

# **Tietosisältöjen semanttinen yhdistäminen toimintakuvausten avulla**

Miikka Junnila

Helsinki 24.2.2006

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Tiedekunta/Osasto — Fakultet/Sektion — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Miikka Junnila			
Työn nimi — Arbetets titel — Title			
Tietosisältöjen semanttinen yhdistäminen toimintakuvausten avulla			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma		24.2.2006	62 sivua
Tiivistelmä — Referat — Abstract			
<p>Tämä tutkimus on pro gradu -työ, joka käsittelee tietosisältöjen semanttista yhdistämistä prosessi- ja toimintakuvausten avulla. Tutkimusongelmana on, toimivatko prosessit ja toiminta erilaisten tietosisältöjen tehokkaina linkittäjinä. Ongelman voisi myös muotoilla toisin: miten prosesseja ja toimintaa tulee kuvata, jotta ne linkittäisivät tehokkaasti yhteen erilaisia tietosisältöjä. Tutkimusongelmaan on pureuduttu käytännössä KulttuuriSampo-nimisen suomalaisen kulttuurin portaalin prototyypin kautta, joka luotiin osana tutkimusta.</p> <p>ACM Computing Classification System (CCS):  J.5 [Arts and humanities],  J.7 [Computers in other systems, publishing],  H.3.5 [Online Information Services, data sharing],  I.2.4 [Knowledge Representation Formalisms and Methods, Representation languages]</p>			
Avainsanat — Nyckelord — Keywords			
Semantic Web, prosessit, tarinat, sisältökeskeinen kuvailu, linkitys, kulttuurisisällöt			
Säilytyspaikka — Förvaringsställe — Where deposited			
Kumpulan tiedekirjasto, sarjanumero C-			
Muita tietoja — övriga uppgifter — Additional information			

# Sisältö

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Tietosisältöjen semanttinen mallintaminen</b>	<b>2</b>
2.1	Semanttinen web . . . . .	2
2.1.1	Metatieto ja RDF . . . . .	3
2.1.2	Ontologiat ja RDFS . . . . .	5
2.1.3	Annotointi . . . . .	6
2.2	Prosessit ja niiden mallintaminen . . . . .	7
2.2.1	Business Process Execution Language for Web Services . . . . .	9
2.2.2	Process Specification Language . . . . .	10
2.2.3	Toiminnan teoria . . . . .	11
<b>3</b>	<b>Tietosisältöjen semanttinen yhdistäminen</b>	
	— <b>teoreettinen malli</b>	<b>12</b>
3.1	Yhdistämisen etuja . . . . .	13
3.2	Prosessikuvaukset yhdistämisessä . . . . .	14
3.3	Tarinat prosesseina . . . . .	16
3.4	Tilanneontologia . . . . .	18
3.4.1	Geneerisen prosessitilanteen ja tarinan tilanteen suhde toisiinsa	20
3.4.2	Tilanneontologian suhde sisällön käsitteet määrittävään ontologiaan . . . . .	20
3.4.3	Toiminnan kuvailu . . . . .	21
3.4.4	Ympäristön kuvailu . . . . .	23
3.4.5	Esimerkki tilanteen kuvailusta . . . . .	23
3.4.6	Monipuolisemman kuvailun mahdollistavia ominaisuuksia . . . . .	26
3.4.7	Ongelmia . . . . .	27
3.4.8	Otteita tilanneontologian koodista . . . . .	28
3.5	Prosessiontologia ja OWL Services . . . . .	30

3.6	Sisällönkuvailun haasteita . . . . .	33
3.6.1	Tulkinnan ongelma . . . . .	33
3.6.2	Totta vai tarua? . . . . .	34
3.6.3	Ajan ja paikan vaikutus ontologioihin ja instansseihin . . . . .	35
<b>4</b>	<b>KulttuuriSampo — case</b>	<b>36</b>
4.1	Kulttuurialan digitaaliset tietosisällöt ja MuseoSuomi . . . . .	36
4.2	KulttuuriSampo — uusi prototyyppi . . . . .	39
4.2.1	Käyttötapauksia . . . . .	39
4.2.2	Tietosisältöjen valinta . . . . .	41
4.2.3	Ontologiat käytännössä . . . . .	43
4.2.4	Aineiston annotointi . . . . .	47
4.2.5	Prototyypin tekninen toteutus . . . . .	49
4.2.6	Haun käyttöliittymä ja assosiaatiosäännöt . . . . .	49
4.3	KulttuuriSampo — jatkokehitysmahdollisuuksia . . . . .	51
4.3.1	Käyttöliittymä ja prosessien visualisointi . . . . .	52
4.3.2	Aineiston laajentaminen . . . . .	52
<b>5</b>	<b>Muuta tutkimusta alalta</b>	<b>54</b>
5.1	CIDOC Conceptual Reference Model . . . . .	54
5.2	ABC Harmony . . . . .	55
<b>6</b>	<b>Arviointia</b>	<b>55</b>
<b>7</b>	<b>Yhteenveto</b>	<b>57</b>
<b>8</b>	<b>Kiitokset</b>	<b>58</b>
	<b>Lähteet</b>	<b>58</b>

# 1 Johdanto

Semanttinen web on nousemassa varteenotettavaksi vaihtoehdoksi pohdittaessa webin tulevaisuutta, ja useita semanttisen webin teknologioita tutkitaan vilkkaasti ympäri maailmaa. Usein tutkimus on kuitenkin niin teoreettista, että ei ehditä kokeilemaan käytännössä, millaisia etuja teknologiasta todella voidaan saada ja toisaalta millaisia käytännön ongelmia syntyy.

Tässä tutkimuksessa käsitellään tietosisältöjen yhdistämistä semanttisen metatiedon avulla. Hypoteesina on, että toimintakuvausten avulla voidaan tuoda yhteen erityyppisiä tietosisältöjä. Tietosisältöjen yhdistäminen on yksi käytännön hyöty, joka semanttisen metatiedon avulla voidaan saavuttaa, kuten on todistettu MuseoSuomi-järjestelmässä [22, 23]. MuseoSuomessa oli kuitenkin käytössä ainoastaan museoesineitä ja muinaiskohteita, joten erityyppisten tietosisältöjen yhdistämistä ei paljokaan voitu tutkia sen yhteydessä.

Tässä tutkimuksessa on tutkittu, miten erilaisia tietosisältöjä voidaan sitoa yhteen ensisijaisesti toiminnan ja prosessien kautta. Käytännössä tämä vaatii tietosisällöiltä, että ne representoivat jonkinlaista toimintaa tai että ne ovat jotenkin osa jotain toimintaa tai prosessia. Esimerkkitapauksen konteksti eli suomalainen kulttuuri ja sen prosessit ja tarinat, ovat tutkimuksessa osoittautuneet soveltuviksi kuvattavaksi toiminnan kautta. Hypoteesi toiminnan keskeisestä osasta kulttuuria kuvattaessa on siis pitänyt paikkansa.

Tutkimusongelmaa lähestyttiin käytännössä toteuttamalla KulttuuriSampo-järjestelmän prototyyppi, jossa suomalaisen kulttuurin merkkiteoksia Kalevalan runoista Gallen-Kallelan maalauksiin kuvailtiin tutkimuksessa kehitetyllä tilanneontologialla ja kokeiltiin, kuinka monipuolisesti ne linkittyvät toisiinsa sisällön perusteella. Lopputulos oli rohkaiseva, sillä metatiedon perusteella generoidut linkit olivat merkityksellisiä ja niitä myös syntyi eri tietosisältöjen välille rajoista välittämättä.

Oleellisimmaksi ratkaistavaksi ongelmaksi nousi prosessin aikana tilanneontologian suunnittelu. Miten kuvata maalausten, tarinoiden, prosessien ja muiden tietosisältöjen oleellinen sisältö? Toiminta valittiin näkökulmaksi jo alussa, mutta toiminnan lisäksi oli löydettävä keinot myös toiminnan kontekstin kuvaamiseen rikkaammin, jottei metatieto jäisi semanttisesti (tai merkityksiltään) köyhäksi. Näin suunniteltiin tilanneontologia, jossa ydinyksiköksi valittiin tilanne, joka koostuu yhdestä toiminnasta ja siihen olennaisesti liittyvistä asioista, kuten toimijasta, paikasta ja ajasta.

Tutkimuksen rakenne on seuraava. Toisessa luvussa kuvataan aiheeseen liittyviä se-

manttisen webin teknologioita, sekä annetaan muutamia esimerkkejä toiminnan ja prosessien mallintamisesta. Kolmannessa luvussa esitellään teoreettinen malli toimintakeskeiseen kuvailuun etuineen ja ongelmineen. Neljännessä luvussa esitellään KulttuuriSampo-portaalin prototyyppi, jossa teoreettista mallia on käytetty. Viides luku käsittelee muuta tutkimusta kulttuurisisältöjen kuvailusta semanttisen metatiedon avulla. Kuudes luku tarjoaa arviointia tutkimuksesta. Viimeisenä tutkimuksessa on yhteenvedo.

## 2 Tietosisältöjen semanttinen mallintaminen

Web on jo nyt täynnä erilaisia tietosisältöjä, mutta metatiedon puuttuessa tietokoneet eivät voi kuin arvailla mistä tiedoissa on kysymys. Jotta tietojen semantiikkaa eli merkitystä voitaisiin avata koneillekin, tarvitaan ontologioita ja niiden mukaista metatietoa. Tässä kappaleessa kuvaillaan semanttisen webin teknologioiden peruspiirteet sekä prosessien mallinnusratkaisuja, jollaiset tässä tutkimuksessa pyritään yhdistämään semanttisten linkkien generoinnin mahdollistamiseksi.

### 2.1 Semanttinen web

Web on kokenut räjähdysmäisen kasvun viimeisen kymmenen vuoden aikana. Se on osoittautunut tehokkaaksi sekä tiedon välityksessä, viestinnässä että palveluiden tarjoamisessa. Teknologian kehittyessä ja käyttäjämäärän kasvaessa on tullut yhä suurempi tarve kehittää webiä sekä käyttäjien että palveluntarjoajien kannalta helppokäyttöisempään ja ihmisläheisempään muotoon [21].

Semanttisen webin teknologiat ovat pyrkimys ratkaista näitä ongelmia semantiikan näkökulmasta. Tällä hetkellä esimerkiksi haut webissä ovat enimmäkseen tekstihakukuja, joissa vertaillaan ainoastaan merkkijonoja toisiinsa. Kun käyttäjä laittaa hakukoneeseen sanat ”hotelli Helsinki”, etsii kone hänelle ainoastaan sellaiset sivut, joilla nämä sanat esiintyvät. Vaikka ihminen ymmärtää heti, että kyse on majoituksen hakemisesta, voi haun tuloksena silti olla esimerkiksi matkakertomuksia. Toisaalta haun ulkopuolelle jäävät esimerkiksi motellit ja hostellit, jotka käyttäjän majoitustarpeen tyydyttämisen kannalta olisivat mahdollisesti varteenotettavia vaihtoehtoja. Tekstipohjaisissa hauissa voidaan tietysti käyttää myös algoritmeja, jotka pyrkivät lisäämään älykkyyttä, mutta tietokoneen kyvyttömyys ymmärtää tekstiä tekee näistä epäluotettavampia kuin ihmisen lajittelemasta tiedosta.

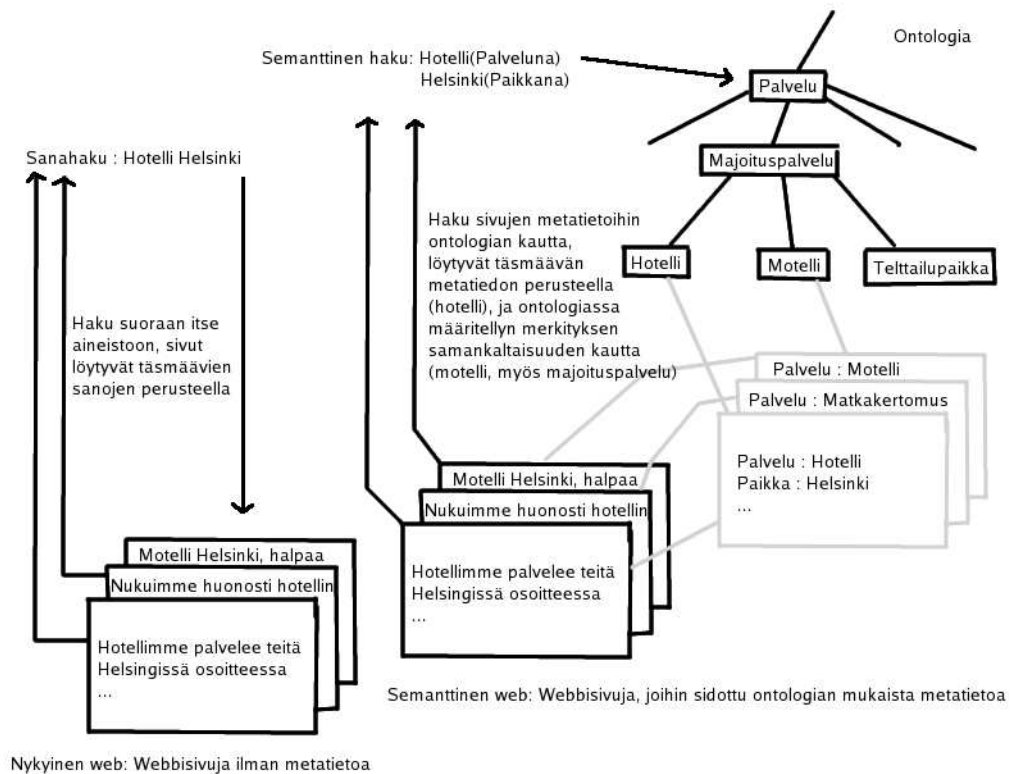
Semanttisen webin perusajatus on lisätä formaalisti määriteltyä metatietoa webissä olevaan tietoon. Tämän metatiedon avulla tietokone voisi ”ymmärtää” asioiden suhteita toisiinsa ja jäsenellä valtavaa informaatiomäärää sen merkitysten mukaan (kuva 1). Näin kone voi esimerkiksi ehdottaa käyttäjälle muitakin majoitukseen liittyviä linkkejä kuin hotelleja. Myös erilaisten ohjelmistoagenttien toiminnan mahdollistamiseen pyritään, ja parhaimmillaan loppukäyttäjälle voitaisiin tarjota hänen mieltymyksiinsä sopivin hotelli ilman, että hän edes itse hakee tietoa webistä. Sopiva agentti voisi käydä läpi kattavasti annotoidut ja oleelliset majoittumisvaihtoehdot tietokoneen tehokkuudella ja näin löytää sopivimman vaihtoehdon. Visio agenttien suorittamista tehtävistä semanttisessa webissä on keskeisellä sijalla nykyisen World Wide Webin kehittäjän Tim Berners-Leen kollegoineen *Scientific American*issa julkaisemassa artikkelissa [3]. Maailmanlaajuisen semanttisen webin toteutuminen tässä artikkelissa kuvatulla tavalla vaatii tosin vielä paljon kehitystyötä.

Tässä tutkimuksessa ei pohdita semanttista linkitystä agenttien näkökulmasta, vaan tutkitaan millaisia suoria etuja ihminen voi saada navigoidessaan tietoa semanttisten linkkien avulla. Tietokone ei pysty ymmärtämään tietoa samoin kuin ihminen, mutta yksinkertainkin merkityksien hahmottaminen metatietoa käyttävien sovellusten kautta mahdollistaa uudenlaisten palveluiden tarjoamisen loppukäyttäjälle.

### 2.1.1 Metatieto ja RDF

Semanttisen webin metatiedon ilmaisemiseen on W3C:n toimesta kehitetty Resource Description Framework eli RDF [2]. Sen päätarkoituksena on toimia yleiskäyttöisenä kielenä ja tietomallina, jolla voidaan esittää tietoa webissä. RDF:n ilmausten pohjana on joukko kolmikoita, jotka koostuvat subjektista, predikaatista ja objektista. Voidaan sanoa esimerkiksi, että Kullervo esiintyy Kalevalassa ja tässä Kullervo on subjekti, esiintyminen predikaatti ja Kalevala objekti. Käytännössä RDF:lla kuvataan suhteita erilaisten resurssien välillä. Esimerkissä Kullervon suhde Kalevalaan ilmaistaan esiintymisen kautta. Erilaiset resurssit yksilöidään Universal Resource Identifier (URI) -tunnisteen kautta. URI:t mahdollistavat esimerkiksi samannimisten ihmisten erottamisen toisistaan, ja toisaalta saman URI:n käyttö eri lähteistä tulevassa metatiedossa mahdollistaa tietojen yhteensopivuuden.

RDF:n syntaksi on yleensä XML:n mukaista, mutta RDF:ssa ei kuitenkaan ole ensisijaisesti kyse rakenteen määrittelemisestä, vaan tietosisältöjen kuvaamisesta — informaation, jonka perusteella voidaan myös koneellisesti päätellä asioita. RDF-kolmikkoja voi myös liittää toisiinsa monimutkaisemmiksi verkoiksi, voidaan esimer-



Kuva 1: Semanttinen metatieto mahdollistaa muun muassa älykkäämmät haut webbiin. Tässä esimerkissä ei ole otettu huomioon hakukoneiden ongelmia suomalaisten sijamuotojen käsittelyssä. Myös sanapohjaisten hakukoneiden omat optimointialgoritmit on jätetty huomiotta.

kiksi laittaa kolmikko Kullervo - esiintyy - Kalevalassa toisen kolmikon objektiksi: Elias Lönnrot - väittää - (Kullervo - esiintyy - Kalevalassa). URI:en lisäksi voidaan myös käyttää literaaleja. Esimerkiksi Elias Lönnrotin tunnisteen URI, mutta tämän tunnisteen kuvaamalla resurssilla on predikaatin *nimi* kautta yhteys literaaliin "Elias Lönnrot".

RDF:a käytetään jo useissa sovelluksissa, esimerkiksi Adobe ja Oracle käyttävät sitä ohjelmistoissaan [31]. RDF:n ongelma on kuitenkin se, että jos käsitteitä, joita kolmikoissa käytetään, ei määritellä, ei tietokone voi päätellä niiden perusteella kuin, että asia A liittyy asia B:hen jotenkin. Ihmisen luontaisesti tekemä merkitysten määrittely jää puuttumaan. Siksi tarvitaan ontologioita.



### 2.1.2 Ontologiat ja RDFS

Ontologiat ovat Gruberin mukaan tiettyjen aihealueiden käsitteistöjen täsmällisiä määrittelyitä [14]. Eli siinä missä ihmisillä on mielessään dynaaminen ja laaja sanojen merkitysverkko, pyritään ontologioilla esittämään koneille tietyn aihealueen sanaston merkitykset ja merkitysten suhteet toisiinsa mahdollisimman eksplisiittisesti. Tarkka määrittely on hyödyksi myös ihmisille, jotka voivat ontologioiden kautta selkeyttää ja yhdenmukaistaa omat mentaaliset mallinsa ja sanastonsa. Kun sanasto ja sen merkitykset on tarkasti määritelty, tiedetään mistä puhutaan. Ontologioiden avulla voidaan myös vertailla käsityksiä käsitteistä ja erottaa aihealueen tieto toiminnallisesta tiedosta (eli ohjelmakoodista) [34].

RDF Schema (RDFS) [5] on RDF:n kanssa käytettäväksi tarkoitettu kieli, jolla voidaan määritellä ontologioita. RDFS määrittelee luokkia ja ominaisuuksia, joilla voidaan kuvata luokkia, ominaisuuksia ja muita resursseja. Näiden luokkien avulla voidaan RDF-metatietoa jäsentää ja yhtenäistää.

Abstraktien tietorakenteiden tasolla RDFS-ontologiat ovat suunnattuja verkkoja, joissa käsitteet on määritelty luokkina. Verkko on luokkien hierarkia, jossa on yläpäässä yleisempiä käsitteitä ja alempana hierarkiassa tarkempia käsitteitä eli aliluokkia (Vertaa Palvelu - Hotelli kuvassa 1). Näitä luokkia voidaan käyttää erilaisien resurssien tyyppin määrittelemiseen. Ontologian ollessa käytössä voimme sanoa, että resurssi Elias Lönnrot on tyyppiä ihminen, eli Ihminen-luokan yksilö. Jos ihminen on ontologiassa määritelty Kädelliset-luokan alaluokaksi, voidaan erilaisia RDF-kolmikkoja käyttää nyt laajempaan päättelyyn. Jos on olemassa kolmikko kädelliset - tarvitsevat - ruokaa, voidaan nyt koneellisesti päätellä, että myös Elias Lönnrot tarvitsee (tai ainakin tarvitsi) ruokaa.

Luokilla on myös ominaisuuksia. Ominaisuus, attribuutti ja RDF-predikaatti ovat käytännössä kaikki sama asia. Ontologiassa on luonnollista puhua ominaisuuksista, koska ne kiinnittyvät luokkiin. Jos kädellisellä on esimerkiksi äiti-ominaisuus, tarkoittaa tämä sitä, että jokaisella resurssilla, joka on tyyppiä Kädellinen tai sen alaluokka, voi olla äiti. RDFS:lla voidaan myös kertoa, että äiti-ominaisuuden arvoksi on tultava jokin Kädellinen-luokan instanssi, jolloin määritelty ontologia myös estää virheellisen metatiedon syöttämisen. Ominaisuuksien arvot voivat olla literaaleja tai resursseja, kuten toisien luokkien ilmentymiä. Olio-ohjelmointiin perehtyneet tunnistavat helposti tämän tavan järjestellä tietoa.

Ominaisuuksien ja luokkien suhteiden eksakti kuvaus auttaa paitsi päättelyssä myös

metatiedon yhtenäistämässä. Kun ontologiassa on määritelty ominaisuus äiti luokalle Ihminen, tulee tämä oleellinen tieto laitettua metatietoihin. Epäoleelliset ominaisuudet voidaan jättää pois. Kun tämän seurauksena metatiedot ovat yhtenäisiä, niiden pohjalle voidaan rakentaa loogisia sääntöjä, jotka käyttävät metatietoa suoraan hyväksi. Mikäli ontologiaa ei ole ja esimerkiksi jokaisesta ihmisestä syötetään aivan erilaisia metatietoja koneelle, ei näillä tiedoilla ole kovinkaan helppo löytää toistuvia kuvioita metatiedosta, joilla semantiikaltaan yhteneviä tapauksia voitaisiin yhdistää automaattisesti.

Kun webin tietoon on lisätty ontologian mukainen metatieto, voidaan kehittää ohjelmia, jotka käyttävät hyväkseen ontologiassa määriteltyjä suhteita, luokkahierarkioita ja ominaisuuksia. Tällainen ohjelma voi esimerkiksi olla selain, jolla pystytään seurailemaan semanttisia linkkejä tiedosta toiseen, kuten esimerkiksi MuseoSuomi-järjestelmässä [23].

### 2.1.3 Annotointi

Ontologiat sinänsä voivat olla joskus hyödyllisiä ilman niiden mukaista metatietoa, esimerkiksi ihmisten mentaalisten käsittemallien selkeyttämisessä, mutta usein ontologiat on tarkoitettu juuri jäsentämään metatietoa. Jos halutaan esimerkiksi kuvata Helsingin kaupunkia semanttisessa webissä, siitä ei normaalisti tehdä ontologian luokkaa, vaan jonkinlaisen paikkaontologian Kaupunki-luokan ilmentymä eli instanssi. Instanssidataan lisätään sitten niiden ominaisuuksien arvot, jotka ontologiassa on Kaupunki-luokalle määritelty. Kaupungin ominaisuuksia voivat olla esimerkiksi maa, jossa kaupunki sijaitsee, asukasluku ja koordinaatit. Näistä ensimmäinen on tyypillisesti arvoltaan toiseen instanssiin viittaava, loput taas literaaliarvoisia. Sekä literaalien että instanssien kautta voidaan luoda automaattisia semanttisia linkkejä. Jos ominaisuuden arvo voi merkkijonona olla moniselitteinen, tulisi ominaisuuden arvo määritellä instanssiksi, jolloin merkkijonoina identtiset arvot voidaan erotella toisistaan URI:n avulla.

Semanttisen metatiedon lisäämistä tietoon kutsutaan annotoinniksi. Annotoinnilla on joissain yhteyksissä tarkoitettu myös yleistä kommenttien lisäämistä sisältöön, mutta tässä tutkimuksessa tarkoitetaan annotoinnilla metatiedon tekemisen prosessia.

Käytännössä annotointia voidaan tehdä monella tavalla. Metatietoa voidaan kirjoittaa käsin, mutta tämä on vaivalloista ja altista virheille. Usein käytetäänkin jotain

työkalua. Protegé-2000-editori [35] on yksi vaihtoehto annotointiin, jolla voidaan myös tehdä ontologioita. On kuitenkin kehitetty myös erilaisia web-sivujen annotointiin tarkoitettuja välineitä, esimerkiksi CREAM ja Ont-O-Mat [15], sekä SHOE Knowledge Annotator [16].

Annotointi on luonteeltaan ihmisymmärrystä vaativaa työtä. Jos tietokone pystyisi itse tekemään kaiken semanttisen metatiedon, olisi kyse siitä, että tietokoneet kertovat tietokoneelle miten asiat ovat. Näin älykkäät tietokoneet tekisivät annotoinnista turhaa, mutta tällä hetkellä asioiden merkitystä joudutaan vielä mallintamaan tietokoneelle ihmisen tekemän metatiedon avulla. Usein annotointityökalut kuitenkin pyrkivät jotenkin auttamaan ihmistä annotointiprosessissa, ja parhaimmillaan annotointia voidaan nopeuttaa huomattavasti erilaisin tekniikoin. Esimerkiksi MuseoSuomi-järjestelmän annotointiprosessista saatiin automatisoitua huomattava osa siirtämällä annotointityö museoesineiden kuvailussa käytettäviin termeihin itse esineiden sijasta, niin että esimerkiksi jokaisen kirveen sijasta annotoitiin merkkijono ”kirves”, jonka jälkeen kaikki kirveet voitiin tunnistaa annotoidun merkkijonon perusteella ontologian ”kirveiksi” [24]. Annotoinnin helpottaminen ohjelmallisesti on siis mahdollista, mutta tapa, jolla se tehdään, riippuu monesta asiasta kuten siitä, millaista tietoa on saatavilla ja missä muodossa. Esimerkiksi vapaata tekstiä on vaikeampi käyttää automaattisen annotoinnin pohjana kuin rakenteista tietoa, mutta vapaa tekstikin on parempaa kuin tekstuaalisen pohjatiedon puuttuminen kokonaan (esimerkiksi kuva). Myös annotoitavan aineiston semanttinen homogeenisuus helpottaa annotoinnin automatisointia.

## 2.2 Prosessit ja niiden mallintaminen

Prosessien mallintamisessa on kyse toiminnan simuloinnista tai tietokoneen ulkopuolisen maailman toimintojen ja prosessien ennakoinnista, kuvaamisesta tai kontrolloinnista. Malleja sovelletaan monentyypisissä sovelluksissa tietokonepelien fyysikkamoottoreista erilaisten bisnessovelluksien kautta myrkkyjen liikkumista meressä kuvaaviin mallinnusohjelmiin. Analysoimalla prosesseja voidaan saada selville prosesseihin liittyviä heikkouksia ja vahvuuksia, joita voitaisiin ehkä myös käyttää joissain toisissa prosesseissa. Tehokkuutta haetaan kaikessa, ja hyvä prosessi voi olla avain entistä parempaan lopputulokseen niin ihmisten kuin koneidenkin suorittamissa töissä.

Yksinkertaisimmillaan prosessi on tietyn ajan kuluessa aina samalla tavalla tapahtuva muutos. Usein prosessit ovat monimutkaisempia tapahtumaketjuja, joissa näen-

näisen yksinkertainen tapahtuma saattaa koostua hyvin monimutkaisista aliprosesseista. Rautaesineen ruostumisprosessi kosteissa olosuhteissa on esimerkiksi yksinkertainen prosessi: laita rautaesine kosteaan, odota ja rauta ruostuu. Kun ruostumiseen johtavia kemiallisia prosesseja aletaan tutkia tarkemmin, voidaan kuitenkin huomata monimutkaisia ja myös epäsäännöllisiä piirteitä. Näin mallintamisessa vallittu abstraktiotaso vaikuttaa huomattavasti prosessin näennäiseen monimutkaisuuteen.

Prosessi-käsitteen määrittely on jo sinänsä vaikea tehtävä, ja siinä voidaan päätyä monenlaisiin lopputuloksiin. Kun koko maailmaa yritetään jäsentää esimerkiksi ontologioin, tulee prosessi usein ylimpien luokkien joukkoon. Intuitiivisesti prosessista tulee mieleen jokin muutosprosessi, mutta esimerkiksi CYC-ontologian kehitysversion prosessiluokka on myös kaiken fyysisesti olemassa olevan yläluokka [38]. Tämän määrittelyn mukaan siis esimerkiksi maapallo on prosessi. Tällainenkin näkökulma on mahdollinen ja kietoutuu olennaisesti kysymykseen asioiden identiteetistä: onko maapallo asia, joka muuttuu, mutta pysyy silti maapallona, vai onko maapallous juuri prosessissa, jossa maapallo muuttuu?

Sowa on määritellyt kirjassaan prosesseja erilaisten ominaisuuksien perusteella [38]. Perustavanlaatuinen jako on jatkuvien (*continuous*) ja irrallisten (*discrete*) prosessien erottaminen toisistaan. Jatkuvat prosessit eivät jakaudu selviin tiloihin, vaan huojuvat erilaisten muuttujien muutosten mukaan suuntaan tai toiseen koskaan alkamatta tai loppumatta. Esimerkki tällaisesta prosessista on sää; se ei ala eikä lopu vaan muuttuu jatkuvasti. Irralliset prosessit taas jakautuvat muutoksiin ja niiden välisiin tiloihin, jolloin muutosta ei tapahdu. Tällainen näkökulma on mallinnuksen kannalta helpompi, kun vaiheet ovat selkeästi erillään toisistaan.

Tietojenkäsittelyn kontekstissa on prosessikuvausten laatiminen usein ikäänkuin algoritmin kirjoittamista. Suorasukaisten prosessien kuvaukset voidaan kirjoittaa tietokoneohjelmaksi, joka suorittaa prosessin tai sen aliprosessin säästäten vaivaa ihmiseltä.

Erilaisia tapoja mallintaa prosesseja on esiintynyt viime vuosikymmenien aikana enenevässä määrin. National Institute of Standards and Technology (NIST) on esimerkiksi käynyt läpi 26 prosessien representaatiota pohtiessaan oman kielensä Process specification language suunnittelua [26].

Useat prosessit ovat niin monimutkaisia, että tietokone ei selviä niiden suorittamisesta yksin. Mitä monimutkaisemmaksi prosessi osoittautuu ja mitä enemmän dynaamisista päätöksentekoa vaativia osuuksia on mukana, sitä todennäköisemmin tar-

vitaan ihmistä kertomaan koneelle, mitä sen pitäisi tehdä. Toisaalta tietokone pystyy ylipäättään vain tietynlaisiin suorituksiin eli niihin, joissa käsitellään tietoa. Siksi prosessikuvaukset tietokoneella eivät aina tähtää siihen, että tietokone myös suorittaisi prosessin, vaan prosesseja mallinnettaessa voidaan sen sijaan yrittää ymmärtää ja analysoida ihmisten käyttäytymistä tai toimintaa. Prosessien mallintaminen saattaa siis eri yhteyksissä tarkoittaa hyvin erilaistakin lähestymistapaa.

Seuraavissa aliluvuissa kuvataan muutamia esimerkinomaisia prosessinmallinnusratkaisuja. Business Process Execution Language for Web Services [39] ja PSL [37] ovat esimerkkejä teollisuuden ja palveluyritysten prosessienmäärittelyistä, joissa pyritään mahdollistamaan yhteistoimintaa ja vähentämään järjestelmien tiedon toisteisuutta. OWL-S [7], joka lopulta valittiin oman mallimme pohjaksi esitellään seuraavassa luvussa, ja se liittyy oleellisesti tämän tutkimuksen ratkaisuun. Toiminnan teoria [40] on valittu esimerkiksi vähemmän teknisestä prosessien analysointimalleista ja erityisesti siksi, että sen lähestymistapa korostaa ihmiskulttuurin merkitystä prosesseihin ja on näin antanut uuden hyödyllisen näkökulman kulttuurialan esimerkkiä työstettäessä.

### **2.2.1 Business Process Execution Language for Web Services**

BPEL4WS on kieli bisnesprosessien ja vuorovaikutusprotokollien formaaliin määrittelemiseen [39]. Jos erilaiset bisnesprosessit pystyttäisiin kuvaamaan formaalisti samalla kielellä, niin niiden yhteistoiminta olisi automatisoitavissa, ja näin palveluiden käyttäminen webissä vaatisi yhä vähemmän ihmisten näkemää vaivaa. BPEL4WS on yksi monista standardeista, joita 2000-luvulla on kehitelty liittyen web-palveluteknikoihin, joilla pyritään laajemmaltikin mahdollistamaan yritysten palveluiden saaminen webiin standardien mukaisesti ja niiden yhteistoiminta.

Koska älykkäiden palveluiden ja maailmanlaajuisen yhteensopivuuden ihanteet ovat samat web-palveluteknikoiden ja semanttisen webin kehittäjille, on näiden yhdistämistä pohdittu monissa tutkimuksissa [6, 28], BPEL4WS:n näkökulmasta esimerkiksi Mandellin ja McIlraithin artikkelissa [30]. Agenttien ja automatisoitavien palveluiden näkökulmasta BPEL4WS on mielenkiintoinen, mutta tietosisältöjen yhdistämiseen näkökulma ei ole paras mahdollinen.

### 2.2.2 Process Specification Language

Process Specification Language eli PSL on NIST:n kehittämä kieli, jolla pyritään kuvailemaan tuotantoprosessien kulkua [37]. Monet sovellukset teollisten yritysten sisällä tuotantoaikataulutuksesta prosessisuunnitteluun ja projektihallintaan käytävät prosessitietoa hyväkseen. Tällä hetkellä on usein niin, että kaikilla näillä ohjelmissa on omat sisäiset tietomallinsa prosessista, ja jos prosessissa tapahtuu muutos, se on päivitettävä jokaisen sovelluksen tietomalliin erikseen. Koska kyse on kuitenkin saman prosessin kuvailusta eri yhteyksissä, olisi yhteinen prosessien mallinnuksen tapa pakollinen näiden erilaisten ohjelmien paremman yhteistoiminnan aikaansaamiseksi. PSL:ää suunnitellaan tällaiseksi yhteiseksi kieleksi. Organisaation sisäisten ohjelmien yhteensopivuuden lisäksi PSL saattaisi myös mahdollistaa isompien yksiköiden, kuten esimerkiksi tietyssä projektissa yhteistyötä tekevien yritysten helpomman yhteistoiminnan, jos molemmat käyttäisivät standardisoitua prosessinmäärittelykieltä. Tällöin kummankaan yrityksen ei tarvitsisi nähdä erillistä vaivaa uusien mallinnustapojen opetteluun ja toteuttamiseen, vaan yhteistoiminta lähtisi yhteiseltä pohjalta.

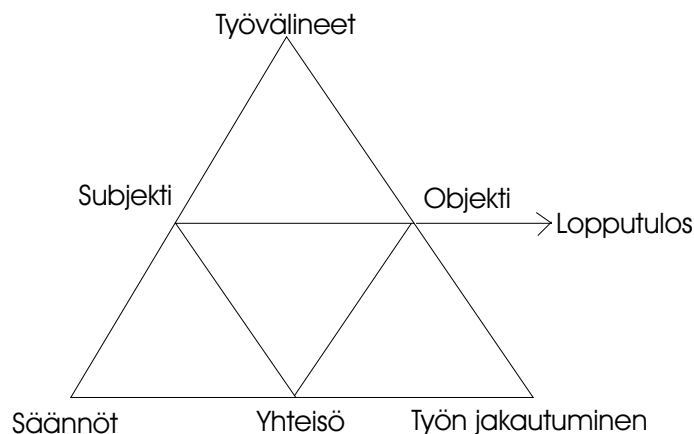
PSL:n kehitys on vielä meneillään, mutta sen selkärankana toimiva ontologia on jo määritelty. PSL:n ydin koostuu muutamista yksinkertaisista peruselementeistä. On olemassa toimintoja (*activity*), toimintojen esiintymiä (*activity occurrence*), hetkiä (*timepoint*) sekä objekteja (*object*). Toimintojen ei tarvitse esiintyä, mutta esiintymisissä toiminnoilla on alku- ja loppuhetki. Objektit ovat mitä tahansa asioita, jotka eivät ole toimintoja tai hetkiä. Ne voivat viitata esimerkiksi prosessin tuloksena syntyviin tuotteisiin tai resursseihin, jotka kuluvat prosessin kuluessa. Myös objekteilla voi näin olla alku- ja/tai loppuhetkiä.

Ytimen lisäksi PSL:ään kuuluu myös erilaisia laajennuksia, jotka koostuvat aksiomista ja määrittelyistä, joilla kielen mahdollisuuksia laajennetaan. Koko ontologia on määritelty Knowledge Interchange Format -kielellä (KIF) [12].

Erilaisten tietosisältöjen semanttiseen yhdistämiseen tarvittavan prosessiontologian pohjaksi PSL ei ole kovinkaan sopiva. PSL:n painotukset nimenomaan teollisiin tuotantoprosesseihin rajoittavat sillä kuvattavia prosesseja. Esimerkiksi toimijan käsitettä ei ole PSL:n ytimessä lainkaan ja laajennusta, jossa sen pitäisi olla, ei saa avatuksi webissä lainkaan. Muutenkin PSL on kokonaisuudessaan sekava joukko erilaisia laajennuksia, joiden keskinäisiä suhteita on vaikea selvittää. Erilaiset ytimen laajennukset voivat olla keskenään ristiriidassa, ja omiin tarpeisiin sopivan paketin kokoamiseen ei ole kunnollista opastusta.

### 2.2.3 Toiminnan teoria

Prosesseja on tutkittu monesta näkökulmasta myös tietojenkäsittelytieteen kontekstin ulkopuolella. Yksi esimerkki tällaisesta teoriasta on Lev Vygotskyn kehittämä toiminnan teoria [40], jota Aleksei Leontev on jatkokehittänyt [29]. Mielenkiintoiseksi teorian tekee se, että sen keskeisin malli on subjektin, objektin ja apuvälineiden keskinäisten suhteiden muodostama kolmio (kuva 2). Näiden asioiden ja suhteiden kautta kuvataan toimintaa, joka teoriassa nähdään tärkeänä itsenäisenä yksikkönä. Tällaisista yksiköistä voi muodostua monimutkaisempia toimintoja.



Kuva 2: Toiminnan teoria liittää toimintaan sosiaalisen ja materiaalsen kontekstin. Vygotskyn mukaan toiminta ei tapahdu suoraan subjektista objektiin, vaan työvälineet ja yhteisö vaikuttavat lopputulokseen oleellisesti, ja subjekti ja objekti taas osaltaan vaikuttavat yhteisön sääntöjen ja työn jakautumisen muotoutumiseen.

Kolmiomallia on laajennettu lisäämällä siihen myös yhteiskunnan säännöt, yhteisö, ja työn jakautuminen, jolloin toiminnan konteksti tulee vahvemmin etualalle. Tällaisen kontekstitiedon käyttäminen voisi avata näkökulmia myös erilaisiin kulttuurisiin tietosisältöihin, koska lopulta kulttuurissa on kyse juuri ihmisistä. Teorian mukaan yhteisö sääntöineen vaikuttaa huomattavasti siihen, millaisia esineitä, taideteoksia tai tarinoita kulttuurin piirissä syntyy, ja toisaalta esimerkiksi esineet vaikuttavat siihen, millaisia prosesseja on olemassa. Näin yhdeksi tärkeäksi asiaksi muodostuu mediaatio eli välittyminen. Toiminta ei tapahdu suoraan subjektista objektiin vaan jonkun tai joidenkin välittävien tekijöiden kautta.

Suorat yhteydet ja toisaalta välittymisen kautta syntyvät yhteydet ovat rinnastettavissa suoriin ja epäsuoriin semanttisiin linkeihin. Toisaalta toiminnan teoria vetää huomion myös enemmän ihmisiin esineiden sijasta, ja tämä näkökulma on tärkeä

muistaa prosesseja kuvattaessa. Yhteisön ja sen sääntöjen vaikutus prosesseihin ja esineisiin olisi erityisen mielenkiintoista materiaalia linkitettäessä esineitä ja asioita eri puolilta maailmaa ja eri aikakausilta, jolloin voitaisiin yhteisömetatiedon avulla tehdä myös antropologista ja sosiologista vertailua erilaisten semanttisten yhteyksien perusteella. Jos tulevaisuudessa semanttinen web toteutuu maailmanlaajuisesti ja sisältää rikkaasti erilaisia tietosisältöjä, jää ihmisten tehtäväksi vain löytää mielenkiintoiset yhteydet asioiden välille, määrittellä ne koneen ymmärtämällä tavalla, ja sen jälkeen voidaan tutkia tietoja erilaisten suhdemallien, esimerkiksi toiminnan teorian, kautta.

### 3 Tietosisältöjen semanttinen yhdistäminen — teoreettinen malli

Semanttisen webin vahvuus tulee sitä selkeämmin ilmi, mitä enemmän ja mitä monipuolisempaa sisältöä saadaan julkaistuksi. Erilaiset tietosisällöt yhdistettynä luovat enemmän mahdollisuuksia mielenkiintoisten yhteyksien löytämiseksi. URI-viitteiden avulla voidaan tunnistaa tietosisältöjen yhteiset tekijät, mutta voisiko erilaisten näkökulmien joukosta löytyä jokin erityisen käyttökelpoinen näkökulma, joka liittyisi jotain kautta kaikkiin tietosisältöihin?

Hypoteesina tässä tutkimuksessa on, että prosessikuvaukset voisivat toimia näkökulmana, joka yhdistää monenlaista tietoa, koska tekeminen ja muutos ovat maailmasamme jatkuvasti esillä olevia asioita. Kulttuuri — kuten yksittäisen ihmisen elämäkin — on koko ajan liikkeessä. Emme koskaan astu samaan jokeen kahta kertaa, uudet vedet siinä jo virtaavat [17]. Koska toiminta ja prosessit aiheuttavat muutosta, on todennäköistä, että ne kietoutuvat monenlaisiin tietosisältöihin — varsinkin niihin, joissa on kyse ihmisestä ja kulttuurista.

Kiinnostus erilaisten prosessien kulkuun on noussut oleelliseksi kysymykseksi myös tieteessä viimeisen sadan vuoden aikana. Eastman kirjoittaa aiheesta seuraavaa: ”*Prosessien rooli erilaisessa älyllisessä toiminnassa on muuttunut dramaattisesti viimeisen vuosisadan aikana. Ennen 1900-lukua ihmisprosessien tutkimus oli erittäin harvinaista. ... Vasta hiljalleen on tietoisuus prosesseista syntynyt 1900-luvulla.*” [10]. On siis syytä tutkia millaisessa suhteessa prosessit ovat ympäröivään maailmaan ja sitä kautta myös niihin moninaisiin tietosisältöihin, joita semanttisessa webissä voidaan julkaista.



Seuraavissa aliluvuissa hahmotetaan aluksi prosessikeskeisen yhdistämisen mahdollisuuksia, sitten kuvaillaan mallin ontologiat, ja lopussa käsitellään vielä joitain mallin käyttämisessä eteen tulevia mielenkiintoisia kysymyksiä.

### 3.1 Yhdistämisen etuja

Tieto, joka ei liity mihinkään, ei ole merkityksellistä. Jos ihminen saa tietää vaikkapa Sunset Boulevard -elokuvan naispääosan esittäjän nimen, muttei ole nähnyt itse elokuvaa tai kuullut siitä eikä myöskään Gloria Swansonista, ei tieto tunnu mielekkäältä vaan täysin turhalta. Ajattelijat Aristoteleesta brittiläisen empiirisen koulukunnan filosofiin (Locke, Hume, Mill, Hartley, Bain) ovat esittäneet, että kaikki tieto on olemassa nimenomaan assosiaatioiden muodossa [11]. Myös opetuksessa pyritään liittämään asioita yhteen: konstruktivistisen oppimisteorian mukaan ihminen oppii uutta rakentamalla tietoa vanhan päälle [18]. Vaikuttaisi siis siltä, että tiedon on aina liityttävä johonkin, selattiin sitä sitten viihtymisen tai oppimisen takia.

Webin luonteeseen on sen alkuaajoista lähtien kuulunut hypertekstuaalisuus. Tiedot on voinut linkittää toisiinsa, ja uutta tekstin navigoimistapaa on tutkittu. HTML-linkit eivät kuitenkaan sisällä minkäänlaista koneymmärrettävää semantiikkaa. Jos linkki on olemassa, sen perusteena on HTML-koodin kirjoittajan käsitys, että linkin kuuluisi olla siinä. Linkkejä on siis saatettu sirotella hypertekstiin hyvin erilaisin perustein ja usein myös ilman sisäistä logiikkaa. Yhdestä kotisivun linkistä saattaa aueta kuva linkkiteksitin henkilöstä, toisesta jonkun kotisivut ja kolmannelta ruokaresepti, joka linkin määrittelijän silmissä kuvaa kyseistä henkilöä parhaiten. Näin hypertekstin lukeminen onkin usein koettu sekavaksi. Conklin on määritellyt hypertekstin suurimmiksi ongelmiksi tiedon sekaan eksymisen ja kognitiivisen painolastin, joka syntyy kun ihmisen täytyy pysähtyä kesken tekstin luvun miettimään, seuratako linkkiä vai ei [8]. Bernstein taas on huomannut, että kun käyttäjää opastetaan linkkien semantiikalla, ei eksymisongelma ole yhtä vakava [4]. Toisaalta rikas semanttinen linkitys voi pahentaa kognitiivisen painolastin määrää, ellei linkkejä pidetä erossa itse informaatiosta.

Näkökulma, jonka kautta tietoja katsotaan, sitoo erilaiset tietosisällöt mielekkääksi kokonaisuudeksi loppukäyttäjän kannalta. Joskus voi olla mielenkiintoista yksinkertaisesti vain surffailta, seurailta linkkiviidakossa polkuja miten sattuu, ja päätyä milloin mihinkin, mutta kun etsitään tietynlaista tietoa, on näkökulma tärkeä. Näkökulman mielekkyyden määrää siis toisaalta loppukäyttäjän tarve, toisaalta näkökulman kyky yhdistää asioita toisiinsa monipuolisesti. Esimerkiksi aakkosjärjestyksen

mukaan voidaan sitoa melkein mitä tahansa tietoa yhteen, mutta loppukäyttäjälle näkökulma tuskin on kovin hedelmällinen. Näkökulman on siis oltava loppukäyttäjälle merkityksellinen, tai linkkejä ei voida kokea mielekkäiksi. Monimutkaisetkin semanttiset yhteydet ovat mielenkiintoisia — ehkä kaikkein mielenkiintoisimpia — mikäli loppukäyttäjä voi selvästi seurata linkkien merkityksellisyyttä ja saada oivalluksia. Jos linkit kuitenkin tuntuvat satunnaisilta, niiden mielekkyys häviää.

## 3.2 Prosessikuvaukset yhdistämisessä

Miksi yhdistäminen pitäisi siis tehdä prosessien kautta? Vaihtoehtoja olisi monia. Prosesseilla on kuitenkin erityisesti kaksi ominaisuutta, joiden perusteella niiden kokeilemista yhdistävänä tekijänä kannattaa kokeilla:

- *Prosessit liittyvät useisiin tietosisältöihin.* Kuten aikaisemmin on todettu, prosessit liittyvät hyvin monenlaisiin tietosisältöihin. Prosessit ovat läsnä kaikkialla yhteiskunnassa ja luonnossa ja usein myös tietoa etsivän ihmisen kiinnostuksen kohteena. Museokontekstissa prosessit ovat monellakin tavalla käyttäjän tiedolle olennaisia. Vanhoista esineistä tietoja selaamalla on sinänsä vielä vaikea saada otetta sen ympärillä pyörineestä elämästä. Jos esineisiin kuitenkin päästään sen prosessin kautta, jossa niitä käytetään, on esineen merkitys selvemmin esillä. Harva esine on käyttäjälleen ollut itseisarvo, vaan oleellista on, mitä jollain esineellä on tehty. Näin prosessikuvaus voi avata oleellisia näkökulmia vanhoihin esineisiin. Samoin esimerkiksi maalauksessa kuvattu tuokio saa laajemman merkityksen tietyn prosessin osan visuaalisena kuvauksena.

Samalla tavalla kuin prosessikuvaukset ovat hyvä välittäjä erilaisten tietosisältöjen välillä, laajentavat myös muut tietosisällöt itse prosessikuvausta. Vaihe vaiheelta -tyyppinen prosessikuvaus saa huomattavasti enemmän todentuntuisuutta ja eloa kun siihen voidaan yhdistää esimerkiksi tarinoita ja maalauksia. Näin vahva linkitys vahvistaa kokonaisuutta, eivätkä erilaiset tiedot jää roikkumaan jonnekin monimutkaisten linkkipolkujen päähän, vaan kaikki tuodaan prosessikuvauksen kautta keskitetysti lyhyen polun päähän.

- *Prosesseilla on itseisarvo.* Prosessikuvaus sinänsä on myös oleellista informaatiota. Erilaisten prosessien dokumentointi ja webiin tuominen laajentavat ihmisten mahdollisuuksia oppia tekemään uusia (tai vanhoja) asioita. Historiallisten, jo katoamassa olevien prosessien säilyttäminen on erityisen tärkeää

kulttuurihistoriallisesti. Vanhojen taitojen häviäminen vie osan vaivalla hankittua tietotaitoa hautaan, ja toisaalta myös ote historiasta ja sen ihmisistä heikentyy, koska ihmisen toiminta määrittää paljon aikakauttaan.

Prosessin käyttäminen kaikkea yhdistävänä asiana ei tietenkään ole ainoa vaihtoehto. Myös esimerkiksi henkilökeskeinen tietojen yhdistäminen voisi olla monessa yhteydessä hyvä lähtökohta. Ihmiset liittyvät useimmiten kulttuurisisältöihin jollain tavalla, jos eivät muuten niin taiteen tekijöinä tai esineiden omistajina. Jos käyttäjä olisi kiinnostunut nimenomaan tietyn ihmisen elämästä, voisi henkilökeskeinen semanttinen yhdistäminen toimia hyvin. Myös esimerkiksi paikkoja ja aikoja voisi ajatella käytettävän yhdistävänä tekijänä. Lopulta kyse on taas siitä, mitä loppukäyttäjä tarkalleen haluaa tehdä. Jos esimerkiksi historioitsija tutkii tietyn aikakauden henkilöitä, voidaan olettaa, että henkilökeskeinen linkitys on hänelle parempi kuin prosessikeskeinen.

Jos tarkoituksena on kuitenkin luoda mahdollisimman monipuolisia yhteyksiä asioiden välille, on prosessikuvauksilla se etu verrattuna esimerkiksi henkilöihin ja aikaan verrattuna, että samoin kuin ihmisen tarpeet ovat pysyneet samoina vuosituhansia, pysyvät monet perustavanlaatuiset prosessit riippumattomina ajasta, paikasta ja yksittäisistä henkilöistä. Esimerkiksi kalastuksen prosessi voidaan löytää jo tuhansien vuosien takaa, joten sen kautta voidaan vetää linkki kivettyneen afrikkalaisen onkivavan fossiilin ja modernin suomalaisen perhokalastusvavan välille. Tällaiset toisistaan ajallisesti ja paikallisesti hyvin etäiset asiat olisi lähes mahdotonta linkittää muuta kautta. Joskus tämä voi olla myös ongelma, koska nämä ajan tai paikan vaihtuessa paljonkin muuttuvat prosessit saattavat joskus myös olla epäolennaisia haun kannalta.

Loppukäyttäjälle voisi tarjota mahdollisuuden valita itse, minkä asian kautta haluaa tietosisältöjä tutkia. Tiedon ja linkkien organisoiminen uudelleen sen asian mukaan, josta loppukäyttäjä on kiinnostunut, olisi varmasti täysin mahdollista, tosin sen verran työlästä, että tämän työn kohdalla kaikkia näitä mahdollisuuksia ei ole voitu tutkia. Jokainen lähestymistapa vaatii kuitenkin tiettyjen tapauskohtaisten kysymysten pohtimista ja ratkaisemista, kuten esimerkiksi sopivan käyttöliittymän suunnittelua ja toteutusta. Jos käyttäjälle antaa liian paljon valinnanvapautta ja liian monimutkaisen käyttöliittymän, tämä saattaa todeta järjestelmän olevan liian vaikea ymmärrettäväksi.

### 3.3 Tarinat prosesseina

Prosessi on hyvin yleisen tason käsite. Joskus on mielekästä yhdistää tietoa juuri yleisellä tasolla, mutta usein on myös mielenkiintoista katsoa asioita jonkin tietyn tapahtuman näkökulmasta: ei siis ainoastaan prosessin yleisen kaavan, vaan tiettyinä aikana tietyssä paikassa tapahtuneen prosessin. Tätä kautta päästään tarinan käsitteeseen.

Tarinat ja kertomukset, runot ja laulut ovat ihmisille ikivanha tapa säilyttää tietoa ja toisaalta myös viihtyä. Ihmiset haluavat kuulla tarinoita [25]. Siksi on oleellista pohtia myös tarinan mahdollisuuksia semanttisen yhdistämisen välineenä. Yleiset prosessikuvaukset jäävät helposti ulkokohtaiseksi eivätkä herätä tunteita. Tarinassa prosessi saa lihaa luittensa ympärille, ja tarina lisää prosessikuvaukseen ne elementit, jotka mahdollistavat ihmisen samaistumisen tarinaan. Yhdestä prosessista voi olla monia tarinoita, ja ne voivat kaikki olla erilaisia — ja mielenkiintoisia. Elämänmaukaiset tositarinat tai myyttiset sankaritarut tuovat prosesseihin erilaisia näkökulmia ja liittyvät erilaisiin kulttuurisiin tietosisältöihin monipuolisesti.

Tarinaa kuvattaessa prosessi on mahdollista rakentaa samoista osista kuin prosessikuvauksessakin. Prosessin osat ovat rinnastettavissa tarinan osiin: molemmissa tapahtuu asioita ajan kuluessa ja asioiden tilat muuttuvat. Runousopissa [1] painoteetaan sitä, että draama on toiminnalla kertomista. Näin toiminta siis on jotain, mikä todella on — ainakin klassisen käsityksen mukaan — draaman ytimessä. Tarina vain vaatii enemmän tietoa kuin prosessikuvaus. Prosessikuvaukseen riittävät tiedot siitä, miten tehdään mitäkin ja missä järjestyksessä. Tarina taas nostaa näiden kysymysten rinnalle uusia kysymyksiä, kuten kuka tekee, millaiset suhteet vaikuttavat tekijöiden välillä ja mikä mahdollisesti epäonnistuu. Tarinaa voidaan pitää prosessin alaluokkana, joka vaatii enemmän tietoja ja mahdollistaa rikkaamman prosessikuvauksen.

Tarinaa voidaan myös ajatella prosessin instanssina siinä mielessä, että tarina ei koostu geneerisistä osista niin kuin prosessi, vaan tarinassa jokainen tilanne on itsenäinen yksilö, joka ei ole yleistettävissä. Prosessin osat päätyvät lopulta johonkin toimintaan, joka voi olla osa myös monia muita prosesseja. Esimerkiksi tislaustoiminta voi olla osa pontikankeittoa tai puhdistusaineen valmistusta. Tarinan osa taas on luonteeltaan ainutlaatuinen, koska vaikka siinäkin on jokin toiminta, erottaa tekijän, paikan ja ajan ym. kontekstin tieto sen toisen tarinan tai prosessin vastaavanlaisesta tapahtumasta.

Jos tarinoita halutaan linkittää ristiin, voidaan annotoida tarinoiden sisällön sijasta myös niiden rakennetta. Strukturalismin ja formalismin jalanjäljissä kulkeva narratologia on suuntaus, jossa tekstit nähdään sääntöjen ohjaamina kokonaisuuksina, joilla ihmiset hahmottavat todellisuutta [13]. Narratologiassa pyritään myös etsimään tekstityyppien välttämättömät ja valinnaiset osat ja tavat, joilla näitä voidaan käyttää. Tällaisten vaadittujen tarinan osien kautta voisi sitoa tarinoita yhteen esimerkiksi loppuhuipennus-käsitteen kautta. Jos jonkinlainen narratologian viisasten kivi löytyisi, voitaisiin tällä tavalla varmasti myös rakentaa tarinoita automaattisesti palasista, jotka vain olisi annotoitu hyvin. Tällaisien generoitujen tarinoiden tuottaminen on alan tutkimuksesta huolimatta edelleen lapsenkengissä. Voidaan kysyä, onko loogista edes olettaa, että tarinat olisivat perimmäiseltä luonteeltaan säännönmukaisia. Tiettyjä elementtejä tarinat tietysti vaativat toimiakseen, minkä jo Aristoteles totesi Runousopissaan [1], mutta silti vaikuttaisi siltä, että mielenkiintoisen tarinan generointi vaatii oleellisesti ihmistä tarinankertojaksi.

Narratologian ja muiden strukturalististen suuntauksien ongelmana on usein se, että jotain hyvin olennaista jää puuttumaan tarinasta, kun kokonaisuus puretaan osiin. Kun tarina kuvataan prosesseina eli peräkkäisinä tapahtumina, ei tarinaa itseään saada metatietoon. Tärkeää on siis säilyttää myös alkuperäinen tarina, eikä erehtyä pitämään metatietoa itse tarinan korvaavana asiana. Peräkkäiset tapahtumat kuvattuna metatiedolla kertovat tietysti juonesta jonkin verran, ja jo sen perusteella voi tehdä jonkinlaisia huomioita tarinasta, mutta paljon jää myös puuttumaan.

Tarinoissa on myös asioita, joiden kuvaaminen toiminnalla on vaikeaa. Esimerkiksi pitkät dialogit voivat olla oleellinen osa monta tarinaa, ja aaltoilevaa dialogia on lähes mahdotonta kuvailla toimintakeskeisistä lähtökohdista. Esimerkiksi Väinämöisen ja Joukahaisen kilpalaulanta on dialogia, jossa miehet laulavat tietojaan julki eivätkä kerro mistään tilanteista, joissa joku tekisi jotain. Myöskään pidemmät kuvailut miljööstä tai muista yksityiskohdista eivät välttämättä sisällä toimintaa ja ovat silti osa tarinaa.

Metatiedon käyttö tarinoiden linkittämiseen ja hakuun ei kuitenkaan estä myös itse tarinan sisällyttämistä tietosisältöön. Näin saadaan yksinkertaistettu metatietomalli, jonka perusteella tarjota uusia polkuja itse tarinan luo, mutta tarina säilyttää oman itsenäisyytensä. Alkuperäisen datan säilyttämisen puolesta vedotaan myös Doerrin CIDOC CRM:ää käsittelevässä artikkelissa, jossa on myös pohdittu kulttuuritietosisältöjen semanttiseen linkittämiseen liittyviä kysymyksiä [9].

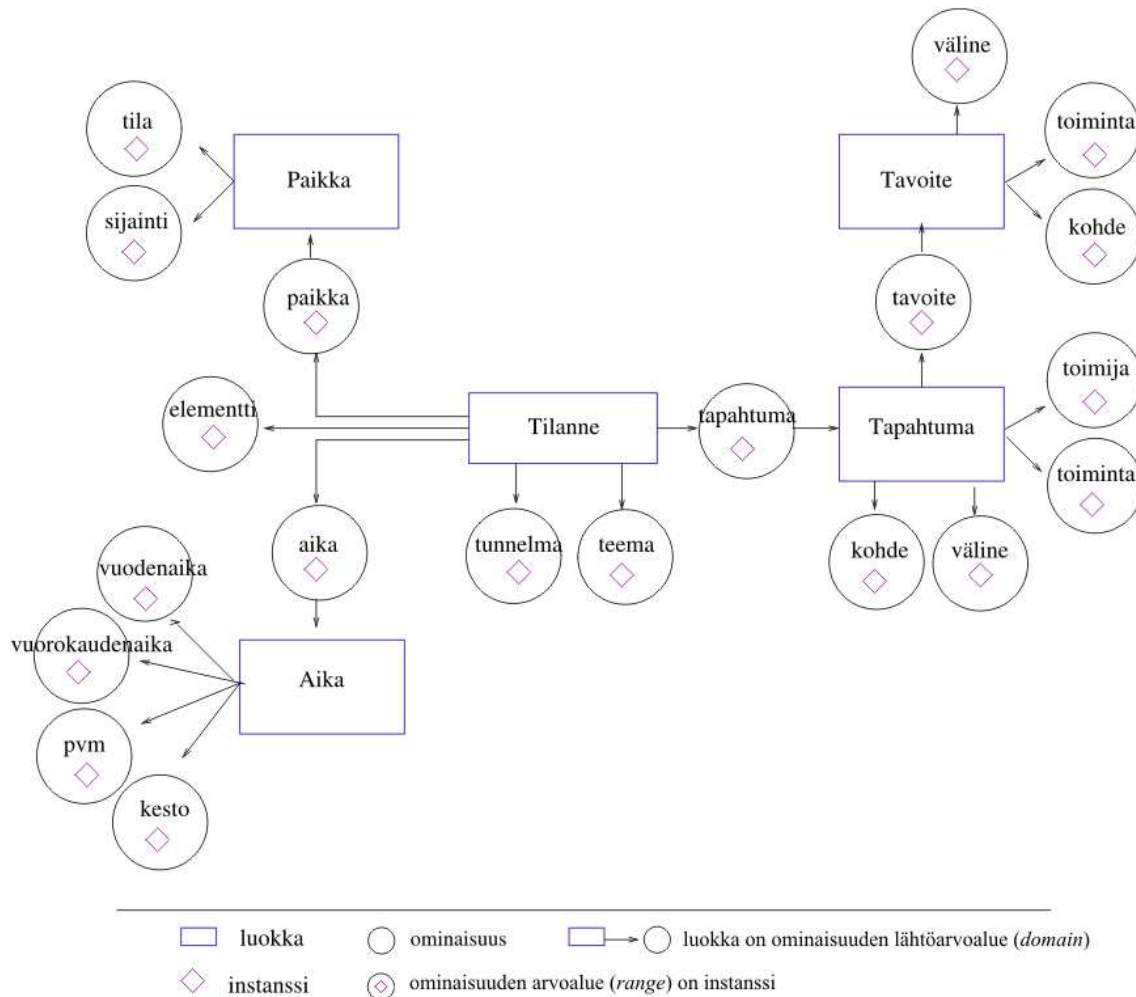
### 3.4 Tilanneontologia

Edeltävän pohdinnan perusteella ja tutkimushypoteesin testaamiseen kehitettiin tässä työssä tilanneontologia, joka on tarkoitettu erilaisten toimintaa sisältävien tietosisältöjen kuvailuun erityisesti kulttuurimateriaalin näkökulmasta. Tilanneontologia pystyy kuvaamaan tilanteita, joissa tehdään jotain, ja prosessiontologian avulla tilanteet pystytään laittamaan oikeaan järjestykseen suhteessa toisiinsa. Prosessiontologiaan palataan seuraavassa luvussa, tässä luvussa keskitytään yhden tilanteen kuvaamisen problematiikkaan. Vaikka tässä luvussa käsitelläänkin tilanneontologiaa yleisesti, on suurin osa esimerkeistä poimittu tutkimuksen teknisen osan materiaalista, koska materiaalin kanssa työskentely myös ohjasi teoreettisen mallin kehittymistä lopulliseen muotoonsa iteratiivisessa prosessissa.

Tilanneontologia ei sisällä hierarkisia luokkasuhteita, ja sikäli se on oikeastaan ennemminkin semanttisen webin teknologialla toteutettu metatietoskeema. Tilanneontologian kohdalla ontologian hierarkia olisi ollut turha, koska tilanneontologia määrittelee vain hyvin yleisen tason kehyksen siitä, millaisia elementtejä reaali maailman (tai kuvitteellisen maailman) prosesseihin liittyy, eikä lähde itsessään määrittelemään sisältöä. Tilanneontologia tarvitsee siis prosessiontologian ohella tuekseen myös sisältöontologian, jota käsitellään myöhemmin lisää.

Tilanneontologia (kuva 3) on ontologia, joka pyrkii mahdollistamaan yhden tilanteen kuvailun, oli se sitten tilanne tarinassa, maalauksessa, prosessissa tai missä tahansa. *Tilanne* on määritelty toiminnan kautta, niin että *toimijan* lopettaessa yhden *toiminnan* ja aloittaessa toisen tilanne vaihtuu. Yhdessä tilanteessa toimija voi kuitenkin tehdä monta asiaa yhtä aikaa ja toimijoita voi olla useita. Tilanteeseen kuuluu mahdollisen toimijan ja toiminnan (eli *tapahtuman*) lisäksi myös *ympäristö*, eli absoluuttinen ja suhteellinen *aika* ja *paikka*, sekä mahdolliset muut tilanteen *elementit*. Myös perusluonteeltaan tulkinnallisempia ominaisuuksia on otettu mukaan, kuten tilanteen *teema* ja *tunnelma*. Tämän luvun lopussa on esimerkki tilanneontologian koodista.

Tilanneontologiaa käytettäessä mallinnukseen on filosofia enimmäkseen se, että jos jotain asiaa ei tiedetä, se jätetään tyhjäksi. Kaikissa asioissa, joita annotoidaan, on kuitenkin myös mahdollisuus merkitä annotaatio tulkinnaksi, jolloin sitä voidaan käsitellä eri tavalla koneellisesti. Tulkinnan monipuolista mahdollistamista tämän prototyypin yhteydessä tukee se, että metatieto on ensisijaisesti tarkoitettu ihmisiä varten, ei niinkään konepäättelyä varten. Monipuolisten ja mielenkiintoisten yhteyksien löytyminen vaatii usein myös tulkintaa, ja siksi se on haluttu sallia. Tulkintaa



Kuva 3: Tilanneontologia. Keskellä olevasta Tilanteesta riippuvat yhden tilanteen kuvailemisen tarvittavat ominaisuudet.

käsitellään tarkemmin seuraavissa aliluvuissa.

Tilanneontologian avulla ei pyritä korvaamaan itse kuvattua teosta/asiaa. Kun esimerkiksi tarinan tai maalauksen sisältöä kuvaillaan tilanneontologialla, on selvää, että hyvin suuri osa itse teoksesta jää vielä kuvaamatta ja kokonaisuus pakenee määrittelyjä. Siitä huolimatta annotointi on mielekästä, kun tarkoituksena ei ole korvata itse teosta, vaan ainoastaan luoda linkkejä ja helpottaa hakua eli käytännössä johdattaa ihmisiä teoksen luo ja tehdä ilmeisemmäksi se kulttuurinen konteksti, johon teos tai asia sijoittuu.

### 3.4.1 Geneerisen prosessitilanteen ja tarinan tilanteen suhde toisiinsa

Tilanneontologiaa käytetään sekä geneeristen prosessin osien että yksilöllisten tarinoiden kuvailemiseen. Vaikka nämä kaksi tapausta ovat joiltain osin luonteeltaan erilaisia, päädyttiin kuitenkin käyttämään samaa ontologiaa. Syynä tälle oli toisaalta se, että sekä geneeriset että yksilölliset tilanteet voivat olla loppukäyttäjän kiinnostuksen kohteina samanaikaisesti ja ne pitää näin saada myös samaan hakuun. Kun ontologia on sama, on molemmantyyppisiä tilanteita helppo hakea. Toisaalta tilanneontologia on suunniteltu siten, ettei erilaisten ominaisuuksien arvoja ole pakko täyttää, ja niiden sallitut arvot on määritelty niin löyhästi, että erityyppisiä tilanteita voidaan kuvata.

### 3.4.2 Tilanneontologian suhde sisällön käsitteet määrittävään ontologiaan

Jotta tietokoneet voisivat ymmärtää myös tilanteeseen liittyvien asioiden suhteita, on kaiken, jolla sisältöä kuvataan, oltava myös määriteltynä ontologiassa. Tilanneontologiassa itsessään ei ole kattavaa sisällönkuvailutermitä, joten prototyypissä olemme ottaneet käyttöön Yleisen suomalaisen ontologian (YSO) alustavan version. Teoreettisen mallin sitominen käytännön ratkaisuun on sikäli ongelmallista, että kyseessä olevan ontologian ratkaisut vaikuttavat myös tilanneontologian rakenteeseen, mutta koska YSO:a suunniteltiin samaan aikaan tilanneontologian kanssa, saatiin tilanneontologian ja sisällön käsitteet määrittelevän ontologian suhteita reflektoitua. Näin erilaisia ratkaisumalleja on tutkittu, ja niihin viitataan myöhemmin.

Kun prosessi koostuu tilanteista, niin tilanteiden toiminnat sekä muut elementit taas koostuvat viittauksista YSO:aan. Esimerkiksi kaskenkaatoprosessissa on tilanne, jossa puut kaadetaan. Tilanteen toiminta-ominaisuuden arvoksi laitetaan viittaus YSO:n käsitteeseen kaataminen, ja tekemisen kohde-ominaisuuden arvoksi viittaus YSO:n käsitteeseen puut. YSO:n luokkien suhteita vertailemalla voidaan sitten luoda mielenkiintoisia linkkejä. Jos esimerkiksi YSO:ssa on määritelty puiden alaluokaksi paju, niin pajupillin valmistusprosessin pajun kaataminen voidaan linkittää kaskiviljelyn puun kaatoon, koska YSO:sta nähdään, että paju on eräänlainen puu. YSO on tällä hetkellä työn alla, ja vielä siinä ei ole kaikkia sanoja, joita sisällönkuvailussa saatetaan tarvita. YSO:aan on prototyypin teon yhteydessä lisätty oleellisia käsitteitä. YSO:n puutteellisuus tietynlaisten sanojen yhteydessä on ongelmallista, mutta jos sen laajentaminen tehdään tulevaisuudessakin helpoksi, syntyvät tärkeät



lisäykset kollektiivisen työn tuloksena, ja kun yhden käsitteen lisää, se on siellä myös seuraavaa annotoijaa varten.

### 3.4.3 Toiminnan kuvailu

Tilanteen tärkein osa-alue on toiminnan kuvailu. Koska tilanteen vaihtumisen vaatimukseksi on määritelty toiminnan vaihtuminen toiseksi, ei ilman toimintaa ole tilannetta. Ainoastaan erikoistapauksissa, joissa kuvataan jokin tilanne erillään muista, kuten esimerkiksi maisemamaalauksen kuvailussa, on tilanne ilman toimintaa mahdollinen.

Yhdessä tilanteessa on mahdollista olla useita tapahtumia. Yksi tapahtuma viittaa yhteen toimintakokonaisuuteen, jossa on toimija, toiminta, toiminnan kohde, ja apuväline. Toimija, toiminta ja apuväline ovat aika selviä semantiikaltaan, ja kaikki viittaavat YSO:n luokkiin tai instansseihin. Toiminnan kohde voi vaihdella tilanteesta riippuen, ollen joskus objekti (esimerkiksi lyödä *palloa*), joskus predikaatti (esimerkiksi yrittää *nousta*) ja joskus adverbiaali (esimerkiksi istua *sohvalla*). Tämä saattaa hieman hämärtää toiminnan kohteen semantiikkaa, mutta johtuen erilaisten sijamuotojen ja toisaalta monentyyppisten verbien ominaislaadusta olisi semantiikan hienomprien vivahteiden mallintaminen hyvin työlästä ja vaikeaa. Ihmisen luontainen taju asioiden suhteista ratkaisee useimmiten semanttisen moniselitteisyyden hänen hakiessaan tietoa, mutta tietokoneen päättelysääntöjä ei voi kovin hyvin rakentaa toiminnan kohteen varaan.

Mikäli YSO:n verbejä ehditään kehittää siten, että ne ovat toimivassa hierarkiassa ja niissä on jokaiselle verbille tyypillisiä ominaisuuksia, voitaisiin toiminnan kohde korvata kulloinkin verbiin sopivalla ominaisuudella. Näin voitaisiin ilmaista, että istumisen paikka on sohva, lyömisen kohde pallo ja puhumisen aihe rakkaus. Tällainen jaottelu toisi hienosyisemmän semantiikan mahdollisuuden toiminnan kohteelle, mutta vaatisi huomattavaa lisätyötä YSO-ontologian parissa.

Toiminnalla voi myös olla tavoite. Tavoite on otettu mukaan ontologiaan, koska se, mitä ihminen tekee, ei aina ole läheskään niin mielenkiintoista, kuin miksi hän tekee. Geneerisissä prosessikuvauksissa tavoitetta ei annotoida, koska prosessikuvauksen semantiikkaan kuuluu, että prosessin vaiheen tarkoitus on viedä prosessia eteenpäin. Kuitenkin heti puhuttaessa ihmiselämästä ymmärretään, että ihmisen tarkoitukset voivat vaihdella hyvinkin paljon. Esimerkiksi Kullervon lähtiessä sotaan voidaan tavoitteeksi lisätä halu kuolla, koska se mainitaan aikaisemmin runossa, ja toisaal-

ta voidaan myös tehdä tulkinta, että tavoitteena on kostaa Untamolle, koska tämä tuhosi Kullervon suvun. Tällaiset tiedot ovat oleellisia sekä haun että linkityksen kannalta, koska esimerkiksi koston teeman toistuminen erilaisissa tietosisällöissä voi olla tiedon etsijän kohteena.

Tulkinnan kysymys on vaikea erityisesti tavoitteen yhteydessä. Jos esimerkiksi maalauksessa näemme miehen ratsastavan, on se jo sinällään oleellinen tieto, mutta jos maalauksen nimestä voimme päätellä, että mies on Kullervo ja matkalla sotaan, niin sotaan lähteminen on myös oleellista ilmaista metatiedolla, jotta maalaus löytyy myös sotaanlähtöön liittyvien sisältöjen joukosta, ei ainoastaan ratsastamisen. Näin olemme mahdollistaneet toiminnan tarkoituksen annotoimisen erikseen, ja silloin on kyse useimmiten tulkinnasta, ainakin jos on kysymys kuvista. Tarinoissa taas saatetaan suoraan sanoa, mikä on toimijan tarkoitus, ja silloin tarkoitus voidaan kertoa ilman tulkinnan tuomaa epävarmuutta. Voi jopa olla, että tulkinnaksi päätyy itse toiminta, jos sitä ei ole tekstissä suoraan sanottu. Kuvasta taas ei voi eksplisiittisesti sanoa, mihin siinä olevat henkilöt pyrkivät, vaan ainoastaan toiminta itse näkyy.

Tavoite on siinä mielessä monitahoinen asia, että tavoitteita voi olla monella eri tasolla. Kuvitellaan tilanne, jossa salamurhaaja piilottelee jätelavalla aseineen. Toiminta on näin ollen jätelavalla istuminen. Selvin tavoite jätelavalla istumiseen on todennäköisesti piilossa pysyminen. Jos katsotaan kuitenkin laajempaa kokonaisuutta, salamurhaajan tavoite on tappaa joku. Jos halutaan kulkea tätä polkua vielä pidemmälle, voidaan sanoa tavoitteeksi myös esimerkiksi oman hengen säilyttäminen, koska salamurhaaja on pakotettu tappamaan oman henkensä uhalla. Tällaista ketjua voidaan usein jatkaa hyvinkin pitkälle. Usein ei ole mielekästä kuvata kovin epäsuorasti itse toimintaan liittyviä tavoitteita, mutta silti muutama eritasoinen tavoite voi olla hyvä annotoida rikkaamman linkityksen aikaansaamiseksi.

Saattaa myös esiintyä tilanteita, joissa samalla toiminnalla on monta tarkoitusta myös samalla abstraktiotasolla, tai monella tilanteessa olevalla toiminnalla sama tarkoitus. Jos samassa tilanteessa Untamo ratsastaa sotajoukon edellä ja sotajoukko marssii perässä, on sekä Untamon ratsastamisella että sotajoukon marssimisella sama tavoite: hyökätä Kalervon väen kimppuun. Monesta erilaisesta tavoitteesta esimerkkinä voidaan mainita Ilmarisen vaimo, joka leipoo kiven leipään. Toisaalta hänen tavoitteensa on ruokkia orja, toisaalta kiusata tätä. Tilanneontologia sallii monen tavoitteen linkittämistä yhteen tapahtumaan ja yhden tavoitteen merkitsemisen monen tapahtuman tavoitteeksi.

Toiminnan lopputulos on tavoitteen lisäksi yksi aspekti toiminnasta. Tätä ei ole mallinnettu, koska esimerkiksi tarinassa tilanne usein johtaa toiseen, ja näin seuraavan tilanteen asetelma on ikään kuin edellisen lopputulos. Tiettyjä rajoituksia lopputuloksen puuttuminen kuitenkin asettaa. Esimerkiksi jossain toiminnassa epäonnistuminen on hieman vaikea mallintaa. Tällä hetkellä sitä voidaan kiertää esimerkiksi sanomalla, että joku yrittää tehdä jotain, josta voi epäsuorasti arvata, ettei lopputulos välttämättä ole juuri se mihin on pyritty.

#### 3.4.4 Ympäristön kuvailu

Vaikka toiminta ja prosessit on valittu tilanteen kuvailun keskipisteeksi, on ympäristöllä myös oma merkityksensä tilanteen kannalta, ja myös sen kautta voi joskus olla mielekästä hakea sisältöjä tai linkittää niitä toisiinsa. Ympäristöä kuvataan tilanneontologiassa ajan ja paikan kautta: missä toiminta tapahtuu ja milloin se tapahtuu. Nämä voivat olla oleellisia tietoja sekä prosessikuvauksissa että tarinoissa.

Sekä ajassa että paikassa on haluttu mallintaa sekä absoluuttista että suhteellista aikaa ja paikkaa. Absoluuttinen aika ja paikka sitovat tilanteen johonkin tiettyyn paikkaan aikatila-avaruudessa, kun taas suhteellinen aika ja paikka kuvaavat aikaa ja paikkaa yleisemmällä tasolla. Absoluuttinen aika ja paikka voidaan antaa esimerkiksi, kun kuvaillaan dokumentaarista materiaalia, kuten lehtijuttua. Jos tiedetään, että lehtijuttu kertoo mitä tapahtui New Yorkissa 9.11.2001, niin tämä voidaan kertoa absoluuttisen ajan ja paikan avulla. Absoluuttista aikaa ei kuitenkaan usein tiedetä, ja monesti se ei ole edes oleellista. Geneerisissä prosessikuvauksissa absoluuttista aikaa ja paikkaa ei edes voida kertoa, koska idea on nimenomaan, että ohjeiden avulla prosessi voidaan suorittaa milloin tahansa ja missä tahansa.

Suhteellinen aika ja paikka kuvaavat tilanteen yleiset puitteet sitomatta sitä mihinkään tiettyyn hetkeen tai paikkaan. Suhteellinen aika jakaantuu vuodenaikaan ja päivän aikaan. Esimerkiksi edellä mainitun lehtijutun suhteellinen aika voisi olla syksyllä, päivällä. Suhteellinen paikka taas olisi kaupunki — tai ehkäpä rakennus, riippuen vähän lehtijutun näkökulmasta. Jos kyse taas on prosessin osasta, voidaan sanoa esimerkiksi, että kaskiviljelyn puunkaato vaihe tapahtuu keväällä metsässä.

#### 3.4.5 Esimerkki tilanteen kuvailusta

Seuraavaksi esitän esimerkin yhdestä tarinatilanteesta ja sen annotoinnista — katso kuva 4. Esimerkki on hieman yksinkertaistettu ja pyrkii esittämään tilanteenkuvai-

lun perustavanlaatuiset piirteet.

### Kalevalan runosta 31:

Kullervo, Kalervon poika,  
hioi siitä kirvehensä;  
päivän kirvestä hiovi,  
illan vartta valmistavi.

Suorihe kasken ajohon  
korkealle korpimaalle,  
parahasen parsikkohon,  
hirveähän hirsikköhön.

Iski puuta kirvehellä,  
tempasi tasaterällä:  
kerralla hyvätki hirret,  
pahat puolella menevi.

Vihoin kaatoi viisi puuta,  
kaiketi kaheksan puuta.  
Siitä tuon sanoiksi virkki,  
itse lausui, noin nimesi:  
"Lempo tuota raatakohon!  
Hiisi hirret kaatakohon!"

### Tilanne

#### Tapahtuma

toimija: Kullervo  
toiminta: Kaataminen  
kohde: Puut  
väline: Kirves  
tavoite:

#### Paikka

tila: Metsä  
sijainti: Untamala

#### Aika

vuodenaika:  
vuorokaudenaika:  
pvm:  
kesto:

elementti:

tunnelma: Aggressiivinen

teema: Epäonnistuminen

vapaaKuvailu: Kullervo  
kaataa kaskea kunnes  
kiroaa koko pellon

Kuva 4: Esimerkki Kalevalan runon yhden tilanteen metatietokuvailusta, vertaa kuva 3.

Tässä esimerkissä voidaan lukea Kalevalasta yksi tilanne. Kullervo on kaatamassa kaskea. Tapahtuma on annotoitu yksinkertaisesti: Kullervo kaataa puita kirveellä. Tekstistä ei käy ilmi Kullervon tavoite, joten tavoitetta ei olla annotoitu. Kaskenkaadon perimmäinen tavoite on pellon raivaaminen, mutta kun Kullervo on orjana pakotettu tähän työhön, hänelle tavoite ei välttämättä olekaan sama kuin käskyn antajalle. Paikan suhteen on käytetty hyödyksi aiemmin tarinassa tullutta tietoa. Tila eli suhteellinen paikka on metsä, koska se mainitaan, mutta absoluuttinen paikka tiedetään aikaisemman tekstin perusteella, koska Kullervo on Untamalassa eli Untamon mailla orjana. Mielenkiintoinen ja monimutkainen sivujuonne tosin aukeaa tästä an-

notoinnista: voidaanko tarinan maailman absoluuttista paikkaa pitää absoluuttisena paikkana? Ja jottei asia olisi liian yksinkertainen, Untamala-niminen paikka tosiaan löytyy Suomesta, Laitilasta. Onko Kalevalan Untamala missään määrin sama kuin Laitilan Untamala? Tähän ongelmaan palataan pian.

Koska tekstistä ei käy ilmi aikaa, ei absoluuttisessa eikä suhteellisessa ajassa ole mitään. Tietenkin voitaisiin tehdä tulkinta, että kyse on keväästä, jos tiedetään että kaski kaadettiin aina keväisin, mutta nyt tätä tulkintaa ei ole haluttu tehdä. Tilanne-elementtejä ei ole myöskään annotoitu, koska tekstissä ei oikeastaan mainita tilanteesta muuta kuin itse toimintaan liittyvät asiat. Jos runossa sanottaisiin, että Kullervo hakkasi puita kahden ison kiven välissä ja että mukana oli myös saha, niin elementteihin voitaisiin laittaa kivet ja saha. Tunnelmaan ja teemaan on esimerkiksi rohkeasti otettu kantaa. Näissähän kyse on tulkinnasta, mutta mitä enemmän tulkintoja saadaan, sitä enemmän potentiaalisesti mielenkiintoisia linkkejä syntyy. Teemassa on myös huomioitu koko Kullervon tarinan teemaa, joka tuntuu olevan orpolapsen elämän välttämätön vaikeus. Kaskan kaadon epäonnistuminen on osa tätä suurta kokonaisuutta. Edellä mainittujen ominaisuuksien lisäksi löytyy vielä vapaa kuvailu, joka on tekstimuotoinen lyhyt kuvaus tilanteesta. Tästä lisää seuraavassa luvussa.

Esimerkin valossa on helppo nostaa esille myös tiettyjä ongelmia, joita mallinnettaessa tulee. Onko paikka esimerkiksi metsä, vai pitäisikö paikan olla pelto? Kun tilanne muuttaa suhteellista paikkaansa tai suhteellista aikaa kuluu, on vaikea yksiselitteisesti ilmaista suhteellinen aika tai paikka. Tässä vaiheessa ratkaisuksi on saanut riittää, että tilanteesta löytyy ainakin jossain vaiheessa kuvailtu asioiden tila. Kun Kullervo tulee paikalle kaatamaan puita, hän on metsässä. Näin väite ei ainakaan ole jokaisessa kohdassa tilannetta epätosi, ja näin olisi myös linkki toisesta metsäaiheisesta asiasta perusteltu.

Palataan nyt ongelmaan Untamalasta. Sama ongelma koskee tietysti myös Kalevalan henkilöitä. Pitäisikö Kullervo jotenkin erottaa lähtökohtaisesti Mannerheimista, vaikka molemmat ovat ihmisiä? Tässä kysymyksessä törmäämme ongelmaan media-lukutaidosta, yleissivistyksestä ja lähdekriittisyydestä. Lapsen saattaa olla vaikea ymmärtää Mannerheimin ja Kullervon eroa, varsinkaan jos hän ei tunne Kalevalaa entuudestaan. Kuitenkin rikkaan linkittämisen ja haun kannalta tuntuisi järkevältä pitää instanssit samantyyppisinä, jotta linkkejä voidaan luoda ja hakutuloksia saada, tai ainakin taruihmisten ja oikeiden ihmisten (tai tarupaikkojen ja oikeiden paikkojen) yhteydet pitäisi saada dokumentoitua johonkin metatietoon.

Toinen ongelma on se, että asioita voi ilmaista monella tavalla. Ontologiat ratkaisevat ongelman, niin kauan kuin näkökulmat ovat suunnilleen samat, mutta jos tapahtumaa lähestytään eri kannalta, voidaan saada hyvin erilaisia lopputuloksia. Jos esimerkiksi yksi henkilö annotoi tapahtuman niin, että Kullervo kaataa puita, ja toinen ilmaisee asian niin, että Kullervo hakkaa mäntyjä, niin puut ja männyt voidaan vielä tulkita samanlaisiksi asioiksi, mutta kaataminen ja hakkaaminen saattavat ontologiassa olla etäämpänä toisistaan, ja näin linkkien generointi jää tekemättä, vaikka olisi kyse samasta toiminnasta. Tämän ongelman takia joitain linkkejä saattaa jäädä syntymättä, koska yksinkertaista ratkaisua ongelmaan ei ole.

Ongelmaan voisi parhaiten puuttua annotointieditorissa, niin että annotoija näkisi muiden tekemät valinnat tiettyyn aiheeseen liittyen. Ohjelma näyttäisi sille, joka on annotoinut tilanteen ”hakata mäntyjä”, että 13 ihmistä on käyttänyt ilmaisua ”kaataa puita” mäntyjen ja muiden puiden yhteydessä, ja ”hakata”-verbin yhteydessä useat ovat käyttäneet ”halkoja”-sanaa. Näin annotoija voisi huomata tehdä valinnan, joka mahdollisimman rikkaasti linkittyy muuhun dataan. Tällä hetkellä käytössä oli kuitenkin vain Protegé-2000, jossa tällaista toimintoa ei ole.

### 3.4.6 Monipuolisemman kuvailun mahdollistavia ominaisuuksia

Jo edellisissä osioissa on kuvattu, miten tietynlaista tulkintaa liittyy kuvailun eri vaiheisiin. Erilaisten yksityiskohtien kuvaileminen ja tulkitseminen on kuitenkin haluttu tehdä mahdolliseksi myös esimerkiksi tapahtuman tavoitteen ulkopuolella, sillä moni muukin asia kuin toiminta voi olla muuta kuin miltä näyttää. Tarinankertajat saattavat käyttää symboliikkaa, joka ei helposti avaudu, mutta joka on oleellista teoksen ymmärtämisen kannalta. Tilanneontologiassa kaikki asiat, joita mallinnetaan, on kääritty tilanne-elementiksi kutsuttuun instanssiin. Sekä Kullervo, metsä, että puukko ovat oma tilanne-elementtinsä. Elementti sisältää linkin YSO:aan, mutta sen lisäksi se sisältää muutaman ylimääräisen ominaisuuden, joita asioihin voidaan sen kautta sitoa.

Elementissä on ensinnäkin attribuutti-ominaisuus, jolla voidaan lisätä adjektiiveja muihin asioihin. Tähän mennessä on vastattu kysymyksiin kuka, mitä, millä, missä ja milloin, mutta attribuutti mahdollistaa myös kysymyksiin miten tai millainen vastaamisen. Attribuutit voivat olla mielenkiintoista linkitysaineistoa, jolla voidaan linkittää esimerkiksi Kullervon rikkinäinen puukko ja palasiksi mennyt Sampo.

Attribuutin lisäksi elementistä löytyy myös symbolointi-ominaisuus. Tällä voidaan

kertoa, mitä tietty tilanteen elementti symboloi. Tällaisen annotoinnin tekeminen vaatii tietysti annotoijalta usein vahvaa ammattitaitoa, mutta mikäli annotoimassa on ihminen, joka on perillä näistä asioista, voi sisältö avautua uudella tavalla myös satunnaiselle tiedon selailijalle. Tässä on suunniteltu, että symbolointi-ominaisuuden pariin tulisi aina selitys, niin että annotoija joutuu perustelemaan sanallisesti, miksi joku asia symboloi toista, ja näin voidaan välttää perusteettomat tai siltä tuntuvat annotoinnit.

Muutenkin tilanneontologiaan on haluttu jättää vapaan kuvailun kenttiä moneen paikkaan, joissa asia voidaan selittää luonnollisella kielellä. Tämä on perusteltua, koska lopullinen tieto on kuitenkin tarkoitettu ensisijaisesti ihmisille, ja ihmisen on helpompi seurata selväsanaista kuvailua kuin ontologian kiemuroihin rajoitettua metatietomallia. Vapaista kuvailuista voidaan siis poimia tekstit käyttöliittymään ihmisen luettaviksi.

### 3.4.7 Ongelmia

Doerr määrittelee ontologioiden rakenteen olevan kompromissi, jossa tasapainoillaan formaalien kyselyiden tavoitettavissa olevan informaation monimutkaisuuden, käyttäjän ymmärrettävissä olevan monimutkaisuuden, monimutkaisuuden, jonka toteuttamiseen käyttäjällä on varaa, ja näiden rakenteiden oppimisen ja käyttämisen (metatiedon luominen) helppouden välillä [9]. Näitä kysymyksiä tuli vastaan myös tässä prosessissa. Tilanneontologian suunnittelu oli toisaalta suoraviivaista, mutta myös monenlaisia filosofisia sekä käytännöllisiä vaikeuksia ilmeni. Suurin ongelma, jonka kanssa oli elettävä, oli mallinnuksen perusongelma: kuinka paljon yksinkertaistaa, kuinka paljon monimutkaisia rakenteita on sallittava. Aluksi rakennettiin ensimmäinen versio ontologiasta, ja kun sitä käytettiin annotointiin, huomattiin puutteita. Myös erilaisten esimerkkien kautta huomattiin usein suunnitteluprosessin edetessä erilaisia erikoistapauksia, joiden kuvaamiseen käytetty malli ei sopinut. Lopulta päädyttiin kompromissiin. Osa ongelmista ratkaistiin. Tässä kohdin käsitellään joitain niistä, jotka jäivät ongelmiksi loppuun saakka.

Tarinoiden mallintaminen sisältää mielenkiintoisia kysymyksiä verrattuna vaikkapa jonkin koneen toiminnan mallintamiseen. Tarina on toisaalta yksi kokonaisuus mutta myös ikkuna toiseen todellisuuteen. Koska tarinaa voidaan kertoa lukemattomilla eri tavoilla, ei ole itsestään selvää, miten tekstiin pitäisi suhtautua. Onko teksti kuvaus toisesta todellisuudesta vai ainoastaan näkökulma siihen? Esimerkki yhdestä tällaisesta ongelmasta löytyy Kullervon tarinasta, jossa alussa kerrotaan, että Untamon

sotajoukko tuli ja tappoi kaiken Kalervon väen Kullervon äitiä lukuun ottamatta. Myöhemmin käy kuitenkin ilmi, että Kalervo perheineen onkin hengissä. Jos edelliseen kohtaan on annotoitu, että Kalervo tapettiin, ja loppupäässä tarinaa Kalervo onkin taas hengissä, niin konepäättely alkaa ainakin hyvin helposti protestoida eikä ihminenkään helposti pysy mukana. Toisaalta, jos annotoidaan, että Untamo yritti tappaa Kalervon, ihmislukijalle tulee ristiriita, jos hän katsoo itse tekstiä, jossa sanotaan, että Kalervo tapettiin. Tiedon pimentäminen lukijalta voi tietysti usein olla myös draaman keino, ja yllätyksen pilaaminen metatiedon avulla ei kuulosta suositeltavalta vaihtoehdolta.

Tällaiset tekstin ristiriitaisuudet ovat mallinnuksen kannalta vaikeita. Jos kaikkia kerronnallisia näkökulmia yritettäisiin saada mukaan metatietoon, voitaisiin konepäättelyä ehkä hieman helpottaa, mutta suurin eli subjektiivisuuden ongelma pysyisi. Jos kirja kertoo jostain subjektiivisesta näkökulmasta, emme voi tietää mitä oikeasti tapahtui. Eli mitä silloin annotoimme? Ja mitä oikeasti tapahtuu...

### 3.4.8 Otteita tilanneontologian koodista

Seuraavaksi esimerkkinä otteita tilanneontologian määrittelyistä RDFS:lla. Protegé-2000-editorilla oli ongelmia ääkkösten kanssa, joten esimerkeistä on kaunisteltu jotain pois.

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY a 'http://protege.stanford.edu/system#'>
  <!ENTITY ns_04_tilanne 'http://www.cs.helsinki.fi/group/seco/ns/2004/10/04-tilanne#'>
  <!ENTITY rdfs 'http://www.w3.org/TR/1999/PR-rdf-schema-19990303#'>
]>

<rdf:RDF xmlns:rdf="&rdf;"
  xmlns:a="&a;"
  xmlns:ns_04_tilanne="&ns_04_tilanne;"
  xmlns:rdfs="&rdfs;">
```

...

Alun määrittelyiden jälkeen tulevat luokat. Ne ovat yleisen RDFS-yläluokan resursin aliluokkia, koska niille ei ole mitään loogista yläluokkaa olemassa. Tapahtuma, tavoite ja tilanne ovat ontologian perusluokkia, kun taas KuvailuElementti on luokka, johon kääritään viittaukset sisällönmäärittelyontologiaan, eli tässä tapauksessa



YSO:aan, ja sinne on laitettu mahdollisuus merkitä asia tulkinaksi sekä mahdollisuus lisätä attribuutteja elementtiin, mikä se sitten onkin.

```
<rdfs:Class rdf:about="&ns_04_tilanne;Tapahtuma"
  rdfs:label="Tapahtuma">
  <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdfs:Class rdf:about="&ns_04_tilanne;Tavoite"
  rdfs:label="Tavoite">
  <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdfs:Class rdf:about="&ns_04_tilanne;Tilanne"
  rdfs:label="Tilanne">
  <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdfs:Class rdf:about="&ns_04_tilanne;KuvailuElementti"
  rdfs:label="KuvailuElementti">
  <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
```

...

Luokkien jälkeen määritellään niiden ominaisuudet. Vapaa kuvaus on ominaisuus, jonka arvona on mikä tahansa literaali, ja sen paikkaa ei ole määritelty, koska vapaan kuvauksen lisääminen on haluttu mahdollistaa kaikille instansseille. Aika-ominaisuus on sen sijaan määritelty kuulumaan Tilanne-luokan instanssille, ja sen arvoksi käy vain Aika-luokan instanssi. Kohde-ominaisuus on sekä Tapahtuman että tavoitteen instansseilla, ja sen arvoksi laitetaan KuvailuElementin instanssi, johon on siis kääritty viittaus sisältöontologiaan, sekä mahdollisia lisätietoja, kuten attribuutti, joka on määritelty kohteen jälkeen tässä esimerkissä.

```
<rdf:Property rdf:about="&ns_04_tilanne;VapaaKuvaus"
  a:maxCardinality="1"
  rdfs:label="VapaaKuvaus">
  <rdfs:range rdf:resource="&rdfs;Literal"/>
<rdf:Property rdf:about="&ns_04_tilanne;aika"
  a:maxCardinality="1"
  rdfs:label="aika">
  <rdfs:range rdf:resource="&ns_04_tilanne;Aika"/>
  <rdfs:domain rdf:resource="&ns_04_tilanne;Tilanne"/>
</rdf:Property>
<rdf:Property rdf:about="&ns_04_tilanne;kohde"
  a:maxCardinality="1"
  rdfs:label="kohde">
  <rdfs:range rdf:resource="&ns_04_tilanne;KuvailuElementti"/>
  <rdfs:domain rdf:resource="&ns_04_tilanne;Tapahtuma"/>
  <rdfs:domain rdf:resource="&ns_04_tilanne;Tavoite"/>
</rdf:Property>
<rdf:Property rdf:about="&ns_04_tilanne;attribuutti"
```

```

    rdfs:label="attribuutti">
    <rdfs:domain rdf:resource="#ns_04_tilanne;KuvailuElementti"/>
    <rdfs:range rdf:resource="#rdfs;Resource"/>
</rdf:Property>
</rdf:RDF>

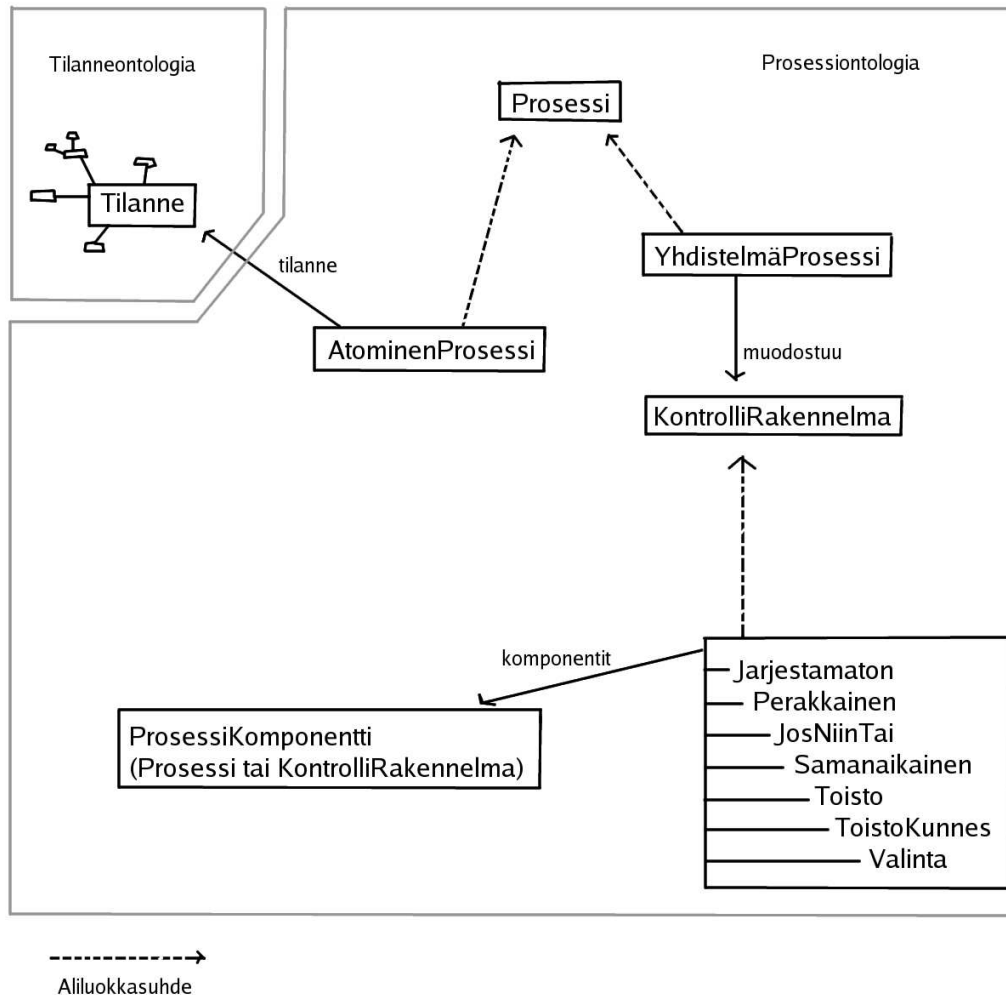
```

### 3.5 Prosessiontologia ja OWL Services

Tilanneontologialla voidaan kuvata yksittäisiä prosessin osia, mutta koska prosessi koostuu usein monesta osasta, oli tehtävä myös ylemmän tason ontologia, joka sitoi tilanteet kokonaisuuksiksi, olivatpa nämä kokonaisuudet sitten prosessikuvauksia tai tarinoita. Erilaisista prosessimalleista OWL-S osoittautui tässä tutkimuksessa hyväksi pohjaksi, koska se mahdollisti atomisten prosessien (eli tämän tutkimuksen tilanteiden) järjestämisen ja oli myös dokumentoitu selkeästi. Näin prosessiontologia (kuva 5) tehtiin suoraan kääntämällä OWL-S:stä ainoastaan pienillä muunnoksilla ja lisäyksillä.

OWL Services (OWL-S) on kehittyvän semanttisen webin palveluiden esittämiseen, tarjoamiseen ja toteuttamiseen kehitetty kieli [7]. Koska yhden webin tälläkin hetkellä tärkeimmistä resursseista ovat niitä, jotka tarjoavat palveluja, olisi myös palveluiden tarjoaminen tehtävä helpommaksi semanttisessa webissä. Tärkeimmät lähtökohdat OWL-S:n kehityksessä ovat olleet seuraavat:

- Automaattinen web-palveluiden löytäminen. Tällä hetkellä ihmiset joutuvat itse etsimään tietoja tekstimuodossa erilaisten firmojen sivuilta. OWL-S:n avulla nämä palvelut kuvattaisiin koneen ymmärtämällä tavalla ja myös niiden oleelliset ominaisuudet niin, että ihmisen tarvitsisi ainoastaan kertoa, mitä ja millaisilla reunaehdoilla haluaa, ja kone löytäisi sen perusteella oikean palvelun.
- Automaattinen web-palveluiden suorittaminen. Koska muun muassa kaupankäynti webissä on samantyyppistä kerrasta toiseen (Kaupat tarvitsevat luottokortin numeron, osoitteet ym.), voisivat agentit suorittaa nämä toimenpiteet automaattisesti. OWL-S pyrkii tarjoamaan ohjelmille ja agenteille käytettäväksi API:t, joiden perusteella palvelu voitaisiin automaattisesti suorittaa.
- Automaattinen web-palveluiden kokoonpano ja yhteistyö. OWL-S:n avulla olisi agenttien pystyttävä kokoamaan useammista pienemmistä operaatioista suurempi kokonaisuus, joka vaaditaan jonkin korkeamman tason käyttäjätarpeen



Kuva 5: OWL-S:n pohjalta tehty prosessiontologia tärkeimpine luokkineen ja ominaisuuksineen. KontrolliRakennelman aliluokat laitettu samaan laatikkoon kuvassa.

tyydyttämiseksi. Siksi erilaisten automaattisten prosessien yhteydet ja järjestykset tulee voida kuvailla OWL-S:llä.

- Automaattinen web-palveluiden seuranta. Isommat prosessit saattavat kuluttaa aikaa ja vaatia odottamista tietyissä väleissä. Käyttäjän on pystyttävä OWL-S:n avulla seuraamaan automaattista prosessin kulkua.

Kaikkia näitä asioita ei tämänhetkisessä versiossa ole vielä toteutettu. Prosessien mallinnus on vain yksi osa OWL-S:ää. Näkökulma on varsin datatransformaatiokeskeinen, koska tietokoneiden prosessit liittyvät useammin vain datan muuttumiseen eikä niinkään koneiden ulkopuolisten asioiden mallintamiseen. Silti OWL-S:n

prosessimalli on hyödyllinen esimerkki prosessin mallintamisesta OWL:lla, joka on semanttisen webin perustekniikoita.

Prosessit jaetaan OWL-S:ssä osiin, kunnes saavutetaan atomisen prosessin taso. Näistä atomisista yksiköistä voidaan rakentaa monimutkaisempia prosesseja, ja edettäessä monimutkaisempiin prosesseihin voidaan alempi taso aina piilottaa yhden osikon alle. Pizzan valmistuksessa on monta vaihetta, mutta kun mallinnetaan vaikkapa pizzan tilauksen eri vaiheita, voidaan pizzan valmistus pitää vain yhtenä osana prosessia, jonka kuitenkin voi halutessaan jakaa pienempiin osiin. Abstraktiotasosta on vain päätettävä tilanne- tai sovelluskohtaisesti.

Erilaiset prosessin osat voivat olla erilaisissa suhteissa toisiinsa. Perusprosessi saattaa kulkea ainoastaan suorittamalla osaprosessit järjestyksessä, mutta OWL-S:stä löytyy myös mahdollisuus toistaa jotain osaa, kunnes jokin ehto täyttyy, valita vaihtoehtoinen prosessi jostain ehdosta riippuen tai tehdä asioita samanaikaisesti määrittelemättömässä järjestyksessä. Nämä ohjelmointikielistäkin tutut mallit tuntuvat järkevältä lähestymistavalta prosesseja kuvatessa, ja vaikka erilaisia operaatioita ei ole paljon, on vaikea keksiä prosessia, jota niillä ei saisi kuvattua. Kaikkea tämän kuvaamiseen tarvittavaa ei tosin sisälly OWL-S:ään, vaan esimerkiksi erilaisten ehtojen ja olosuhteiden kuvaaminen on jätetty OWL-S:stä toisille kielille. Esimerkiksi ehtojen kuvaamiseen on ehdotettu käytettäväksi Semantic Web Rules Language -kieltä (SWRL) [19].

Käytännössä OWL-S:n pohjalta tehtyä ontologiaa käytettiin siten, että tilanneontologian instansseista tehtiin atomisia prosesseja. Atomisia prosesseja voitiin sitten kasata suuremmiksi osakokonaisuuksiksi. Tärkeintä oli pystyä kuvaamaan atomisten prosessien järjestys toisiinsa nähden, mutta tämän mallintamiseen ControlConstruct-luokan aliluokkavalikoima OWL-S:stä sopi hyvin. Ongelmaksi muodostui Protegé-2000-editorin kyvyttömyys luoda järjestettyjä listoja, vaikka ne kuuluvat RDF:n perusmäärittelyihin. Prototyypin ensimmäisessä versiossa ongelma kierrettiin lisäämällä järjestysnumero atomisiin prosesseihin, niin että järjestys voitiin generoida numeron perusteella. Uusiin versioihin olisi syytä kehittää älykkäämpi järjestelmä, etteivät kovakoodatut numeroarvot estä atomisten prosessien kierrätystä silloin, kun sama atominen prosessi on usean ylemmän tason prosessin osa.

## 3.6 Sisällönkuvailun haasteita

Tässä tutkimuksessa lähtökohdaksi otettu tapa kuvata tietosisältöjä niiden sisällön kautta (sisällön myös tulkinnallisessa mielessä) aiheuttaa uusia ongelmia. On helpompaa kuvata konkreettisia asioita kuten tuolin materiaalia ja mittoja kuin jotain abstraktimpaa asiaa kuten tarinan kohtausta tai maalauksen tunnelmaa. Seuraavissa aliluvuissa on nostettu esiin oleellisimpia kysymyksiä, joihin työssä törmättiin.

### 3.6.1 Tulkinnan ongelma

Kuinka tulkintaa vaativia ominaisuuksia otetaan mukaan ontologiaan itsestäänselvien ominaisuuksien lisäksi? Tällaisia tulkintaa vaativia asioita ovat esimerkiksi jonkin kuvan tunnelma. Jonkun mielestä kuva on surumielinen, jonkun toisen mielestä taas rauhallinen ja lohdullinen. Tämä kysymys oli yksi oleellisimpia ontologiaa suunniteltaessa. Ihminen tulkitsee asioita usein huomattavasti pidemmälle, kuin jokin asia eksplisiittisesti antaa ymmärtää. Näiden epäsuorempien ominaisuuksien kautta on mahdollista yhdistää asioita semanttisesti monipuolisemmin ja kiertää tietokoneiden eksplisiittistä perusluonnetta.

Tulkintaa vaativat ominaisuudet tuovat mukanaan ongelmia. Koska asiat voidaan tulkita monella tavalla ja usein on vielä lopulta mahdotonta tarkistaa, kuka todella on oikeassa, muuttaa tulkintaa vaativien ominaisuuksien lisääminen ontologiaan metatiedon subjektiivisemmaksi. Tulkinnanvaraisten ominaisuuksien lisääminen ontologiaan on valinta, jota tehdessä on jälleen hyvä pohtia, mihin ontologiaa ensisijaisesti tullaan käyttämään. Jos ontologiaa ja sen mukaista dataa käyttävä sovellus on ohjelmoitu ottamaan jotenkin huomioon tulkinnallisten ominaisuuksien epävarmuus, voi niitä lisätä huolettomammin, ja itse asiassa joidenkin käyttäjryhmien mielenkiinto saattaa suuntautua nimenomaan tulkinnallisiin ominaisuuksiin. Jos tietoa taas tulevat käyttämään pelkät koneet eli agentit tai muut vastaavat, on tulkinta ongelmallisempaa.

Tulkinnan ongelman kohdalla tulevat myös esiin kysymykset arvottamisen oikeellisuudesta. Siinä vaiheessa kun erilaisten tietosisältöjen taiteellista tai tiedollista arvoa aletaan määrittellä ja lisätä metatietoon, saa annotoija yhä suuremman vallan ja sitä kautta vastuun. Arvottajan ammattitaidon pitäisi ainakin olla mahdollisimman hyvä, ennen kuin arvottamista vaativaa metatietoa edes harkitaan otettavaksi mukaan ontologiaan. On kuitenkin selvää, ettei ammattitaitoisenaan arvottajan arvostelukyky vastaa kaikkien ihmisten tarpeita. Kulttuuriset ja elämäkatsomuk-

selliset seikat vaikuttavat niin paljon siihen, miten ihminen asioita tulkitsee, että kaikki arvostelmat ovat aina vajavaisia jollekin yleisölle.

Mitä suurempia kokonaisuuksia semanttinen web tulee tulevaisuudessa kattamaan, sitä suurempia ongelmia syntyy siitä, että yhä heterogeenisempi ihmisjoukko on yhden rajoitetun näkökulman vankina. Tällaisen tilanteen estämiseksi voidaan tehdä ontologioista monipuolisempia, niin että ne mahdollistavat tulkintojen perusteluiden lisäämisen, joka saattaisi hieman helpottaa näkökulman rajoittuneisuutta kertomalla eksplisiittisesti arvostelijan lähtökohdat. Monien erilaisten arvostelujen lisääminen metatietoon vähentäisi arvostelijan vastuuta, ja ihmiset saisivat itse tulkita arvosteluita. Samalla tavoin voimme lukea monet arvostelut elokuvasta ja tehdä niiden perusteella omat johtopäätöksemme.

### 3.6.2 Totta vai tarua?

Yksi ongelmallinen seikka maailmaa mallinnettaessa on mallinnettavan asian suhde todellisuuteen. Jos ontologiat tehdään ainoastaan reaalimaailman asiat huomioon ottaen, voivat monenlaiset kiinnostavat ja oleellisetkin asiat jäädä kokonaan metatiedon ulottumattomiin. Jotkut asiat ovat helppoja määritellä, mutta jos puhutaan esimerkiksi leikkilentokoneesta tai Väinämöisestä, on luokittelu jo huomattavasti vaikeampaa. Leikkilentokone ei ole lentokone, mutta jos ollaan kiinnostuneita esimerkiksi ilmailun historiasta, on sääli jos lelulentokone ei yhdisty mitenkään aiheeseen vain siksi, että se on leikkikalu. Ja miten määrittelemme Kalevalan tarinoiden toimijoita ja prosesseja, jos niitä ei voi annotoida samojen prosessien ja henkilöiden sekaan kuin oman maailmamme asiat?

Sekä tarinat että leikit ovat luonteeltaan monilta osin samanlaisia kuin meidän maailmamme. Niihin on kopioitu suuria osia todellisuudesta, mutta kaikki säännöt eivät enää päde. Tarinoiden kirjoittajien tai leikkijöiden mielikuvitus saattaa tehdä joistain reaalimaailman asioille määritellyistä ominaisuuksista tai rajoitteista pilkkaa, mutta silti asioiden välillä on myös selvä yhteys. Mielikuvitus voi tosin aiheuttaa ongelmia luokittelijoille reaalimaailman sisälläkin. Jos kehitysmaihin vietyä traktoria on käytetty vuosikaudet lasten kiipeilytelineenä, pitäisikö se annotoida traktoriksi vai kiipeilytelineeksi — vai mahdollisesti linnaksi, jollaiseksi siinä leikkivät lapset ovat sen mieltäneet?

Näennäisten loogisten konfliktien kanssa eläminen tulisi jotenkin mahdollistaa ontologioita käytettäessä, mikäli niiden perusteella pitäisi yhdistellä monipuolisia tietosi-

sältöjä. Luokittelu ei saisi rajoittaa ihmisen assosiaatiokykyä. Jokaiselle leikkimaailmalle ei voida tehdä omia ontologioita, mutta assosiaatiot voidaan mahdollistaa esimerkiksi lisäämällä ontologioiden luokkiin ominaisuus, josta voidaan määritellä onko kyseessä oikea asia vai ainoastaan leikki.

### 3.6.3 Ajan ja paikan vaikutus ontologioihin ja instansseihin

Jos samalla ontologialla pyritään kattamaan instansseja hyvin laajalta aikaväliltä, tulee helposti ongelmia asioiden määrittelyn kanssa. Jos puhutaan vaikkapa museoesineistä ja mietitään millaisia ominaisuuksia kirveelle pitäisi määritellä, on esimerkiksi sotakirveen ominaisuuksia turha annotoida nykyaikaisille kirveille, koska niitä ei enää käytetä (ainakaan moderneissa) sodissa. Näin esineen suhteet vaikkapa sotakalustoon ovat relevantteja vain tietyllä aikavälillä, mutta eivät enää sen jälkeen.

Kirves on siinä mielessä helppo käsite, että kirveet ovat kuitenkin kautta aikojen olleet suunnilleen samanlaisia esineitä. Tiettyjen käsitteiden merkitys taas muuttuu totaalaisesti ajan kuluessa. Tämä pitää erityisesti paikkansa prosessien kanssa, joiden päämäärät pysyvät usein samana, mutta tekniikat ja aliprosessit voivat muuttua radikaalisti esimerkiksi uuden teknologian tai innovaation vaikutuksesta. Jos puhutaan tietyn sairauden hoidosta, niin kappaus saattoi olla oleellinen osa prosessia vielä sata vuotta sitten, mutta nyt annetaankin potilaalle lääkettä. Näin yhtenä aikana tärkeäksi katsottu osa prosessia on saattanut muuttua jopa haitalliseksi päämäärän saavuttamisessa.

Ajan lisäksi myös paikka (tai kulttuuri) voi vaikuttaa käsitteiden sisältöön huomattavasti. Viljelyprosessi on hyvin erilainen eri puolilla maailmaa. Jo aikaisemmin mainitussa toiminnan teoriassa keskitettiin näkökulma tähän: kulttuuri, yhteisö ja välineet vaikuttavat prosessien syntyyn ja ihmisen toimintaan. Jos metatietoon on määritelty jokin prosessi, se on totta ainoastaan yhden ajan ja paikan suhteen. Se voi olla tietysti universaalimpikin asia, mutta näin ei voida olettaa, ellei asiaa ole tutkittu.

Prosessien lisäksi myös paikat, henkilöt ja tarinat voivat muuttua. Jos näitä muutoksia halutaan tutkia, on erilaisten versioiden sisällyttäminen aineistoon tärkeää, mutta mitä monipuolisemmin asioita halutaan katsoa ajan vaihtelun näkökulmasta, sitä enemmän on metatietoa ja sen taustalla olevaa osaamista oltava. Jos tällaiseen kattavaan metatietoon on mahdollisuus, voisi semanttisten linkkien tarkkuus olla huomattavasti parempi eikä linkeistä tulisi liian hämmentäviä suurten aikahyppäys-

ten tai paikka-/kulttuuriloikkien takia. Toisaalta joskus juuri nämä yhteydet voivat olla mielenkiintoisimpia.

## 4 KulttuuriSampo — case

Jotta prosessien soveltuvuus erilaisten tietosisältöjen semanttiseen yhdistämiseen voitaisiin todeta oikeaksi, sitä on kokeiltava käytännössä. Alkuperäinen ajatus prosessien käyttämisestä on syntynyt nimellä MuseoSuomi<sup>1</sup> kulkevan suomalaisen kulttuurin portaalin kehittämisen yhteydessä, ja tämän tutkimuksen yhteydessä toteutettiin prototyyppi MuseoSuomen laajennuksesta, jossa on monipuolisempia tietosisältöjä ja prosessikuvauksia mukana. Tämän prototyypin työnimenä on KulttuuriSampo, ja sen visiona on pyrkiä kuvaamaan semanttisia yhteyksiä erilaisten suomalaisten kulttuurisisältöjen välillä. Tätä varten on myös tutkittu muita olemassa olevia digitaalisia tietosisältöjä, jotta lähestymistapamme hyvät ja huonot puolet suhteessa muihin saataisiin kartoitettua.

### 4.1 Kulttuurialan digitaaliset tietosisällöt ja MuseoSuomi

Web on avannut erilaisille kulttuuritahoille uuden kanavan aineistonsa esittelyyn. Uuden teknologian käyttö kulttuurialalla on lisääntynyt kuten webin käyttö muillakin yhteiskunnan aloilla. Yksi webin ominaisuuksista on sen mahdollistama multicasting, eli tiedon julkaiseminen on mahdollista useille ja julkaistu materiaali on useiden ulottuvilla. Erilaisten kulttuuristen sisältöjen julkaiseminen onkin yleinen webin hyödyntämistapa. Niin museot kuin yksityiset ihmisetkin saavat töitään esille helposti, ja itse esineet eivät ole vaarassa vaurioitua tai joutua varastetuiksi.

Tavallisesti webissä julkaistuun laajempaan materiaaliin on lisätty hakujärjestelmä. Myös erilaisten medioiden käyttö yhdessä on yksi yleinen ja vaihtelevasti toimiva ratkaisu. Parhaimmillaan videot, äänet ja animaatiot tukevat kokonaisuutta hyvin ja antavat mahdollisuuden rikkaampaan kokemukseen ja monipuolisempaan tiedon saantiin. Pahimmillaan erilaiset päälle liimatut medialeikkeet vain sotkevat järjestelmän selkeyttä eivätkä sisällä mitään todella hyödyllistä tietoa. Myös interaktiivisuuden mahdollisuutta on käytetty hyväksi joskus paremmalla, joskus huonommalla menestyksellä.

Erilaisten näkökulmien käyttö vaikuttaa myös olevan yleinen lähestymistapa, eli

---

<sup>1</sup><http://www.museosuomi.fi>



tietoon annetaan erilaisia polkuja aina tietyn näkökulman mukaan. Tämä on mahdollisuus semanttiselle yhdistämiselle, koska erilaisten linkkien generoitua automaattisesti voidaan näistä valita tietyt linkit tiettyyn näkökulmaan, ja näin linkkien suunnitteluun ei tarvitse käyttää niin paljon aikaa, kunhan vain valitaan oikeat linkkisäännöt jokaiseen näkökulmaan.

### **MuseoSuomi — semanttinen portaali**

MuseoSuomi on semanttinen portaali, jonka tarkoitus on julkaista museokokoelmia semanttisessa webissä [23]. Ensimmäiseen versioon julkaistuna materiaalina on museoesineitä Suomen kansallismuseon, Espoon kaupunginmuseon sekä Lahden kaupunginmuseon kokoelmista, yhteensä noin 4500 esinettä. MuseoSuomen tarkoituksena on toimia sekä esimerkkinä semanttisen webin tekniikoiden mahdollisuuksista että näyttää suuntaa museoille, joissa pohditaan digitaalitekniikan mahdollisuuksia. MuseoSuomessa on myös kokeiltu mobiilitekniikan toimivuutta museoesineiden verkkoselaamiseen ja kokeiltu paikannustietojen ja erilaisten tietosisältöjen — tässä tapauksessa muinaismuistotietojen — yhdistämistä esinetietoihin monipuolisempien palveluiden tarjoamiseksi.

MuseoSuomen tärkeimmiksi päämääriksi on määritelty seuraavat [23]:

- Globaali näkymä hajautettuihin aineistoihin. Vaikka eri museoiden materiaalit ovat omissa tietokannoissaan ja erilaisilla logiikoilla kuvailtu, MuseoSuomessa ne on yhdistetty syntaktisesti ja semanttisesti ja mahdollistettu eri kokoelmien tietojen linkittäminen ja vertailu.
- Aineistopohjainen tiedonhaku. MuseoSuomessa voidaan hakea tietoa ontologian käsitteiden avulla älykkäämmin kuin perinteisissä sanahakujärjestelmissä, ottamalla huomioon myös haettavan sanan paikka ontologisessa hierarkiassa, mikäli se löytyy ontologiasta.
- Semanttinen linkitys esineiden välillä. Erilaiset suorat tai epäsuorat semanttiset yhteydet nostetaan materiaalista esiin luomalla käyttöliittymään linkkejä. Nämä linkit määritellään vain yleisellä tasolla logiikkasääntöinä, minkä jälkeen linkit syntyvät materiaalin pohjalta itsestään.
- Helppo aineistojen julkaiseminen. Portaalin tulisi olla helppo ja halpa tapa museoille julkaista aineistojaan webissä.

Näille periaatteille rakennettu järjestelmä on toiminut hyvin, vaikka kyseessä onkin vasta ensimmäinen versio. Toisaalta myös lähestymistavan ongelmia on löydet-

ty prosessin kuluessa. Yksi keskeinen ongelma on ollut alkuperäisten museotietojen epämääräisyys: tietokannoista saatavasta datasta puuttuu usein tietoja, ja olemassa olevat ovat epäkonsistentisti luetteloituja, niin että ne on vaikea lokeroida tietokoneen ymmärtämiin luokkiin. Myös erilaiset käytetyt terminologiat ovat aiheuttaneet tiettyjä yhteentörmäyksiä ja vaatineet toimenpiteitä.

### **MuseoSuomen ongelmia**

Hyvästä käyttäjäpalautteesta huolimatta myös ongelmia ja puutteita on ilmennyt. Satunnaisten virheellisten tietojen ja vastaavien virheiden lisäksi on myös tullut palautetta kokonaisuudesta. Yhdessä palautepostissa todetaan muun muassa seuraavaa (suora lainaus):

*”Onko suunnitelmissa tarjota esineistölle tarkempia taustatietoja? Nyt esineet esitellään: siinä ne ovat — katselkaa vaan. Mutta ei kerrota, mihin, miten ja miksi niitä on käytetty. Vaikka nyt voin kuvitella, mitä kynttiläsaksilla tehtiin, en ole aivan sataprosenttisen varma siitä, että olen oikeassa. Ja mitä varten berlokkeja oli ja kuka sellaisia tarvitsi ja missä he sellaista pitivät, kaulassa vai taskussa vai työpöydällä vai ihan jossain muualla? Sellainen museo-oppaan selitys palvelisi ainakin minua. Osa esineistä ja niiden käyttötarkoitus tietysti on tuttua, mutta eivät kaikki esineet.”*

Myös museoalan ammattilaiset ovat etsimässä ratkaisuja näihin kysymyksiin. Trilce Navarrete kirjoittaa seuraavasti [33]:

*”Kun museot laajentavat ’opettamisen’ määritelmää, on syytä ottaa huomioon vaihtoehtoisia selitysmalleja, kuten suullinen historia, mytologia, kansanperinne ja muita tapoja, jotta voitaisiin luoda konteksti, jossa tarinoita kerrotaan — jokaisen museon tarina. Miten museot vaalivat yhteisönsä kertomuksien rakentumista? Kysymys museoissa ei ole ainoastaan se, miten selittää tietyn esineen tarina, vaan myös se, miten museot voisivat paremmin luoda ja herättää kontekstin yleisölle niin että nämä voivat rakentaa kiinnostuksen paneutua tarinoihin.”*

Näissä puheenvuoroissa on kerrottu pähkinänkuoressa se ongelma, jota uudessa prototyypissä pyritään ratkaisemaan. Itse museoesineiden tiedot eivät yksin tyydytä ihmisten mielenkiintoa. Museoesineen valmistaja ei tunnu merkitykselliseltä, koska pelkkä nimi ei kerro tästä henkilöstä mitään. Valmistusaika ei kerro paljoakaan siitä, millaista elämä oli siihen aikaan. Materiaali herättää vain pienen aavistuksen siitä, miltä esine on saattanut tuntua kädessä. Entä jos valmistajasta itsestään löytyisi kuva ja tietoa, ja vuodesta pääsisi sivulle, jossa kerrotaan millaista elämä oli siihen

aikaan? Ja esineestä olisi vielä linkki jonkun lapsuustarinaa, jossa tämäkin esine näyttelee osaansa. Tällaisilla lisäyksillä historiasta voisi saada paljon konkreettisemmän ja rikkaamman kuvan. Tällä hetkellä museoesineet ovat tosin linkitettyjä toisiinsa eri tavoin, mutta tiedot ovat kuitenkin sen verran yksipuolisia, että on vaikeaa muodostaa kokonaiskuvaa siitä ajasta ja maailmasta, josta esineet ovat.

Kulttuurisen kontekstin avaaminen kertomuksilla ja prosessikuvauksilla ihmisten elämästä olisi siis erittäin oleellinen lisä tällaiseen järjestelmään. Myös audiovisuaalisessa muodossa tarjottu sisältö helpottaisi menneen ajan elämän hahmottamista ja avaisi kulttuurista kontekstia.

## 4.2 KulttuuriSampo — uusi prototyyppi

Koska MuseoSuomi on osoittautunut mielenkiintoiseksi ja lupaavaksi järjestelmäksi, sitä on syytä jatkokehittää. Uuden prototyypin tarkoitus on toimia MuseoSuomen seuraajan — KulttuuriSammon — ensimmäisenä demoversiona, jolla voidaan kokeilla MuseoSuomen laajentamisen mahdollisuuksia. KulttuuriSammon visio on saada laajamittainen läpileikkaus suomalaisesta kulttuurista semanttiseen webiin. Prototyypissä keskitytään ensisijaisesti erilaisten tietosisältöjen sisällönkuvailuun niin, että erilaiset mediat toimisivat yhteen ja monipuolisten ja tarkoituksenmukaisten semanttisten linkkien generointi onnistuisi. Tulevaisuudessa edessä on laajemman aineistomassan lisääminen.

### 4.2.1 Käyttötapauksia

Käyttötapaukset ovat hyödyllisiä toisaalta kuvattaessa järjestelmän mahdollisuuksia, toisaalta pohdittaessa järjestelmän toimivuutta. Mikäli järjestelmä vastaa johonkin käyttäjän tarpeeseen, siitä on hyötyä. Mikäli taas suunnitellut käyttötapaukset eivät onnistu järjestelmällä, on syytä kysyä miksi.

#### Ensimmäinen käyttötapaus — Villen koulutyö

Ville on kuudesluokkalainen koululainen Espoosta. Opettaja on antanut Villelle ja hänen kaverilleen Sampsalle tehtäväksi parityön, jossa heidän pitäisi kertoa maalaiselämästä 1800-luvun Suomessa. Työ on poikien viimeinen iso koulutyö historiasta ala-asteella, ja sen pitäisi olla suhteellisen laaja-alainen ja kertoa joistain ilmiöistä tarkemminkin. Myös kuvat ja muu materiaali olisivat eduksi.

Villen tavoite on löytää monipuolista tietoa historiallisesta maanviljelyksestä Suo-

nessa.

### *Skenaario*

Sampsu haluaa kirjoittaa maatalojen elämistä. Vilelle jää sitten maanviljelystä ja ihmisten elämästä kirjoittaminen. Opettaja antaa Vilelle KulttuuriSammon osoitteen vinkiksi tiedonhakua varten. Vile menee sivulle ja katselee hetken erilaisia vaihtoehtoja. Aluksi hän valitsee kategorian Maatalous ja karjanhoito ja sen jälkeen vielä vuosisadoista 1800-luvun. Hänelle aukenee kuva sirpistä, jota on käytetty aikanaan Espoossa. Vile ottaa kuvan talteen. Hän huomaa myös linkin kaskiviljelyprosessiin, koska sirppiä on käytetty kaskiviljelyssä. Linkkiä seuraamalla Vile pääsee tutustumaan kaskiviljelyn eri vaiheisiin. Vile toteaa kaskiviljelyn olevan hyvää materiaalia parityöhön ja kirjoittaa siitä. Hän päätyy myös lukemaan Kalevalaa, jossa Kullervo kaataa kaskea sen jälkeen, kun hänet on yritetty tappaa. Viljelyprosessista löytyy myös linkki Kalevalan runoon, jossa Sampsu Pellervoinen kylvää metsää. Vile pistää heti linkin Sampsalle sähköpostissa. KulttuuriSammon tietojen pohjalta pojat saavat tehtyä monipuolisen esitelmän, johon he saavat myös kuvamateriaalia netistä.

### **Toinen käyttötapaus — Veeran perinneleipäprojekti**

Veera on 47-vuotias kirjastonhoitaja, joka on juuri muuttanut vanhemmiltaan perimäänsä maalaistaloon. Hän muistaa, kuinka isoäiti valmisti talon isossa uunissa erityistä leipää. Veera haluaisi nyt oppia tekemään samanlaista leipää saatuaan naapurilta tuoreita ohrajauhoja, joita isoäitikin aina käytti.

Veeran tavoite on oppia leipomaan perinteistä suomalaista ohraleipää.

### *Skenaario*

Veera on kuullut ystävältään Marjatalta, että KulttuuriSampo-sivuilta löytyy kuvauksia vanhan ajan elämästä. Hän avaa sivut ja laittaa sanahakuun sanan ohraleipä. Ohraleivän valmistusprosessi löytyy kuin löytyykin, ja Veera katselee erityövaiheita KulttuuriSammosta verraten niitä siihen, mitä muistaa isoäitinsä tehneen. Veera huomaa myös prosessiin linkitetyn asioita ja esineitä, joista tulee lapsuus mieleen, ja hän katselee leivänteko-ohjeen lisäksi myös vanhojen esineiden tietoja. Laittamalla paikaksi kotipaikkakuntansa hän löytää serkkunsa museolle lahjoittamaa materiaalia. Tarpeekseen katseltuaan hän printtaa vielä muutaman kohdan leivänteko-ohjeista ja rupeaa suunnittelemaan leipomista samaksi illaksi.

### **Kolmas käyttötapaus — Samin Kalevala-näyttelysuunnitelma**

Sami on töissä Espoon kaupunginmuseossa. Museossa on suunnitteilla pienimuotoinen Kalevala-näyttely, jonka kohderyhmänä olisivat lähinnä koululuokat. Samin teh-

täväksi on annettu ottaa selvää, millaisia aiheeseen liittyviä esineitä heidän museostaan löytyy ja millaisiin Kalevalan tarinoihin ne voisi kytkeä.

Samin tavoite on tutkia Kalevalan tarinoita ja siihen jotenkin liittyvää oman museon materiaalia.

### *Skenaario*

Sami on kuullut kollegaltaan KulttuuriSampo-portaalista, jonne heidän tärkeimpiä esineitään on laitettu. Webin kautta hakeminen sopii Samille varsin mainiosti, koska sieltä näkee myös esineiden kuvat, ja näin on helpompi hahmottaa, millaista materiaalia on saatavilla. Sami valitsee Espoon kaupunginmuseon historiallisen museokoelman kategoriaksi ja rajoittaa ajaksi 1800-luvun, koska sitä vanhempia esineitä ei näytä paljoa olevan ja 1900-luku taas tuntuu liian modernilta.

Sami selailee erilaisia esineitä ja tekee muistiinpanoja. Valitessaan esineitä hän huomaa, että KulttuuriSammosta löytyy myös Kalevalan runoja. Hän alkaa selaillla runoja ja katselee samalla erilaisia Kalevalaan linkitettyjä esineitä. Kullervon tarinan varrelta löytyy monia yhtymäkohtia Espoon kaupunginmuseon aineistoihin, ja tarina on myös jännittävä. Ajatus näyttelyn kokonaisuudesta alkaa hiljalleen aueta Samin mielessä.

## **4.2.2 Tietosisältöjen valinta**

KulttuuriSammon prototyyppiin on pyritty ottamaan mukaan materiaalia, joka on oleellista suomalaisen kulttuurin viitekehäksessä ja joka myös liittyy mahdollisimman monipuolisesti muuhun aineistoon. Myös erilaisten medioiden saaminen mukaan on etu, koska niiden yhdistäminen merkityksellisesti on huomattavasti mielenkiintoisempaa kuin ainoastaan hyvin samantyyppisten aineistojen yhdistäminen. Käytännössä myös tietosisältöjä käyttöön luovuttaneet tahot ovat vaikuttaneet valittuihin materiaaleihin, koska kaikkea mahdollista mieleen tulevaa materiaalia ei ole vapaasti saatavilla. Tutkimusprojektin yhteistyökumppanit ovat kuitenkin antaneet hyvän läpileikkauksen digitaalisesta kulttuurisisällöstä.

*Prosessit* ovat oleellinen tietosisältö tutkimuksessa, kun puhutaan toimintakeskeisestä linkittämisestä. Prototyyppiin valittiin esimerkkiprosesseiksi kaskiviljely ja nuotakalastus, joiden kautta ihmiset voivat sekä tutustua prosessiin että saada mielenkiintoisia linkkejä asioiden välille. Tärkeää valittaessa oli, että prosessi on ollut yleinen Suomen historiassa mutta ei enää nykyihmisille itsestäänselvä. Tärkeää prosesseissa oli myös, että niissä käytettäisiin työkaluja, joita löytyisi MuseoSuomesta.

*Tarinat* ovat yksi osa-alue, joka valittiin mukaan, koska tarinat kiinnostavat ihmisiä, kuten jo MuseoSuomen palautteestakin on voitu päätellä. Kalevala valittiin mukaan prototyyppiin, koska se on suomalaisen kulttuurin ja historian kannalta olennainen teos. Koska Kalevalassa liikutaan vuosisatojen päässä nykyajasta, linkittyvät Kalevalan prosessit ja tapahtumat hyvin vanhoihin suomalaisiin esineisiin, joita MuseoSuomessa on. Koska koko Kalevalan annotointi olisi ollut työlästä, valittiin prototyyppiin Kullervon tarina. Se on suhteellisen itsenäinen kokonaisuus Kalevalan sisällä, ja myös siinä mielessä mielenkiintoinen, että se sisältää paljon mainintoja erilaisista prosesseista ja toimista, kuten sotimisesta, kalastamisesta, paimentamisesta ja kaskenkaadosta. Jotta saataisiin mukaan linkkejä eri runojen välille, otettiin Kalevalasta myös toinen runo, jossa Väinämöinen oppii viljelemään maata.

*Maalauksia* saatiin mukaan Ateneumista. Ne on valittu teeman mukaan, siten että erityisesti Kalevala- ja Kullervo-aiheisia maalauksia on muutama ja toisaalta maalauksia, jotka kuvaavat valittuja prosesseja, kuten kaskenkaatoa ja nuottakalastusta. Maalauksissa on siis toivottu olevan aina joku elementti, joka sitoo ne muihin tietosisältöihin.

*Valokuvia* on myös mukana aineistossa. Käytännössä valokuvat on valittu pitkälti samoin perustein kuin maalauksetkin. Koska sopivaa materiaalia oli yllättävän vaikea saada haltuun, käytettiin webistä löytyviä laadukkaita ja teemaan sopivia I.K. Inhan valokuvia Karjalasta. Valokuvien teknisiä ominaisuuksia ei saatu tähän versioon mukaan, koska webin tiedot eivät näistä kertoneet, mutta koska prototyypin ensimmäinen versio keskittyy joka tapauksessa sisällönkuvailuun, tämä ei haitannut kovinkaan paljon.

*Videot* pääsivät ensimmäiseen versioon mukaan hyvin yksinkertaisesti annotoituina, ja taas teemoina olivat kaskenkaato ja kalastus.

*Ihmiset* olivat myös yksi tietosisältö, jota haluttiin kokeilla osana prototyyppiä. Koska toimintaan liittyy useimmiten toimija ja kulttuurin muodostavat viime kädessä ihmiset, on myös ihmisistä kertominen luonnollinen lisä tietosisältöihin. Mukana on lähinnä muiden tietosisältöjen luoja, kuten taiteilijoita. Tämän lisäksi mukana on myös henkilöitä Kalevalasta.

*Esineet* otettiin luonnollisesti mukaan myös prototyyppiin. Kaikkia MuseoSuomen esineitä ei kuitenkaan saatu vielä mukaan, joten tässä vaiheessa valittiin jonkin verran vanhoja maatalon esineitä, jotka sopivat hyvin tematiikkaan ja pystyvät demonstroimaan, miten esineet linkittyvät prosessien ja tarinoiden kautta muuhun tietosisältöön.

### 4.2.3 Ontologiat käytännössä

KulttuuriSampo-prototyypissä tarvittiin kaikkia ontologioita, joita teoreettisessa mallissa esitettiin, joten kaikkia päästiin testaamaan käytännössä. Seuraavaksi lyhyt selostus niiden käytöstä esimerkkikoodin kanssa ja selostus siitä, miten YSO ja Dublin Core otettiin käyttöön prototyypissä.

#### Yhden tilanteen kuvailu - esimerkki maalauksesta

Tilanneontologia ja prosessionontologia toteutettiin prototyypissä hyvin pitkälti luvussa 3 esitellyn teoreettisen mallin mukaan. Teoreettista mallia rakennettiin paljolti tämän tapauksen pohtimisen ohessa, ja näin se sopi käytännössä projektin tarpeisiin ja oli luonnollista, ettei sitä tarvinnut paljoakaan muuttaa. Mielenkiintoista olisi ollut myös kokeilla teoreettisen mallin käyttöönottoa jossain muussa yhteydessä, mutta projektin puitteissa siihen ei ollut resursseja.

Seuraavaksi esimerkki instanssidatasta, jossa yhden maalauksen tilanne on kuvattu kahden tapahtuman kautta, joissa toiset naiset kantavat saavia ja toiset seisoskelevat.

```
<ns_04_tilanne:Tilanne rdf:about="&ns_04_tilanne;tilanne_Instance_100005"
  ns_04_tilanne:vapaaKuvailu="naiset rappusilla"
  rdfs:label="naiset rappusilla">
  <ns_04_tilanne:tapahtuma rdf:resource="&ns_04_tilanne;tilanne_Instance_100006"/>
  <ns_04_tilanne:tapahtuma rdf:resource="&ns_04_tilanne;tilanne_Instance_100009"/>
</ns_04_tilanne:Tilanne>
<ns_04_tilanne:Tapahtuma rdf:about="&ns_04_tilanne;tilanne_Instance_100006"
  ns_04_tilanne:vapaaKuvailu="naiset kantavat saavia"
  rdfs:label="naiset kantavat saavia">
  <ns_04_tilanne:toiminta rdf:resource="&ns_04_tilanne;tilanne_Instance_100007"/>
  <ns_04_tilanne:kohde rdf:resource="&ns_04_tilanne;tilanne_Instance_100008"/>
  <ns_04_tilanne:toimija rdf:resource="&ns_04_tilanne;tilanne_Instance_66620008"/>
</ns_04_tilanne:Tapahtuma>
<ns_04_tilanne:KuvailuElementti rdf:about="&ns_04_tilanne;tilanne_Instance_100007"
  ns_04_tilanne:vapaaKuvailu="kantaminen"
  rdfs:label="kantaminen">
  <ns_04_tilanne:viitattavaElementti rdf:resource="&YSO;kantaminen"/>
</ns_04_tilanne:KuvailuElementti>
<ns_04_tilanne:KuvailuElementti rdf:about="&ns_04_tilanne;tilanne_Instance_100008"
  ns_04_tilanne:vapaaKuvailu="saavi"
  rdfs:label="saavi"/>
  <ns_04_tilanne:viitattavaElementti rdf:resource="&YSO;saavit"/>
<ns_04_tilanne:Tapahtuma rdf:about="&ns_04_tilanne;tilanne_Instance_100009"
  ns_04_tilanne:vapaaKuvailu="naiset seisovat"
  rdfs:label="naiset seisovat">
  <ns_04_tilanne:toimija rdf:resource="&ns_04_tilanne;tilanne_Instance_66620008"/>
  <ns_04_tilanne:toiminta rdf:resource="&ns_04_tilanne;tilanne_Instance_66622"/>
</ns_04_tilanne:Tapahtuma>
```

Koodin luettavuus ei ole paras mahdollinen, koska URI:t on generoitu satunnaisluvuilla. Jos eri elementtien viittauksia toisiin seuraa, muodostuu kuitenkin järkevä kokonaisuus. Tässä voidaan myös huomata, kuinka KuvailuElementti-instansseilla on ominaisuus viitattavaElementti, jossa on viittaus sisällön määrittelyontologiaan eli tässä tapauksessa YSO:aan.

Muutokset teoreettiseen malliin olivat pieniä tekniseen toteutukseen liittyviä säätöjä, jotka eivät ole kovin oleellisia. Varsinkin Protegé-2000-työkalun rajoitteet pakottivat tekemään joitain ratkaisuja epätarkoituksenmukaisesti, mutta näiden apuohjelman rajoittuneisuudesta johtuvien ratkaisujen listaaminen ei ole hyödyllistä. Semanttisen webin kehitys kuitenkin johtaa ajan myötä toivottavasti myös parempiin ohjelmiin, jotka toteuttavat spesifikaatiot juuri niin, kuin ne on määritelty, eivätkä laita mukaan omia ratkaisujaan, jotka eivät ole spesifikaation kanssa yhteensopivia.

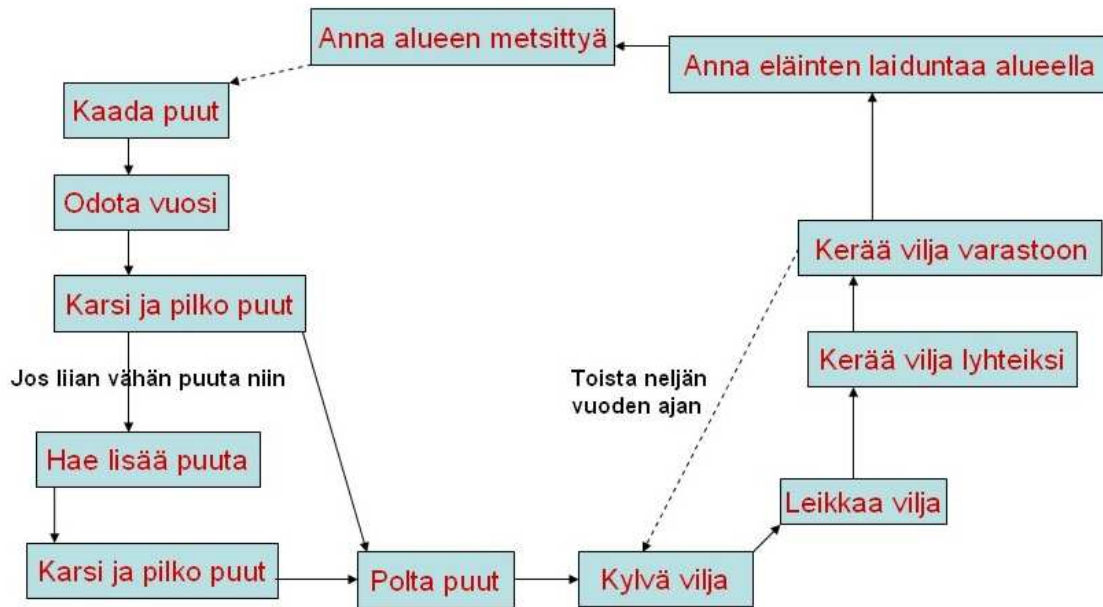
### **Esimerkki prosessi- ja tarinakuvauksesta — kaskiviljelyprosessi ja Kullervon tarina**

Tilanneontologian mukaisista tilanteista rakennettiin myös prosesseja ja tarinoita. Kaskiviljelyprosessi koottiin kahdestatoista osasta (kuva 6). Teoreettisen mallin mukaisesti jokainen vaihe jaoteltiin toiminnan muuttumisen mukaan. Käytännössä voidaan havaita, että kestollisesti pidempiä prosesseja kuvattaessa (kuvan esittämä prosessi kestää kymmeniä vuosia) ajan kulumisen on asia, johon pitäisi kiinnittää mallissa enemmän huomiota. Voidaan kysyä, onko ”odota vuosi” toimintaa, mutta peräkkäinkään puiden kaatamista ja pilkkomista ei voida laittaa, koska muuten ajan kulumisen jää huomiotta. Muuten malli tuki prosessin kuvaamista hyvin. Eri tilanteet pystyttiin kuvaamaan, ja niiden ominaisuuksien kautta myös monet muista tietosisällöistä linkittyivät prosessiin. Prosessiontologia taas salli tilanteiden laittamisen oikeaan järjestykseen suhteessa toisiinsa. Kaskiviljelyprosessin kuva käyttöliittymässä löytyy kuvasta 7.

Kullervon tarina toteutettiin myös mallin mukaan. Tilanteita tuli enemmän kuin kaskiviljelyprosessiin, mutta enimmäkseen malli toimi hyvin. Käytännössä tarinan kuluessa joutui välillä jättämään jotain epäolennaisia toimintoja pois, niin että esimerkiksi kirveen hiominen laskettiin osaksi kaskenkaatamistilannetta, kuten jo teoreettisen mallin esimerkkikuvassa 4 näytettiin. Kun näitä valintoja teki, sai tarinan jaettua pariinkymmeneen osaan, jotka kuvaavat tarinan olennaisimpia käänteitä. Kuvassa 8 esimerkki yhdestä tarinan osasta KulttuuriSammon käyttöliittymässä.

Yksi vaikea runonpätkä tarinassa oli. Tarinan kerronta lakkaa pitkäksi aikaa Untamon vaimon alkaessa loitsia ennen kuin tämä lähettää Kullervon paimeneen. Loit-





Kuva 6: Kaskiviljelyprosessi ja tilanteet joista se koostuu

su jatkuu pitkään, eikä varsinaisesti sisällä toimintaa, joten sen jakaminen osiin ei toimintapohjaisesti onnistu. Tämä on erikoistapaus ongelmasta joka mainittiin teoreettisen mallin yhteydessä luvussa Tarinat prosesseina: dialogin kuvaamiseen malli ei ole paras mahdollinen, eikä näin myöskään monologin.

### Yleinen suomalainen ontologia ja Dublin Core sisällön pohjana

Yleinen suomalainen ontologia eli YSO on Yleisen suomalaisen asiasanaston pohjalta rakennettu ontologia, jolla pyritään määrittämään yleisellä tasolla asiasanaston käsitteet suhteessa toisiinsa. KulttuuriSammon prototyypissä YSO oli luonnollinen valinta sisällön kuvailuun, koska se sisältää paljon käsitteitä ja se on tehty suomalaisesta näkökulmasta, joten myös kulttuuriimme erityisesti liittyvät käsitteet sopivat tähän ontologiaan. Koska YSO:n ontologisointi ei kuitenkaan prototyypin valmistumisen aikaan ollut vielä täydellinen, käytettiin siitä versiota, johon oltiin valittu ne käsitteet, joita tarvittiin esimerkkimateriaalin annotoimiseen. Tämä ontologia määritteli siis käytetyt käsitteet suhteessa toisiinsa ja muuhun maailmaan.

Koska YSO:n kehitys ei vielä ollut myöskään ehtinyt ominaisuuksienmäärittelyvaiheeseen, määriteltiin tarvittuihin YSO:n luokkiin perusominaisuuksia, joiden avulla pystyttiin sitten myös antamaan lisätietoa järjestelmän käyttäjälle. Dublin Core otettiin perusominaisuuksien pohjaksi, koska se sisältää oleelliset perustiedot, joita

The screenshot shows a web browser window displaying the KulttuuriSampo website. The page title is "Kaskiviljely". The main content area is divided into several sections:

- Kuvaus:** Kaskiviljely on viljelyteknikka jossa metsää polttamalla luodaan ravinteellisesti rikas maaperä viljelykasveille.
- Tarkoitus:** Kaskiviljelyn tarkoitus on saada hyvä sato.
- Kesto:** Yksi kierros metsästä viljelyn kautta metsäksi kestää noin 30-50 vuotta.
- Prosessi:**
  - immateriaalinen tuotos > menetelmät > viljelymenetelmät > kaskiviljely
  - immateriaalinen tuotos > tieto > viljelymenetelmät > kaskiviljely
- Prosessin vaiheet:**
  - Kaskipuiden kaato.
  - Odotetaan vuosi niin että puut ehtivät hieman lahota.
  - Karsitaan vuosi sitten kaadetut puut.
  - Levitetään kaadetut puut tasaisesti tulevan pellon alueelle.
  - Haetaan lisää puuta muualta.
  - Tulevalta pellolta poltetaan puuaines.
  - Kylvetään vilja tai muu viljeltävä.
  - Leikataan vilja sen kypsyttyä.
  - Kerätään leikattu vilja yhteyksi.
  - Aitetaan vilja varastoon.
  - Annetaan karjan laidunnaa kaskipelloilla.
  - Odotetaan metsän kasvavan takaisin.

Kuva 7: Kaskiviljelyprosessin sivu KulttuuriSampon käyttöliittymässä. Oikealla linkit kaikkiin prosessin vaiheisiin.

tutkimuksessa tarvittiin ja on lisäksi laajalti käytetty.

Dublin Core on metatietostandardi, joka on pyrkinyt määrittelemään sellaiset peruselementit, joita millä tahansa informaatioresurssilla voi olla, siten että informaatioresurssi on määritelty ”miksi tahansa, jolla on identiteetti”. Dublin Core sisältää elementit esimerkiksi otsikon, aiheen ja tekijän tiedoille. Yhteensä viisitoista peruselementtiä on määritelty, ja ideana on ollut pitää standardi mahdollisimman geneerisenä, yksinkertaisena, ja mahdollistaa myös sen käyttäminen osana monimutkaisempia ja laajempia metatietoskeemoja.

Dublin Coren lähtökohdat ovat hyvät siinä mielessä, että ominaisuudet kuten tekijä ja otsikko/nimi ovat asioita, jotka sopivat kuvaamaan melkein mitä tahansa, ja näin yhteinen standardi edistää eri kohteissa käytetyn metatiedon yhteensopivuutta. Dublin Core ei tosin pakota käyttämään minkäänlaisia URI-tunnisteita, joten tekstimuotoisen metatiedon moniselitteisyyden ongelma tulee helposti vastaan. URI:en

The screenshot shows a web browser window displaying the KulttuuriSampo website. The page title is "Runo 31 - osa 5". The main content is a poem by Elias Lönnrot, titled "Kullervo laitetaan kaskea kaatamaan". The poem text is displayed in a large font, with semantic annotations (links) overlaid on it. The annotations are organized into several sections: "Aineistotyyppi:", "Tapahtuma:", "Toimija:", "Toiminnan kohde:", and "Paikka:". Each section contains a list of related terms and their counts in parentheses. For example, under "Aineistotyyppi:", there are links for "kirjallisuus (32)", "kaunokirjallisuus (32)", "lyriikka (32)", "runot (32)", "runot (32)", "sanataide (32)", "lyriikka (32)", "runot (32)", "runot (32)", "immateriaalinen tuotos (46)", and "runot (32)".

On the right side of the page, there are several navigation and information panels: "Linkki prosessiin" (with a link to "Kaskivijely"), "Linkki tarinaan" (with a link to "Kullervon tarina"), "Seuraava tarinan osa" (with a link to "Runo 31 - osa 6"), "Edellinen tarinan osa" (with a link to "Runo 31 - osa 4"), "Linkki esineeseen" (with a link to "Kirves"), "Sama tilanne" (with links to "Runo 2 - osa 4" and "Kaskipuiden kaato"), "Työn tekijä" (with a link to "Elias Lönnrot"), and "Sama elementti" (with links to "Kirves" and "Karsitaan vuosi sitten kaadetut puut").

Kuva 8: Yhden tarinan osan sivu KulttuuriSammon käyttöliittymässä

avulla Dublin Core voi kuitenkin luoda hyvän pohjan monelle digitaaliselle kulttuurisovellukselle.

Prosessi- ja toimintakeskeisen kuvailun kannalta Dublin Core ei tarjoa paljoakaan, koska sitä ei ole suunniteltu kovinkaan yksityiskohtaiseen sisällönkuvailuun. Asia-sanoja voidaan Dublin Coren yhteydessä käyttää, mutta niiden semantiikan määrittelymättömyys ei mahdollista semanttisten linkkien luomista tai älykästä hakua. Erilaisten tietosisältöjen perustietojen kuvailuun Dublin Core on kuitenkin varten otettava vaihtoehto.

#### 4.2.4 Aineiston annotointi

Semanttista webiä kritisoidessa otetaan usein esille annotoinnin työläys. On totta, että annotointi ei ole kovinkaan nopeaa ja käytännöllistä, ja se on näin myös esteenä uuden teknologian leviämiseksi. Osittain ongelmana on kuitenkin vielä sen verran tuore teknologia, että työkalujen käytettävyyteen ei vielä ole ehditty panostaa.

Toinen näkökulma on se, että erilaista metatietoa on käytetty monissa yhteyksissä jo vuosisatoja (esimerkiksi kirjastoissa), eikä annotoinnin työläys ole estänyt työn tekemistä. Kysehän on siitä, säästääkö metatieto työtä muualla ja saadaanko siitä muuten hyötyä enemmän, kuin sen tekemisestä on vaivaa. Jos semanttisen webin teknologia nousee uudeksi metatietostandardiksi, saadaan hyvää metatietoa semanttisen webin sovellusten käyttöön ikään kuin ilmaiseksi, koska metatietoa tarvitaan monessa yhteydessä joka tapauksessa.

Alkuperäisessä MuseoSuomessa materiaalia oli sen verran paljon, että aineiston annotointi museoesine kerrallaan ei olisi ollut tarkoituksenmukaista. Siksi kehiteltiin aineistolle sopivat työkalut, joilla annotointi saatiin tehtyä puoliautomaattisesti [24]. Tämän prototyypin aineistotilanne oli kuitenkin erilainen. Oli hyvin monentyyppisiä aineistoja, joista tarvittiin kustakin lähinnä yhtä tai muutamaa esimerkkiä. Näillä voitiin testata, miten haku ja linkitys toimii yhteen. Tässä tapauksessa ei katsottu aiheelliseksi ryhtyä kehittämään massa-annotoinnin mahdollistavia työkaluja, vaan päädyttiin tekemään vaadittavat annotoinnit käsin. Käsin annotointi oli myös sikäli tärkeää, että sitä tehtäessä heräsi uusia kysymyksiä itse ontologioista, ja näin annotoitaessa tehtyjen havaintojen perusteella pystyttiin ontologioita muuttamaan toimivampaan suuntaan.

Annotointi tapahtui Protegé-2000-editorilla. Se tuki tyydyttävästi työskentelyä, mutta ajoittain tietyt käytettävyysongelmat vaikeuttivat työn etenemistä. Monien sisäkkäisten instanssien luonti oli vaivalloista, ja eri instanssien erottaminen toisistaan oli myös hankalaa, koska tietystä ohjeellisesta arvosta, jonka mukaan instanssi yritettiin identifoida, oli vaikea pitää mielessä mistä todella on kyse. Vakavampia puutteita Protegé-2000:ssa olivat RDF-listojen tai muiden kokoelmien tuen puute. Olisi suotavaa, että editori toteuttaisi käyttämänsä spesifikaation kokonaan.

Annotoinnin kuluessa huomattuja ongelmia on käsitelty tarkemmin edellisessä luvussa, joka koskee tilanneontologiaa ja sen suunnittelua, joten niihin ei mennä tässä yhteydessä syvemmälle. Yleisesti on sanottava, että käsin annotointi on työlästä, mutta käytetyssä sisältölähtöisessä kuvailussa automatisoinnin mahdollisuudet ovat rajalliset. Tietysti tietosisällöt ovat prototyypissä eriluonteisia, ja niiden automaattinen annotointi olisi näin myös keskenään erilaista. Tekstimuodossa olevasta tarinasta voitaisiin ehkä jo poimia automaattisesti tapahtumia, kun taas valokuvasta ilman metatietoja ei tietokoneella ole paljoakaan mahdollisuuksia ymmärtää mitä siinä tapahtuu.

#### 4.2.5 Prototyypin tekninen toteutus

Koska prototyypin ensisijainen tarkoitus on ollut kokeilla uudenlaisen ontologian ja sen kautta uuden sisällönkuvailulogiikan toimivuutta uudenlaisissa aineistoissa eikä niinkään uusia teknisiä ratkaisuja, oli luonnollista käyttää prototyypin toteuttamisessa vahvasti hyväksi olemassa olevaa MuseoSuomen tekniikkaa. En kuvaile tässä kovinkaan tarkkaan tekniikkaa, koska sitä on kuvailtu toisaalla kattavasti [32, 20]. Suppea kuvaus tekniikasta on kuitenkin paikallaan, jotta pääpiirteet selviävät.

MuseoSuomen tekninen arkkitehtuuri koostuu kolmesta osasta. OntoDellaksi nimetty Prolog-osa käsittelee RDF-dataa ja ontologioita, ja OntoDellaan voidaan määrittellä sääntöjä, joiden kautta hakukoneen toimintaa voidaan säätää sekä tehdä semanttisia suosituksia. Monifasettihaun mahdollistaa Javalla toteutettu Ontogator. Käyttöliittymän piirtämisestä ja interaktion toteuttamisesta vastaa Cocoonin pohjalle rakennettu OntoViews.

#### 4.2.6 Haun käyttöliittymä ja assosiaatiosäännöt

KulttuuriSammon prototyypin käyttöliittymään ei ehditty tehdä suuria muutoksia, vaikka suunnitelmia näistäkin oli. Näihin asioihin palataankin seuraavassa luvussa Jatkokehitysmahdollisuuksia. Tässä esitellään kuitenkin muutamalla kuvalla ja lauseella millainen käyttöliittymä prototyyppiin tuli ja millaisia assosiaatiosääntöjä käytettiin.

KulttuuriSammon perushaku toimii samalla tavalla kuin MuseoSuomessakin eli monifasettihaaku toimii erona vain, että tällä kertaa pohjalla olevana ontologiana on ensisijaisesti Tilanneontologia. Muutama fasetti on valittu myös muuten hyödyllisten tietojen lisäämiseksi, niin että esimerkiksi esinetyypin kautta pystyy myös katselemaan aineistoa. Koska YSO käytettynä sisällönkuvailuontologiana ei ollut vielä aivan valmis prototyypin kehittämisen aikaan, jäi hakukäyttöliittymään monia kummallisen oloisia luokkia, mutta koska tämä tutkimus keskittyy ensisijaisesti tietojen linkittämiseen eikä niiden hakuun, kannattaa prototyyppiä katsoa enemmän semanttisten linkkien näkökulmasta. Assosiaatiosäännöt olivat siis olennaisempia kuin YSO:n rakenne tässä vaiheessa KulttuuriSammon kehittämisen prosessia.

KulttuuriSammon assosiaatiosäännöt tehtiin pitkälti samalla logiikalla kuin MuseoSuomen, vaikka uusi ontologia ominaisuuksineen ja uudet tietosisällöt toivat muutoksiakin (esimerkit maalauksesta, kuva 9 ja valokuvasta, kuva 10). Rakennetta monimutkaistettiin hieman, niin että jos esimerkiksi MuseoSuomessa housuilla ja

The screenshot shows a web browser window with the URL `http://localhost:8081/Item?&l-fi&m=0&sb=http%3A%2F%2Fwww.cs.helsinki.fi%2Fgroup%2Fseco%2Fns%2F2004%2F11%2Fannottioid%23taulu_instance_21`. The page title is "KulttuuriSampo". The main content area is titled "Kullervon kirus" and features a painting of a man on a cross. To the right of the painting, there is a list of metadata: "Tekijä: Axel Waldemar Gallen-Kallela", "ID: A.II 765", "Valmistumisvuosi: 1899", "Mitat: leveys 102,5 cm", "Mitat: korkeus 184 cm", and "Tekniikka: kiinnitetty tukikankaalle vahahartsilla". Below this, there are several sections of semantic links:

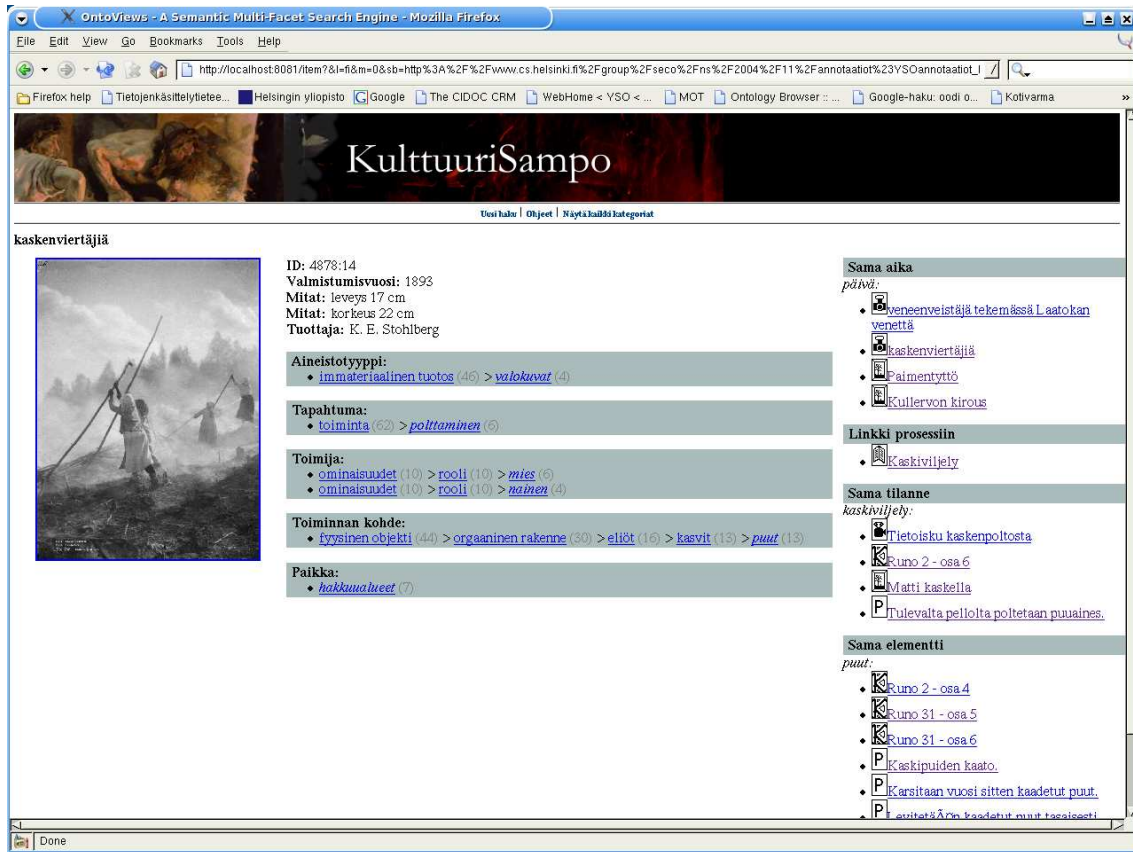
- Aineistotyyppi:**
  - maalaukset (3) > [maalaukset](#) (3)
  - immateriaalinen tuotos (46) > [maalaukset](#) (3)
- Tapahtuma:**
  - toiminta (62) > [kiroaminen](#) (3)
  - toiminta (62) > [nuhkiminen](#) (1)
  - toiminta (62) > [seisominen](#) (3)
- Toimija:**
  - fyysinen objekti (54) > fyysinen kokonaisuus (54) > eliökunta (54) > eläinkunta (54) > selkälänteiset (54) > selkärangaiset (54) > nisäkkäät (53) > kädelliset (51) > ihminen (51) > [Kullervo](#) (19)
  - fyysinen objekti (54) > fyysinen kokonaisuus (54) > eliökunta (54) > eläinkunta (54) > selkälänteiset (54) > selkärangaiset (54) > nisäkkäät (53) > petoeläimet (3) > koiraeläimet (3) > [koira](#) (2)
  - fyysinen objekti (54) > orgaaninen rakenne (54) > eliökunta (54) > eläinkunta (54) > selkälänteiset (54) > selkärangaiset (54) > nisäkkäät (53) > kädelliset (51) > ihminen (51) > [Kullervo](#) (19)
  - fyysinen objekti (54) > orgaaninen rakenne (54) > eliökunta (54) > eläinkunta (54) > selkälänteiset (54) > selkärangaiset (54) > nisäkkäät (53) > petoeläimet (3) > koiraeläimet (3) > [koira](#) (2)
  - fyysinen objekti (54) > orgaaninen rakenne (54) > eliöt (3) > eläimet (3) > eläimet käytön mukaan (3) > lemmikkieläimet (2) > [koira](#) (2)
- Toiminnan kohde:**
  - fyysinen objekti (44) > fyysinen kokonaisuus (13) > eliökunta (13) > eläinkunta (13) > selkälänteiset (13) > selkärangaiset (13) > nisäkkäät (13) > kädelliset (12) > ihminen

Other sections include "Sama aika" (kesä, Paimenttyttö), "Linkki esineeseen" (reput: kontti, tuohikontti), "Työn tekijä" (Axel Waldemar Gallen-Kallela), "Sama toiminta" (kiroaminen: Runo 31 - osa 5, Runo 33 - osa 2; seisominen: Vormasten naisia, Matti Kaskella), and "Sama elementti" (Pohjan tytär: Runo 33 - osa 3, Runo 22 - osa 4).

Kuva 9: Esimerkki maalauksesta ja sen semanttisista linkeistä

paidalla oli sama käyttäjä, ne linkitettiin toisiinsa, mutta koska KulttuuriSammossa on voitu kuvailla myös henkilöä, tuleekin esineistä linkki henkilöön ja henkilöstä sekä paitoihin että housuihin. Eli MuseoSuomessa ainoastaan linkittävänä tekstirivinä (jonka takana tietysti URI) toimineet asiat saavat nyt lisää tietoa itsestään järjestelmään, ja niitäkin voidaan katsoa. Myös samuuden kanssa on tehty joitain linkkejä. Jos esimerkiksi jokin tilanne on kahdessa tapauksessa sama, vaikkapa kaskenkaato esiintyy runossa ja maalauksessa, niistä tulee linkki kaskenkaatoprosessiin, mutta myös suoraan toisiinsa saman tilanteen kautta. Myös MuseoSuomen kaltaisia ainoastaan teksti- ja URI-muodossa olevia linkkejä löytyy, esimerkiksi sääntö, joka yhdistää samaan vuodenaikaan sijoittuvia tapahtumia sisältävät kaksi tietosisältöä suoraan eikä esimerkiksi kesällä ole omaa sivua, koska tämä tuskin olisi tarkoituksenmukaista.

Myös prosessien ja tarinoiden kausaalisuus on otettu assosiaatiosäännöissä huomioon, niin että prosessin osasta tai tarinan osasta pääsee aina myös suoraan edelliseen ja seuraavaan osaan. Näin esimerkiksi Kullervon tarinan voi lukea seuraamalla



Kuva 10: Esimerkki valokuvasta ja sen semanttisista linkeistä

linkejä alusta loppuun. Tämä on oleellista, jotta prosessit ja tarinat voidaan pitää kokonaisuuksina ristiinlinkityksistä huolimatta. Linkeistä voi lähteä ja saakin lähteä kesken tarinan sivupoluille, mutta on tärkeää, että käyttäjä pääsee taas takaisin jatkamaan tarinaa linkejä pitkin palattuaan takaisin kohtaan, jossa poikkesi polulta.

### 4.3 KulttuuriSampo — jatkokehitysmahdollisuuksia

MuseoSuomen kaltaista järjestelmää kehitettäessä tulee mieleen monenlaisia suuntia, joihin järjestelmää voisi viedä. Koska kaikkea ei pystytä toteuttamaan resurssien ja ajan puutteessa, ne on hyvä kuitenkin kirjata muistiin, jos tulevaisuudessa huomataan resursseja jatkokehitykseen taas olevan. Jo pelkästään tähän tutkimukseen liittyviä аспектеja, joita ei ehditty tutkia niin perusteellisesti kuin olisi voinut toivoa, on monia, mutta ideoita on syntynyt myös tämän työn aihepiirin ulkopuolelta.

### 4.3.1 Käyttöliittymä ja prosessien visualisointi

Tilanneontologian suunnittelun ja MuseoSuomen teknisen puolen muokkaaminen uusille aineistoille sopivaksi priorisoitiin tutkimuksessa, ja käyttöliittymään ei saatu tehtyä kovinkaan suuria parannuksia, vaikka uudenlaiset tietosisällöt asettavatkin uusia haasteita myös käytettävyydelle. Hakunäkymässä ei enää voidakaan näyttää kaikkia ominaisuuksia, koska esimerkiksi valokuvalla ja runolla on erilaisia media-kohtaisia ominaispiirteitä. Tulevaisuudessa olisikin hyvä mahdollistaa haussa joko tietyn mediatyyppin valitseminen, jolloin saisi tämän median ominaispiirteet näkyviin hakuehdoiksi, tai sitten pelkän sisällön perusteella hakemisen, jossa ei otettaisi kantaa mediaan vaan ainoastaan sisältöön eli tilanteeseen. Jälkimmäinen näistä ehdittiin toteuttaa jo ensimmäisessä versiossa.

Haussa on myös muita ongelmia. Koska tilanneontologiassa erilaiset asiat liittyvät nyt oleellisella tavalla toisiinsa, kuten esimerkiksi toimija tai toiminta, näiden koostaminen yhteen metatietokuvailulaatikoissa olisi etu. Tulkintojen erottaminen niin sanotusta varmasta tiedosta graafisesti puuttuu myös tämänhetkisestä versiosta. Ontologian hyponymiapolkujen pituudet ja moniperintä ovat myös lisänneet käyttöliittymään epäoleellista ja osittain toisteista tietoa (esimerkiksi sana ”koira” kuvassa 9).

Prosessien visualisointi metatiedon pohjalta olisi oleellinen ominaisuus, jota tulevisissa versioissa olisi hyvä ruveta kehittämään. Prosessien visualisoinnissa voisi käyttää hyväkseen GraphViz-nimistä ohjelmaa, joka piirtää erilaisia kaavioita saatuaan tietyllä kielellä tehdyn kuvauksen. Tällä hetkellä prototyypissä on demonstrointitaroitukseen tehty vastaavat kaaviot käsin, mutta tällaista ratkaisutapaa ei voi käyttää tulevaisuudessa aineiston ollessa laajempi.

### 4.3.2 Aineiston laajentaminen

Tällä hetkellä prototyypin aineisto on käsin valittua ja käsin annotoitua ja aineiston määrä on pieni. Laajempaan versioon olisi saatava lisää materiaalia, jotta nähtäisiin, miten toimintakeskeinen linkitys toimii isommilla aineistoilla. Annotaation automatisointi mahdollisimman pitkälle olisi erittäin toivottavaa mutta myös haastavaa. Käytännössä ongelmaksi saattavat muodostua myös tekijänoikeudet, sillä jos nyt oli ajoittain vaikeuksia saada pieniäkään aineistoja käyttöön, ja osa otettiin webistä ilman sen kummempia oikeuksia, koska ensimmäinen versio on epävirallinen ja suljetulle yleisölle tarkoitettu. On sääli, mikäli hyvää materiaalia ei saada verkkoon



tekijänoikeuksien takia, koska käytännössä tämä johtaa useimmiten siihen, että materiaali pölyttyy käyttämättömänä jossain eikä siitä näin hyödy oikeuksien omistajakaan. Jonkinlaisen sopuratkaisun ja kultaisen keskitien löytyminen julkaisu-oikeuksien kanssa olisi siis tärkeää KulttuuriSammon tulevaisuuden kannalta. Virallisten tahojen materiaalin lisäksi on kuitenkin myös toinen vaihtoehto.

Digitaalisten medioiden olennaisimpia piirteitä verrattuna niin sanottuun vanhaan mediaan on interaktion mahdollisuus. Webille ovat viime aikoina osoittautuneet ominaisiksi myös erilaiset yhteisölliset palvelut, joissa ihmiset ympäri maailmaa pääsevät tekemään jotain yhdessä ja vaihtamaan ajatuksia. KulttuuriSammolla voisi olla mahdollisuus kehittyä portaaliksi, jossa ihmiset itse pystyvät lisäämään omia tarinoita tai esineitään järjestelmään ja saisivat näin itsensä osaksi suurta semanttista webiä. Jos olisi olemassa työkalu, jolla esimerkiksi omaelämäkerrallisen tarinan tai oman ”museoesineen” voisi tarpeeksi helposti lisätä järjestelmään metatietoineen, voitaisiin parhaimmillaan saada hyvää materiaalia, joka tukisi kokonaisuutta ja sitoisi kulttuuriset artefaktit ja suuret suomalaiset taideteokset suomalaisten ihmisten elämään. Käytännössähän museoesineet ja taideteokset ovat usein juuri sitä: asioita joiden kautta voimme oppia kulttuuristamme ja nähdä arkemme toisenlaisessa valossa. Webin kautta kuka tahansa voisi lisätä tietoa taidoista ja prosesseista, joita asiantuntijat eivät välttämättä osaisi löytää mistään.

Tällaisen interaktiivisuuden toteuttaminen toimivalla tavalla olisi varmasti suuri haaste, koska varsinkin ne vanhat ihmiset, jotka eniten tietäisivät menneen ajan elämästä, eivät todennäköisesti käytä tietokoneita eivätkä varsinkaan, jos se ei ole todella helppoa. Voi myös olla että ihmiset eivät lopulta kokisi mielekkääksi tehdä tällaista työtä, vaikka sillä saisikin omat tarinansa maailman tietoon. Todennäköisesti vaadittaisiin melkoista mainostusta ja ehkä jonkinlainen yhteisö, joka tietäisi ainakin joidenkin tuttujen seuraavan verkon kehittymistä, jolloin motivaatio olisi aivan toisenlainen. Täytyisi olla myös mahdollisuus tutkia virallista kautta järjestelmään tulleita sisältöjä erikseen, ja jonkun olisi valvottava, mitä sisältöä ihmiset järjestelmään laittavat.

Itse asiassa Google on toteuttanut tämän idean: juuri vähän aikaa sitten avattu Google Base -palvelu<sup>2</sup> antaa ihmisten lisäillä omia esineitään globaaliin tietokantaan ja lisätä myös avainsanoja, joiden avulla niitä voi etsiä.

---

<sup>2</sup><http://base.google.com>

## 5 Muuta tutkimusta alalta

Metatietojen käyttö on kulttuuripiireissä vanha keksintö, ja näin myös digitaalisen metatiedon tutkimusta on tehty kulttuuritietosisältöjen kontekstissa. Esittelen seuraavissa aliluvuissa muutamaa ontologiaa, joita on kehitelty kulttuuritietosisältöjen annotointiin. Nämä ontologiat pyrkivät kuitenkin kuvaamaan esineitä ja niiden elinkaarta eivätkä niinkään niihin liittyviä tarinoita tai kulttuurillisia prosesseja, joten näkökulma on hieman erilainen kuin tilanneontologiassa.

### 5.1 CIDOC Conceptual Reference Model

CIDOC Conceptual Reference Model (CRM) pyrkii määrittelemään ja antamaan formaalin muodon kulttuuriperinteeseen liittyvän materiaalin suhteiden ja käsitteiden kuvailulle. Se pyrkii siis standardiksi, jolla voitaisiin kuvata erilaisia kulttuuriin liittyviä asioita eri yhteyksissä ja järjestelmien yhteensopivuuden säilymiseen. Käytännössä pyrkimys on siis sama kuin MuseoSuomessa mutta vielä laajemmalla mittakaavalla.

CIDOC CRM:llä pystytään tosin kuvailemaan monenlaisia asioita ja niiden suhteita, mutta käytännössä se on vaikuttanut rakenteeltaan monimutkaiselta, ja sillä mallinnettua metatietoa on vaikea hahmottaa. Sen rinnalla Dublin Coren yksinkertainen perusmalli tuntuu toimivammalta lähestymistavalta. CIDOC CRM:n keskiöön on valittu esineiden lisäksi myös toimijat ja tapahtumat, joten monimutkaisempaa semantiikkaa voidaan odottaa. CRM:n käyttöä vaikeuttaa käytännössä monimutkaisuuden lisäksi se, ettei koko CRM:ää löydy RDFS-muodossa, vaan ainoastaan osa määrittelystä.

CIDOC CRM:n määrittelyissä on nimenomaan kerrottu, että sillä ei kuvata tarinoita tai erilaisia psykologisia tulkintoja vaan ainoastaan (mahdollisesti keskenään ristiriidassa olevia) faktoja [36]. Tämä valinta voi olla CIDOC:n kontekstissa perusteltu, sillä ammattilaisten tieteellisen tiedon hakuun tähtäävässä järjestelmässä tulkinnat ja tarinat saattavat tehdä kokonaisuudesta epäluotettavan oloisen ja vaikeammin käytettävän. Tästä valinnasta huolimatta ei objektiivisuuden ongelmasta päästä, sillä tieteellisillä teorioilla ja tarinoilla on usein paljon yhteistä — niillä pyritään hahmottamaan ja selittämään maailmaa. Rationaalisten perusteluiden olemassaolo ei estä hyvinkin vastakkaisia tieteellisiä teorioita ja tulkintoja syntymästä. Siksi CRM:n valinta sallia ristiriitaista tietoa on järkevä ja käytännöllinen, vaikka päättelykoneille saattaakin tulla vaikeuksia.

## 5.2 ABC Harmony

ABC Harmony [27] on yksinkertainen ontologia, jonka tarkoituksena on toimia erilaisten aihealueiden ontologioiden yhteisenä pohjana, joka sekä edesauttaa ontologioiden yhteensopivuutta että säästää työtä muilta ontologioiden suunnittelussa. Erityispainotuksena ABC Harmonyssa on esineiden luomisen, kehittymisen ja siirtymisen mallintaminen ajan kuluessa. Kun ontologioita tehdään tiettyyn aihealueeseen liittyen ja sovelluksissa kaivataan usein nimenomaan tiettyä sanastoa, tulee esiin se ongelma, että lähes missä tahansa kontekstissa esiintyy esimerkiksi toimijoita ja tapahtumia, ja näin tällaisia perusasioita mallinnetaan aina uudestaan. ABC Harmonyn pohjalle voidaan siis rakentaa millaisia ontologioita tahansa ilman, että joudutaan tekemään kaikkea alusta.

Tämä lähestymistapa on sinänsä hyvä ja oikeastaan filosofialtaan hyvin lähellä Dublin Corea. Perusasioiden mallinnus kun on usein loppujen lopuksi kaikkein vaikeinta. Siksi tarkasti mietittyä pohjaa käyttäisi mielellään. Ongelmaksi nousee kuitenkin se, että tietyn aihealueen näkökulma vaikuttaa myös näkökulmaan, joka pitäisi ottaa ontologian ylimpiinkin luokkiin. Ja ABC Harmonyn näkökulma on — kuten muissakin kuvailluissa metatietomalleissa — esine-/asiakeskeinen, ei prosessikeskeinen. ABC Harmonyllä pystyy kuvailemaan esineen elinkaareen liittyviä tapahtumia ja tiloja, joissa se on tapahtumien välillä. Mutta kun tapahtuma ja toiminta otetaan keskipisteeksi, käy vaikeaksi kuvailla erilaisten toimintaan liittyvien asioiden muutoksia ja niiden katoamista prosessista taas seuraavassa vaiheessa.

Käsitellyistä metatietomalleista ABC Harmony tarjosi kuitenkin selkeiden esimerkkien kautta eniten ajatuksia, ja ainakin esineiden ja niihin liittyvien tapahtumien kuvaamiseen malli tuntui olevan toimiva ja silti tarpeeksi yksinkertainen.

## 6 Arviointia

KulttuuriSammon prototyypin ensimmäinen versio valmistui tutkimusprojektin johtoryhmän kokoukseen 21.12.2004. Siinä on mukana 2 prosessia, jotka koostuvat 21 atomisesta prosessista, 2 tarinaa 32:ssa runon osassa Kalevalan runoista 2 ja 31–36, 8 maalausta, 4 valokuvaa, 2 videota ja 14 museoesinettä.

Sovellus on aina hyvä arvioida suhteessa ennen ohjelmaa suunniteltuihin käyttötapauksiin. Käytännössä prototyypin tämänhetkistä versiota arvioitaessa näin voidaan sanoa, että hakuun liittyvät käyttäjätavoitteet eivät vielä täysin toteudu, koska YSO

ei ole vielä täysin siinä kunnossa, että sen luokkahierarkia näyttäisi selkeältä ja hyvältä. Toisaalta tässä tutkimuksessa ei ensisijaisesti olekaan keskitytty hakuun vaan tietojen linkittämiseen.

Tietojen linkittämisen suhteen käyttötapausten kuvaamat tapaukset on tehty mahdolliseksi. Varsinaista läpivientiä ei vielä voi tehdä siksi, ettei materiaalia ole esimerkiksi perinneleivän valmistuksesta, mutta koululaisten esitelmä kaskiviljelystä ja siihen liittyvä tiedonhaku ja semanttinen samoilu aiheeseen liittyvien tietosisältöjen välillä voidaan todeta toimivaksi jo tässä versiossa.

Erilaiset aineistot yhdistyivät sääntöjen avulla hyvin, mikä johtui toki paljolti siitä, että tietosisältöjen esimerkit oli valittu samantyyppisestä aihepiiristä. Todella suuressa mittakaavassa alkaisi varmasti tulla esiin enemmän ongelmia liittyen eri ihmisten erilaiseen tapaan annotoida tietosisältöjä, mutta toisaalta silloin saatettaisiin saada enemmän myös odottamattomia yhteyksiä, joita tässä vaiheessa ei päässyt syntymään, kun kokonaistietosisällön määrä oli suhteellisen pieni.

Sisältökuvailun keskiöön valittu tilanne osoittautui keskeiseksi asiaksi erilaisten tietosisältöjen kohdalla, ja sen avulla pystyttiin kuvailemaan erityyppisiä aineistoja. Luonnollisesti erilaiset mediat asettivat uusia kysymyksiä, ja jotkut ongelmat olivat toisten tietosisältöjen kohdalla vaikeampia kuin toisten. Lopulta tilanneontologia pystyi kuitenkin erottamaan sisällön mediasta, ja erilaisia versioita ontologiasta ei tarvittu erilaisten tietosisältöjen kuvaamiseen, vaan mediakohtaiset erityistarpeet voitiin pitää toimintakeskeisen sisällönkuvailun ulkopuolella.

Koska tilanneontologia toimi sekä tekstimuodossa olevien tarinoiden, kuvien ja videoiden kuvaamiseen, voidaan olettaa sen toimivan myös muunlaisten kerronnallisten tekstuaalisten tai audiovisuaalisten medioiden kuvaamiseen. Ainakin äänitettyä puhetta, runonlaulua tai kuunnelmia, animaatiota, interaktiivista multimediaa ja tietokonepelejä voisi kuvailla tilanneontologian avulla, mikäli niissä vain on edes jossain määrin konkreettista kerrontaa. Tosin interaktiivisten teosten kanssa tulisi ehkä vielä miettiä, miten interaktiivisten hetkien tilanteet kuvattaisiin — siinähan on kyse taas toiminnasta, mutta tällä kertaa teoksen käyttäjän toiminnasta. Abstraktien teosten kanssa tilanneontologiaa on lähes mahdoton käyttää. Myös hyvin monimerkityksinen kerronta, kuten syvälliset allegoriat, tajunnanvirta tai modernit runot voivat olla hyvin vaikeita annotoida ja linkittää toimintakeskeisesti. Tässä mielessä Kalevala-teema oli osuva valinta, koska Kalevalaan ja menneeseen aikaan liittyvät teokset ovat olleet kietoutuneita enimmäkseen konkreettisiin tapahtumiin ja esineisiin.

Syvämmän symboliikan kuvaaminen on prototyypin ensimmäisessä versiossa jäänyt oikeastaan kokonaan pois, vaikka attribuutit, tulkinnat ja symboliikan mallintaminen on ontologiassa mahdollistettu. Käytännössä tämä johtui asiantuntemuksen puutteesta. Olisi kuitenkin mielenkiintoista nähdä, millaisia linkkejä syntyisi, jos ammattimainen taiteentuntija voisi annotoida osan teoksista myös symboliikan ja allegoriat huomioon ottaen. Tällainen monikerroksisempi linkitys olisi tavoiteltavaa, ettei taiteen kerroksellisuus jäisi täysin huomioimatta metatiedossa. Kuitenkin korostan edelleen sitä, että metatiedon tehtävä ei ole korvata itse teosta. Vasta teoksen äärellä saa jokainen kokija tehdä omat tulkintansa ja johtopäätöksensä, ja semanttisten linkkien tarkoitus on tuoda aiheesta kiinnostunut ihminen kosketuksiin kulttuurin kanssa.

## 7 Yhteenveto

Tässä tutkimuksessa tutkittiin erilaisten tietosisältöjen linkittämistä toimintakeskeisesti. Erilaisten kulttuuristen tietosisältöjen keskeiseksi yhdistäväksi tekijäksi valittiin näin tilanne, joka sisältää toimintaa. Tämän tilanne-käsitteen pohjalta voitiin kuvailla sekä prosesseja että tarinoita, jotka koostuvat tilanteista suhteessa toisiinsa. Tilanteen kuvaamiseen kehitettiin tilanneontologia, joka määrittelee, mitkä tiedot ovat oleellisimpia erilaisten tietosisältöjen linkittämiseen sisältöpohjaisesti ja miten ne kuvataan konsistentisti. Tilanneontologian rinnalle otettiin myös prosessiontologia, jonka avulla voidaan määritellä tilanteiden suhteet aikajatkumossa.

Tutkimuksen aikana kävi myös selväksi, että toiminta- ja sisältökeskeistä kuvailua kulttuuritiedoille ei ole paljoakaan tutkittu, sillä olemassa olevat kulttuurisisältöjen ontologiat olivat lähes poikkeuksetta nimenomaan esineen eivätkä niiden prosessien ja tarinoiden näkökulmasta, joihin esineet tai teokset viittaavat. Tutkimus siis oli hyödyllinen siinä mielessä, että se kartoitti aihetta uudesta näkökulmasta, mutta toisaalta sitä on vaikea suoraan verrata muihin tutkimuksiin, koska sellaisia ei löydy.

Tilanneontologiaa kehiteltäessä tutkittiin myös ongelmia, joita liittyy tilanteiden kuvaamiseen metatiedon avulla. Kun taideteoksen fyysisten mittojen tai materiaalin sijasta tutkittiin sen sisältöä, oli jo alusta alkaen selvää, että tehtävä ei ole suoraviivainen. Siksi oli tärkeää sitoa tutkimus käytännön tilanteisiin, jotteivät vaikeat teoreettiset kysymykset jäisi leijumaan ilmaan vailla muotoa. Tilanneontologia otettiinkin käyttöön KulttuuriSampo-portaalin ensimmäisessä prototyypissä, joka toteutettiin tutkimuksen aikana. KulttuuriSammon idea on esitellä suomalaista kulttuuria mah-

dollisimman monipuolisesti, useasta näkökulmasta. Se on rakennettu MuseoSuomi-portaalin tekniikan varaan, mutta sisältää paljon monipuolisemman kirjon erilaisia tietosisältöjä.

Käytännön sovellus osoitti tilanneontologian ja toimintakeskeisen kuvailun pystyvän sitomaan yhteen tietosisältöjä maalauksista runoihin ja kulttuurisista prosesseista esineisiin. Myös ongelmiin, joita teoreettisen mallin suunnittelussa oli tullut vastaan, jouduttiin ottamaan kantaa, ja joihinkin saatiin myös ratkaisut. Ensimmäisen version suppeahkon aineiston takia ei täydellistä kuvaa suuren mittakaavan toiminnasta vielä voitu saada, mutta automaattisesti metatiedon avulla luodut linkit rohkaisivat tutkimaan aihetta lisää, sillä yhtäläisyyksiä todella löytyi. Museoesineistä löytyi juuri samanlainen tuohilaukku kuin Kullervolla on vieressään kuuluisassa maalauksessa, ja kaskenpolttoprosessi satoi yhteen materiaalia valokuvista Kalevalan runoihin. Toivottavasti tällainen tutkimus antaa motivaatiota museoille panostaa monipuolisen ja hyvin suunnitellun metatiedon tuottamiseen, sillä tämän tiedon perusteella luotavat merkitykselliset linkit voisivat auttaa tulevaisuudessa luomaan mielenkiintoisia kokonaisuuksia esimerkiksi web-näyttelyyn automaattisesti.

## 8 Kiitokset

Haluan kiittää Semantic Computing -tutkimusryhmää, jossa tutkimustani toteutin, ja erityisesti taistelupariiani Mirva Salmista. Hyvästä ilmapiiristä ja usein annettua teknisestä ja henkisestä tuesta kiitos koko ryhmälle. Olette hyvän T-paitatekstin arvoisia. Tutkimusta tukeneille tahoille kiitokset materiaalin käyttöoikeuksista. Kiitokset myös Tuuli Linterille ja vanhemmilleni suuresta avusta tutkimuksen oikolu-  
kemisessä, sekä työn ohjaajille palautteesta ja neuvoista.

## Lähteet

- 1 Aristoteles. *Aristoteles IX : Retoriikka, Runousoppi*. Gaudeamus, Helsinki, Finland, 1997.
- 2 D. Beckett. *RDF/XML Syntax Specification (Revised)*, February 2004. <http://www.w3.org/TR/rdf-syntax-grammar/>.
- 3 T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.

- 4 M. Bernstein. Hypertext and technical writing. In *Proceedings of DEXA '90, International Conference on Databases and Expert Systems Applications, Vienna, Austria*, August 1990.
- 5 D. Brickley and R. Guha. *RDF Vocabulary Description Language 1.0: RDF Schema*, February 2004. <http://www.w3.org/TR/rdf-schema/>.
- 6 L. Cabral, J. Domingue, E. Motta, T. Payne, and F. Hakimpour. Approaches to Semantic Web Services: An Overview and Comparisons. In *The First European Semantic Web Symposium. Heraklion, Greece*, May 2004.
- 7 The OWL Services Coalition. *OWL-S: Semantic Markup for Web Services*, November 2003. <http://www.daml.org/services/owl-s/1.0/owl-s.pdf>.
- 8 J. Conklin. Hypertext: An introduction and survey. *IEEE Computer*, 20(9):17–41, 1987.
- 9 M. Doerr. The CIDOC CRM — an Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, 2003.
- 10 C. Eastman. Representation of design processes, April 1999. Conference on Design Thinking, MIT.
- 11 M. Eysenc and M. Keane. *Cognitive Psychology A Students Handbook*. Psychology Press, Exeter, UK, 2002.
- 12 M. Genesereth and R. Fikes. Knowledge Interchange Format Version 3.0 — Reference Manual. Technical report, Stanford University, 1992.
- 13 M. Groden and M. Kreiswirth. *The Johns Hopkins Guide To Literary Theory and Criticism*. The Johns Hopkins University Press, 1997.
- 14 T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
- 15 S. Handschuh, S. Staab, and F. Ciravegna. S-CREAM — Semi-automatic CREAtion of Metadata. In *Proceedings of EKAW 2002, LNCS*, pages 358–372, 2002.

- 16 J. Hefflin, J. Hendler, and S. Luke. SHOE: A Knowledge Representation Language for Internet Applications. Technical report, Dept. of Computer Science, University of Maryland at College Park, 1999.
- 17 Herakleitos. *Peri Phuseos (On Nature)*. Artemis Temple Press, Athens, Greece, 535–475 BC.
- 18 B. Holmes, B. Tangney, A. FitzGibbon, T. Savage, and S. Meehan. Communal constructivism: Students constructing learning for as well as with others. In *Proceedings of SITE 2001, Florida*, 2001.
- 19 I. Horrocs, P. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, and M. Dean. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*, November 2003. <http://www.daml.org/2003/11/swrl/>.
- 20 E. Hyvönen, S. Saarela, and K. Viljanen. Ontogator: combining view- and ontology-based search with semantic browsing. In *Proceedings of the XML Finland 2003 conference. Kuopio, Finland*, 2003. <http://www.cs.helsinki.fi/u/eahyvone/publications/xml-finland2003/yomXMLFinland2003.pdf>.
- 21 E. Hyvönen, K. Viljanen, and A. Hättinen. Yellow pages on the semantic web. Number 2002-03 in HIIT Publications, pages 3–14. Helsinki Institute for Information Technology (HIIT), Helsinki, Finland, 2002. <http://www.hiit.fi/publications/>.
- 22 E. Hyvönen, E. Mäkelä, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula. MuseumFinland — Finnish Museums on the Semantic Web. *Journal of Web Semantics*, 3(2):224–241.
- 23 E. Hyvönen, S. Saarela, K. Viljanen, E. Mäkelä, A. Valo, M. Salminen, S. Kettula, and M. Junnila. A semantic portal for publishing museum collections on the web. In *Proceedings of ECAI/PAIS 2004, Valencia, Spain*, August 2004.
- 24 E. Hyvönen, M. Salminen, and M. Junnila. Annotation of heterogeneous database content for the semantic web. In *Proceedings of SemAnnot2004, Hiroshima, Japan*, November 2004.
- 25 L. Kelly. Developing access to collections through assessing user needs, May 1999. Museums Australia Conference, Albury.



- 26 A. Knutilla, C. Schlenoff, S. Ray, S. Polyak, A. Tate, S. C. Cheah, and R. Anderson. Process Specification Language: An analysis of Existing Representations, 1998. NISTIR 6160, National Institute of Standards and Technology, Gaithersburg, MD.
- 27 C. Lagoze and J. Hunter. The ABC Ontology and Model. *Journal of Digital Information*, 2(2), 2001.
- 28 M. Laukkanen, K. Viljanen, M. Apiola, P. Lindgren, and E. Hyvönen. Towards ontology-based yellow page services. In *Proceedings of WWW2004 Workshop, Application Design, Development, and Implementation Issues, New York*, 2004.
- 29 A. Leontev. *Activity, Consciousness and Personality*. Prentice-Hall: Englewood Cliffs, NJ, 1978.
- 30 D. Mandell and S. McIlraith. Adapting BPEL4WS for the Semantic Web: The Bottom-Up Approach to Web Service Interoperation. In *The Proceedings of the Second International Semantic Web Conference (ISWC2003), Sanibel Island, Florida*, 2003.
- 31 E. Miller, R. Swick, D. Brickley, B. McBride, J. Hendler, G. Schreiber, D. Wood, and D. Connolly. *Semantic Web*, 2005. <http://www.w3.org/2001/sw/>.
- 32 E. Mäkelä, E. Hyvönen, S. Saarela, and K. Viljanen. OntoViews — A Tool for Creating Semantic Web Portals. In *Proceedings of 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan*, November 2004.
- 33 T. Navarrette. Museums in the street: Cultural creation in the community, October 2002. INTERCOM Conference Leadership in Museums: Are our core values shifting? Dublin, Ireland.
- 34 N. Noy and D. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University, Medical Informatics, 2001.
- 35 N. F. Noy, M. Sintek, S. Decker, M. Crubezy, R. W. Fergerson, and M. A. Musen. Creating semantic web contents with Protege-2000. *IEEE Intelligent Systems*, 16(2):60–71, 2001.

- 36 Delos Working Group on Ontology Harmonization. *Building Core Ontologies*, October 2003.  
<http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Ontologies.pdf>.
- 37 C. Schlenoff, M. Gruninger, F. Tissot, J. Valois, J. Lubell, and J. Lee. The Process Specification Language(PSL): Overwiev and Version 1.0 Specification, 2000. NISTIR 6459, National Institute of Standards and Technology, Gaithersburg, MD.
- 38 J. Sowa. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole, 2000.
- 39 T. Andrews and F. Curbera and H. Dholakia and Y. Goland and J. Klein and F. Leymann and K. Liu and D. Roller and D. Smith and S. Thatte and I. Tricovic and S. Weerawarana. Business process execution language for web services — version 1.1, 2003.
- 40 L. Vygotsky. *Mind in Society, The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978.