viittaus voidaan esittää yksilöivänä URI-tunnusteena. Myös asiasana voi olla yksilöivä, mikäli tiedetään mihin asiasanastoon se kuuluu ja mikäli sama asiasana tarkoittaa sanastossa vain yhtä asiaa.

Yksilöviittaukset kuten organisaatiot, henkilöt ja julkaisukokoelmat esitetään viitteinä tietueisiin. Viittaus voidaan tehdä joko yksilöivällä URI-tunnusteella tai viittauksen kohteen yksilöivällä nimellä, joka on esitetty tässä dokumentissa määritellyllä tavalla.

2.3 Metatietokenttien pakollisuus ja toistettavuus

Kukin kenttä voi olla joko pakollinen tai vapaaehtoinen. Lisäksi se voi olla toistettavissa eli esiintyä joko kerran tai useammin. Jos samasta kentästä, kuten nimekkeestä tai kuvauksesta, on samaa tarkoittavia eri kielisiä versioita, lasketaan samaa tarkoittavat versiot kentän esiintymiseksi vain kerran.

Dublin Core määrittelee kaikki metatietokentät vapaaehtoisiksi ja toistettaviksi, mutta joidenkin kenttien osalta toisteisuusrajoituksia on tiukennettu metatiedon käsittelyn helpottamiseksi ja epäselvien tilanteiden (esim. mitä tarkoittaa jos dokumentilla on kaksi nimekettä) välttämiseksi.

⁴<http://www.ietf.org/rfc/rfc3986.txt>

2.4 Metatiedon laajennettavuus

Tämä dokumentti määrittelee TerveSuomi-portaalin edellyttämän metatiedon määrän ja muodon, mutta ei sulje pois muun kuin tässä esitetyn metatiedon lisäämistä aineistolle. Esimerkiksi organisaatiokohtaiset lisäkentät ovat mahdollisia ja sallittuja, vaikka TerveSuomi ei niitä hyödynnäkään.

Myös organisaatiokohtaisten käsitteistöjen, sanastojen ja luokitusten käyttö metatiedoissa on mahdollista, kunhan nämä metatiedot on esitetty niin, ettei sekaannuksen vaaraa ole. Käsiteviittausten osalta tämä tarkoittaa sitä, että viittauksen kohteena oleva käsitteistö on aina ilmaistu käsitteistötunnuksin.

2.5 Kieli- ja formaattiversiointi

Tyypillisiä tilanteita dokumenttikokoelmissa on se, että samasta dokumentista on useita käännöksiä eri kielillä (kieliversioita) tai että sama dokumentti on esitetty useassa teknisessä formaatissa (esim. HTML ja PDF).

Dokumentin eri versioilla voi olla sekä yhteisiä metatietoja (esim. julkaisija, aihe) että toisistaan eriäviä metatietoja (esim. julkaisuajankohta, kieli, formaatti). Julkaisujärjestelmän sisällä kannattaa noudattaa normalisointiperiaatetta eli tallettaa eri versioille yhteiset metatiedot vain kertaalleen. TerveSuomi-metatiedon kannalta dokumentin eri versiot ovat kuitenkin erillisiä ja itsenäisiä dokumentteja.

Metatiedon kieliversiointiin liittyy myös kysymys saman tiedon (esim. dokumentin nimeke tai julkaisijaorganisaation nimi) esittämisestä eri kielillä. Yksittäisten vapaatekstimuotoisten kenttien monikielinen esittäminen on hyvin tuettu sekä RDF- että HTML-kielissä ja näiden esittäminen käsitellään tarkemmin myöhemmin tässä dokumentissa asianomaisten kenttien kohdalla.

3 Metatietokentät

TerveSuomi-portaalin metatietokentät sekä niiden toisto ja pakollisuus esitetään taulukossa 2. Kukin metatietokenttä käsitellään tarkemmin seuraavissa aliluvuissa.

3.1 Identifointitunnus (dc:identifier)

Dokumentin tunniste ilmaistaan URI-muodossa. Tunnisteen tulee yksilöidä dokumentti, ts. sama tunniste vastaa vain yhtä dokumenttia, mutta dokumentilla voi olla useita tunnisteita. Jos tunnisteita on useita, tulee tiedon julkaisijan päättää mikä tunnisteista on se, jolla dokumentti ensisijaisesti yksilöidään.

URI voi perustua esimerkiksi ISBN, ISSN⁵, ISO OID⁶ ja muihin yleisesti käytettyihin tunnisteikäytäntöihin, joille on varattu IANA:ssa oma URI⁷- tai URN-skeematunnus. Se voi myös perustua organisaation sisäiseen arkistointitunnukseen tai vastaavaan. Jälkimmäisessä tapauksessa URI tulee muodostaa siten, että se on maailmanlaajuisesti yksilöivä.

⁵<http://www.ietf.org/rfc/rfc3044.txt>

⁶<http://www.ietf.org/rfc/rfc3061.txt>

⁷<http://www.iana.org/assignments/uri-schemes.html>

Taulukko 2: Metatietokentät. Pakolliset kentät on lihavoitu. Sallittujen arvojen määrä eli pakollisuus ja toisteisuus on ilmaistu P/T-sarakkeessa UML-merkintätavan mukaisesti. Esimerkiksi 0..* tarkoittaa valinnaista kenttää joka voi saada useita arvoja ja 1 tarkoittaa, että kenttä on pakollinen mutta saa vain yhden arvon.

	Nimi	Tunniste	P/T	Arvotyyppi	Arvojoukko
perustiedot	Identifointitunnus	dc:identifier	1	URI	
	URL-osoite	ts:url	0..1	URL	
	Nimeke	dc:title	1 ^a	Vapaateksti	Ei-tyhjät merkkijonot
	Kuvaus	dc:description	1 ^a	Vapaateksti	Ei-tyhjät merkkijonot
	Kieli	dc:language	1	Rakenteinen merkkijono	RFC4646
aikamääreet	Julkistamisaika	dcterms:issued	1	Rakenteinen merkkijono	W3CDTF (ISO 8601)
	Hyväksymisaika	dcterms:dateAccepted	0..1	Rakenteinen merkkijono	W3CDTF (ISO 8601)
	Muokkaus aika	dcterms:modified	0..1	Rakenteinen merkkijono	W3CDTF (ISO 8601)
tekijät	Julkaisija	dc:publisher	1..*	Yksilö	foaf:Organization
	Tekijä	dc:creator	0..*	Yksilö	foaf:Organization, foaf:Person tai foaf:Group
	Kuuluu kokoelmaan	dcterms:isPartOf	0..1	Yksilö	ts:PublicationCollection
	Oikeudet	dc:rights	0..*	Vapaateksti ja/tai dokumentti	URI tai sanallinen kuvaus
	Tallennuskielto	ts:noindex	0..1	Boolean	“true”, “false”, “1” tai “0”
luokittelut	Julkaisun laji	ts:genre	1	Käsite	Julkaisun laji -käsitteet
	Esitysmuoto	dc:type	1 ^b	Käsite	DCMI Type -käsitteet
	Formaatti	dc:format	1	Rakenteinen merkkijono	IANA-rekisteröidyt MIME-tyypit
sisällönkuvailu	Aihe	dc:subject	1..*	Käsite	TERO-, YSO-, MeSH-, TESA- ja STAMETA-käsitteet
	Avainsana	ts:keyword	0..*	Vapaateksti	Ei-tyhjät merkkijonot
	Kohdeyleisö	dcterms:audience	1	Käsite	TerveSuomi-portaalin osioluokitus

^aErikieliset versiot sallittu, kunhan kullakin kielellä on annettu vain yksi arvo.

^bPakollinen, jos arvo voi olla muuta kuin Teksti.

Esimerkkejä URI-tunnisteista:

```
http://www.ktl.fi/doc/123456
urn:isbn:951-611-441-5
urn:issn:1234-5678
```

Dokumentin erikieliset versiot sekä eri formaatissa olevat versiot samasta dokumentista (esim. HTML ja PDF) tulkitaan kuitenkin erilliseksi dokumenteiksi, mikäli niistä kaikista tuotetaan metatietoa.

TerveSuomen metatiedoissa suositellaan käytettäväksi URI-tunnisteita, jotka toimivat tarvittaessa myös URL-osoitteina. Esimerkiksi

```
http://www.ktl.fi/doc/123456
```

Jos sisällönhallintajärjestelmä on rakennettu niin, että dokumenttien URL-osoitteet ovat luonteeltaan pysyviä (esim. osoitteet eivät muutu vaikka sivuston rakenteeseen tai osiojakoon tehdään muutoksia), voi dokumentin tunnisteena toimia sen URL-osoite. Jos taas dokumentin URL voi vaihtua, dokumentin URI muodostetaan jollakin toisella tavalla (esim. julkaisujärjestelmän sisäisestä dokumenttitunnisteesta) siten, että se pysyy samana dokumentin URL:n muuttuessa. Pysyvien URL-osoitteiden suunnittelusta kertoo enemmän W3C:n dokumentti *Cool URIs don't change*⁸.

3.2 URL-osoite (ts:url)

Kenttä kertoo dokumentin URL-osoitteen, jonka kautta se on ladattavissa esimerkiksi WWW-selaimella. Sitä tarvitaan, jos dokumentin varsinaista tunnistetta (dc:identifier) ei voi käyttää dokumentin lataamiseen.

Suosittelavaa on joko käyttää pysyviä URL-osoitteita kaikkialla (dokumentin tunnisteena ja URL-osoitteena), tai mikäli tunniste poikkeaa sivustolla käytössä olevasta URL-osoitteesta, rakentaa tunniste niin, että sen kautta kuitenkin pääsee käsiksi (esim. HTTP redirect-mekanismilla) dokumenttiin. Näissä tapauksissa tämä kenttä on tarpeeton.

3.3 Nimeke (dc:title) ja kuvaus (dc:description)

Kullakin dokumentilla tulee olla nimeke ("otsikko") ja kuvaus, kumpiakin täsmälleen yksi. Nimekkeen tulisi liittyä yksinomaan dokumenttiin eli siihen ei saa sisältyä vakionuotoisia alku- tai loppuliitteitä, esimerkiksi julkaisijaorganisaation nimeä tai sivuston osaa johon se kuuluu. Suositeltavaa on, että nimeke vastaa dokumentissa näkyvää pääotsikkoa.

Kuvauksen ohjeellinen maksimipituus on 300 merkkiä ja sen tulee kiteyttää dokumentin keskeinen sisältö ja auttaa näin tiedonhakijaa näkemään yhdellä silmäyksellä, onko kyseinen dokumentti hänen kannaltaan kiinnostava. Suositeltavin tapa tuottaa kuvaus on tehdä se käsin, jotta tekstin laatu olisi mahdollisimman hyvä ja esittelisi mahdollisimman hyvin dokumentin ydinsisällön. Muita vaihtoehtoja ovat kuvauksen automaattinen tuottaminen dokumentin ingressin tai leipätekstin alkuosan perusteella.

Jos dokumentin otsikko on käännetty useille kielille (esim. väitöskirjat), voidaan kentässä antaa useita erikielisiä arvoja, kunhan samaa tarkoittavia käännöksiä on vain yksi. Myös tiivistelmästä voidaan vastaavaan tapaan antaa useita erikielisiä versioita.

3.4 Kieli (dc:language)

Dokumentin kieli annetaan RFC 4646 -standardin mukaisena kielikoodina [RFC 4646, 2006]. Esimerkiksi **fi** on suomi, **sv** ruotsi ja **en** englanti.

⁸<http://www.w3.org/Provider/Style/URI>

Tarvittaessa kielikoodi voidaan ilmaista kaksiosaisena, esim. `sv-FI` on suomenruotsi ja `en-US` on amerikanenglanti.

Suosittelavaa on kirjoittaa kielikoodin alkuosa pienillä kirjaimilla ja kaksiosaiset koodit kuten edellä, mutta kentän sisältö tulkitaan kuitenkin kirjainkoosta riippumatta.

3.5 Julkistamis-, hyväksymis- ja muokkausaika (`dc:terms:issued`, `dc:terms:dateAccepted` ja `dc:terms:modified`)

Aikamäärekentät ilmaisevat dokumentin julkistamis-, hyväksymis- ja muokkausajankohdan. Näistä ainoastaan julkistamisaika on pakollinen.

Julkistamisaika kertoo milloin dokumentti ensimmäistä kertaa julkaistiin yleisön saataville. Mikäli ajankohta ei ole tiedossa (esimerkiksi julkaisujärjestelmän vaihtamisen vuoksi), annetaan julkistamisajaksi varhaisin ajankohta, jolloin dokumentin tiedetään olleen julkaistu.

Hyväksymisaika kertoo milloin dokumentin asiasisältö on viimeksi käyty läpi ja siinä esitettyjen tietojen ajantasaisuus on varmistettu.

Muokkausaika kertoo ajankohdan, jolloin dokumentin asiasisältöä on viimeksi muokattu. Suositus on, että kentän arvoa ei päivitetä sellaisten pienten muutosten (esim. kirjoitusvirheiden korjaus) takia, jotka eivät vaikuta dokumentin asiasisältöön.

Aikamääreet esitetään W3C:n määrittelemässä W3CDTF-muodossa [[W3C-DTF, 1997](#)], joka määrittelee osajoukon ISO 8601 -standardin sallimista päivämääräformaateista [[ISO 8601, 1988](#)]. ISO 8601 mahdollistaa hyvin monta hiukan toisistaan poikkeavaa esitystä päiväyksille, joten rajoittuminen W3CDTF-muotoihin helpottaa päiväysten käsittelyä.

Päiväysten tarkkuus voi vaihdella: ne voidaan ilmaista vuoden, kuukauden, päivän, minuutin, sekunnin tai sekunnin murto-osan tarkkuudella. Silloin kun päiväykseen sisältyy kellonaika, on myös aikavyöhyke ilmaistava W3CDTF:n edellyttämällä tavalla. Suositeltavaa on käyttää päivän tarkkuutta.

Esimerkkejä hyväksyttävistä päiväyksistä:

2006	vuoden tarkkuudella
2006-11	kuukauden tarkkuudella
2006-11-18	päivän tarkkuudella
2006-11-18T16:49+02:00	minuutin tarkkuudella (ilmaistu Suomen talviajassa)
2006-11-18T14:49Z	minuutin tarkkuudella (sama hetki ilmaistu UTC-ajassa)
2006-06-12T12:34:56+03:00	sekunnin tarkkuudella (ilmaistu Suomen kesäajassa)

3.6 Julkaisija ja tekijä (`dc:publisher`, `dc:creator`)

Dokumentin julkaisija ja tekijä (organisaatio, henkilö tai ryhmä) voidaan ilmaista joko vapaatekstinä tai semanttisesti rikkaammin yksilöviittausten kautta. Kummassakin tapauksessa yksilöt kuvataan FOAF-ontologian luokkien `foaf:Person`, `foaf:Organization` ja `foaf:Group` avulla, mutta **vapaatekstimuotoiset** yksilömääritykset eivät sisällä täsmällistä tietoa siitä, milloin kysymys on samasta yksilöstä, kun taas **yksilöviittaukset** mahdollistavat mm. sen, että saman kirjoittajan artikkelit voidaan automaattisesti liittää toisiinsa, vaikka samassa organisaatiossa työskentelisi esimerkiksi kaksi Matti Virtanen -nimistä henkilöä.

Yksilön nimissä on noudatettava yhdenmukaista kirjoitusasua. Organisaatioiden nimet kirjoitetaan auki käyttämättä lyhenteitä (esim. *Terveysten ja hyvinvoinnin laitos* eikä *THL*). Henkilöiden nimet annetaan muodossa *Sukunimi, Etunimi*. Organisaatioiden nimistä voidaan tarjota useita erikielisiä versioita, kunhan ne on asianmukaisesti merkitty kielikoodein.

3.7 Kuuluu kokoelmaan (dcterms:isPartOf)

Julkaisu voi kuulua julkaisukokoelmaan, joka käsittelee esimerkiksi tiettyä aihepiiriä ja/tai on useiden organisaatioiden jäsenistä koostuvan työryhmän tuotos. Metatiedoissa julkaisukokoelmaan kuulumisen ilmaistaan *Kuuluu kokoelmaan* -suhteella. Viittauksen kohteena on kokoelma eli *PublicationCollection*-tyyppinen yksilö. Viittaus voidaan tehdä joko kokoelman URI-tunnisteella tai sen yksilöivällä nimellä.

3.8 Oikeudet (dc:rights)

Kenttä kertoo dokumenttiin liittyvistä tekijänoikeuksista ja käyttöoikeuksista. Suositeltavaa on antaa kentän arvoksi URI, joka viittaa dokumenttiin, jossa kuvataan käyttöoikeudet (esim. monilta sivustoilta löytyvä sivu *Käyttöoikeudet* tai *Tietoja sivustosta*). Myös kolmannen osapuolen julkaisemaan käyttöoikeustietoon, esim. Creative Commons⁹ -lisensseihin, voidaan viitata niiden URIa käyttämällä.

Kun kentän arvona on dokumentin URI, suositus on, että myös viittauksen kohteena olevasta dokumentista on olemassa tämän ohjeen mukaiset metatiedot.

Vaihtoehtoisesti kentän arvona voidaan antaa sanallinen kuvaus oikeuksista, esimerkiksi *Copyright UKK-instituutti. Kaikki oikeudet pidätetty..*

3.9 Tallennuskielto (ts:noindex)

Tällä kentällä voidaan ilmaista, että vaikka sisältö on kuvailtu metatiedoin, sitä ei tulisi kerätä TerveSuomi-portaaliin esimerkiksi siksi, että se on suunnattu vääränlaiselle kohderyhmälle tai julkaisija ei muusta syystä halua, että se tuodaan näkyville portaalissa.

Tallennuskielto on voimassa, mikäli tämä kenttä on olemassa ja sisältää arvon "true" tai "1". Muussa tapauksessa sisällön saa kerätä TerveSuomen sisältöjen joukkoon ja julkaista portaalissa.

3.10 Julkaisun laji (ts:genre)

Julkaisun laji kuvaa dokumentin käyttötarkoitusta tai genreä. Mahdolliset arvot on määritetty julkaisun laji -sanastossa. Mahdollisia arvoja ovat mm. *artikkeli*, *opas* ja *uutinen*.

3.11 Esitysmuoto (dc:type)

Esitysmuoto kuvaa kohteen rakennetta tai muotoa. Mahdollisia arvoja ovat DCMI-tyyppisanastossa määritellyt käsitteet kuten *Teksti* ja *Kuva*.

Kenttä on määritelty pakolliseksi, mutta kentän puuttuessa TerveSuomi-portaali käyttää oletusarvoa *Teksti* (Text). Lähes kaikki verkossa julkaistavat aineistot ovat Dublin Coren määritelmän mukaan tekstiä (vaikka niihin sisältyisi kuvitusta), mutta poikkeuksia ovat mm. videot ja sovellusohjelmat.

3.12 Formaatti (dc:format)

Formaatti-kentässä kerrotaan dokumentin MIME-tyyppi [[MIME, 2007](#)], esimerkiksi *text/html* tai *application/pdf*.

⁹<http://creativecommons.org/>

3.13 Aihe (dc:subject) ja Avainsana (ts:keyword)

Aihe-kenttä kertoo dokumentin käsittelemistä aiheista (vrt. asiasanat). Aihe-kentän arvoina käytetään aiheontologioiden ja -sanastojen (Tero, YSO, FinMeSH, TESA, STAMETA) käsitteitä. Suositus on, että jokaiselle dokumentille määritellään useita mahdollisimman kuvaavia aiheita.

Aiheet voidaan kuvailla pelkästään Terveyden edistämisen ontologialla (Tero), joka on portaalin käyttämä sanasto ja joka sisältää suurimman osan muista sanastoista (kaikki muut sanastot kokonaan paitsi FinMeSH:stä vain portaalin kannalta keskeiset osat). Vaihtoehtoisesti voidaan käyttää muita sanastoja tai niiden yhdistelmiä, mikä saattaa helpottaa integraatioita muihin järjestelmiin jotka eivät tunne Tero-ontologiaa, mutta vaikeuttaa sisältöjen kuvailua koska aiheita on haettava monesta eri sanastosta.

Avainsana-kenttä on tarkoitettu sellaisille aihetta kuvaaville asiasanoille, joita ei löydy portaalin käyttämistä sanastoista eli ns. vapaille asiasanoille. Kentän tarkoituksena on kerätä tietoa erityisesti Terveyden edistämisen ontologian laajennustarpeiden selvittämiseksi. Usein esiintyvät vapaat asiasanat voidaan lisätä Teroon, jos ne sisällöllisesti sinne sopivat. Kukin avainsana annetaan omalla arvonaan. Kielikodeja tulisi käyttää, jotta erikieliset, samalla tavoin kirjoitettavat avainsanat eivät sekaantuisi keskenään.

3.14 Kohdeyleisö (dcterms:audience)

Kohdeyleisö-kenttä kertoo, kenelle dokumentti on käsittelytavaltaan ja kirjoitustyyliältään suunnattu eli kuka on sen ajateltu lukija. Kohdeyleisöt valitaan käyttäen TerveSuomi-portaalin osioluokitusta, joka on muodostettu portaalin keskeisten kohderyhmien ja käyttötilanteiden mukaan. Julkaisulla on täsmälleen yksi kohdeyleisö/osio.

4 Metatiedon esittäminen RDF/XML-muodossa

RDF (Resource Description Framework) -tietomalli sopii erittäin hyvin metatietojen esittämiseen [RDF, 2009]. Allaoleva ohjeistus RDF/XML-muotoisen metatiedon esittämiseen perustuu Dublin Coren suosituksiin [DC-RDF, 2008].

Tässä esiteltyä RDF-muotoista metatietoa voi tuottaa joko RDF-työkaluin (esimerkiksi Jena-kirjasto¹⁰ Javalle) tai melko suoraviivaisesti myös XML-työkaluin, jotka eivät ymmärrä RDF:n erityisominaisuuksia.

Oli tiedon tuottamistapa mikä hyvänsä, on syytä kiinnittää erityistä huomiota siihen, että kenttien arvoissa esiintyvät erikoismerkit kuten kulmasulkeet ja lainausmerkit on korvattu XML-standardin mukaisilla koodilla (esim. `"` ja `>`). RDF-työkalut ja useimmat XML-työkalut huolehtivat tästä automaattisesti, mutta mikäli tietoja joudutaan tuottamaan merkkijono-operaatioin on vaara, että lopputulos ei noudata XML-standardia eikä näin ollen ole myöskään portaalin tulkittavissa.

Myös merkistöasiat on huomioitava RDF-muotoisen tiedon tuottamisessa. Yleensä XML-muotoiset dokumentit käyttävät UTF-8-merkistöä, mutta muun merkistön käyttö on mahdollista ilmoittamalla siitä XML-tiedoston otsakkeessa. Metatieto saa sisältää vain ilmoitetun merkistön mukaisia merkkejä. Muut merkit tulkitaan virhetilanteiksi.

4.1 RDF/XML-dokumentin rakenne

Metatietoa sisältävät RDF-dokumentit koostuvat XML-otsakkeesta, `rdf:RDF`-elementistä nimiavaruusesittelyineen sekä yhdestä tai useammasta `ts:Publication`-elementistä. Kukin `ts:Publication`-elementti sisältää yhden dokumentin metatiedot.¹¹ Kukin metatietokenttä esitetään omalla elementillään, jonka nimi vastaa metatietokentän tunnustetta.

RDF-dokumentin suositeltu rakenne on seuraava:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:ts="http://www.yso.fi/onto/tervesuomi-schema/">
  ... ts:Publication-elementtejä ...
</rdf:RDF>
```

4.2 Dokumentin tunniste ja URL-osoite

Kukin dokumentti kuvataan yhdellä `ts:Publication`-elementillä, jonka attribuutti `rdf:about` kertoo dokumentin identifiointitunnuksen eli URI:n. Sama URI esitetään myös `dc:identifier`-kentässä seuraavasti:

```
<ts:Publication rdf:about="http://www.ktl.fi/portal/8131">
  <dc:identifier rdf:resource="http://www.ktl.fi/portal/8131"/>
  ... muita metatietoja tästä dokumentista...
</ts:Publication>
```

Mikäli metatietoihin kuuluu dokumentin URL-osoite (`ts:url`), se esitetään `ts:url`-elementillä samaan tapaan kuin `dc:identifier` ylläolevassa esimerkissä.

¹⁰<http://jena.sourceforge.net>

¹¹Näin ei välttämättä tarvitse olla; RDF/XML mahdollistaa saman tietosisällön esittämisen hyvin monella eri tavalla. Yksinkertaisuuden vuoksi tämä dokumentti kuvaa sellaisen RDF/XML-dokumentin, joka on helppo tuottaa XML-työkaluin ja myös melko helposti ihmisen tulkittavissa. Muut, RDF-tietosisällöltään yhtäpitävät muodot ovat toki myös sallittuja.

4.3 Vapaatekstikentät

Vapaatekstikentät esitetään elementteinä, joiden tekstisisältönä on kentän sisältö. Mikäli kentän arvon kieli on tiedossa, se merkitään `xml:lang`-attribuutilla.

Esimerkki:

```
<dc:title>Rokottajan käsikirja 2005</dc:title>

<dc:description xml:lang="fi">Rokotteet kehittyvät ja tietomme niiden vaikutuksista lisääntyy. Infektiotautien epidemiologinen tilanne maailmassa ja Suomessa muuttuu, markkinoille tulee entistä parempia rokotteita</dc:description>
```

4.4 Rakenteiset merkkijonot

Rakenteiset merkkijonot tulee esittää siten, että elementin sisältönä on kentän arvo ja käytetty koodaustapa ilmaistaan `rdf:datatype`-attribuutilla. Esimerkiksi päivämäärä esitetään seuraavasti:

```
<dcterms:issued
  rdf:datatype="http://purl.org/dc/terms/W3CDTF">2008-12-17</dcterms:issued>
```

Esitysmuodon tunnisteet (URI) ovat seuraavat

Tyyppi	Esitysmuoto
Päiväys	http://purl.org/dc/terms/W3CDTF
Formaatti (MIME-tyyppi)	http://purl.org/dc/terms/IMT
Kieli	http://purl.org/dc/terms/RFC4646

4.5 Käsiteviittaukset

Käsiteviittaukset voidaan ilmaista kahdella tavalla: joko käsitteen yksikäsitteisellä tunnukseella (URI) tai ilmaisemalla sekä käsitteen yksilöivä termi että käsitteistö, johon se kuuluu.

Käsiteviittaukset tehdään ensisijaisesti viittaamalla käsitteen tunnisteeseen eli URIin. Tämä on standarditapa ilmaista käsiteviittaukset RDF:llä, ja näin ollen mikä tahansa RDF-muotoista tietoa käsittelevä sovellus ymmärtää sen.

Esimerkki (viittauksen kohteena YSO-käsite "lastenvaunut"):

```
<dc:subject rdf:resource="http://www.yso.fi/onto/yso/p12345"/>
```

Tällaisten käsiteviittausten tekeminen siis edellyttää, että käsitteen URI on tiedossa. TerveSuomi-portaalin tukemat käsitteistöt julkaistaan ONKI-palvelimella, jonka kautta käsitteitä voidaan hakea esimerkiksi niihin liittyvien termien perusteella, ja tätä kautta selvittää käsitteen URI. ONKI-palvelimelta voidaan selvittää käsitteiden URI-tunnisteita myös ohjelmallisesti mm. Web Service -rajapinnan kautta¹².

Jos käsitteen URI:n selvittäminen on mahdotonta esimerkiksi siksi, että tietojärjestelmään talletetaan vain termit, voidaan käsiteviittaus esittää myös termin ja käsitteistötunnuksen yhdistelemällä seuraavaan tapaan:

```
<dc:subject
  rdf:datatype="http://www.yso.fi/onto/tervesuomi-schema/YSO">lastenvaunut</dc:subject>
```

¹²Tarkempia tietoja esim. <http://www.yso.fi/onkiwebservice/?o=tero>

Tällöin termi ilmaistaan elementin sisältönä ja käsitteistö ilmaistaan `rdf:datatype`-attribuutilla, joka samalla kertoo termin kielen. Käsitteistötunnisteet käsiteviittauksille ovat seuraavat:

Käsitteistö	Käsitteistön tunnistee	Termin kieli
Julkaisun laji	http://www.yso.fi/onto/tervesuomi-schema/GENRE	suomi
Julkaisun laji	http://www.yso.fi/onto/tervesuomi-schema/GENRE-SV	ruotsi
Julkaisun laji	http://www.yso.fi/onto/tervesuomi-schema/GENRE-EN	englanti
DCMI Type Vocabulary	http://purl.org/dc/terms/DCMIType	englanti
TERO	http://www.yso.fi/onto/tervesuomi-schema/TERO	suomi
TERO	http://www.yso.fi/onto/tervesuomi-schema/TERO-SV	ruotsi
TERO	http://www.yso.fi/onto/tervesuomi-schema/TERO-EN	englanti
YSO	http://www.yso.fi/onto/tervesuomi-schema/YSO	suomi
YSO	http://www.yso.fi/onto/tervesuomi-schema/YSO-SV	ruotsi
YSO	http://www.yso.fi/onto/tervesuomi-schema/YSO-EN	englanti
FinMeSH	http://www.yso.fi/onto/tervesuomi-schema/FINMESH	suomi
SweMeSH	http://www.yso.fi/onto/tervesuomi-schema/SWEMESH	ruotsi
MeSH	http://purl.org/dc/terms/MESH	englanti
TESA	http://www.yso.fi/onto/tervesuomi-schema/TESA	suomi
TESA	http://www.yso.fi/onto/tervesuomi-schema/HPMULTI-SV	ruotsi
TESA	http://www.yso.fi/onto/tervesuomi-schema/HPMULTI	englanti
STAMETA	http://www.yso.fi/onto/tervesuomi-schema/STAMETA	suomi
STAMETA	http://www.yso.fi/onto/tervesuomi-schema/STAMETA-SV	ruotsi
STAMETA	http://www.yso.fi/onto/tervesuomi-schema/STAMETA-EN	englanti
TerveSuomi-osiot	http://www.yso.fi/onto/tervesuomi-schema/SECTION	suomi
TerveSuomi-osiot	http://www.yso.fi/onto/tervesuomi-schema/SECTION-SV	ruotsi
TerveSuomi-osiot	http://www.yso.fi/onto/tervesuomi-schema/SECTION-EN	englanti

4.6 Yksilöt

Yksilöt (henkilöt, organisaatiot, ryhmät ja julkaisukokoelmat) voidaan ilmaista vapaatekstityyppisesti esittämällä ne RDF-muodossa `foaf:Person`, `foaf:Organization`, `foaf:Group` tai `ts:PublicationCollection`-luokkien ilmentyminä seuraavaan tapaan:

```
<dc:creator>
  <foaf:Person>
    <foaf:name>Virtanen, Matti</foaf:name>
  </foaf:Person>
</dc:creator>

<dc:publisher>
  <foaf:Organization>
    <foaf:name xml:lang="fi">Kansanterveyslaitos</foaf:name>
    <foaf:name xml:lang="sv">Folkhälsoinstitutet</foaf:name>
    <foaf:name xml:lang="en">National Public Health Institute</foaf:name>
  </foaf:Organization>
</dc:publisher>

<dcterms:isPartOf>
  <ts:PublicationCollection>
    <dc:title xml:lang="fi">Pieni Päätös Päivässä</dc:title>
  </ts:PublicationCollection>
</dcterms:isPartOf>
```

Henkilön, organisaation tai ryhmän nimi ilmaistaan `foaf:name`-ominaisuudella. Julkaisukokoelman nimi ilmaistaan `dc:title`-ominaisuudella. Useita erikielisiä versioita nimistä voidaan antaa, kunhan ne on merkitty kielikoodein.

Vaihtoehtoisesti voidaan viitata erikseen määriteltymiin (ks. kohta 4.9) yksilöihin niiden URIn avulla:

```
<dc:publisher rdf:resource="http://www.ktl.fi/organization/1234"/>
```


4.7 Dokumentti- ja yksilöviittaukset

Dokumentteihin ja yksilöihin viitataan samaan tapaan kuin käsitteisiin viittaamalla niiden URIn. Esimerkki:

```
<dc:rights rdf:resource="http://www.ktl.fi/portal/suomi/palaute/tietoa_sivuista/">
```

4.8 Esimerkki RDF/XML-muotoisesta metatiedosta

Alla on esimerkki RDF/XML-muotoisesta metatietodokumentista, jossa ei ole käytetty organisaatiokohtaisia yksilömäärittäjiä). Esimerkki perustuu todelliseen dokumenttiin. Lisäksi XML-kommentein on kerrottu joidenkin käsitteiden päätermit, jotka helpottavat esimerkin ymmärtämistä.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:ts="http://www.yso.fi/onto/tervesuomi-schema/">
  <ts:Publication rdf:about="http://www.ktl.fi/portal/11945">
    <dc:identifier rdf:resource="http://www.ktl.fi/portal/11945"/>
    <ts:url rdf:resource="http://www.ktl.fi/portal/11945"/>

    <dc:title xml:lang="fi">Liukuesteet ja nastakengät antavat talvisin
    liikkumiseen varmuutta</dc:title>

    <dc:description xml:lang="fi">Useat tutkimukset Pohjoismaissa ja muualla
    ovat osoittaneet vuodenaikojen ja sääolojen vaihtelun merkityksen
    liukastumis- ja kaatumistapaturmien määrään ja sattumistiheyteen.
    Suomalaisen tutkimuksen mukaan liukastumisten...</dc:description>
    <dc:language rdf:datatype="http://purl.org/dc/terms/RFC4646">fi</dc:language>

    <dcterms:issued rdf:datatype="http://purl.org/dc/terms/W3CDTF">2006-11-27</dcterms:issued>

    <dc:publisher>
      <foaf:Organization>
        <foaf:name xml:lang="fi">Kansanterveyslaitos</foaf:name>
        <foaf:name xml:lang="sv">Folkhälsoinstitutet</foaf:name>
        <foaf:name xml:lang="en">National Public Health Institute</foaf:name>
      </foaf:Organization>
    </dc:publisher>
    <dc:creator>
      <foaf:Person>
        <foaf:name>Grönqvist, Raoul</foaf:name>
      </foaf:Person>
    </dc:creator>
    <dc:creator>
      <foaf:Person>
        <foaf:name>Juntunen, Pauliina</foaf:name>
      </foaf:Person>
    </dc:creator>
    <dc:creator>
      <foaf:Person>
        <foaf:name>Mattila, Susanna</foaf:name>
      </foaf:Person>
    </dc:creator>
    <dc:rights xml:lang="fi">http://www.ktl.fi/portal/208</dc:rights>

    <ts:genre rdf:resource="http://www.yso.fi/onto/tervesuomi-genre/article"/> <!-- artikkeli -->
    <dc:type rdf:resource="http://purl.org/dc/dcmitype/Text"/> <!-- Teksti -->
    <dc:format rdf:datatype="http://purl.org/dc/terms/IMT">text/html</dc:format>

    <dc:subject rdf:resource="http://www.yso.fi/onto/tero/p6784"/> <!-- talvi -->
    <dc:subject rdf:resource="http://www.yso.fi/onto/tero/D000368"/> <!-- iäkkäät -->
    <dc:subject rdf:resource="http://www.yso.fi/onto/tero/D000369"/> <!-- yli 80v iäkkäät -->
    <dc:subject rdf:resource="http://www.yso.fi/onto/tero/p3664"/> <!-- jalkineet -->
    <dc:subject rdf:resource="http://www.yso.fi/onto/tero/p14545"/> <!-- tapaturmantorjunta -->
    <dc:subject rdf:resource="http://www.yso.fi/onto/tero/T00392"/> <!-- kaatumiset -->
    <dcterms:audience rdf:resource="http://www.yso.fi/onto/tervesuomi-section/s1"/> <!-- Voi hyvin -->
  </ts:Publication>
</rdf:RDF>
```

4.9 Organisaatiokohtaisten yksilöiden esittäminen

Organisaatiokohtaiset yksilöt esitetään siten, että kullekin yksilölle annetaan oma yksilöivä URI-tunniste ja lisäksi yksilöstä kerrotaan metatietoa RDF-muodossa. Organisaatio vastaa siitä, että yksilöiden URI-tunnisteet ovat pysyviä ja yksiselitteisiä.

Yksittäisen henkilön tiedot esitetään FOAF-skeemaa [Brickley and Miller, 2005] hyödyntäen seuraavasti:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <foaf:Person rdf:about="http://www.ktl.fi/person/1234">
    <foaf:name>Nohynek, Hanna</foaf:name>

    <!-- ... mahd. lisää tietoa henkilöstä FOAF:in avulla ilmaistuna ... -->

  </foaf:Person>
</rdf:RDF>
```

Yksittäisen organisaation tiedot esitetään FOAF-skeeman [Brickley and Miller, 2005] avulla seuraavasti:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <foaf:Organization rdf:about="http://www.ktl.fi/organization/1234">
    <foaf:name>Kansanterveyslaitos</foaf:name>

    <!-- Hanna Nohynek kuuluu ao. organisaation resursseihin (henkilökuntaan) -->
    <foaf:member rdf:resource="http://www.ktl.fi/person/1234" />

    <!-- ... mahd. lisää tietoa organisaatiosta FOAF:in avulla ilmaistuna ... -->

  </foaf:Organization>
</rdf:RDF>
```

FOAF on vakiintunut tapa ilmaista henkilötietoja RDF-muodossa, mutta erityisesti ryhmien ja organisaatioiden kuvaamista ajatellen se ei välttämättä sisällä kaikkia tarvittavia tietokenttiä. Tämän vuoksi yksilöiden määrittelytapaa saatetaan tulevaisuudessa laajentaa.

Julkaisukokoelman tiedot esitetään seuraavasti:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ts="http://www.yso.fi/onto/tervesuomi-schema/">

  <ts:PublicationCollection rdf:about="http://www.kaypahoito.fi/collection/khp">
    <dc:title xml:lang="fi">Käypä hoito, potilasversiot</dc:title>
    <dc:title xml:lang="sv">God medicinsk praxis, patientversionerna</dc:title>

    <!-- ... mahd. lisää tietoa kokoelmasta esim. DC:n avulla ilmaistuna... -->

  </ts:PublicationCollection>
</rdf:RDF>
```

5 Metatiedon esittäminen osana HTML-sivua

Mikäli edellä kuvattua RDF-muotoista metatietodokumenttia ei voida tai haluta tuottaa erillisenä dokumenttina, voidaan metatieto myös sisällyttää HTML-sivun `head`-osioon. Tämä ratkaisu noudattaa Dublin Coren suositusta [DC-HTML, 2008] metatiedon esittämiseen HTML-sivun sisällä.

Menetelmän hyvänä puolena on se, että monet julkaisujärjestelmät tukevat Dublin Core -metatiedon sisällyttämistä HTML-sivuihin jo valmiiksi. Menetelmän rajoituksena on, ettei sitä voi soveltaa muihin kuin HTML/XHTML-muotoisiin dokumentteihin. Esimerkiksi PDF-dokumenttien metatiedot on kuvattava muulla tavalla. Menetelmän avulla ei myöskään voida määritellä tietoja yksilöistä kuten henkilöistä ja organisaatioista, vaikka muualla RDF-muodossa (ks. kohta 4.9) määriteltyihin yksilöihin voidaan viitata.

Useimmat ohjeet koskevat sekä HTML- että XHTML-dokumentteja. Mikäli näiden esitysmuoto eroaa, se on mainittu erikseen. Esimerkit ovat XHTML-syntaksin mukaisia. HTML-muoto eroaa näistä siten, että elementin sulkemista merkitsevä kauttaviiva `</>` jää pois.

HTML-tiedostossa tulee esitellä käytetty merkistö (esim. UTF-8 tai ISO-8859-1) käytössä olevan HTML-version mukaisella standarditavalla. HTML-dokumentin sisältämä metatieto ilmaistaan käyttäen samaa merkistöä kuin itse dokumenttikin. Muut merkit tulkitaan virhetilanteiksi.

5.1 Kenttätunnisteet

Kenttätunnisteet HTML-muodossa poikkeavat jonkin verran RDF-muodosta. Tunnistemuoto perustuu Dublin Coren suositukseen [DC-HTML, 2008]¹³. HTML-muodossa käytettävät kenttätunnisteet on kuvattu taulukossa 3.

¹³Tunnisteista on olemassa myös vanhempi muoto aiemmasta Dublin Coren suosituksesta [DCQ-HTML, 2000], joka on käytössä myös Dublin Coren suomalaisessa versiossa [SFS 5895, 2001]. Aiemmassa muodossa esimerkiksi Julkistamisaika-kentän tunniste on `DC.Date.issued` eikä `DCTERMS.issued`. Tämä dokumentti ei ota kantaa siihen, hyväksyykö TerveSuomi-portaali muita kenttätunnistemuotoja kuin tässä dokumentissa kuvatut muodot.

Taulukko 3: Metatietokenttien HTML-tunnisteet

Kenttä	Tunniste	HTML-muoto
Identifiointitunnus	dc:identifier	DC.identifier
URL-osoite	ts:url	TS.url
Nimeke	dc:title	DC.title
Kuvaus	dc:description	DC.description
Kieli	dc:language	DC.language
Julkistamisaika	dcterms:issued	DCTERMS.issued
Hyväksymisaika	dcterms:dateAccepted	DCTERMS.dateAccepted
Muokkusaika	dcterms:modified	DCTERMS.modified
Julkaisija	dc:publisher	DC.publisher
Tekijä	dc:creator	DC.creator
Kuuluu kokoelmaan	dcterms:isPartOf	DCTERMS.isPartOf
Oikeudet	dc:rights	DC.rights
Tallennuskielto	ts:noindex	TS.noindex
Julkaisun laji	ts:genre	TS.genre
Esitysmuoto	dc:type	DC.type
Formaatti	dc:format	DC.format
Aihe	dc:subject	DC.subject
Avainsana	ts:keyword	TS.keyword
Kohdeyleisö	dcterms:audience	DCTERMS.audience

5.2 Profilimäärittely

HTML-dokumentin HEAD-elementissä ilmaistaan käytetty metatietoprofiili `profile`-elementillä. Koska TerveSuomen käyttämä metatiedon esitystapa on Dublin Coren suosituksen [DC-HTML, 2008] mukainen, viitataan Dublin Coren profiiliin seuraavasti:

```
<head profile="http://dublincore.org/documents/2008/08/04/dc-html/">
```

5.3 Nimiavaruusmäärittelyt

Kenttätunnisteissa käytettävät nimiavaruudet esitellään `link`-elementein seuraavasti:

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<link rel="schema.TS" href="http://www.yso.fi/onto/tervesuomi-schema/" />
```

Tämä esittely on vakio eli toistuu sellaisenaan kaikissa metatietoa sisältävissä dokumenteissa.

5.4 Dokumentin tunniste ja URL-osoite

Dokumentin tunniste (`dc:identifier`) ja URL-osoite (`ts:url`) esitetään `meta`-elementillä seuraavasti:

Pysyvä tunniste:

```
<meta name="DC.identifier" scheme="DCTERMS.URI" content="http://www.ktl.fi/portal/8131" />
```

Tämänhetkinen URL-osoite:

```
<meta name="TS.url" scheme="DCTERMS.URI"
content="http://www.ktl.fi/portal/suomi/julkaisut/oppaat_ja_kirjat/rokottajan_kasikirja/" />
```

5.5 Vapaatekstikentät

Vapaatekstikentät esitetään `meta`-elementillä. Mikäli tekstin kieli on tiedossa, ilmaistaan se erityisellä kielimääreellä. HTML-muodossa kielimääreenä toimii `lang`-attribuutti. XHTML Strict -muodoissa käytetään `xml:lang`-attribuuttia. XHTML Transitional -muodoissa käytetään molempia attribuutteja. Esimerkki:

HTML:

```
<meta name="DC.title" lang="fi" content="Rokotteiden hävittäminen">
```

XHTML Strict:

```
<meta name="DC.title" xml:lang="fi" content="Rokotteiden hävittäminen" />
```

XHTML Transitional:

```
<meta name="DC.title" lang="fi" xml:lang="fi" content="Rokotteiden hävittäminen" />
```

Kielikoodit ovat RFC 4646 -standardin mukaisia, kuten Kieli-kentän (`dc:language`) arvot [RFC 4646, 2006].

5.6 Rakenteiset merkkijonot

Rakenteiset merkkijonot esitetään `meta`-elementillä. Kentän tunnisteeseen ja arvon lisäksi kerrotaan `schema`-attribuutilla kentän arvojoukon tunniste.

Esitysmuotojen tunnisteet rakenteisille merkkijonoille ovat seuraavat:

Tyyppi	scheme-tunniste
Päiväys	DCTERMS.W3CDTF
Formaatti (MIME-tyyppi)	DCTERMS.IMT
Kieli	DCTERMS.RFC4646

Esimerkki rakenteisista merkkijonoista:

```
<meta name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2004-12-17" />
<meta name="DC.language" scheme="DCTERMS.RFC4646" content="fi" />
```

5.7 Tallennuskielto

Tallennuskielto-kenttä (ts:noindex) esitetään tarvittaessa seuraavasti:

```
<meta name="TS.noindex" content="true" />
```

Mikäli tallennuskielto ei ole voimassa, jätetään meta-elementti kokonaan pois tai asetetaan content-attribuutin arvoksi false.

5.8 Käsite-, dokumentti- ja yksilöviittaukset URI-tunnisteella

Käsite-, dokumentti- ja yksilöviittaukset esitetään link-elementillä seuraavaan tapaan:

```
<link rel="DC.subject" href="http://www.yso.fi/onto/yso/a12345" />
<link rel="DC.rights" href="http://www.ktl.fi/portal/suomi/palaute/tietoa_sivuista/" />
```

Itse yksilöä koskevat tiedot (esim. henkilön tai organisaation nimi ja mahdolliset yhteystiedot) tulee määrittellä URI-tunnisteen osoittamassa URL-osoitteessa, RDF-muodossa. Tämä johtuu siitä, että HTML:n metatietomäärittelyjen kohde on aina dokumentti itse, joten dokumentista riippumattomaa yksilöä koskevan metatiedon antaminen ei HTML:ssä onnistu.

5.9 Yksilöviittaukset vapaatekstimuodossa

Yksilöt (henkilöt, organisaatiot, ryhmät ja julkaisukokoelmat) voidaan ilmaista vapaatekstityyppisesti meta-elementillä, jonka arvona on yksilön nimi. Tällöin scheme-attribuutti kertoo yksilön tyyppin seuraavien tunnisteiden avulla:

Tyyppi	scheme-tunniste
Henkilö	TS.FOAFPerson
Organisaatio	TS.FOAFOrganization
Ryhmä	TS.FOAFGroup
Julkaisukokoelma	TS.TSPublicationCollection

Useita erikielisiä versioita nimistä voidaan antaa, kunhan ne on merkitty kielikoodein samaan tapaan kuin vapatekstikenttien kohdalla.

Esimerkki yksilöistä:

```
<meta name="DC.creator" scheme="TS.FOAFPerson" content="Virtanen, Matti" />
<meta name="DC.publisher" scheme="TS.FOAFOrganization" xml:lang="fi"
content="Kansanterveyslaitos" />
<meta name="DC.publisher" scheme="TS.FOAFOrganization" xml:lang="sv"
content="Folkhälsoinstitutet" />
<meta name="DC.publisher" scheme="TS.FOAFOrganization" xml:lang="en"
content="National Public Health Institute" />
<meta name="DCTERMS.isPartOf" scheme="TS.TSPublicationCollection"
```

5.10 Käsiteviittaukset termin avulla

Käsiteviittaukset voidaan URI-viittauksen sijaan tai lisäksi esittää **meta**-elementillä. Jokaista käsitettä varten on tällöin oma **meta**-elementtinsä. Tällaisia viittauksia saatetaan hyödyntää myös muissa sovelluksissa kuin TerveSuomessa, joten niiden lisääminen URI-käsiteviittausten rinnalle voi olla järkevää. TerveSuomi-portaali pystyy hyödyntämään molemmin tavoin esitettyjä viittauksia, eikä ole haitaksi, jos sama tieto esitetään kahteen kertaan.

Kun käsiteviittaukset tehdään **meta**-elementeillä, tulkitaan tämä TerveSuomessa hyödyntäen annetun käsitteistön viimeisintä versiota ja muunnetaan URI-muotoon. Näin tehdyt viittaukset voivat lakata toimimasta, jos esimerkiksi käsitteiden nimet muuttuvat. Tiedontuottaja vastaa siitä, että metatieto hyödyntää aina viimeisintä käsitteistöä ja että käsitteet joihin viitataan ovat sillä hetkellä olemassa.

Esimerkki **meta**-elementillä esitetyistä käsitteistä:

```
<meta name="DC.subject" scheme="TS.TESA" content="lapset" />
<meta name="DC.subject" scheme="TS.TESA" content="nuoret" />
<meta name="DC.subject" scheme="TS.TESA" content="aikuiset" />
```

Kentän tunniste (name) ja arvon (content) lisäksi kerrotaan **scheme**-attribuutilla sen käsitteistön tunniste, johon kentän arvo kuuluu. Kentän arvo on käsitteeseen liittyvä ensisijainen termi (preferred term) täsmälleen siinä muodossa, jossa se esiintyy käsitteistössä. Joihinkin kenttiin voi soveltaa useita käsitteistöjä tai samoihin käsitteisiin voi viitata erikielisten termien kautta (englanninkielinen MeSH vs. suomenkielinen FinMeSH).

Käsitteistötunnisteet käsiteviittauksille ovat seuraavat:

Kenttä	Käsitteistö	Käsitteistön tunniste	Termin kieli
Julkaisun laji	Julkaisun laji	TS.GENRE	suomi
Julkaisun laji	Julkaisun laji	TS.GENRE-SV	ruotsi
Julkaisun laji	Julkaisun laji	TS.GENRE-EN	englanti
Esitysmuoto	DCMI Type Vocabulary	DCTERMS.DCMIType	englanti
Aihe	TERO	TS.TERO	suomi
Aihe	TERO	TS.TERO-SV	ruotsi
Aihe	TERO	TS.TERO-EN	englanti
Aihe	YSO	TS.YSO	suomi
Aihe	YSO	TS.YSO-SV	ruotsi
Aihe	YSO	TS.YSO-EN	englanti
Aihe	FinMeSH	TS.FINMESH	suomi
Aihe	SweMeSH	TS.SWEMESH	ruotsi
Aihe	MeSH	DCTERMS.MESH	englanti
Aihe	TESA	TS.TESA	suomi
Aihe	TESA	TS.HPMULTI-SV	ruotsi
Aihe	TESA	TS.HPMULTI	englanti
Aihe	STAMETA	TS.STAMETA	suomi
Aihe	STAMETA	TS.STAMETA-SV	ruotsi
Aihe	STAMETA	TS.STAMETA-EN	englanti
Kohdeyleisö	TerveSuomi-osiot	TS.SECTION	suomi
Kohdeyleisö	TerveSuomi-osiot	TS.SECTION-SV	ruotsi
Kohdeyleisö	TerveSuomi-osiot	TS.SECTION-EN	englanti

5.11 Esimerkki metatiedosta HTML-dokumentissa

Alla on esimerkki XHTML Strict -muotoisesta dokumentista, joka sisältää metatietoa. Esimerkki perustuu todelliseen dokumenttiin, jonka metatietoja on kuitenkin täydennetty ja muokattu vastaamaan tämän dokumentin suosituksia sekä XHTML Strict -muotoa.

Käsiteviittaukset on esimerkinomaisesti tehty vaihdellen joko meta- tai link-elementillä tai molemmilla. Esimerkiksi dc:type-kenttä on esitetty ainoastaan meta-elementillä kun taas ts:genre-kenttä on esitetty ainoastaan link-elementillä.

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head profile="http://dublincore.org/documents/2008/08/04/dc-html/">

    <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
    <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
    <link rel="schema.TS" href="http://www.yso.fi/onto/terveysuomi-schema/" />

    <meta name="DC.identifier"
        content="http://www.kaypahoito.fi/kh/kaypahoito?suositus=khp00003" />

    <meta name="DC.title" content="Astma" xml:lang="fi" />

    <meta name="DC.description" content="Astma on keuhkoputkien
tulehdussairaus. Perimä altistaa sen puhkeamiselle. Hoitona on
hengitysoireiden aiheuttajan välttäminen ja riittävä lääkitys."
xml:lang="fi" />

    <meta name="DC.language" content="fi" scheme="DCTERMS.RFC4646" />

    <link rel="TS.genre" href="http://www.yso.fi/onto/terveysuomi-genre/recommendation" />
    <meta name="DC.type" content="Text" scheme="DCTERMS.DCMIType" />
    <meta name="DC.format" content="text/html" scheme="DCTERMS.IMT" />

    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p10851" /> <!--lääkehoito-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p17099" /> <!--siedätyshoito-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p6194" /> <!--ympäristötekijät-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p14134" /> <!--diagnoosi-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p2414" /> <!--oireet-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p19552" /> <!--ennaltaehkäisy-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/T00843" /> <!--passiiv. tupakointi-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p13277" /> <!--riskitekijät-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p19733" /> <!--keuhkoputket-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p824" /> <!--hoito (huolenpito)-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p4354" /> <!--lapset (ikään l.r)-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p10017" /> <!--tupakointi-->
    <link rel="DC.subject" href="http://www.yso.fi/onto/tero/p19468" /> <!--astma-->

    <link rel="DCTERMS.audience" href="http://www.yso.fi/onto/terveysuomi-section/s2" />
    <!-- Sairaudet ja oireet -->

    <meta name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2007-01-03" />

    <meta name="DC.publisher" scheme="TS.FOAFOrganization" content="Duodecim" xml:lang="fi" />

    <meta name="DC.creator" scheme="TS.FOAFPerson" content="Riikola, Teija" />
    <meta name="DC.creator" scheme="TS.FOAFPerson" content="Kaila, Minna" />

    <meta name="DCTERMS.isPartOf" scheme="TS.TSPublicationCollection"
        content="Käypä hoito, potilasversiot" />

    <link rel="DC.rights"
        href="http://www.kaypahoito.fi/kotisivut/sivut.nayta?p_navi=2146&p_sivu=15021" />

    <title xml:lang="fi">Astma</title>

    <!-- stylesheet ym. määrityksiä -->
</head>
<body>
    <!-- varsinainen dokumentin sisältö -->
</body>
</html>
```

6 Sanastot ja ontologiat

TerveSuomen hajautettu sisällöntuotanto perustuu yhteisten sanastojen hyödyntämiseen, joissa määritellään sisällön kuvailuun käytettävät käsitteet. Tässä luvussa esitellään kaikki käytettävät sanastot ja ontologiat. Ne löytyvät myös ontologiapalvelimelta www.yso.fi.

Ontologiapalvelin tarjoaa ontologioihin pääsyn sekä selailtavan ONKI-web-käyttöliittymän kautta, Web Service -rajapintojen kautta (kts. web-käyttöliittymän sivu Tietoa Onkista) että koneluettavassa RDF-muodossa.

Kunkin ontologian URL-osoite on muotoa

<http://www.yso.fi/onto/tunniste/>
missä *tunniste* on annettu taulukossa 4.

Taulukko 4: TerveSuomi-portaalin käyttämät ontologiat

Nimi	Tunniste	Kuvaus
Terveystiedon edistämisen ontologia (Tero)	tero	Terveystiedon edistämisen ontologia on koostettu Yleisestä suomalaisesta ontologiasta, Terveystiedon edistämisen sanastosta, osasta MeSHiä sekä muista terveysalan sanastoista.
Yleinen suomalainen ontologia (YSO)	yso	Yleinen suomalainen ontologia (YSO) sisältää suuren joukon jokapäiväiseen elämään liittyviä käsitteitä. Se toimii TerveSuomessa ns. yläontologia- na, johon muut ontologiat, kuten MeSH ja TESA, kytketään. Yhteinen yläontologia edistää ontologioiden välistä yhteentoimivuutta, mikä edistää mm. loppukäyttäjälle näkyvän käyttöliittymän sisällöllistä yhtenäisyyttä.
MeSH, FinMeSH ja SweMeSH	mesh	MeSH ^a (Medical Subject Headings) on kansainvälinen lääketieteellinen käsitteistö. FinMeSH on sen suomenkielinen käännös ja SweMeSH vastaavasti ruotsinkielinen käännös.
Terveystiedon edistämisen sanasto (TESA/HPMULTI)	hpmulti	Terveystiedon edistämisen sanasto ^b on Euroopan komission julkaisema sanasto terveyden edistämiseen liittyvän materiaalin luokitteluun.
STAMETA	stameta	STAMETA on Stakesin (nykyinen THL) julkaisema sosiaalialan asiasanasto.
TerveSuomi-portaalin osioluokitus	tervesuomi-section	TerveSuomi-portaalin osiot on valittu siten, että ne vastaavat portaalin keskeisiä kohderyhmiä.
Julkaisun laji -sanasto (Genre)	tervesuomi-genre	Julkaisun laji -sanastolla kuvataan julkaisun pääasiainen sisällöllinen laji tai genre.
DCMI-tyyppisanasto	dcmitype	DCMI-tyyppisanasto (DCMI Type Vocabulary) on Dublin Coren julkaisema yleiskäyttöinen, sovelsalarippumaton luokitus metatiedon kohteiden luokitteluun. TerveSuomi-portaali käyttää DCMI-tyyppisanastoa sellaisenaan. Ontologiapalvelimella olevaan versioon on lisätty suomenkieliset käännökset termeistä ja määritelmistä.

^a<http://www.nlm.nih.gov/mesh/>

^b<http://www.hpmulti.net>

7 Apuvälineet metatiedon tuottamiseen

Tämän määrittelyn mukaisen metatiedon tuottamisen helpottamiseksi SeCo-tutkimusryhmässä on kehitetty apuvälineitä ja -ohjelmistoja. ONKI-palvelin tarjoaa pääsyn ajantasaisiin sanastoihin ja ontologioihin ja tavoitteena sen kehittämisessä on, että sisällöntuottajien ei tarvitse toteuttaa sanastoihin liittyviä toiminnallisuuksia omissa järjestelmissään, vaan he voivat yksinkertaisesti ottaa käyttöön ONKIn tarjoamat toiminnallisuudet.

Metatietoa voidaan lisäksi tuottaa Saha-metatietoeditorilla, jolloin tiedontuottajan omaa järjestelmää ei tarvitse muuttaa. Tällöin kuitenkin yhteys varsinaisen sisällön ja metatiedon välillä jää löyhäksi, joten sisällöntuottajan on erikseen huolehdittava siitä, että metatietoa päivitetään sisällön muuttuessa.

7.1 ONKI-palvelimen toiminnallisuudet

ONKI-palvelimella säilytetään TerveSuomi-portaalin käyttämiä asiasanastoja, luokituksia ja ontologioita sekä muita kansallisia tai kansainvälisiä sanastoja. Lisätietoja ONKista löytyy ONKI-hankkeen kotisivulta¹⁴. Itse sanastot löytyvät osoitteesta <http://www.yso.fi>.

7.1.1 ONKI-selain

Yksinkertainen tapa käyttää ONKIn sanastoja on *ONKI-selain* eli selainkäyttöliittymä, jolla sanastoja voi tutkia ja etsiä niistä käsitteitä. ONKIn selainkäyttöliittymä aukeaa kunkin ontologian URI-tunnisteen mukaisessa osoitteessa; esimerkiksi Terveystieteen edistämisen ontologiaa voi katsella osoitteessa <http://www.yso.fi/onto/tero/>. Selainkäyttöliittymä sopii hyvin asiansanoitus- ja luokittelutyön apuvälineeksi erityisesti silloin, kun sisällönhallintajärjestelmään ei ole integroitu muita ONKI-toiminnallisuuksia.

7.1.2 ONKI-valitsin

Uuseimmissä sisällönhallintajärjestelmissä metatietojen syöttö tapahtuu web-lomakkeella. Tällaisen lomakkeen kenttiin voidaan pienellä määrällä JavaScript-koodia liittää *ONKI-valitsin*, valmis JavaScript-komponentti joka muodostaa yhteyden ONKI-palvelimeen ja näkyy käyttäjälle siten, että hänen kirjoittaessaan kenttään asiansanojen ensimmäisiä kirjaimia hänelle esitetään mahdollisia asiansanoja, joista hän voi valita oikean (kuva 1). Lisäksi ONKI-valitsimeen kuuluu nappi, joka avaa ONKI-selaimen erilliseen ikkunaan. Kun ONKI-selain avataan valitsimen kautta, selaimesta voi myös poimia yksittäisen käsitteen ja palauttaa sen Nouda käsite -napilla takaisin metatietolomakkeelle. ONKI-valitsimen avulla sisällönhallintajärjestelmään kuuluva metatietojen syöttölomake saadaan siis tietoiseksi sanastoista ilman, että järjestelmään tarvitsee tehdä isoja muutoksia.

Lisätietoja ONKI-valitsimesta ja sen käytöstä löytyy ONKI-wikistä¹⁵, Terveystieteen edistämisen ontologian ONKI-valitsin-sivulta¹⁶ sekä SeCo-ryhmän ONKIä käsittelevistä tieteellisistä artikkeleista [Viljanen et al., 2008, Tuominen et al., 2009].

7.1.3 ONKI Web Service -rajapinta

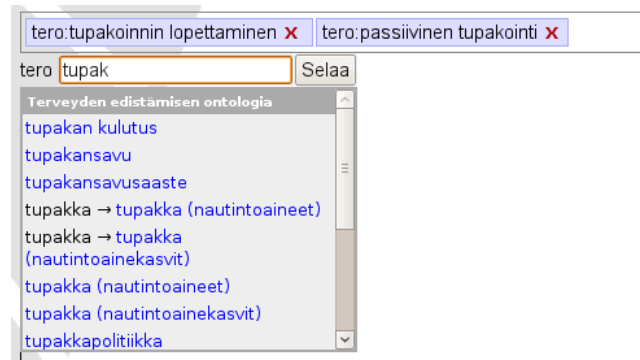
ONKI tarjoaa myös Web Service -tyyppisen rajapinnan käsitteiden hakuun ja tarkasteluun. Rajapintaa voidaan käyttää sellaisiin integraatioihin, joihin ONKI-valitsin ei syystä tai toisesta sovellu. Lisätietoja Web Service -rajapinnasta löytyy Terveystieteen edistämisen ontologian Web Service -rajapintasivulta¹⁷.

¹⁴<http://www.seco.tkk.fi/services/onki/>

¹⁵<http://www.yso.fi/wiki/>

¹⁶<http://www.yso.fi/onkiselector/?id=tero&l=fi>

¹⁷<http://www.yso.fi/onkiwebservice/?o=tero&l=fi>



Kuva 1: ONKI-valitsin täydentää automaattisesti käyttäjän syöttämiä merkkejä ehdottaen mahdollisia ONKI-palvelimen tuntemia käsitteitä.

7.2 Saha-metatietoeditori

Saha-metatietoeditori tarjoaa vaihtoehdoisen tavan tuottaa TerveSuomen tarvitsemaa metatietoa. Sitä voidaan käyttää tilanteessa, jossa metatiedon tuottaminen varsinaisella sisällönhallintajärjestelmällä ei onnistu esimerkiksi kustannussyistä. Saha-editoria käytetään tavallisella selaimella. Lisätietoja Saha-editorista löytyy sen kotisivulta¹⁸.

7.3 Palauteraportit

TerveSuomi-portaalissa tavoitteena on tiedon ja metatiedon laadun parantaminen sen alkulähteillä. Tiedon julkaisijat saavat käyttöönsä säännöllisesti päivittyvän raportin omissa metatiedoissa esiintyvistä ongelmista, jotka haittaavat metatietojen käyttöä TerveSuomessa tai estävät sen kokonaan. Palauteraportit ovat muodoltaan www-sivustoja.

¹⁸<http://www.seco.tkk.fi/services/saha/>

8 Kenttien vastaavuudet JHS 143- ja Sähke2-malleissa

TerveSuomen metatietokentät on valittu siten, että ne vastaavat läheisesti kansallisia JHS 143- [JHS 143, 2004] ja Sähke2-metatietomalleja [SÄHKE2, 2008]. Kaikkia näissä malleissa esiintyviä kenttiä TerveSuomi ei kuitenkaan tarvitse ja toisaalta TerveSuomen metatietomallissa on valinnaisia kenttiä, joille ei aina löydy vastinetta JHS 143:sta tai Sähke-mallista. Vastaavuudet mallien välillä on esitetty taulukossa 5.

Vaikka TerveSuomen käyttämä metatietokenttä vastaisi jotakin olemassaolevan JHS 143:n tai Sähke2-mallin mukaan tehdyn järjestelmän kenttää, tämä ei välttämättä tarkoita, että kentän arvot suoraan soveltuisivat TerveSuomen käyttöön. Kenttien arvojoukot tai niihin liittyvät muut rajoitteet eivät välttämättä ole yhteneväiset. Sekä JHS 143 että Sähke2 ovat kehyksiä, jotka jättävät tilaa sovelluskohtaisille ratkaisuille, joten edellä mainittuja malleja käyttävien sovellusten yhteensopivuus on aina varmistettava tarkastelemalla kenttien sovelluskohtaisia määrittelyjä. TerveSuomen tapauksessa sovelluskohtaiset rajoitteet, arvojoukot ja käytännöt on kuvattu edempänä tässä määrittelydokumentissa.

Taulukko 5: Metatietokenttien vastaavuudet JHS 143- ja Sähke2-malleissa

TerveSuomi-kenttä	JHS 143 kenttä/tarkenne	Sähke2 kenttä/tarkenne	Huomautuksia
Identifointitunnus	Identifointitunnus	Identifointitunnus	Vaihtoehtoinen tunnistemuoto
URL-osoite	Identifointitunnus	Identifointitunnus	
Nimeke Kuvaus Kieli	Nimeke Kuvaus Kieli	Nimeke Kuvaus Kieli	
Julkistamisaika Hyväksymisaika Muokkausaika	Aikamääre/Julkistamisaika Aikamääre/Hyväksymisaika Aikamääre/Muokkaamisaika	Aikamääre/Julkistettu Aikamääre/Hyväksytty Aikamääre/Muokattu	
Julkaisija Tekijä Kuuluu kokoelmaan	Toimija/julkaisija Toimija/tekijä Suhde/Kuuluu	Toimija/julkaisija Toimija/laatija Suhde/Kuuluu	Sisältää/Kuuluu-tarkenteet ilmeisesti määritelty ristiin JHS 143:ssa. TerveSuomi käyttää tätä suhdetta isPartOf-mielessä. JHS 143 Julkisuus- ja Sähke2 Julkisuusluokka -kenttien käyttötarkoitus on hiukan samankaltainen.
Oikeudet Tallennuskielto	Oikeudet -	Oikeudet -	
Julkaisun laji Esitysmuoto Formaatti	Asiakirjan tyyppi Laji Formaatti/Tiedostomuoto	Tyyppi - -	
Aihe Avainsana	Aihe Aihe	Aihe Aihe	Muuten kuten normaali Aihe-kenttä, mutta ilman määriteltyä merkintäjärjestelmää (sanastoa).
Kohdeyleisö	Kohdeyleisö	Kohdeyleisö	

Viitteet

- [Brickley and Miller, 2005] Brickley, D. and Miller, L. (2005). FOAF vocabulary specification. <http://xmlns.com/foaf/0.1/>.
- [DC-HTML, 2008] DC-HTML (2008). Expressing Dublin Core metadata using HTML/XHTML meta and link elements. <http://dublincore.org/documents/dc-html/>.
- [DC-RDF, 2008] DC-RDF (2008). Expressing Dublin Core metadata using the Resource Description Framework (RDF). <http://dublincore.org/documents/dc-rdf/>.
- [DC-SFS, 2002] DC-SFS (2002). Dublin Core metadataformaatin suomalainen versio. http://www.lib.helsinki.fi/dublin_core/dc-sfs.html.
- [DCES, 2008] DCES (2008). Dublin Core metadata element set, version 1.1. <http://dublincore.org/documents/dces/>.
- [DCQ-HTML, 2000] DCQ-HTML (2000). Recording qualified Dublin Core metadata in HTML meta elements. <http://dublincore.org/documents/2000/08/15/dcq-html/>.
- [DCTERMS, 2008] DCTERMS (2008). DCMI metadata terms. <http://dublincore.org/documents/dcmi-terms/>.
- [GRDDL, 2007] GRDDL (2007). Gleaning resource descriptions from dialects of languages (grddl). <http://www.w3.org/TR/grddl/>.
- [ISO 8601, 1988] ISO 8601 (1988). ISO 8601: Data elements and interchange formats - information interchange - representation of dates and times.
- [JHS 143, 2004] JHS 143 (2004). JHS 143: Asiakirjojen kuvailun ja hallinnan metatiedot. <http://www.jhs-suositukset.fi/suomi/jhs143>.
- [MIME, 2007] MIME (2007). Mime media types. <http://www.iana.org/assignments/media-types/>.
- [RDF, 2009] RDF (2009). Resource description framework (RDF). <http://www.w3.org/RDF>.
- [RFC 4646, 2006] RFC 4646 (2006). RFC 4646: Tags for identifying languages. <http://www.ietf.org/rfc/rfc4646.txt>.
- [SFS 5895, 2001] SFS 5895 (2001). SFS 5895: Dublin Core -metadataformaatin suomalainen versio. Suomen Standardoimisliitto.
- [SÄHKE2, 2008] SÄHKE2 (2008). SÄHKE2 liite 2: Metatietomalli. Arkistolaitos.
- [Tuominen et al., 2009] Tuominen, J., Frosterus, M., Viljanen, K., and Hyvönen, E. (2009). Onki skos server for publishing and utilizing skos vocabularies and ontologies as services. In *Proceedings of the 6th European Semantic Web Conference (ESWC 2009)*. Springer-Verlag.
- [Viljanen et al., 2008] Viljanen, K., Tuominen, J., and Hyvönen, E. (2008). Publishing and using ontologies as mash-up services. In *Proceedings of the 4th Workshop on Scripting for the Semantic Web (SFSW2008), 5th European Semantic Web Conference 2008 (ESWC 2008)*.
- [W3C-DTF, 1997] W3C-DTF (1997). Date and time formats. <http://www.w3.org/TR/NOTE-datetime>.