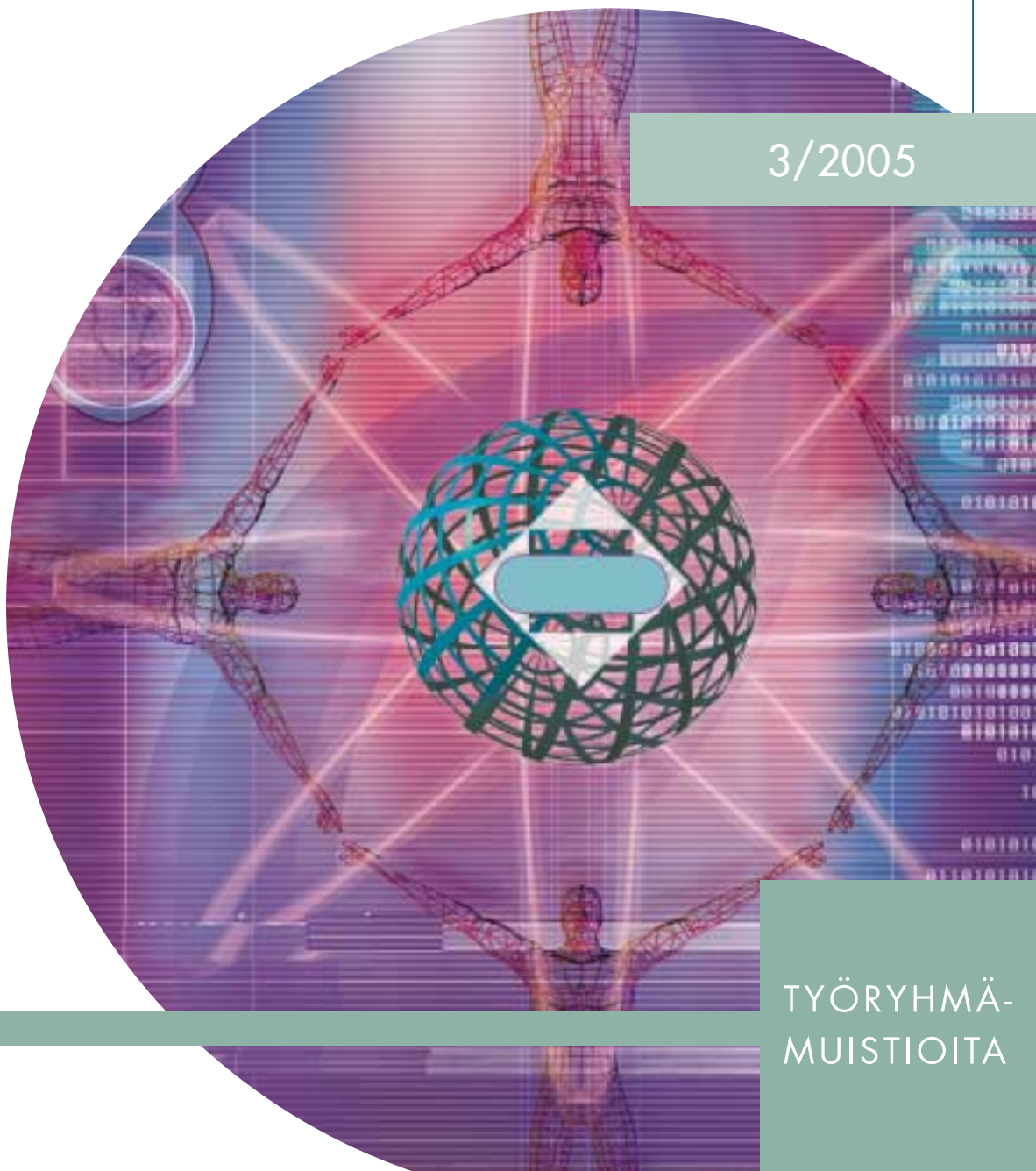




VALTIOVARAINMINISTERIÖ

SANASTOT SILTANA SAUMATTOMALLE TIEDONVAIHDOLLE

3/2005



TYÖRYHMÄ-
MUISTIOITA

SANASTOT SILTANA SAUMATTOMALLE TIEDONVAIHDOLLE

Tietojärjestelmien semanttisen
yhteentoimivuuden kehittämismalli

3/2005

VALTIOVARAINMINISTERIÖ
HALLINNON KEHITTÄMISOSASTO

TYÖRYHMÄ-
MUISTIOITA

VALTIOVARAINMINISTERIÖ

Snellmaninkatu 1 A
PL 28
00023 VALTIONEUVOSTO

Puhelin

(09) 160 01

Telefaksi

(09) 160 33123

Internet

www.vm.fi

Julkaisun tilaukset

Puh. (09) 160 33287

ISSN 0788-6322

ISBN 951-804-489-9 (nid)

ISBN 951-804-490-2 (pdf)

Edita Prima Oy

HELSINKI 2005

KUVAILULEHTI

Julkaisija ja julkaisu-aika	Valtiovarainministeriö, helmikuu 2005
Tekijät	Riku Moisio, Valtiovarainministeriö, pj. Juhani Koivunen, Valtiovarainministeriö, pj. (- 31.10.2004) Jukka Ahtikari, WM-data Oy Virpi Kantola, Tullihallitus Seija Karevaara, Verohallitus Väinö Malin, Ympäristöministeriö Timo Salovaara, Väestörekisterikeskus Henri Ståhl, Avain Technologies Oy Timo Tarhonen. Tietotarha Oy, HL7 -yhdistys Riia Töyrylä, Maa- ja metsätalousministeriö Lasse Akselin, TietoEnator Oyj Riitta Alkula, TietoEnator Oyj
Julkaisun nimi	Sanastot siltana saumattomalle tiedonvaihdon - Tietojärjestelmien semanttisen yhteentoimivuuden kehittämismalli
Julkaisun osat	Julkaisu on saatavissa Internetistä osoitteesta www.vm.fi/julkaisut . Em. osoitteesta voi ladata myös työryhmän laatiman taustamateriaalin ”Eri alojen sanastotyö”, jota ei ole painettu.
Tiivistelmä	Julkishallinnon tietojärjestelmien kehittämistyön päämääränä on julkisen hallinnon laadun ja palvelevuuden sekä julkishallinnon tarjoamien palvelujen tuottavuuden edistäminen. Parhaiten päämäärän saavuttamista tukee sellainen asiakaskeksinen sähköinen hallinto, jossa julkishallinnon ja sen keskeisten sidosryhmien tietojärjestelmät ovat yhteentoimivia ja mahdollistavat monilta osin automatisoidut asiankäsittely- ja palveluprosessit. Julkishallinnon XML-strategiatyössä on havaittu, että teknisistä ratkaisuista sopiminen ei yksinomaan ratkaise tietojärjestelmien yhteentoimivuutta. Järjestelmien on toimittava yhteen myös semanttisella tasolla, jotta tietojärjestelmät voivat muodostaa saumattoman kokonaisuuden. Semanttisen yhteentoimivuuden mahdollistamiseksi toiminnan ja tietojärjestelmien kehittäjien on yhdessä sovittava käytettävistä käsitteistä ja sanastoista. Raportissa esitellään, mitä on semanttinen yhteentoimivuus ja mikä sen rooli on tietojärjestelmien yhteentoimivuuden edistäjänä. Myös semanttisen yhteentoimivuuden kehittämisen keskeisin väline sanastotyö esitellään. Raportti antaa vastauksia kysymyksiin: mitä sanastotyö on ja mitä sillä pyritään saavuttamaan. Raporttiin on kuvattu työryhmän laatima semanttisen yhteentoimivuuden kehittämismalli, joka sisältää vision ja kehityspolun, menetelmiä konkreettiseen sanastotyöhön sekä valtionhallinnon IT-toiminnan uudistukseen (ValtIT) sovitettun koordinoitiehdotuksen.

Avainsanat	JUHTA, VATI, tietohallinto, sähköinen asiointi, sähköiset palvelut, tietotekniikka, yhteentoimivuus, XML		
Sarjan nimi ja numero	Työryhmämuistioita, 3/2005		
Julkaisun tunnistetiedot	ISSN 0788-6322	ISBN 951-804-489-9 (nid) 951-804-490-2 (pdf)	
	Sivuja 73	Kieli Suomi	Hinta
Julkaisun myynti/jakaja	Valtiovarainministeriö		
Painopaikka ja -aika	Edita Prima Oy, Helsinki 2005		

PRESENTATIONSBLAD

Julkaisija ja julkaisuaika	Finansministeriet, februari 2005
Författare (i organet: organets namn, ordförande)	Riku Moisio, Finansministeriet, ordf. Juhani Koivunen, Finansministeriet, ordf. (- 31.10.2004) Jukka Ahtikari, WM-data Oy Virpi Kantola, Tullstyrelsen Seija Karevaara, Skattestyrelsen Väinö Malin, Miljöministeriet Timo Salovaara, Befolkningsregistercentralen Henri Ståhl, Avain Technologies Oy Timo Tarhonen. Tietotarha Oy, HL7 -yhdistys Riia Töyrylä, Jord- och skogsbruksministeriet Lasse Akselin, TietoEnator Oyj Riitta Alkula, TietoEnator Oyj
Publikationens titel	Ordlistorna som broar för ett smidigt informationsutbyte – Utvecklingsmodell för den semantiska interoperabiliteten av datasystem
Publikationens delar	Publikationen är tillgänglig på Internet på adressen www.vm.fi/julkaisut . Från den före nämnda adressen kan man också ladda det av arbetsgruppen utarbetade bakgrundsmaterialet “Ordlistearbetet inom olika branscher” som inte har tryckts.
Sammandrag	Syftet för utvecklandet av den offentliga förvaltningens datasystem är att främja den offentliga förvaltningens kvalitet och serviceförmåga samt produktiviteten av de tjänster som den offentliga förvaltningen erbjuder. Uppnåendet av syftet stöds bäst av en sådan kundkoncentrerad elektronisk förvaltning där den offentliga förvaltningens och dess centrala intressentgruppers datasystem är interoperabla och möjliggör till många delar automatiserade ärendegenomgångs- och serviceprocesser. I den offentliga förvaltningens XML-strategiarbete har observerats att enbart ett avtal om tekniska lösningar inte löser datasystemens interoperabilitet. Systemen skall vara interoperabla också på semantisk nivå för att datasystemen skall kunna bilda en smidig helhet. För att möjliggöra semantisk interoperabilitet skall de som utvecklar verksamheten och datasystemen tillsammans avtala om de begrepp och ordlistor som används. I rapporten presenteras vad semantisk interoperabilitet är och vilken dess roll är vid främjandet av datasystemens interoperabilitet. Också det centralaste redskapet vid utvecklandet av semantisk interoperabilitet ordlistearbetet presenteras. Rapporten ger svar på frågorna: vad är ordlistearbete och vad eftersträvar man att uppnå med det. I rapporten beskrivs den av arbetsgruppen utarbetade utvecklingsmodellen för semantisk interoperabilitet som innehåller visionen och utvecklingsbanan, metoderna för ett konkret ordlistearbete samt koordineringsförslaget som anpassats till reformen av statsförvaltningens IT-verksamhet (ValtiT).

Nyckelord	JUHTA, VATI, dataadministration, uträtta ärende på elektronisk väg, elektroniska tjänster, informationsteknik, Samverkan, XML		
Publikationsseriens titel och nummer	Finansministeriets arbetsgruppspromemorior 3/2005		
Antal sidor, totalt	ISSN 0788-6322	ISBN 951-804-489-9 (nid) 951-804-490-2 (pdf)	
	Sidor 73	Språk Finska	Hinta
Distributör/ Förläggare	Finansministeriet		

FACT SHEET

Julkaisija ja julkaisu-aika	Ministry of finance, February 2005
Authors (name, chairman of the body)	Riku Moisio, Ministry of Finance, Chairman Juhani Koivunen, Ministry of Finance, Chairman (- 31.10.2004) Jukka Ahtikari, WM-data Oy Virpi Kantola, National Board of Customs Seija Karevaara, National Board of Taxes Väinö Malin, Ministry of the Environment Timo Salovaara, Population Register Centre Henri Ståhl, Avain Technologies Oy Timo Tarhonen, Tietotarha Oy, HL7 Association Riia Töyrylä, Ministry of Agriculture and Forestry Lasse Akselin, TietoEnator Corporation Riitta Alkula, TietoEnator Corporation
Name of the publication	Vocabularies as a bridge for seamless information exchange - Development model for the semantic interoperability of information systems
Sections of the publication	Julkaisu on saatavissa Internetistä osoitteesta www.vm.fi/julkaisut . Em. osoitteesta voi ladata myös työryhmän laatiman taustamateriaalin "Eri alojen sanastotyö", jota ei ole painettu.
Summary	The goal of the development of the information systems of the public sector is to promote the quality and service-orientation of the public sector as well as the productivity of services provided by it. Attainment of the goal is best supported by customer-oriented electronic administration where the information systems of the public sector and its stakeholders interoperate and allow largely automated services and service processes. The Working Group of the XML Strategy of the Public Sector has noticed that technical interoperability alone does not solve the interoperability of information systems. The systems have to be interoperable also on the semantic level in order for information systems to form a seamless entity (entity ei tässä kuulosta hyvältä). In order to make semantic interoperability possible, those developing operations and information systems will have to jointly agree on the concepts and terminology used. The report discusses the concept of semantic interoperability and its role in promoting the interoperability of information systems. It also presents the central tool in the development of semantic interoperability, vocabulary work. The report answers the questions: what is vocabulary work and what are its aims. The report describes a development model for semantic interoperability drafted by the Working Group, which contains a vision and a development path, methods for tangible vocabulary work and a co-ordination proposal adapted to the reform of the IT activities of State government (ValtiT).

Keywords	JUHTA, VATI, data administration, electronic conduct of business, electronic services, information technology, interoperability, XML		
Serial name and number	WORKING PAPERS, 3/2005		
	ISSN 0788-6322	ISBN 951-804-489-9 (nid) 951-804-490-2 (pdf)	
	Pages 73	Language Finnish	Finnish Price
Distributed by/ Published by	Ministry of Finance		

Sisällys

1	TIIVISTELMÄ.....	11
2	JOHDANTO.....	13
2.1	EIF-viitekehys.....	15
2.2	Lisääntyvä semanttisen yhteentoimivuuden tarve	16
2.3	Semanttinen yhteentoimivuus ja sanastotyö	17
2.4	Sähköisen hallinnon vaatimuksia semanttiselle yhteentoimivuudelle	18
3	HALLINNON SEMANTTISEN YHTEENTOIMIVUUDEN KIPUPISTEITÄ... ..	19
1	Sanastotyön organisoimattomuus	19
2	Yhteisistä peruskäsitteistä ei ole sovittu	19
3	Yhdellä käsitteellä on useita käyttötarkoituksia tai yhteen käyttötarkoitukseen käytetään useita eri käsitteitä	20
4	Tietojen luokitukset (ja kooditukset) vaihtelevat.....	20
5	Käsitteet ovat teknisiä ja tietokantalähtöisiä	20
6	Sanastot eivät ole koneellisesti hyödynnettävissä.....	21
7	Inhimillinen valinta sanaston termiä valittaessa, luokiteltaessa tai kooditettaessa	21
4	SANASTOTYÖN YLEISPIIRTEITÄ.....	23
5	SEMANTTISEN YHTEENTOIMIVUUDEN LÄHESTYMISTAPOJA	25
5.1	Kooditukset ja luokitukset	25
5.2	Käsitteiden yleistämiseen perustuvia malleja	26
5.3	Semanttinen web...27	
6	SEMANTTISEN YHTEENTOIMIVUUDEN VISIO JA KEHITYSPOLKU	29
7	Sanastotyö prosessina.....	35
7.1	Liikkeellelähtö sanastotyössä.....	35
7.2	Sanastotyö osana järjestelmäkehitystä	37
7.3	Sanastojen yhtenäistämisen prosessi	38
8	SANASTOTYÖN JA SEMANTTISEN YHTEENTOIMIVUUDEN KEHITTÄMISEN KOORDINOINTI VALTIONHALLINNOSSA	41
9	TERMINOLOGIA.....	45
10	LÄHTEET	51
	LIITE: Semanttisen yhteentoimivuuden kipupisteitä – hallinnon esimerkkejä.....	53

1 TIIVISTELMÄ

Tämä raportti on tehty osana valtiovarainministeriön ”Julkishallinnon XML-strategian toteuttaminen” -hanketta. Hankkeessa asetettu työryhmä sai tehtäväkseen selvittää tietojärjestelmien semanttisen yhteentoimivuuden ongelmia sekä ratkaisuja näihin ongelmiin.

Tietojärjestelmien semanttisen yhteentoimivuuden kehittämisellä ja sanastotyöllä saavutettavia tärkeimpiä hyötyjä ovat:

- Julkishallinnon palvelujen laatu paranee ja niiden tuottaminen tehostuu, koska tietojärjestelmät toimivat paremmin yhteen.
- Asiakaslähtöinen tietojärjestelmien tukema asiakokonaisuuksien hoito yli hallinnon organisaatiorajojen tulee mahdolliseksi.
- Tiedon laatu paranee, koska saman tiedon keräämisen ja tallettamisen tarve virastoissa vähenee.
- Maan rajat ja kielirajat ylittävä tietojärjestelmien yhteentoimivuus erityisesti toisten EU-maiden kanssa tehostaa hallinnon toimintaa.
- Kehitystyön keskitetty koordinointi säästää taloudellisia ja henkilöresursseja.

Tässä projektissa on luotu edellytyksiä semanttisen yhteentoimivuuden kehittämiselle. Tätä varten on tunnistettu semanttisen yhteentoimivuuden luonne sekä luotu viitekehys sitä tukevalle sanastotyölle. Jatkokehittämisessä viitekehys on pohja yhteistyössä tapahtuvalle konkreettiselle sanastotyölle ja sanastojen luomiselle.

Julkishallinnon tietojärjestelmien kehittämistyön päämääränä on julkisen hallinnon laadun ja palvelevuuden sekä julkishallinnon tarjoamien palvelujen tuottavuuden edistäminen. Parhaiten päämäärän saavuttamista tukee sellainen asiakaskeskeinen sähköinen hallinto, jossa julkishallinnon ja sen keskeisten sidosryhmien tietojärjestelmät ovat yhteentoimivia ja mahdollistavat monilta osin automatisoidut asiankäsittely- ja palveluprosessit.

Tietojärjestelmien yhteentoimintaa voidaan kehittää useilla eri tasoilla. Kehittämisen painopiste on tähän asti ollut teknisen yhteentoimivuuden parantamisessa eli sopimisessa tiedonsiirto- ja rajapintatekniikoista. Pelkkä tekninen yhteentoimivuus ei kuitenkaan riitä. Saumattomat, asiakaskeskeiset ja organisaatiorajat ylittävät sähköisen hallinnon palve-

luprosessit edellyttävät myös tietosisällöllistä eli semanttista yhteentoimivuutta. Semanttinen yhteentoimivuus tarkoittaa, että tietojärjestelmä pystyy yhdistelemään eri lähteistä vastaanottamaansa tietoa ja käsittelemään sitä tavalla, jossa tietojen merkitys säilyy. Yhteentoimivuusongelman aiheuttavat palveluprosessiin osallistuvien tietojärjestelmien ja organisaatioiden erilaiset tietosisällöt ja käsitteistöt.

Keskeisin väline semanttisen yhteentoimivuuden kehittämisessä on sanastotyö. Sanastolla tarkoitetaan luetteloa jossain kielessä sallituista sanoista luokitteluineen, määriteltymineen, kuvauksineen ja esimerkkeineen. Sanastoilla kuvataan käsitteiden merkityksiä siten, että eri tietojärjestelmät voivat ”ymmärtää” käsittelemäänsä tietoa. Sanasto voi pohjautua viitemalliin, jossa kuvataan käsitteiden välisiä suhteita.

Raportissa esitellään semanttisen yhteentoimivuuden visio sekä kehityspolku, jolla vision mukainen tavoitetilä voidaan saavuttaa. Vision mukaisessa tilassa semanttisen yhteentoimivuuden merkitys organisaatioiden välisessä yhteentoimivuudessa on tiedostettu, sen kehittäminen ja ylläpitäminen on koordinoitu ja resursoitu hallinnon eri tasoilla, ja sanastoja kehitetään ja hyödynnetään suunnitelmallisesti. Lisäksi itse sanastotyön tarve vähenee, koska päällekkäisen tiedon tallennusta on vähennetty samansisältöisiä tietojärjestelmiä yhdistämällä.

Kehityspolussa on kuvattu karkean tason askeleet, joilla pyritään ensin tietoisuuden kasvattamiseen ja myöhemmin työn riittävään resurssointiin ja vastuutukseen hallinnon eri tasoilla.

Valtion tietohallintoa ollaan kehittämässä ns. TIME-työryhmän esitysten mukaisesti yhtenäisempään suuntaan. Valtion tietohallintoa kehittämään perustetaan IT-toiminnan johtamisyksikkö. Yksikön tehtäviin kuuluu IT-toimintaan liittyvien standardien ja arkkitehtuurien kehittäminen. Työryhmä ehdottaa, että IT-toiminnan johtamisyksikkö ottaa päävastuun semanttisen yhteentoimivuuden ja sanastotyön kansallisesta kehittämisestä. Lisäksi alemmilla hallinnon hierarkiatasoilla on semanttisella yhteentoimivuudella oltava selkeät omistajat, jotta yksittäisissä hankkeissa osataan käyttää hyödyksi semanttisen yhteentoimivuuden kehittämiseksi tehtyä työtä.

Työryhmä esittää seuraavia jatkotoimenpiteitä:

- Työryhmän työtä jatkamaan perustetaan jatkotyöryhmä, jonka tehtävänä on tuottaa sanastotyöstä JHS-suositus.
- Aloitetaan semanttisen yhteentoimivuuden tärkeydestä ja sanastotyöstä tiedottaminen sekä menetelmien ja välineiden kehittäminen.
- Järjestetään sanastotyön koordinointi ja resurssointi sekä valtion IT-johdossa että vi-rastoissa.
- Raportti ja siinä esitetyt toimenpide-ehdotukset luovutetaan perustettavalle valtion IT-toiminnan johtamisyksikölle, jonka tehtäviin kuuluu valtion IT-toiminnan standardointi- ja arkkitehtuurityö. Semanttisen yhteentoimivuuden kehittäminen on osa em. tehtäväkokonaisuutta.

2 JOHDANTO

Tietotekniikan käytön yleistyessä ja nykyisten tietoverkkojen mahdollistaessa välittömät yhteydet tietojärjestelmien välille on keskeiseksi haasteeksi noussut tietojärjestelmien yhteentoimivuuden muiden esteiden vähentäminen. Julkishallinnon järjestelmien tulee pysyä toimimaan yhteen, paitsi keskenään, tarvittaessa myös mm. yksityisen sektorin ja muiden EU-maiden järjestelmien kanssa.

Tässä raportissa keskitytään tietojärjestelmien semanttiseen yhteentoimivuuteen, joka nousee keskeiseen asemaan yhdistettäessä toisiinsa erillisiä tietojärjestelmiä hallinnon sisällä sekä hallinnon ja sen sidosryhmien välillä.

Semanttisen yhteentoimivuuden kehittäminen edellyttää sekä teknisten asiantuntijoiden että toiminnan kehittäjien panosta. Tämä raportti onkin kohdennettu sekä virastojen sisäisestä ja niiden välisestä toiminnasta vastaaville että IT-kehittämisestä vastaaville henkilöille.

Raportissa esitellään, mitä on semanttinen yhteentoimivuus ja mikä sen rooli on tietojärjestelmien yhteentoimivuuden kehittämisessä. Myös semanttisen yhteentoimivuuden kehittämisen keskeisin väline sanastotyö esitellään. Raportti antaa vastauksia kysymyksiin: mitä sanastotyö on, mitä sillä pyritään saavuttamaan ja millaisia tapoja sanastotyön tekemiseen on olemassa. Myös eri yhteisöjen erilaisia näkemyksiä sanastoista on selvitetty.

Raporttia valmistelleessa työryhmässä on ollut kattava edustus valtionhallinnosta. Työryhmän keskusteluissa on havaittu monia tietojärjestelmien yhteentoimivuutta haittaavia ongelmia, jotka on voitu paikallistaa semanttisen tason ongelmiksi. Tyypillisimmät ja yleisimmät kipupisteet on listattu raportissa. Raportin liitteenä on lisäksi muutamia käytännön esimerkkejä tällaisista järjestelmäkehityksessä vastaan tulleista ongelmista.

Semanttinen yhteentoimivuus käsitteenä sekä sen kehittämisen menetelmät ovat melko uusia. Teknisiä apukeinoja tietojärjestelmien semanttisen yhteentoimivuuden lisäämiseksi ovat muun muassa:

- yhteisesti sovittavien kooditusten ja luokitusten käyttäminen,

- sellaiset mallit, joissa kohdealueiden erikoiskäsitteitä pyritään yleistämään yhteentoimivuuden saavuttamiseksi sekä
- semanttisen webin teknologiat, joilla pyritään määrittelemään ja siirtämään koneiden käsiteltäväksi ja ”ymmärrettäväksi” sellaisia asioiden välisiä suhteita, mihin on aiemmin tarvittu ihmistä.

Sanastojen kehittäminen on tärkeää suunnata tietojärjestelmä- ja organisaatiokohtaisista sanastoista ensin organisaatioiden yhteisiin sanastoihin ja edelleen kohti kansallisia ja kansainvälisiä toimialakohtaisia ja toimialariippumattomia sanastoja. Raportissa on kuvattu tätä kehityssuuntaa sanastotyön matriisina (taulukko 1, luku 6), jota tietojärjestelmäkehittäjät voivat käyttää hyväkseen sanastotyön merkitystä arvioidessaan sekä sanastotyön oikean tason valitsemisessa.

Työryhmän mielestä semanttisen yhteentoimivuuden kehittäminen on erittäin tärkeää tietojärjestelmien yhteentoimivuuden lisäämiseksi. Työn jatkuvuuden varmistamiseksi työryhmä on määritellyt semanttisen yhteentoimivuuden vision sekä kehityspolun, jolla vision mukainen tavoitetilä voidaan saavuttaa. Visiossa korostetaan, että semanttisen yhteentoimivuuden merkitys organisaatioiden välisessä yhteentoimivuudessa tulee tiedostaa, sen kehittäminen ja ylläpitäminen tulee koordinoida ja resursoida hallinnon eri tasoilla ja sanastoja tulee kehittää ja hyödyntää suunnitelmallisesti. Lisäksi itse sanastotyön tarve vähenee, koska päällekkäisen tiedon tallennusta on vähennetty samansisältöisiä tietojärjestelmiä yhdistämällä.

Raportissa esitellään suuntaviivoja, millä tavalla sanastotyötä tulisi tehdä prosessina sekä mikä on sanastotyön rooli yleisessä järjestelmäkehityksen elinkaareissa. Lisäksi ehdotetaan uuteen valtion IT-toiminnan organisointimalliin sovitettua sanastotyön organisointimallia. On tärkeää, että sanastotyö on selkeästi vastuutettu eri hallinnon tasoilla ja työlle on varattu asianmukaiset resurssit.

Semanttisen yhteentoimivuuden kehityspolussa ehdotetaan, että ensimmäisenä toimenpiteenä on perustaa tämän työryhmän työtä jatkamaan jatkotyöryhmä, jonka tehtävänä on tuottaa sanastotyöstä JHS-suositus. Välttömästi on myös aloitettava semanttisen yhteentoimivuuden tärkeydestä ja sanastotyöstä tiedottaminen sekä menetelmien ja välineiden kehittäminen. Erittäin tärkeänä työryhmä pitää, että jatkossa järjestetään sanastotyön koordinointi ja resursointi sekä valtion IT-johdossa että virastoissa. Raportti ja siinä esitetyt toimenpide-ehdotukset luovutetaan perustettavalle valtion IT-toiminnan johtamisyksikölle, jonka tehtäviin kuuluu valtion IT-toiminnan standardointi- ja arkkitehtuurityö. Semanttisen yhteentoimivuuden kehittäminen on osa tuota tehtäväkokonaisuutta.

2.1 EIF-viitekehys

IDA:n (Interchange of Data between Administrations) julkaisema EIF-viitekehys (European Interoperability Framework) tarjoaa tavan jäsentää tietojärjestelmien yhteentoimivuutta. EIF pyrkii linjaamaan sähköisen hallinnon (e-government) ja sähköisten asiointipalvelujen yhteentoimivuutta EU:n jäsenmaiden välillä. EIF:n mukaan tietojärjestelmien yhteentoimivuuden vaatimukset voidaan jakaa kolmeen eri luokkaan:

- tekninen
- semanttinen ja
- organisatorinen yhteentoimivuus.

Teknisellä yhteentoimivuudella tarkoitetaan sopimista avoimista rajapinnoista sekä standardeista tietoformaateista ja protokollista. Internet-pohjaiset standardit ja tekniikat ovat parantaneet viime vuosina tietojärjestelmien teknisen yhteentoimivuuden tasolle, joka ei ole aiemmin ollut mahdollista. Muun muassa standardit tiedonsiirto- ja sähköpostiprotokollat, kuten TCP/IP, HTTP ja S/MIME, sekä standardi tietoformaatti XML ovat luoneet pohjaa yhtenäisille teknisille ratkaisuille.

Semanttinen yhteentoimivuus mahdollistaa, että järjestelmä pystyy yhdistelemään eri lähteistä vastaanottamaansa tietoa ja käsittelemään sitä tavalla, jossa tietojen merkitys säilyy. Semanttinen yhteentoimivuus tukee myös halutun tiedon löytämistä. Tässä hankkeessa ja raportissa keskitytään semanttiseen yhteentoimivuuteen siltä osin kuin se tukee automaattisesti (tietojärjestelmissä) käsiteltävän tiedon (uudelleen)käytettävyyttä.

Organisatorisella yhteentoimivuudella tarkoitetaan organisaatioiden toimintaprosessien sekä digitaalisten ja automaattisten prosessien kuvaamista siten, että ne tukevat sähköisen hallinnon tarpeita ja mahdollisuuksia. Tämä edellyttää organisaatioiden sisäisten ja niiden välisten toimintaprosessien yhteistoimintaa sekä tietoarkkitehtuurien rakentamista vastaamaan organisaatioiden toiminnallisia tavoitteita.

Tekninen yhteentoimivuus takaa tiedon esteettömän kulun tietojärjestelmien välillä ja organisatorinen yhteentoimivuus mahdollistaa eri osapuolien digitaalisten toimintaprosessien nivoutumisen yhteen. Semanttisen yhteentoimivuuden ydinkysymys on, millä tavalla sekä ihmiset että koneet voivat ymmärtää sellaisen tiedon merkityksen, jonka joku toinen osapuoli on luonut. Semanttisen yhteentoimivuuden haaste on siinä, että tietojärjestelmien tulisi ymmärtää automaattisesti teknisellä tasolla ihmisten ja organisaatioiden toiminnasta lähtöisin olevia käsitteitä.

2.2 Lisääntyvä semanttisen yhteentoimivuuden tarve

Tietojen vaihto tietojärjestelmien välillä on perinteisesti perustunut määrämuotoisten tiedostojen ja sanomien siirtoon. Siirtotapoja ovat olleet yhteisesti sovitut tietueet sekä standardimmat tavat kuten EDI-standardeihin pohjautuva tiedonsiirto. Viime aikoina XML on yleistynyt siirrettävän tiedon koodaustapana. Yhteistä näille tiedonsiirtotavoille on ollut, että lähettäjä on ollut aktiivisessa roolissa. Myös kerralla siirrettävän tiedon määrä on ollut varsin suuri.

Perinteinen tiedonsiirto on perustunut pitkälti teknisen yhteentoimivuuden saavuttamiseen. Semanttiseen yhteentoimivuuteen on pyritty kahdenvälisillä sopimuksilla sekä hyvin homogeenisissa yhteisöissä sopimalla yhteisistä tietokentistä ja niiden sisältöjen merkityksistä. Usein semanttista tasoa vastaavat sopimukset on esitetty suoraan teknisinä kuvauksina, XML:n tapauksessa DTD- tai XML Schema -määrittelyinä.

Käytännössä monilla alueilla tiedot ovat hajallaan viranomaisen eri järjestelmissä tai jopa eri viranomaisten järjestelmissä. Esimerkiksi elintarvikkeita koskevia tietoja on niin eri viranomaisten kuin muidenkin toimijoiden tietojärjestelmissä. Siirryttäessä kahdenvälisistä sopimuksista ja homogeenisista ympäristöistä useamman, sisällöllisesti erilaiseen kohdealueeseen kuuluvan, osapuolen väliseen tiedonvaihtoon, tekninen yhteentoimivuus ja skeemat eivät enää takaa semanttista yhteentoimivuutta.

Tärkeä trendi tietojärjestelmien kehittämisessä tällä hetkellä on palvelupohjainen arkkitehtuuri (Service Oriented Architecture, SOA). Perusajatuksena on julkaista tietojärjestelmän palvelujen julkiset rajapinnat näiden palveluiden toteutuksesta riippumattomalla tavalla. Palvelut avautuvat tarjolle yhä useammille potentiaalisille käyttäjille ja toisille tietojärjestelmille. Palvelupohjaisen arkkitehtuurin käytännön toteutuksen on suositeltavaa perustua avoimiin löyhää kytkentää tukeviin standardeihin ja tekniikoihin (kuten Web Services). Näin voidaan varmistaa järjestelmien ja verkkopalvelujen tekninen yhteentoimivuus. Semanttinen yhteentoimivuus on tässäkin tapauksessa kuitenkin vielä ratkaisematta.

Sekä perinteisempi tiedonvaihto heterogeenisessä ympäristössä että palvelupohjaisuuteen siirtyminen asettavat suuria haasteita semanttiselle yhteentoimivuudelle. Kokemus osoittaa, että teknisen tason yhtenäisyys ei toimi mm. seuraavista syistä (Burkett, 2004):

- Yksikäsitteisesti nimetyt tietokentät eivät määrittele yksikäsitteisesti kenttien sisällön semantiikkaa.
- Eri tietojärjestelmissä tulkitaan yksikäsitteiseksi tarkoitettun tietokentän sisältö poikkeavasti sopimuksista huolimatta.
- Eri tietojärjestelmissä samaa tarkoittava käsite halutaan nimetä eri tavoin ja sille annetaan erikoismerkitys.

2.3 Semanttinen yhteentoimivuus ja sanastotyö

Eri tietojärjestelmissä käytetyt käsitteet ovat tyypillisesti syntyneet yhden toimialan ja sen sisällä olevan tietyn sovellusalueen tarpeista ja vieläpä yksittäisten tietokantojen ja sovelusten teknisten tarpeiden mukaan. Tietojärjestelmiä rakennettaessa ei siis tavallisesti ole varauduttu siihen, että niissä käytetyt käsitteet olisivat ymmärrettävissä ja käsiteltävissä toisissa tietojärjestelmissä. Lisäksi tietojen siirrossa, käsittelyssä ja tulkinnessa tarvittavaa tietoa ei ole kokonaisuudessaan edes tietojärjestelmien tietokannoissa ja käsittelysäännöissä, vaan näitä tietojärjestelmiä käyttävien ihmisten mielessä ja organisaatioiden toimintakulttuurissa. Tällainen tieto on saatava kuvattua eksplisiittisesti tavalla, jota tietokoneetkin pystyvät ymmärtämään. Näin tietojärjestelmät pystyvät toimimaan abstraktimmalla tasolla kun nykyään.

Kun tiedon löytämisen ja järjestelmien integroimisen tarve kasvaa, joudutaan kehittämään menetelmiä ja välineitä, jotka mahdollistavat käsitteiden yhdenmukaistamisen ja niiden välisten suhteiden kuvaamisen tavalla, joka edesauttaa yhteentoimivuutta semanttisella tasolla. Tietojärjestelmien semanttisen yhteentoimivuuden kehittämisellä tarkoitetaan sellaisia toimenpiteitä ja kehitystyötä, joilla järjestelmien välistä semanttista yhteentoimivuutta voidaan lisätä.

Sanastotyötä yleisesti voidaan pitää välineenä, jolla semanttista yhteentoimivuutta pyritään kehittämään. Sanastolla tarkoitetaan tässä yhteydessä luetteloa jossain kielessä salituista sanoista (elementeistä) luokitteluineen, määritelmineen, kuvauksineen ja esimerkkeineen. Sanastoilla kuvataan käsitteiden merkityksiä ja niiden välisiä suhteita siten, että eri tietojärjestelmät voivat ”ymmärtää” käsittelemäänsä tietoa. Sanaston pohjana voi olla myös käsitteiden välisiä suhteita kuvaava viitemalli, kuten HL7:n RIM (Reference Information Model).

Tietojärjestelmien käyttöön määriteltäville sanastoille on olennaista, että sanat (tai käsitteet) ja niiden väliset suhteet on määriteltä tarkasti. Eri kohdealueiden käsitteistöjä määriteltäessä on varottava liiallista nojautumista teknisiin käsitteisiin, joita käytetään muun muassa tietokannoissa. Sanastot tulee määritellä ensi sijassa toteutusteknologiasta riippumattomalla tavalla. Käyttötilanteen ja teknologioiden kypsyysasteen mukaan sanastot voidaan implementoida standardeilla kuvaustavoilla kuten XML-pohjaiset DTD, XML Schema, RDF Schema tai OWL.

Perinteisesti sanastotyön varsinaisia asiantuntijoita ja hyödyntäjiä ovat olleet kohdealueen toiminnan tuntevat henkilöt. Jatkossa sanastotyön hyödyntäjiä ovat aiempien yhteisöjen lisäksi myös tietojärjestelmien kehittäjät. Sanastojen kehittämisessä tarvitaan molempien asiantuntijaryhmien panosta. Tietopalveluammattilaiset ovat hyvä esimerkki ammattiryhmästä, jolla on kokemusta sekä sanastojen sisällöllisestä kehittämisestä että niiden hyödyntämisestä tietojärjestelmissä. Kohdealueen asiantuntijat pitävät huolta siitä, että syntyvä sanasto vastaa kyseisen yhteisön asiantuntemusta ja siinä vakiintuneita merkityksiä. Tietotekniikan asiantuntijat taas pitävät huolta, että tuloksena syntyvä sanas-

to on sovellettavissa myös tietojärjestelmissä. Näiden kahden asiantuntijayhteisön tarpeiden yhdistäminen on haasteellista. Jos sanastotyötä tekevät pelkästään tietojärjestelmäkkehittäjät, ajaututaan helposti skeema- ja kenttätason määrittäisiin. Jos sanastotyötä tekevät vain kohdealueen toiminnan asiantuntijat, sanastotyö keskittyy ammattisanastojen ja termistöjen määrittelyyn.

Sanastotyössä määritellään eri käyttötarkoituksiin sekä sovellusalue- ja toimialakohdaisia että toimialariippumattomia sanastoja. Sanastot voivat olla sovellettavuudeltaan paikallisia, kansallisia, EU-tasoisia tai maailmanlaajuisia.

Semanttiseen yhteentoimivuuteen liittyvän sanastotyön varsinaisia käytännön hyödynnäittäjiä ovat tietojärjestelmien ja verkkopalvelujen kehittäjät. Tietojärjestelmien ja verkkopalveluiden käyttäjille sanastotyön hyödyt ilmenevät parempana ja kokonaisvaltaisempaan palveluna.

2.4 Sähköisen hallinnon vaatimuksia semanttiselle yhteentoimivuudelle

Semanttisen yhteentoimivuuden kehittämisen tarvetta ja hyötyjä voidaan arvioida kysymällä, mitkä sähköisen hallinnon kehitystä ohjaavat tekijät asettavat vaatimuksia semanttiselle yhteentoimivuudelle.

Lähtökohtaisena tavoitteena kaikessa kehittämistyössä tulee olla julkisen hallinnon laadun ja palvelevuuden sekä sen tarjoamien palvelujen tuottavuuden edistäminen. Semanttinen yhteentoimivuus kokonaisuudessaan on yksi tärkeimmistä osa-alueista koko sähköisen hallinnon kehittämisessä. Sähköinen hallinto asettaa muun muassa seuraavia vaatimuksia semanttisen yhteentoimivuuden lisäämiselle:

- Asiakaslähtöinen asiakokonaisuuksien hoito yli hallinnon organisaatorajojen
- Hallinnon sekä sen keskeisten sidosryhmien, mukaan lukien elinkeinoelämä, yhteiset tietojärjestelmien tukemat prosessit
- Maan rajat ja kielirajat ylittävä tietojärjestelmien yhteentoiminta erityisesti toisten EU-maiden kanssa
- Tietosisältöjen ja verkkopalvelujen tehokkaampi tuottaminen ja kevyempi ylläpito
- Tietosisältöjen ja verkkopalvelujen tarjoaminen julkishallinnon yleisportaaaleissa sekä toisten viranomaisten ja muiden sidosryhmien portaaaleissa
- Perusrekisterien ja vastaavien tietokantojen primäärisen tiedon hyödyntäminen palvelulähtöisesti.

Semanttisen yhteentoimivuuden merkitys voidaan kiteyttää seuraavasti:

Tietojärjestelmien semanttisen yhteentoimivuuden tulee osaltaan mahdollistaa asiakaskeskeinen sähköinen hallinto, jossa julkishallinnon ja sen keskeisten sidosryhmien tietojärjestelmät toimivat sekä teknisellä että semanttisella tasolla yhteen mahdollistaen organisaatioiden tehokkaan yhteistoiminnan automatisoituina asian käsittely- ja palveluprosesseina.

3 HALLINNON SEMANTTISEN YHTEENTOIMIVUUDEN KIPUPISTEITÄ

Alla on esitelty yhteenvedona tyypillisimpiä hallinnossa havaittuja semanttisen yhteentoimivuuden ongelmia ja kipupisteitä. Yksityiskohtaisempia hallinnonala- ja organisaatiokohtaisia kuvauksia löytyy liitteestä.

1 Sanastotyön organisoimattomuus

Pahin semanttisen yhteentoimivuuden kipupiste on, että sanastotyötä ei ole resursoitu eikä vastuutettu eikä sitä tehdä systemaattisesti. Sen sijaan ongelmat ratkaistaan tapauskohtaisesti jokaisessa projektissa erikseen. Kun kukaan ei omista sanastotyötä eikä vastaa sen kehittämisestä, yhteistä linjaa ei synny ja samaa työtä tehdään eri tahoilla ja projekteissa uudelleen.

Syynä ongelmaan on myös, että semanttisen yhteentoimivuuden tarvetta ja ongelmia ei ole laajasti tiedostettu. Täten myöskään keinoja ongelmien ratkaisemiseksi ei ole kehitetty.

2 Yhteisistä peruskäsitteistä ei ole sovittu

Semanttisen yhteentoimimattomuuden perusongelma on, että peruskäsitteistä, niiden kuvauksista ja merkintätavoista ei ole sovittu (esimerkiksi nimi, osoite, asiakirjan tunniste). Ongelma esiintyy sovellusalueen sisällä tietojärjestelmien välillä, toimialan sisällä, kansallisesti toimialojen välillä sekä kansainvälisesti monilla tasoilla (semanttisen yhteentoimivuuden eri tasot ks. taulukko 1, luku 6). Peruskäsitteistöä ei ole olemassa eikä sellais-

ta pystytty lyhyellä aikavälillä luomaan. Ratkaisu ongelmaan löytyy pienistä korjauksista vallitsevaan tilanteeseen. Perusrekistereitä ja vastaavia tietokantoja tulisi osaltaan hyödyntää peruskäsitteiden määrittelyssä.

3 Yhdellä käsitteellä on useita käyttötarkoituksia tai yhteen käyttötarkoitukseen käytetään useita eri käsitteitä

Ongelma on syntynyt tietojärjestelmien ja niiden tietomallien ja sanastojen kehittämisen paikallisen luonteen seurauksena. Ongelma ilmenee käytännössä muun muassa seuraavilla tavoilla:

- Toimialan sisällä on käytössä useita järjestelmiä ja sanastoja, joissa samasta käsitteestä käytetyt termit ja merkintätavat ovat erilaisia.
- Epätarkka tai väärä käsite vakiintuu käyttöön, koska sen merkitystä yhteentoimivuudelle ei ole otettu huomioon.
- Puutteellisten kuvausten ja dokumentaation vuoksi käsitteiden merkitys on ainoastaan tietojärjestelmien vastuuhenkilöiden tiedossa. Ulkopuolisten tahojen on vaikeaa tietää tällaisten käsitteiden täsmällistä merkitystä.
- Tietojen yhdistäminen esimerkiksi verkkopalveluissa tai julkaisujärjestelmissä on vaikeaa käsitteiden erimittaisuuden vuoksi.

4 Tietojen luokitukset (ja kooditukset) vaihtelevat

Tyypillinen ongelma on, että eri tietojärjestelmissä käytetään eri luokituksia saman tiedon arvoina. Toinen ongelma on, että silloin kun käytetään samaa luokitusta, tietojärjestelmissä annetaan luokitetuille arvoille erilaiset tulkinnat. Tietojärjestelmien yhteentoimivuus edellyttää, että osapuolet sopivat käytetystä luokituksesta ja käyttävät sitä säännönmukaisesti yhdenmukaisella tavalla.

5 Käsitteet ovat teknisiä ja tietokantalähtöisiä

Tavallisesti tietoa siirrettäessä järjestelmien välillä lähettäjän ja vastaanottajan tietokantojen tietorakenteet ja tietomallit poikkeavat toisistaan. Tietokantatason kentät ovat lisäksi toteutusnäkökulmasta luotuja ja tavallisesti semanttisella tasolla heikosti dokumentoituja. Usein järjestelmien yhteentoimivuus pyritään kuitenkin rakentamaan näiden teknisen tason käsitteiden varaan. Tämä taso ei kuitenkaan riitä, vaan yhteistoimintaa tukevista käsitteistä on sovittava ylemmällä ei-teknisellä tasolla.

6 Sanastot eivät ole koneellisesti hyödynnettävissä

Monissa hallinnon organisaatioissa on tehty sanasto- tai sanakirjatyötä sellaisessa muodossa, joka ei ole koneellisesti hyödynnettävissä. Vaikka tämä työ hyödyttää työntekijöitä asioiden käsittelyssä, sitä ei voida hyödyntää suoraan tietojärjestelmien yhteentoimivuuden automatisoinnissa. Näitä sanastoja voidaan kuitenkin hyödyntää määriteltäessä koneille tarkoitettuja kuvauksia, kuten ontologioita ja skeemoja.

Koneelliseen muotoon viedyt sanastokuvaukset ovat usein kehittymättömiä. Yleensä niissä ei ole kuvattu käsitteiden välisiä suhteita monipuolisesti esimerkiksi hyödyntämällä semanttisen webin teknologiaa. Tämä aiheuttaa vakavan ongelman tietojen yhdistämiselle toimialojen sisällä ja erityisesti toimialojen välillä.

7 Inhimillinen valinta sanaston termiä valittaessa, luokiteltaessa tai kooditettaessa

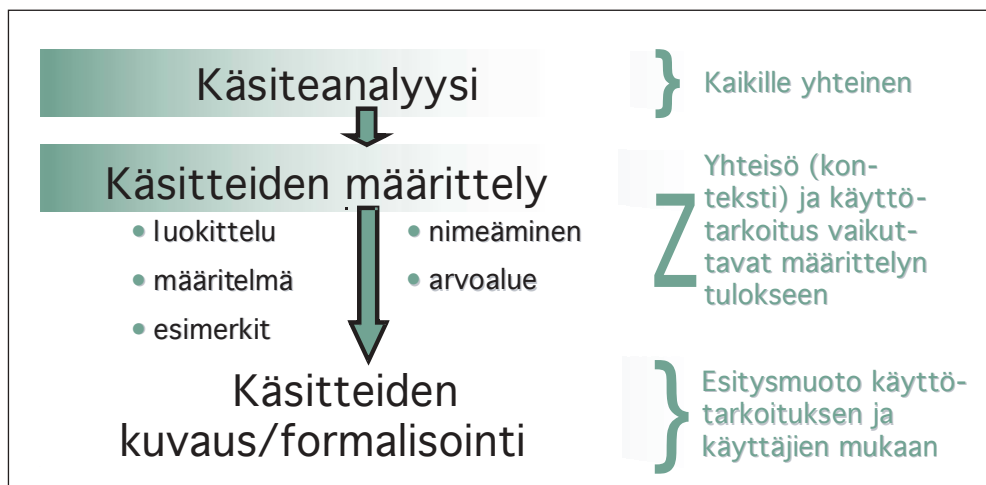
Semanttisesta yhteentoimivuudesta puhuttaessa ei voida unohtaa inhimillistä arviointia ja valintatilannetta tiedon kuvaamisessa, luokittelussa tai koodituksessa. Valintoihin vaikuttavia asioita ovat muun muassa sanastoja hyödyntävien henkilöiden asiantuntemus, mieltymykset ja perusteellisuus. Tietojärjestelmien välinen semanttinen yhteentoimivuus on parhaassakin tapauksessa vain niin hyvää kuin ihmisten tekemät valinnat ja yhdenmukaisissa käytännöissä pysyminen mahdollistavat.

4 SANASTOTYÖN YLEISPIIRTEITÄ

Luvussa 3 esitetyt hallinnon semanttisen yhteentoimivuuden ongelmat eivät rajoitu julki-
shallintoon, vaan pätevät mihin tahansa yhteisöön, jossa on tarvetta sopia yhteisistä ter-
meistä ja niiden käytöstä. Sanat tai termit ovat sopimuksia siitä, miten jokin yhteisö halu-
aa jotain kohdetta nimittää, olipa kohde sitten fyysinen reaalimaailman esine (ämpäri) tai
abstrakti idea (vapaus). Sanastotyön yleinen päämäärä on yhdenmukaistaa ja täsmentää
käsitteiden määrittelyä, jotta käsitteiden ja termien epäyhtenäinen ja vaihteleva tulkinta
ei haittaisi eri osapuolten välistä kommunikointia.

Kuva 1 kuvaa sanastotyön kohdealue riippuvuutta. Sanastotyö alkaa aina käsiteanalyysillä, jossa tutkitaan käsitteiden merkityssisältöä ja suhteita toisiin käsitteisiin. Tämä analyysi on periaatteessa kohdealue riippumatonta eli siinä määritellään elementit (merkityspiirteet), joista käsite rakentuu.

Seuraavalla tasolla käsitteiden määrittelyn tulos riippuu siitä, mitä merkityspiirteitä käsitettä määrittelevä yhteisö pitää tärkeinä ja mitä vähemmän tärkeinä. Esimerkiksi tiedonhaussa voidaan olla väljempää sen suhteen, viittaavatko jotkin ilmaukset samaan käsitteeseen eli ovat toistensa synonyymejä, kun taas terminologiatyössä pyritään täsmällisempään määrittelyyn. Tietojärjestelmille tehtävissä sanastoissa taas voidaan edellyttää, että yhtä käsitettä voi vastata vain yksi termi.



Kuva 1. Käsitteiden analysoinnin ja määrittelyn eri tasot ja asiayhteyden (kontekstin) vaikutus tuloksiin

Alin taso on toteutus- eli implementointitaso: ihmisille tarkoitetut tiedonhakutesau-
rukset voivat olla manuaalisesti käytettäviä kirjoja, tietojärjestelmiä varten taas laaditaan
konelukuinen skeema tai ontologia jne.

Sanastotyötä on monissa yhteisöissä tehty jo pitkään – esimerkkeinä tällaisista ovat
sanakirjojen laatijat, terminologit ja muistiorganisaatiot (kirjastot, museot ja arkistot).
Toisilla alueilla taas sanastotyön historia on lyhyt, koska itse alakin on uusi (ainakin ver-
rattuna edellä mainittuihin aloihin); tällaisia ovat tietokantojen ja tietojärjestelmien käsit-
teellinen mallinnus ja tiedonsiirto eri tietojärjestelmien välillä.

Koska kukin yhteisö noudattaa sanastotyössä oman alueensa käytäntöjä ja terminolo-
giaa ja koska sanastojen käyttötarkoitukset poikkeavat toisistaan, eri yhteisöt eivät aina
osaa hyödyntää toisten alojen sanastotyön tuloksia, kun käsitteiden määrittelyjen tulok-
set ja yleensä lähestymistapa semanttisuuteen ja sanastotyöhön on toisenlainen kuin mi-
hin omalla alalla on totuttu.

5 SEMANTTISEN YHTEENTOIMIVUUDEN LÄHESTYMISTAPOJA

Erilaisia semanttisuuden lisäämiseen pyrkiviä viitemalleja, käytäntöjä ja standardeja on jo ehtinyt syntyä. Merkittävän haasteen tietojärjestelmien semanttisen yhteentoimivuuden lisäämiselle aiheuttaakin se, että eri tahoilla tehtävä kehitystyö menee osittain päällekkäin. Tosin eri puolilla tehty kehitystyö myös osittain täydentää toistaan. Lähestymistavat voivat kuitenkin myös poiketa niin paljon toisistaan, että niiden vertailua on vaikea saada yhteismitalliseksi. Alla on lueteltu muutamia keskeisimpiä pyrkimyksiä semanttisen yhteentoimivuuden kasvattamiseksi.

5.1 Kooditukset ja luokitukset

Luokitusten ja kooditusten pääasiallinen tarkoitus on lisätä kontrollia tiedon esittämiseen. Käytännössä näiden hyöty perustuu yhdessä tehtyihin sopimuksiin kontrolloiduista arvoista sekä kurinalaiseen ja yhdenmukaiseen käyttöön.

Esimerkki kansainvälisestä luokitteluun ja kooditukseen pohjautuvasta työstä on mm. kaupan, kuljetuksen ja tullialalla käytettävä UN/TDED (United Nations/Trade Data Elements Directory), joka on ISO:n standardi. Koska tulliala on luonteeltaan ja lähtökohdiltaan kansainvälistä, on tietoluettelo käytössä myös EU:ssa. Standardi on käytännössä luettelo sallituista elementeistä määrittelyineen. Luettelo perustuu tiedon koodiin (neljä numeroa), sen nimeen, kuvaukseen, muotoon sekä huomioihin (esim. suositeltava koodiluettelo). Luettelossa olevista tiedoista on myös tehty luokkakaavio (UML).

Koodien yhdenmukaiseen ja kontrolloituun käyttöön voidaan käyttää apuna ns. koodistopalvelinta, joka on keskitetty hakemisto, joissa koodistoa ylläpidetään. Palvelimelta on tyypillisesti saatavissa ajantasaiset koodit sekä niiden kuvaukset sekä ihmiselle että

koneelle tarkoitettussa muodossa. Koodistoja voidaan tietojärjestelmäkäyttöä ajatellen hakea ajantasaisesti tai kopioimalla tietoja tietyin aikavälein paikallisiin järjestelmiin. Suomessa STAKES ylläpitää sosiaali- ja terveydenhuoltoalan koodistopalvelinta osoitteessa: <http://koodistopalvelu.stakes.fi>.

5.2 Käsitteiden yleistämiseen perustuvia malleja

Käsitteiden yleistämiseen perustuvien mallien lähtökohtana on yleisten ja yhteisten kontekstista riippumattomien ydintietomallien -käsitteiden ja -komponenttien määrittely. Malleissa on kehitetty mekanismi, jolla yleisistä rakenteista voidaan johtaa tiettyyn toimialaan tai käyttötarkoitukseen erikoistettuja asiayhteyteen sidottuja rakenteita.

Ydintietomallit, -käsitteet ja -komponentit kuvaavat kaikille yhteisiä osia, jotka ovat periyttävissä erikoistettuihin komponentteihin. Näin pyritään välttämään samojen kaikille yhteisten osien määrittelemistä eri tavoin useassa paikassa.

Abstrakteimman tason mallinnuskielenä käytetään tavallisesti UML-kieltä, jolla mallinnetaan käsitteet ja niiden väliset suhteet, mukaan lukien periytyvyysuhteet. Malleissa on myös tavallisesti mekanismi tai systematiikka, jolla abstraktin tason käsitelmalleista johdetaan skeema- ja sanomatason konkreettiset kuvaukset.

ebXML core components –määrittely on osa ebXML-viitekehystä. Sen tavoitteena on määrittellä prosessi, jolla tietokomponentteja voidaan löytää ja luokitella systemaattisesti siten, että niistä voidaan analysoida ydinkomponentteja. Syntyneen ydinkomponentti-luettelon tarkoitus on mahdollistaa yhteentoimivuus eri toimialojen välisessä sähköisessä kaupankäynnissä.

UBL (Universal Business Language) on XML-pohjainen määrittely liiketoimintasanomien kuvaamiseen yli toimialarajojen. UBL pyrkii mahdollistamaan erityisesti sanomanvälitykseen perustuvan liiketoiminnan toimialojen välillä. UBL pohjautuu, samalla tavalla kuin ebXML core components, ydinkomponenttien ja niistä johdettujen standardielementtien määrittelyyn sekä toimialakohtaisten komponenttien ja elementtien erikoistamiseen näistä.

Terveydenhuoltoalan HL7-standardi on hyvä esimerkki pitkäkestoisesta ja laajamittaisesta toimialakohtaisesta sanastotyöstä. HL7:n yleinen viitemalli on UML-kielillä kuvattu RIM (Reference Information Model). RIM:ssä määritellään ydinluokat, jotka kuvaavat ”koko maailman” abstraktilla, yleisellä tasolla.

Eri sovellusalueet erikoistavat RIM:stä sovellusaluekohtaiset abstraktit mallit sekä edelleen konkreettisemmat hierarkkiset ja sanomatason kuvaukset. Malli pyrkii mahdollistamaan sovellusalueiden välisen semanttisen yhteentoimivuuden määrittelemällä kaikki käsitteet abstraktilla toiminnallisella tasolla, josta kaikki sovellusaluekohtaiset ja konkreettiset kuvaukset johdetaan.

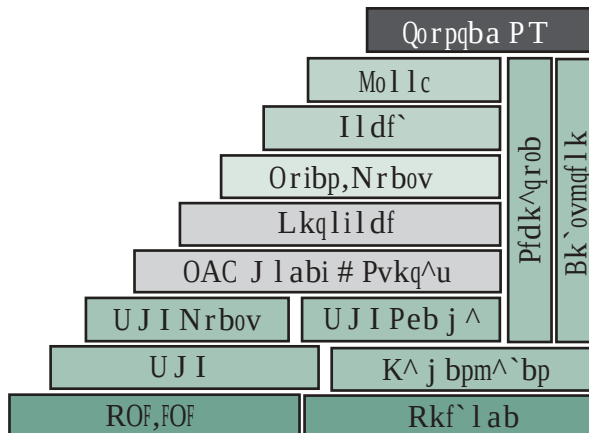
5.3 Semanttinen web

Semanttinen web on W3C:n (World Wide Web Consortium) mukaan nykyisen webin laajennus, jossa tiedolle pyritään antamaan automaattisesti tulkittavissa oleva merkitys. Tavoitteena on, että semanttisen webin käyttäjiä ovat ihmisten ohella koneet. Käytännössä W3C kehittää XML:ään pohjautuvia määrittämiä, joiden avulla WWW:n sisältöjä ja verkkopalveluja voidaan esittää monipuolisemmalla tavalla kuin nykyään. Tällä tavoin internet-sovelluksista pyritään tekemään ”älykkäämpiä”. Älykkyydellä tarkoitetaan koneiden kykyä päätellä asioita, mihin on perinteisesti tarvittu ihmistä.

Semanttisen webin perustietomalli on suunnattu graafi, joka on tarkoitettu resurssien (käsitteiden) ja niiden välisten suhteiden kuvaamiseen. Tässä Resource Description Framework (RDF) -mallissa käsitteiden välisten suhteiden esittäminen on XML:n hierarkista esitystapaa ilmaisuvoimaisempi. RDF:n tietomallin mukaisesti kuvatut käsitteiden väliset suhteet ovat välittömästi koneiden käsiteltävissä.

Semanttisen webin standardointityö etenee pinomaisesti uusia määrittämiä kehittämällä (kuva 2). Pinon alempia tasoja voidaan pitää ylempien kerroksien mahdollistajina; esimerkiksi RDF-tietomallin mukaisia kuvauksia voidaan koodata XML-syntaksin mukaisesti.

Semanttisen webin toteutumisen kannalta keskeisessä asemassa ovat ontologiat. Ontologia on formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä. Ontologiassa määritellään jonkin kohdealueen termit ja käsitteet sekä niiden välisiä suhteita. Käytännössä näin syntyy kohdealueen terminologinen käsittehierarkia. Ontologioiden avulla voidaan esittää esimerkiksi julkishallinnon toimialojen (kuten verotus, tulli, terveydenhuolto) käsitteistöjä. Käsitteiden ja ontologioiden formaali ja yksiselitteinen esitysmuoto mahdollistaa sen, että tietojärjestelmät ja sovellukset voivat käsitellä niitä automaattisesti.



Kuva 2. Semanttinen web – teknologiapino (Lähde: W3C, <http://www.w3.org/2003/Talks/0624-BrusselsSW-IH/>, kalvo 26)

Semanttisessa webissä semanttinen yhteentoimivuus toteutetaan kytkemällä käsitteitä eksplisiittisesti ontologiaan ja määrittelemällä ontologiassa kyseisen käsitteen konteksti (esimerkiksi onko Nokia paikkakunta vai yritys). Ontologioissa määriteltyjen käsitteiden välisten suhteiden avulla sovellukset pystyvät tulkitsemaan vastaanottamansa tiedon käsitetasolla (esimerkiksi että puhutaan maantieteellisestä paikkakunnasta, joka on kytketty paikkaontologiaan eikä yritysongologiaan).

Merkittävä semanttisesta webistä saatava hyöty on mahdollisuus mallintaa ja synnyttää uutta tietämystä. Ontologioihin kuvattavilla käsitteiden välisillä suhteilla sekä sääntö- ja päättelykielillä voidaan tehdä loogisia päätelmiä. Näiden päätelmien avulla voidaan eri lähteistä yhdistetystä tiedosta luoda sellaista uutta tietämystä, jota tiedon tuottaja ei ole alun perin ajatellut tai eksplisiittisesti luonut.

Semanttisen webin ontologiat kuvataan ontologiakielillä (RDF Schema ja OWL). Ontologioiden teknisessä kuvaamisessa voidaan käyttää apuna ontologiaeditoreja. On kuitenkin huomattava, että nämä ovat vain teknisiä apuvälineitä. Käytännössä ontologioiden määrittely edellyttää huolellista sanastotyötä. Tietojärjestelmien semanttista yhteentointaa tukevien ontologioiden määrittelyssä on siis huomioitava tässä raportissa esitetyt menetelmät ja periaatteet sanastojen laatimisessa.

6 SEMANTTISEN YHTEENTOIMIVUUDEN VISIO JA KEHITYSPOLKU

Kun tietojärjestelmän kehittäjä aloittaa uuden järjestelmän rakentamisen, ei ole järkevää laatia järjestelmää varten tarvittavaa sanastoa tyhjästä ikään kuin rakennettaisiin ainutkertaista järjestelmää ensimmäistä kertaa. On enemmän kuin todennäköistä, että organisaatio on jo aiemmin tehnyt toisia tietojärjestelmiä (tietokantoja yms.), joita varten on pitänyt tehdä käsiteanalyyskejä ja määritellä käsitteitä. Toisaalta toimialalla saattaa olla käytössä yhteisiä sanastoja, joita kannattaa hyödyntää.

Lisäksi lisääntyvä EU-tasoinen yhteistyö ja raportointivelvollisuus Euroopan unioniin edellyttävät EU-tasoisten yhteentoimivuuden viitemallien huomioimista. Jos jollakin toimialalla on voimakasta EU-tasoista tai maailmanlaajuisia sanastotyötä käynnissä, siihen kannattaa osallistua ja sitä tulee hyödyntää mahdollisuuksien mukaan. Esimerkkejä tällaisista toimialoista ovat Tulli ja terveydenhuolto.

Taulukko 1 havainnollistaa, miten tietojärjestelmäkehittäjän tulisi sovellusalue- ja tietojärjestelmäkohtaisten valintojen sijasta pyrkiä hyödyntämään yleisempiä sanasto- ja käsittemäärittelyitä.

Taulukko 1. Tietojärjestelmien kehittämisen ja sanastojen laatimisen tasorakenne (sanastotyön matriisi)

Q fj f^i^* dfmr j ^q k			VP>	RK,BAR> @Q bL œbdfkd	R? I ^ Ar _ifk @ db
Q fj f^i^* hl eq^fkbk				BI > * pq^ka^œafq	RK,QABA) EI 4
PI sbiir p* ^ir b** hl eq^fkbk					
H^œ^sr r p	Qfbq * gœbpfij Ć	L œ^kfp^œfi	H^kp^iifkbk	BR7k i^g fkbk	J ^fij ^k* i^g fkbk

Alimpana vasemmalla on ruutu, johon sijoittuisivat yksittäiset tietojärjestelmät, joilla ei ole mitään kytkentöjä ko. järjestelmän ulkopuolelle. Jos järjestelmästä siirretään tietoa organisaation muihin tietojärjestelmiin, on järkevää käyttää yhtenäistä käsitteistöä muun organisaation kanssa. Jos tietoja vaihdetaan usean organisaation välillä, on syytä tarkistaa, onko olemassa jotain kansallisen tason standardeja, joita voi hyödyntää käsitteiden määrittelyssä (ja yhtenäistämässä). Standardeja hyödyntämällä sanastotyö on myös vaivatompaa eikä kaikkia kysymyksiä tarvitse ratkaista itse.

Semanttinen yhteentoimivuus on järkevää rakentaa yhteistyössä sellaisten toimijoiden kanssa, jotka toimivat samalla toimialalla tai joiden toimialoilla on paljon yhteistä toimintaa oman toiminnan kanssa. Ja vastaavasti - jos kahdella toimialalla ei ole käytännössä juurikaan vuorovaikutusta, ei ole järkevää kuluttaa resursseja yhtenäistämällä sanastoja varmuuden vuoksi varastoon. Esimerkiksi suomalaisessa TEKES-rahoitteisessa RUBIC-projektissa on tavoitteena rakentaa eri toimijoiden välinen yhteistyöverkosto. Sen lähtökohtana on yritysmaailman toimijoiden yhteentoimivuuden lisääminen, mutta myös julkinen sektori on kehitystyössä mukana (<http://www.tieke.fi/rubic>).

Taulukko 1 kuvaa myös sanastotyön eri tasoja. Sanastotyö on lähtökohtaisesti toimialakohtaista ja vieläpä sovellusala- ja järjestelmäkohtaista toimialan sisällä. Sanastoja laadittaessa tulisi pyrkiä siihen, että tunnistetaan käsitteet, jotka ovat yleiskäyttöisempiä kuin kullakin tasolla (taulukon solussa) käytettävät käsitteet. Esimerkki tällaisesta on 'osoite', joka on toimialariippumaton tieto ja jolle on tarvetta kautta maailman. Käsitteiden määrittelyn tulisi tapahtua niin yleisellä tasolla kuin mahdollista, esimerkiksi siirtyä

toimialakohtaisesta tasosta toimialariippumattomalle tasolle (taulukossa siis suunnata ylhäälle ja/tai oikealle).

Yhteisten toimialarajat ylittävien sanastojen tehtävä on käyttötarpeen mukaan tarjota yhteinen kieli tietojärjestelmien tai verkkopalvelujen yhteentoiminnan mahdollistamiseksi. Tämä puolestaan edellyttää vastaavuuskuvausten (mappausten) kehittämistä toimialakohtaisten ja yhteisten sanastojen välille. Yksi tekninen toteutustapa voi olla erilliset toimialakohtaiset ja yhteistoimintaa tukevat nimiavaruudet. Tietojärjestelmien yhteentoimivuuden voidaan sanoa tällöin rakentuvan ns. semanttisten siltojen rakentamiselle.

Semanttista yhteentoimivuutta tukevien siltojen rakentaminen saa vielä uuden ulottuvuuden, kun joudutaan ottamaan huomioon viranomaisten kansainväliset yhteydet. Tällöin ratkaistavaksi tulee sekä erilaisissa kulttuureissa kehittyneet erilaiset merkitykset samoille käsitteille että eri luonnolliset kielet käsitteiden nimeämisessä ja kuvaamisessa.

Sanastotyön toteuttaminen edellyttää erilaisia etenemis- ja organisointimalleja tilanteen mukaan. Joillakin toimialoilla on käytettävissä melko hyvin resursseja sanastotyöhön sekä kansallisesti että kansainvälisesti. Toisaalta on toimialoja, joilla on varsin rajalliset resurssit käytettävissä.

Yksi tapa organisoiua ovat ns. ”Communities of Practice” -yhteisöt. Tällainen yhteisö muotoutuu jonkin ”käytännön” ympärille ja sitoutuu yhteiseen toimintaan esimerkiksi yhteisen asiantuntemuksen perusteella. Luonteensa mukaisesti CoP voi olla eri toimialojen välistä. Siihen voi osallistua osapuolia sekä julkiselta että yksityiseltä sektorilta. Silällä on yhteinen tavoite, jonka ympärille muotoutuu oma kulttuurinsa, menetelmänsä, tekniikat, kielenkäyttö ja tapa jakaa tietoa. ”Communities of practice” -yhteisöt perustuvat vapaaehtoisuuteen ja epäviralliseen organisoitumiseen ja ne lakkaavat olemasta, kun tarvetta yhteistyöhön ei enää ole.

Koska sanastotyöhön käytettävissä olevat resurssit ovat rajalliset, on tärkeää suunnata työ oikeisiin kohteisiin. Tämä edellyttää, että kansallisesta sanastotyöstä tiedotetaan avoimesti, jotta valmiit sanastot ja käytännöt niiden kehittämiseksi ovat muiden hyödynnettävissä. On myös tärkeää tuntee kansainvälinen yleinen ja toimialakohtainen sanastotyö. Jos kansainväliset sanastot ovat sovellettavissa Suomessa, voidaan välttyä määrittelemästä samoja asioita uudelleen.

Tärkeässä osassa sanastojen yhteisessä kehittämisessä on tunnistaa tietojen käyttötarve – mm. prosessi- ja käyttäjäryhmälähtöisesti. Siten pystytään priorisoimaan käytännön tasolla olennaiset tiedot ja niiden tietolähteet (mukaan lukien perusrekisterit) sekä niitä hyödyntävät järjestelmät. Tällainen näkökulma tukee sanastotyötä ja sen yhteistä kehitystä sekä tietojärjestelmien palvelulähtöisyyden kasvattamista, myös toimialariippumattomasti. Yhtenä esimerkkinä tästä on yhteistyöhanke Väestörekisterikeskuksen (VRK) ja Vantaan terveystoimen välillä. Kyseisessä hankkeessa luotiin terveyskeskuksen ajanvarauksen prosessilähtöisen tietotarpeen pohjalta ajantasainen XML-pohjainen kysely VRK:n väestötietorekisteristä. Tällä rajatulla projektilla haettiin kokemuksia laajempien palvelukokonaisuuksien kehittämiseksi ja sitä kautta sanastotyön yhtenäisen perustan luomista käytännössä.

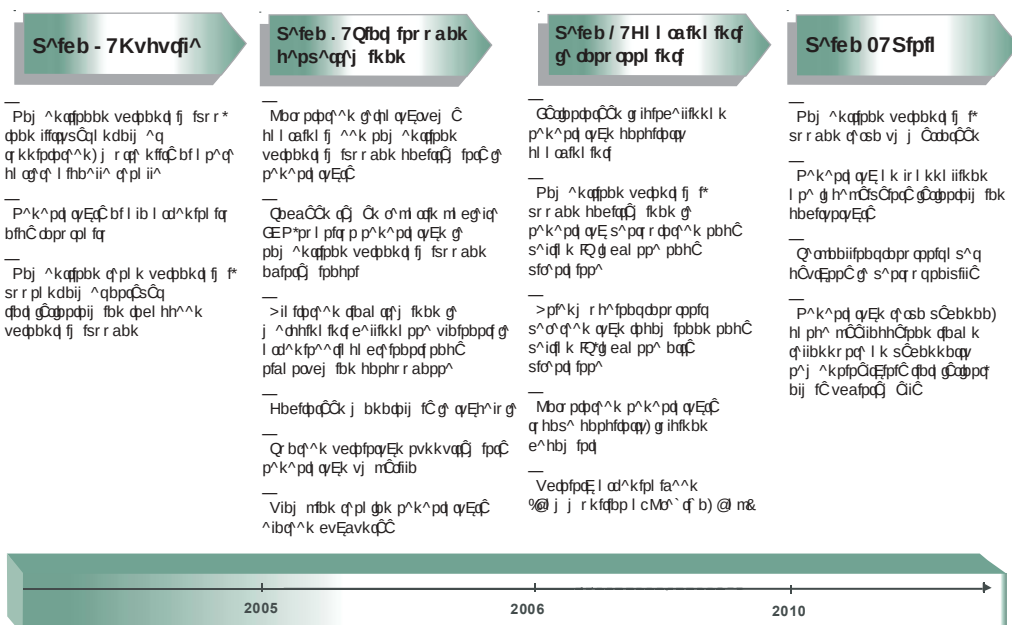
On olemassa riski, että semanttisen yhteentoimivuuden ongelmaa ei tunnisteta tai sitä ei pidetä merkittävänä. Tällöin sanastotyöhön ei panosteta riittävästi, vaan jatketaan kuten tähänkin asti. Toteutuessaan riski tarkoittaa, että resursseja kulutetaan ratkottaessa semanttisen yhteentoimivuuden ongelmia teknisen yhteentoimivuuden tasolla ja samalla hukataan mahdollisuuksia useita osapuolia hyödyttävän yhteentoimivuuden edistämiseksi.

Toinen riski on, että sanastotyö ylimitoitetaan. Sanastotyöstä saattaa tulla itsetarkoitus tai laajoja kansainvälisiä sanastoja sovelletaan kansallisesti ilman todellisen tarpeen analysointia. Tällöin sanastot kasvattavat tarpeettomasti byrokratiaa ja ne tekevät järjestelmäkehityksestä joustamatonta. Näin ne eivät myöskään luo pieniä käytännöllisiä siltoja järjestelmien välille, mikä on tehokkaan sanastotyön edellytys.

Semanttisen yhteentoimivuuden ja sanastotyön kehittämistä ohjaamaan on määritelty *semanttisen yhteentoimivuuden visio*:

- Semanttisen yhteentoimivuuden merkitys organisaatioiden välisessä yhteentoimivuudessa on tiedostettu.
- Semanttisen yhteentoimivuuden kehittäminen ja ylläpitäminen on koordinoitu ja resurssoitu hallinnon eri tasoilla.
- Sanastoja kehitetään ja hyödynnetään suunnitelmallisesti.
- Sanastotyön tarve vähenee, koska päällekkäisen tiedon tallennusta on vähennetty samansisältöisiä tietojärjestelmiä yhdistämällä.

Kuvassa 3 on esitetty semanttisen yhteentoimivuuden kehityspolku kohti vision mukaisia tavoitetilaa. Keskeytymättömän kehityksen kannalta on tärkeää, että välittömästi tämän raportin julkaisemisen jälkeen käynnistetään JHS-suosituksen laatiminen sekä semanttisen yhteentoimivuuden merkityksestä ja sanastotyöstä tiedottaminen laajasti mutta kohdennetusti. Tietoisuuden kasvattamisen rinnalla tulee järjestää semanttisen yhteentoimivuuden kehittämisen ja sanastotyön koordinointi sekä keskitetysti että virastokohtaisesti. Tässä vaiheessa työ vastuutetaan ja varataan riittävät resurssit työlle.



Kuva 3. Semanttisen yhteentoimivuuden ja sanastotyön kehityspolku

Vision mukaisessa tilanteessa sanastotyö on osa jokapäiväistä ja projekteihin luonnollisesti kuuluvaa tietojärjestelmien kehitystyötä. Toisaalta sanastotyön tarve vähenee, koska päällekkäisiä tietojärjestelmiä on voitu yhdistää ja tietoja käytetään laajasti niiden ensisijaisista tuottamis- ja ylläpitopaikoista.

7 SANASTOTYÖ PROSESSINA

Tässä luvussa esitetään, mitä kysymyksiä sanastotyön aloittamiseen liittyy sekä mikä on sanastotyön rooli yleisessä järjestelmäkehityksessä. Lisäksi esitellään menetelmällinen tapa, sanastojen yhtenäistämisprosessi, kohdealuekohtaisten ja yhteisten sanastojen määrittelyyn.

7.1 Liikkeellelähtö sanastotyössä

Nykytilassa semanttisen yhteentoimivuuden tarve usein tunnistetaan, mutta tarpeeseen vastaamiseksi ei ole tarpeeksi tietoa ja välineitä. Käytännön järjestelmäkehityksessä esiin nousevia tyypillisiä kysymyksiä ovat muun muassa:

- Mikä on sanastotyön tarve tässä tilanteessa?
- Mistä tiedän, että pitää tehdä sanastotyötä eli ongelma liittyy semanttiseen yhteen-
toimivuuteen?
- Mistä pitäisi lähteä liikkeelle?
- Onko valmiita sanastoja käytettävissä?
- Miten tunnistan toisella nimellä tehtyjä valmiita sanastoja (metatietomäärittely, ku-
vailuopas, asiasanasto, viestirakenne)?
- Mitä työkaluja ja menetelmiä on käytettävissä?
- Miten kerron omasta työstäni ja kehitetyistä sanastoista?

Seuraavia kysymyksiä voi pitää eräänlaisena järjestelmäkehittäjän tai projektipäällikön tarkistuslistana siitä, millaisia asioita pitää arvioida ja selvittää, jotta semanttisen yhteen-
toimivuuden tarve tunnistetaan ja osataan lähteä oikealla tavalla liikkeelle. Listaa lukies-
sa on hyvä käyttää apuna sanastotyön matriisia (Taulukko 1, luku 6) valmiiden sanasto-
jen löytämiseksi ja oikean etenemissuunnan valitsemiseksi.

- Onko tarve integroitua organisaation muihin tietojärjestelmiin?
- Onko tarve integroitua organisaation ulkopuolisiin tietojärjestelmiin?
- Onko tarve EU-tasoiseen integraatioon?
- Onko omassa organisaatiossa valmiita sanastoja?
- Onko oman organisaation sanasto kehitetty itse vai onko käytetty yleisempää sovellusalue- tai toimialakohtaista viitemallia?
- Onko keskeisillä sidosryhmillä valmiita sanastoja?
- Löytyykö sovellusalueelta tai toimialalta kansallisia tai kansainvälisiä sanastoja?
- Kenen kanssa pitäisi lähteä yhtenäistämään sanastoja?

Sanastotyön liikkeellelähdössä on kaksi päävaihtoehtoa:

1. valmiiden sanastojen hyödyntäminen (top-down)
2. uusien sanastojen luominen ja yhtenäistäminen (bottom-up).

Pääsääntöisesti pyritään aina valmiiden sanastojen hyödyntämiseen (vaihtoehto 1), mikäli niitä on käytettävissä.

Tilanteesta riippuen joudutaan kuitenkin harkitsemaan, kumpi lähestymistapa on sopivampi. Silloin kun tietojärjestelmän on lähtökohtaisesti oltava kansainvälisesti yhteensopiva, vakiintuneita käsitteitä on löydettävissä, eivätkä vanhat järjestelmät sido liikaa kehitystä, top-down-periaatetta kannattaa suosia.

Sen sijaan sellaisissa kansallisissa tilanteissa, joissa olemassa olevat järjestelmät asettavat kehitystyölle reunaehdoja ja tehdään pienempimuotoista sovitustyötä, bottom-up-menettely soveltuu paremmin. Bottom-up-menettelyssä tulee edetä yhtenäisesti ja systemaattisesti ja siten kehittää ”alhaalta päin” yleiskäyttöisempiä sanastoja, jotka mahdollistavat puolestaan top-down-tyylinen etenemisen jatkossa.

Sanastotyössä on usein lähdettävä liikkeelle siitä huolimatta, että kansainvälisiä tai kansallisia valmiita sanastoja ei olisikaan olemassa. Käytännön tarve ohjaa tekemään sanastotyötä pienin askelin siellä missä tarve on ajankohtainen. Tarvelähtöisyyden pitääkin olla ehdottomasti tärkein sanastotyötä ohjaava voima. Sanastotyön organisointi pelkääntään yleisten sanastojen luomiseksi on tavallisesti resurssien tuhlaamista.

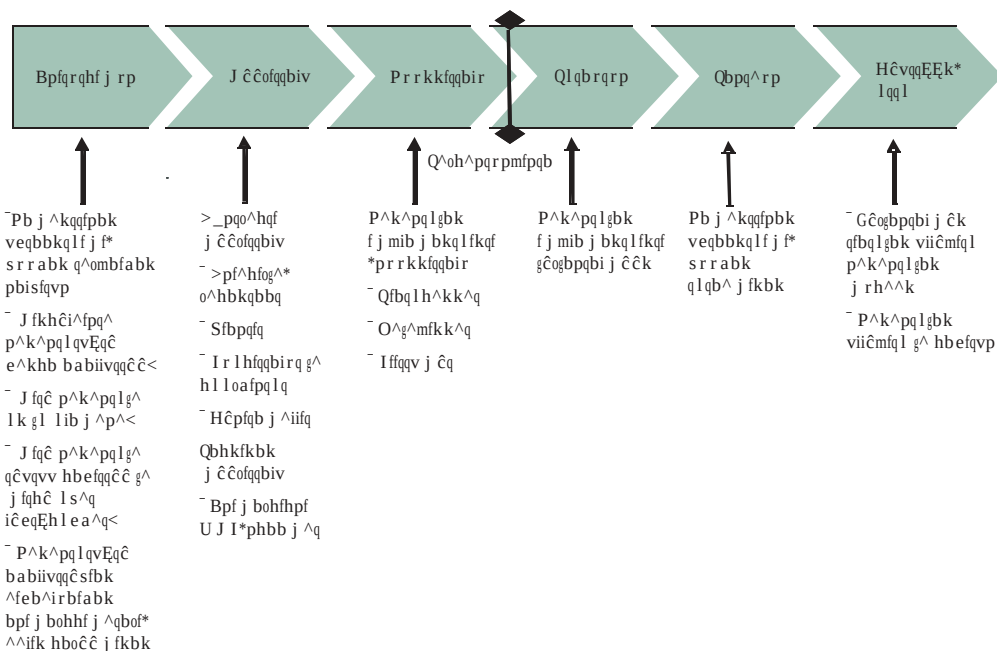
Käsitteiden yleistäminen kannattaa tehdä vain sellaisten käsitteiden osalta, jotka edistävät tietojärjestelmien yhteentoimintaa. Tapauskohteisesti on analysoitava mikä on käsitteiden harmonisoinnin tarve. Käytännön sanastojen määrittely tapahtuu siten projekteissa ja osana järjestelmäkehitystä. Sanastotyön koordinointi edellyttää kuitenkin työn hallinnan osittaista keskittämistä. Tarvitaan keskitetty julkinen paikka, josta kehitetyt kohdealuekohtaiset ja yhteiset sanastot ovat löydettävissä. Lisäksi tarvitaan koordinoitiryhmä, jonka tehtävänä on tiedottaa olemassa olevista sanastoista, edistää niiden käyttöä ja toimia päällekkäisen sanastotyön välttämiseksi.

Erityisesti julkishallinnossa poikkihallinnollisten sanastojen luonnissa on huomioitava, että sanastotyön organisointi riippuu toimialojen ja sillä toimivien organisaatioiden mahdollisuuksista osallistua sanastotyöhön. Kansallisten poikkihallinnollisten sanastojen

yhtenäistämistyön tulisi perustua joustavaan ja kevyeen menettelyyn, koska näin varmistetaan parhaiten pienistä askelista oppiminen ja resurssien edullinen kohdentaminen.

7.2 Sanastotyö osana järjestelmäkehitystä

Sanastotyön tulisi olla osa jokapäiväistä järjestelmäkehitystä. Kuvassa 4 on esitetty sanastotyön sijoittuminen tyypillisiin järjestelmäkehityksen vaiheisiin. Olennaista on, että vältetään sanastotyön tekeminen toteutusvaiheessa kiireellä ja ohjelmistosuunnittelu-vetoisesti.



Kuva 4. Sanastotyö osana järjestelmäkehitystä

Esitutkimuksessa voidaan käyttää apuna sanastotyön matriisia sekä yllä esitettyä tarkistuslistaa. Kehityspolussa esitetään sanastotyötä tukevan keskitetyn julkisen hakemiston perustamista. Tällainen hakemisto olisi hyödyllinen valmiiden sanastojen etsimisessä esitutkimusvaiheessa.

Järjestelmän määrittelyvaiheessa voidaan hyödyntää valmiita analysointimenetelmiä sanastojen luomisessa ja yhtenäistämisessä. Seuraavassa aliluvussa esiteltävää sanastojen yhtenäistämisprosessia voidaan hyödyntää sanastojen määrittämisessä. On tärkeää hu-

mata, että määrittely jakaantuu abstraktiin ja tekniseen määrittelyyn. Sanastojen tulee olla lähtökohtaisesti teknologiariippumattomia. Abstraktissa määrittelyssä käsitteet kuvataan teknologiariippumattomalla tavalla. Teknisessä määrittelyssä sanastoista voidaan johtaa tekniset kuvaukset, esimerkiksi XML-skeemat. Teknisten kuvausten yhdenmukaista tuottamista varten tulee luoda tekniset toteutusohjeet.

Suunnittelu- ja toteutusvaiheissa suunnitellaan ja rakennetaan tekniset ratkaisut sanastojen käytölle tapauskohtaisesti järjestelmän tarkoituksen mukaan. Suunnittelusta ei saa edetä toteutusvaiheeseen, jos sanastotyö on edelleen kesken.

Testausvaiheessa on tärkeää huomata, että teknisen yhteentoimivuuden testaaminen ei riitä, vaan on pystyttävä testaamaan myös semanttinen yhteentoimivuus. Käytännössä tämä merkitsee sitä, että toimintaprosessiin osallistuvat järjestelmät toimivat yhteen oikein todellista toimintaa simuloivilla testitapauksilla.

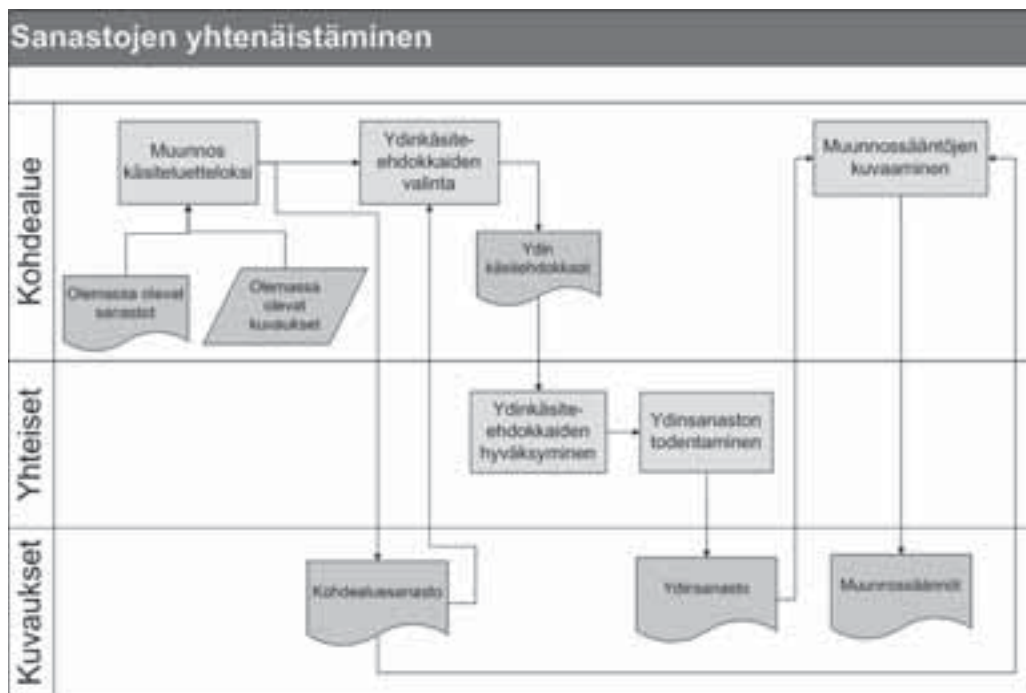
Periaatteessa semanttisen yhteentoimivuuden tason käsitteiden pitäisi pysyä melko vakaina ajan mittaan. On kuitenkin varauduttava uusiin käsitteisiin ja sen kautta laajentuviin sanastoihin. Käyttöönotto- ja ylläpitovaiheessa on huolehdittava järjestelmän tietojen ylläpidosta sekä sanastojen kehityksestä ja ylläpidosta.

7.3 Sanastojen yhtenäistämisprosessi

Nykytilanteessa sanastotyö on monilla alueilla vasta alussa. Tämä tarkoittaa sitä, että tyyppillinen sanastotyön lähtötilanne on sanastotyön matriisin (ks. taulukko 1, luku 6) vasen alareuna: valmiita kohdealuekohtaisia tai yhteisiä sanastoja ei ole ehditty määritellä tai ainakaan valmiita sanastoja ei tunneta tarpeeksi hyvin. Järjestelmäkehityksen kannalta on kuitenkin tarve yhteisten sanastojen avulla päästä parempaan yhteentoimivuuteen toisten tietojärjestelmien ja organisaatioiden kanssa.

Näissä tapauksissa menetelmällinen sanastojen yhtenäistämisprosessi auttaa ensimmäisten askelten ottamisessa kohti kohdealuekohtaisia ja yhteisiä sanastoja. Systemaattinen prosessi mahdollistaa yhdessä kokemuksesta oppimisen. Sekä kohdealuekohtaisten että yhteisten sanastojen laajenemisen myötä on mahdollista synnyttää sanastotyön ympärille ”Communities of Practice”-tyyppisiä yhteisöjä, joilla on yhteinen intressi sanastojen kehittämiseen.

Sanastojen yhtenäistämisprosessi on Rekisteripoolin sanastotyöhankkeissa sovellettu menetelmä, jota voidaan käyttää järjestelmä- ja organisaatiokohtaisten käsitteistöjen kuvaamisessa sekä yhteisten yhteentoimivuuden mahdollistavien käsitteistöjen tuottamisessa. Prosessin tavoitteena on tehdä verkkopalveluiden ja järjestelmien yhteistoiminta mahdolliseksi auttamalla yhteisten ydinkäsitteiden löytämisessä eri kohdealueiden tietojärjestelmistä. Tavoitteena on ollut kehittää formaali prosessi, joka helpottaa sanastojen määrittelyä, nopeuttaa liikkeellelähtöä sekä auttaa osallistujia hahmottamaan sanastojen määrittelyyn kuuluvat tehtävät ja tulokset. Kuvassa 5 on esitetty kaaviokuva prosessista.



Kuva 5. Sanastojen yhtenäistämisprosessi

Prosessin täsmällisempi kuvaus löytyy Rekisteripoolin sivuilta osoitteesta: http://www.rekisteripooli.fi/sana04/Sanastoprosessi_luonnos_v09.doc.

Sanastojen yhtenäistämisprosessi on kehitetty erityisesti julkishallinnon kohdealueille. Prosessissa on huomioitu julkishallinnon budjetointimalli, joka lähtökohtaisesti aiheuttaa sen, että eri osapuolilla on erisuuret resurssit osallistua sanastotyöhön. Prosessi mahdollistaa sen, että organisaatiot voivat osallistua työhön resurssiensa mukaan ja tulla mukaan joustavasti työn edetessä. Sanastotyöprosessi on suunniteltu täyttämään seuraavat vaatimukset:

- teknologiariippumattomuus
- joustavuus ja laajennettavuus
- kustannustehokkuus.

Sanastojen yhtenäistämisprosessia on sovellettu Rekisteripoolin kahden sanastopilotin lisäksi Museoviraston Suomen Museot Online -selainjärjestelmän (<http://www.suomenmuseotonline.fi>) määrittelyssä.

8 SANASTOTYÖN JA SEMANTTISEN YHTEENTOIMIVUUDEN KEHITTÄMISEN KOORDINOINTI VALTIONHALLINNOSSA

Valtion tietohallintoa kehitetään ns. TIME-työryhmän esitysten mukaisesti yhtenäisempään suuntaan. Valtion tietohallintoa kehittämään perustetaan IT-toiminnan johtamisyksikkö. Yksikön tehtäviin kuuluu muun muassa IT-toimintaan liittyvien standardien ja arkkitehtuurien kehittäminen.

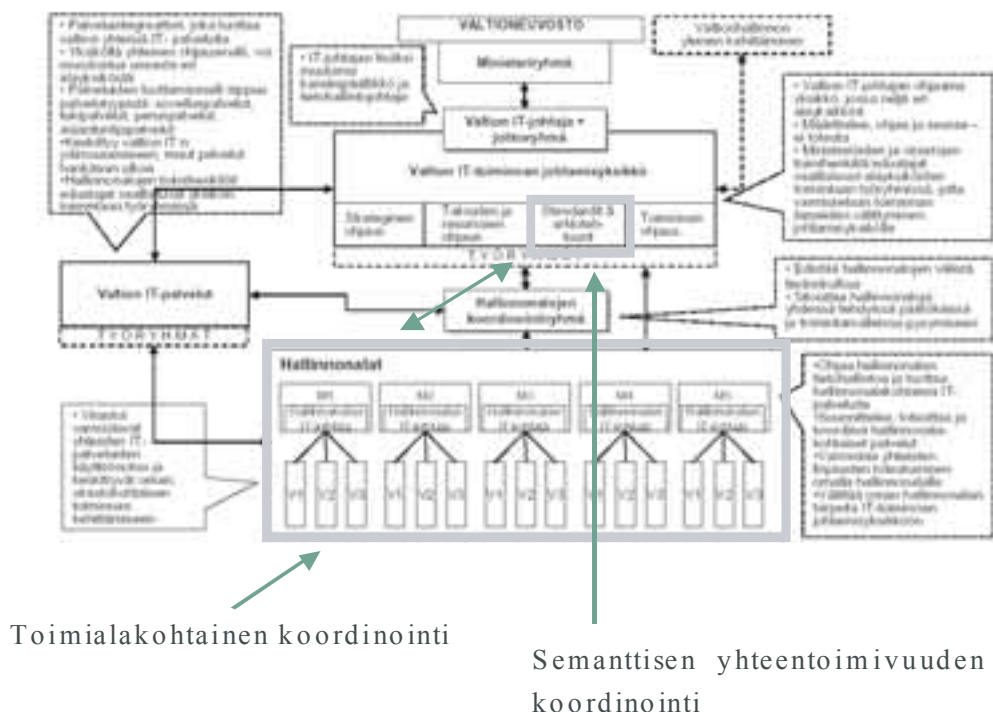
Yhteentoimivuuden kaikki osa-alueet kuuluvat edellä mainittuun tehtäväkenttään, joten on luonnollista, että kyseiseen tehtäväalueeseen liitetään myös semanttisen yhteentoimivuuden kehittäminen. Tehtäväalue tarvitaan myös, jotta EU:n tekemälle kehittämiselle saadaan kansallinen kontaktipiste. IT-johtamisyksikön on myös koordinoitava kansallisia semanttisen yhteentoimivuuden tukitietojärjestelmiä ja muunnosliittymiä yleiseurooppalaisiin järjestelmiin. Tehtäviin kuuluu lisäksi yleisen semanttisen yhteentoimivuuskehityksen seuraaminen ja yhteistyöverkostojen ylläpitäminen. Kuvassa 6 on esitetty ehdotus, miten semanttisen yhteentoimivuuden kehittäminen voidaan koordinoida valtion tietohallinnossa.

Kun IT-toiminnan organisointi hallinnonaloillakin pyritään saamaan konsernimaiseksi, on jokaisella hallinnonalalla resurssoitava semanttisen yhteentoimivuuden kehittäminen. Hallinnonalojen sanastojen ei tarvitse olla täsmälleen samoja kuin valtion yhteisten sanastojen. Riittää, että ne ovat keskenään yhteentoimivia.

Useiden tietojärjestelmien sisältämien tietojen hyödyntäminen edellyttää sanastojen kehittämisen ja ylläpidon vastuutusta sekä asianmukaisten sanastojen käyttämistä. Esimerkiksi perusrekisterien tulisi vastata kohdealueidensa sanastojen ylläpidosta ja tarjota ajantasaisinta ja luotettavinta tietoa yhteiskunnan perusyksiköistä (esim. VTJ: henkilötiedot, KTJ: kiinteistötiedot). Järjestelmien rakentamisessa on huomioitava tietojen ja

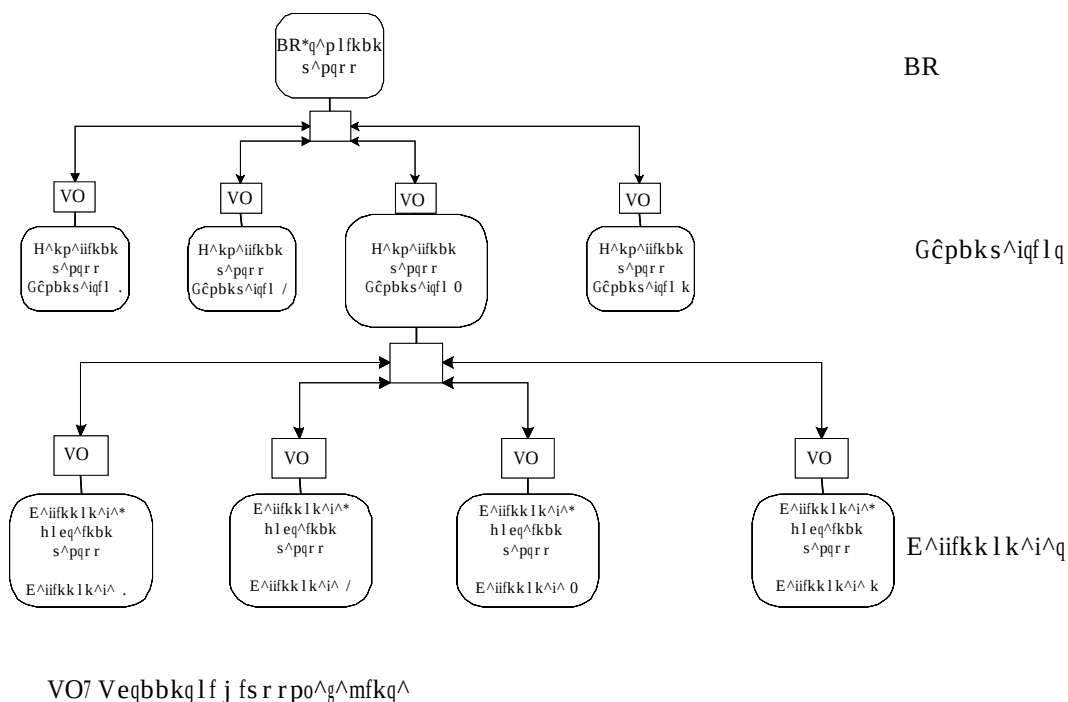
sanastojen omistajat, jotta ensisijaista tietoa hyödynnetään mahdollisimman tehokkaasti ja päällekkäistä sanastotyötä pystytään välttämään.

Alemmilla hierarkiatasoilla semanttisella yhteentoimivuudella on oltava selkeät omistajat, jotta yksittäisissä hankkeissa osataan käyttää hyödyksi semanttisen yhteentoimivuuden kehittämiseksi tehtyä työtä.



Kuva 6. Sanastojen ja semanttisen yhteentoimivuuden koordinointi valtion tietohallinnossa

Semanttisen yhteentoimivuuden kehittämisen toimialakohtaiset vastuukysymykset kuuluvat kunkin toimialan/vastuualueen vastuuviranomaiselle, joka huolehtii yhteydenpidosta ja yhteentoimivuudesta EU:n ja kansainvälisiin järjestelmiin. Kotimaassa vastuuviranomainen koordinoi vastuualueensa osalta semanttista yhteentoimivuutta. Kuva 7 havainnollistaa, miten semanttisen yhteentoimivuuden suhteet ja vastuut jakautuvat hierarkkisesti EU-tasolla, jäsenvaltioissa sekä hallinnonaloittain valtion sisällä. EU-tasolla määritelty hallinnon toimialojen jäsenyys on esitetty kuvassa 8.



**Kuva 7. Semanttisen yhteentoimivuuden suhteet ja vastuutus hallinnon eri tasoilla (mu-
kaillen: EU e-Government metadata framework)**

Kehittämistyössä on syytä tehdä yhteistyötä julkishallinnon sidosryhmien kanssa. Si-
dosryhmiä ovat ainakin julkishallinnolle tietojärjestelmiä toimittavat tahot sekä julkishal-
linnon kanssa yhteistoiminnassa toimivat organisaatiot.

European Union: Activities of the administration

- Law and justice
- External relations
- Trade – competition
- Economy – finance
- Social
- Information – education – culture – sport
- Agriculture – forestry – fisheries
- Industry – enterprise – services
- Energy
- Transport
- Environment – ecology
- Scientific and technical research
- Statistics

Kuva 8. Hallinnon toimialoja (EU)

9 TERMINOLOGIA

Communities of practice, CoP

- Ryhmä ihmisiä tai organisaatioita, jotka kokoontuvat jakaakseen tietoja ja oppiakseen toisilta omassa työssään tarvitsemiaan tietoja. Tällaiset ryhmät tarjoavat sosiaalisen ympäristön tietojen jakamiseen ja perustuvat asiantuntemukseen ja vapaaehtoisuuteen. Esimerkiksi terveydenhuoltoalan HL7.

DTD - Document Type Definition/Document Type Declaration

- Formaali määrittelytapa XML-sanomien ja -dokumenttien rakenteen ja sisällön kuvaukseen. Alunperin Standard Generalized Markup Language (SGML) –kielen määrittelytapa.

Dublin Core

- Kansainvälinen, erityisesti digitaalisten julkaisujen kuvailuun tarkoitettu metatietformaatti.

ebXML- Electronic Business using eXtensible Markup Language

- UN/CEFACT:n ja OASIS:n rahoittama standardointityö, jossa luodaan määrittelyksiä yritysten väliseen elektroniseen kaupankäyntiin. ebXML koostuu modulaarisesti määrittelyksistä, jotka ovat itsenäisiä ja riippumattomia toisistaan.

EIF- European Interoperability Framework

- IDA:n (Interchange of Data between Administrations) julkaisema viitekehys, joka pyrkii linjaamaan sähköisen hallinnon ja sähköisten asiointipalvelujen yhteentoimivuutta EU:n jäsenmaiden välillä. Viitekehyksessä annetaan sähköisten asiointipalvelujen kehittämiseen suosituksia ja ohjeistuksia, joihin perustuen julkishallinto, yritykset ja kansalaiset voivat käyttää näitä palveluja yli kansallisten rajojen.

ELA-standardit

- ELA:n (European Logistics Association) määrittelemiä logistiikka-alan käsitteitä.

eOrdering

- EU:n laajuinen sähköisten hankintojen tilausvaiheen käsitteiden määrittely.

HL7- Health Level 7

- Alunperin terveydenhuollon sovellusten väliseen tiedonsiirtoon tarkoitettu sovellustason sanomaprotokolla.
- Terveydenhuoltoalan tietojärjestelmien yhteentoimivuuden standardoinnista vastaava organisaatio.

HTTP - Hyper Text Transfer Protocol

- World Wide Webin (WWW) tiedonsiirtoprotokolla

Käsite (concept, reference)

- Ihmisen mielessä (tai tietojärjestelmässä tai tietomallissa) oleva mielikuva tai ajatusmalli, joka vastaa tiettyä reaali maailman kohdetta (tarkoitetta).
- Käsite kuvataan määritelmän avulla, esimerkiksi luettelemalla sen ominaisuudet (merkityspiirteet; attributes, attribute values) ja kuvaamalla sen suhteet (relations) toisiin käsitteisiin
- Käsite voidaan ilmaista esimerkiksi luonnollisen kielen sanalla (puhuttuna tai kirjoitettuna) tai jollain symbolilla (asiasana, luokituskoodi) tai terminologisessa sanastotyössä ➤ termillä

Käsitteellinen malli (conceptual model)

- Jonkin aihe- tai kohdealueen (domain) malli, joka määrittelee ko. kohdealueen oliot (entities) ja niiden väliset suhteet, ominaisuudet (attribute/property) sekä olioiden ja niiden ominaisuuksien arvot, ja näiden lisäksi mahdollisesti myös säännöt, joilla oliot, suhteet ja ominaisuudet yhdistetään toisiinsa

Luokitus (classification)

- Arkielämässä kohteiden ryhmittely (usein hierarkkisiin) ryhmiin, joista voi tehdä yleistyksiä; muuten luokitukselle on ominaista systemaattinen esitystapa ja luokkien kuvaaminen symboleilla (notaatioilla)

Metatieto (metadata)

- ”Tietoa tiedosta”; kuvaa kohteena olevan tiedon tarkoituksen tai merkityksen jollain sovitulla kuvaustavalla; kuvailutapa riippuu kuvailun tekävän yhteisön tarpeista ja metatiedon käyttötarkoituksesta (esimerkiksi tietokannan kentän nimi, XML-tunnisteen nimi); metatietoa voidaan käyttää esimerkiksi kohteen tunnistamiseen, kuvailuun, hallintaan, käyttöön, tallennukseen ja metatietojen hallintaan.

OASIS - Organization for the Advancement of Structured Information Standards

- Kansainvälinen voittoa tuottamaton yhteisö, jonka tavoitteena on tuottaa, kehittää ja soveltaa erilaisia sähköisen kaupankäynnin standardeja; <http://www.oasis-open.org>

Ontologia (ontology)

- Formaali, eksplisiittinen määrittely yhteisestä käsitteistöstä. Ontologiassa määritellään jonkin kohdealueen termit ja käsitteet sekä niiden välisiä suhteita.

Organisatorinen yhteentoimivuus (organisational interoperability)

- Liiketoimintatavoitteiden ja –prosessien määrittely ja mallintaminen siten, että organisaatiot voivat vaihtaa keskenään tietoa siitä huolimatta, että niiden organisaatio-rakenne ja toimintatavat voivat poiketa toisistaan. Käyttäjäyhteisön tarpeet pyritään täyttämään tarjoamalla palveluita ja tekemällä ne helposti löydettäviksi ja helppo-käyttöisiksi.

➤ Yhteentoimivuus

OWL – Web Ontology Language

- ➤ RDF Schema –kieltä laajentava määrittely, jonka avulla pystytään monipuolisem-paan luokkien ja ominaisuuksien kuvaamiseen

RDF / RDF Schema – Resource Description Framework

- RDF määrittelee tietomallin, jonka avulla voidaan kuvata (konelukuisessa muodos-sa) minkä tahansa resurssin (asiakirja, kuva, verkkopalvelu jne.) metatietoja.

Sanasto (vocabulary)

- Luettelo jossain kielessä sallituista sanoista (elementeistä) luokitteluineen, määritel-mineen, kuvauksineen ja esimerkkeineen

Semantiikka, semanttinen (semantics, semantic)

- Merkitys, (sanojen) merkityksen määritteleminen

Semanttinen yhteentoimivuus (semantic interoperability)

- Semanttisen yhteentoimivuuden tavoitteena on varmistaa, että tietoja siirrettäes-sä mikä tahansa sovellus voi ymmärtää saamansa tiedon täsmällisen merkityksen, vaikka tietoa ei olisi alun perin laadittu kyseistä sovellusta varten. Semanttisen yh-teentoimivuuden avulla tietojärjestelmä voi yhdistää saamaansa tietoa toisista tieto-lähteistä saamiinsa tietoihin ja käsitellä sitä järkevästi.

➤ Yhteentoimivuus

Skeema (schema)

- Määrittely(kieli), jonka sääntöjen mukaisesti elementeistä muodostetaan yksittäisiä (hierarkkisia) dokumentteja tai tietokantarakenteita (esimerkiksi sääntöjen mukaisia XML-dokumentteja tai tietokantatauluja).

S/MIME - Secure/Multipurpose Internet Mail Extensions

- Standardiperhe, joka määrittelee menetelmät, joilla MIME-muotoisia sähköpostiviestejä salataan ja allekirjoitetaan (varmenteiden avulla)

SOA - Service-Oriented Architecture (Palvelulähtöinen arkkitehtuuri)

- SOA on arkkitehtuurityyli, jonka tavoitteena on mahdollistaa vuorovaikutuksessa olevien sovellusten välinen löyhä kytkentä. Tyypillinen SOA:n toteutustapa perustuu WWW:n tekniikoihin ja tiedonsiirtoprotokolliin, kuten XML ja SOAP.

Syntaksi, syntaktinen (syntax, syntactic)

- (Kielen) rakenne, määrittelee miten kielen elementtejä (sanoja) yhdistetään kielen sääntöjen mukaisiksi rakenteiksi (lauseiksi, lausumiksi)
- määrittelee elementtien (sanojen) välisiä suhteita ja tehtäviä

Tarkoite (referent)

- Maailmassa oleva kohde, johon ihmisen mielessä oleva mielikuva tai ajatus (ä käsite) viittaa; tarkoitteet voivat olla konkreettisia (pöytä, auto) tai abstrakteja (vapaus, aika)

TCP/IP - (Transmission Control Protocol/Internet Protocol)

- Internetin perustavan tason määritys: tiedonsiirtoprotokolla, joka kehitettiin, jotta erilaiset tietojärjestelmät voivat vaihtaa keskenään tietoa Internetin välityksellä
- Tekninen yhteentoimivuus (technical interoperability)
- Tekninen yhteentoimivuus kattaa tietojärjestelmien ja palveluiden yhdistämiseen liittyvät tekniset kysymykset, kuten avoimet rajapinnat, tiedon integroinnin, middle-waren, tiedon esittämis- ja siirtotavat sekä käytettävyyden ja tietoturvakysymykset.
- Yhteentoimivuus

Termi (term)

- Käsitteen esittämistapa/nimitys jossain (terminologisessa) sanastossa; esimerkiksi Tietohuollon sanastossa suositellaan tietystä käsitteestä käytettäväksi termiä emojulkaisu eikä termiä isäntäjulkaisu

UBL - Universal Business Language

- OASIS:n tekemä XML-pohjainen määritys sähköiseen kaupankäyntiin. Tavoitteena määritellä avoin kirjasto, joka sisältää yleisiä standardeja liiketoimintasanomien kuvauksia, kuten tilaus tai lasku.

UML - Unified Modelling Language

- Graafinen kuvauskieli, jolla voidaan määritellä, kuvata visuaalisesti ja dokumentoida (hajautetun oliopohjaisen) tietojärjestelmän toimintoja ja rakennetta.

UN/EDIFACT

- YK:n Euroopan talouskomissiossa (ECE) kehitettävä ja ylläpidettävä määrittäminen sanojen yhteiseksi esitystavaksi EDI-järjestelmissä.

UN/TDED (United Nations/Trade Data Elements Directory)

- Muun muassa kaupan, kuljetuksen ja tullialalla käytettävä ISO:n standardi. Luokitteluun ja kooditukseen pohjautuva tietoluettelo.

W3C – World Wide Web Consortium

- Tim Berners-Leen 1994 perustama yritysten välinen yhteisö, jonka tarkoituksena on kehittää WWW:tä varten standardeja; <http://www.w3c.org>

WWW – World Wide Web

- Yksi Internetin peruspalveluista. Web-palvelimille tallennettujen dokumenttien hakeminen selaimella à HTTP-tiedonsiirtoprotokollan avulla.

XML – eXtensible Markup language

- ➤ W3C:n suositus rakenteisen tiedon esittämiseen sähköisessä muodossa.

Yhteentoimivuus (interoperability)

- Tietojärjestelmien ja liiketoimintaprosessien kyky vaihtaa ja jakaa tietoa (dataa) ja tietämystä keskenään (European Interoperability Framework, EIF). Yhteentoimivuus voidaan jakaa kolmeen osa-alueeseen:
- ➤ organisatorinen, à semanttinen ja à tekninen yhteentoimivuus

YSA – Yleinen suomalainen asiasanasto

- Yksikielinen, kaikki tieteen- ja tiedonalat kattava yleissanasto, johon on koottu eri alojen yleisin terminologia. Se on tarkoitettu käytettäväksi kirjojen, artikkeleiden, elektronisen aineiston ja muiden aineistotyyppien indeksointiin sekä tiedonhakuun..

10 LÄHTEET

- Burkett, W. (2004). The myths of “standard” data semantics. URL: <http://www.sys-con.com/xml/articleprint.cfm?id=524>
- CEN/ISSS Workshop Metadata for Multimedia Information – Dublin Core EU e-Government METADATA FRAMEWORK, draft (2004).
- Daconta, M. C., Obrst, L. J. & Smith, K. T. (2003). The Semantic Web. Indianapolis: Wiley Publishing.
- Friesen, N. (2002). Semantic Interoperability and Communities of Practice. URL: <http://www.cancore.ca/documents/semantic.html>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, Vol. 5, 199-220. URL: http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html
- Handbook for Standardization by the Danish XML Committee, Version: 1.0. URL: http://www.oio.dk/files/Standardization_Handbook_v1_20030203.pdf
- Hyvönen E. (2001). Semantic Web - kohti uutta merkitysten Internetiä. URL: <http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/SemanticWebVisio.PDF>
- Interchange of Data Between Administrations, European Interoperability Framework for pan-European e-government services, Version 1.0, November 2004, European Communities. URL: <http://europa.eu.int/ida/servlets/Doc?id=18101>
- Julkishallinnon XML-strategian toteuttaminen: XML-pohjaiset verkkopalvelutekniikat (2004). Helsinki: Valtiovarainministeriö. (Työryhmämuistioita 7/2004.) URL: <http://www.vm.fi/tiedostot/pdf/fi/86188.pdf> (31.5.2004)
- Karlsson, F. (1994). Yleinen kielitiede. Helsinki: Yliopistopaino.
- Nickols, F. (2000). Communities of Practice: Definition, Indicators & Identifying Characteristics. URL: <http://home.att.net/~discon/KM/CoPCharacteristics.htm> (30.11.2004)
- Sanastojen yhtenäistämismenettely (2004). TietoEnatorin ja Profiumin yhdessä kehittämä sanastotyömenetelmä.
- Tietohuollon sanasto (1993). Helsinki: Tekniikan sanastokeskus ja Tietopalveluseura. (TSK 20)

Semanttisen yhteentoimivuuden kipupisteitä – hallinnon esimerkkejä

1 TERVEYDENHUOLTO

Terveystenhuollossa on paljon tietoa, josta suuri osa on dokumenttimuodossa. Tietoa tarvitaan toiminnan ohjaukseen ja hoidon järjestämiseen, varsinaiseen potilaan hoitoon, tutkimustyöhön ja seurantaan (laskutus, tilastointi, hoidon toteutuminen).

Kaikkea tietoa ei voi eikä kannata kuvata rakenteisessa ja semanttisesti yhteensopivassa muodossa, osa on taas pakko kuvata. Osa tietomäärityksistä tulee viranomaismääräyksinä. Näitä ovat etenkin seurantaan käytettävät tiedot. Toiminnan ohjaukseen ja hoidon järjestämiseen liittyvää semanttista yhteistyötä terveydenhuolto on tehnyt etenkin Kuntaliiton ja Stakesin ohjauksessa.

Viime aikoina on keskitytty potilaan hoitoon liittyvien tietojen semanttiseen yhteensopivuuteen. Ongelmana on, että eri koulukunnilla terveydenhuollossa on hiukan eri käsitys siitä, mitkä tiedot pitäisi olla rakenteisena vapaamuotoisen tekstin sijaan. Tämän vuoksi sosiaali- ja terveysministeriö on kansallisessa terveysprojektissa asettanut projektin, joka pyrkii kansallisesti sopimaan potilaan hoitoon liittyvät ydintiedot (core components). Kun tiedot saadaan sovittua, ne on mahdollista kuvata kansainvälisesti yhteensopivalla HL7-standardilla.

Toinen ongelma ovat käytettyjen luokitusten sisällölliset erot eri sairaanhoitopiireissä ja tietojärjestelmissä. Tämä ongelma korostuu etenkin kommunikoitaessa muun julkishallinnon sekä palveluyritysten kanssa. Lisäksi hallinnolliset semanttiset peruskäsitteet poikkeavat eri tahoilla toisistaan.

Erityisongelman semanttisille määrityksille terveydenhuollossa muodostaa lakisääteinen tietojen pitkäaikaissäilytys. Ei riitä, että ollaan semanttisesti yhteensopivia eri osapuolien kanssa tällä hetkellä, vaan yhteensopivuuden pitää toimia nykyisillä määrityksillä osittain myös tulevaisuudessa. Lyhyemmällä aikavälillä sama problematiikka tulee esiin normaalissa potilaan hoidossa, sillä potilaan hoidossa tarvitaan aina myös potilaan sairauskeromushistoriaa. Sama ongelma tulee esiin myös lääketieteellisessä tutkimustyössä.

2 TULLI

Suomen tullissa on sanastotyötä tehty melko hajanaisesti järjestelmäkehityksen, kääntämisen, tietopalvelun tai muun käyttötarkoituksen mukaan. Mm. tietojärjestelmiä uusittaessa on sanastoja pyritty harmonisoimaan, mutta se on toisinaan ollut hankalaa esimerkiksi ulkopuolisia toimittajia käytettäessä.

Tullialalla yleisesti on jo pitkään tehty tietosisältöjen ja sen myötä myös sanastojen harmonisointityötä. Pyrkimys on käyttää samoja käsitteitä kuljetuksen ja kaupan alan sanastojen kanssa, joista tärkein on UN/TDED. Esimerkkinä sanaston käytöstä on Maailman tullijärjestön (WCO) tietomalli, jonka odotetaan tulevaisuudessa yhtenäistävän tullille annettavien ilmoitusten tietosisältöjä.

Alla on kuvattu tyypillisiä semanttiseen yhteentoimivuuteen ja sanastotyöhön liittyviä ongelmia tullialalla.

1 Käytössä ei ole sanastoa, johon voisi liittää viittauksia muihin sanastoihin

Tullialalla lainsäädännön ja tekniikan käyttämät termit ja merkintätavat ovat usein erilaisia. Pitäisi esimerkiksi tietää, että allaolevan luettelon eri kohdissa on kyse samasta asiasta:

- Aktiivisen kuljetusvälineen tunnus ja kansallisuus saavuttaessa
- SAD box 21
- otsikko.KULJETUS.Aktiivisen kuljetusvälineen kansallisuus rajalla / otsikko.KULJETUS.Aktiivisen kuljetusvälineen tunnus rajalla
- TDT[17](8051=11).C222.8453 / TDT[17](8051=11).C222.8212
- borderNationality code / borderIdentity identifier (luokassa TransportMeans)
- identifier identity / identifier registrantionNationality (luokassa TransportMeans)

2 Kaksi samaa tarkoittavaa käsitettä rinnakkain & eri tasoissa

Melkein ikuisuusongelmaksi on muodostunut käsitteiden ”consignment” ja ”shipment” määrittely. Tullimaailmassa on melko yhtenäisesti käytetty käsitettä ”consignment” kuvaamaan sellaista tavaraerää, joka sisältyy yhteen tullille annettavaan ilmoitukseen. Se voi olla osa isompaa lähetystä eli esim. kymmenen paketin lähetyksestä tullataan kolme ja loput seitsemän tullataan myöhemmin tai vaikkapa palautetaan. Se voi myös olla yhdistelmä eri lähetystyyppejä, esim. eri aikoina saapuneet lähetykset tullataan samalla kertaa (jaksoituuksella). Kauppa ja kuljetus ovat käyttäneet tästä käsitteestä nimeä ”customs consignment”. Kaupalla on ollut toisaalta tarve eritellä yhden kaupan (tilaus-toimitus-laskutus) sisältämät tavarat ja kuljetuksella yhden kuljetussopimuksen sisältämät tavarat. Nyt nämä kaksi osapuolta ovat sopineet, että ”consignment” tarkoittaa yhden kuljetussopimuksen sisältämiä tavaroita ja ”shipment” yhden kaupan (/kauppatapahtuman) sisältämiä tavaroita. Tulli (WCO-tietomalli) on päättänyt käyttämään käsitteitä samalla tavalla eli toiveita käsitteiden yhtenäisestä käytöstä on.

3 Ei tarpeeksi suomenkielisiä vastineita

Käännettäessä on joskus tarpeen käyttää rinnalla jotain toista kieltä, esim. englantia. Esimerkkinä vaikkapa pakkauslajikoodit (YK:n talouskomission UN/ECE:n suositus 21), joita tullin asiakkaat käyttävät tulli-ilmoituksissa. Kun niitä käännetään suomeksi, ei voida tehdä eroa käsitteiden bucket, pail tai tub välillä - kaikki ovat ämpäreitä. Sen vuoksi niiden viereen lisätään englanninkielinen nimi, mutta sekään ei anna tarpeeksi lisävalaistusta, koska kyse ei ole oman kielen käsitteestä. Koodiluettelon pitäisi siis sisältää myös kuvia tai ainakin jonkinlaisia tarkempia selitteitä, toisaalta pitkän luettelon läpikäyminen veisi silloin kohtuuttomasti aikaa.

4 Epätarkka/väärä käsite vakiintuu

Uusia käsitteitä valittaessa voidaan ottaa vahingossa epätarkka ilmaisu käyttöön. Kun eurooppalaista tulli-ilmoituslomaketta suunniteltiin 1980-luvun lopussa, päädyttiin käyttämään lomakkeen kohdassa 36 ilmaisua etuuskohdeltu. Ko. kohtaan merkitään koodi sen mukaan, millaista tullin tai muun veron alennusta / poistoa tullattavalle tavaralle sen alkuperämaahan mukaan haetaan. Toisaalta on mahdollista, ettei tavara voi saada mitään etuutta tai sitä ei haeta ja silloinkin lomakkeen kohta on täytettävä asianomaisella koodilla. Oikeampi käsite olisikin tullikohtelu, jota muukin lainsäädäntö käyttää. Etuuskohdeltu on kuitenkin ehtinyt vakiintua eikä sitä saada muutettua vielä pitkään aikaan.

5 Sama asia on nimetty ja kooditettu eri tavalla

Esimerkiksi 'edeltävä asiakirja tuonnissa' ja 'edeltävä toiminta passituksessa' on nimetty erikseen, mutta kyseessä on periaatteessa sama asia (eli missä menettelyssä tavara on ollut ennen siirtymistään seuraavaan). Toisaalta näillä kahdella tiedolla on omat koodistonsa (molemmat on jopa harmonisoitu yhteisötasolla) ja tiedot ovat eri pituisia.

Vanhojen ja uusien ohjelmien erilaisuuksien vertailu on hankalaa, koska vanhoista vieläkin käytössä olevista ohjelmista ei välttämättä löydy minkäänlaista dokumenttia tai ei ainakaan käsitelistää.

6 Samalla sanalla useita eri käyttötarkoituksia

Tullilla on paljon erilaisia ohjelmia, kuten tuonti, passitus, vienti, autoverotus ja valmisteverotus, joilla on samoja termejä. Joskus samoilla nimillä kutsuttavat asiat voivat olla todellisuudessa kaksi eri asiaa.

Esimerkiksi sana tulopäivä saattaa tarkoittaa tavarán (tullialueelle) saapumisen päivää tai sitä koskevan tullí-ilmoitussanomán saapumisen päivää. Järjestelmiä suunniteltaessa on oltava tarkka, kumpaa missäkin tilanteessa tarkoitetaan.

7 Sama tarkoite saa uuden nimen

Siitä materiaalista, johon kuljetettava tavara on pakattu, on mm. tullí-lomakkeilla käytetty ilmaisua "kolli". Eräässä projektissa päädyttiinkin käyttämään sanaa "pakkaus". Tämän lopputuloksena lainsäädännössä ja muussa materiaalissa on pieniä ristiriitaisuuksia.

3 VEROHALLINTO

Verohallinto on käynnistänyt sisäisen sanastotyön edistämään tietojärjestelmien semanttista yhteentoimivuutta. Alla on kuvattu Verohallinnon sanastoryhmän havaitsemia yhteentoimivuutta haittaavia ongelmia.

Käsitteistön ja sanaston hallinta perustuu ajan mittaan muotoutuneisiin käytäntöihin ja vastuuhenkilöiden tietämykseen. Koska tietämys on sovellusalueiden vastuuhenkilöillä, on sovellusten ulkopuolisten tahojen vaikea saada sitä käyttöönsä. Esimerkiksi tietopalvelun on vaikea hahmottaa riittävästi käsitteistöön ja tietovarastoihin liittyviä yksityiskohtia. Myöskään ulkopuolisille tiedontarvitsijoille ei voida antaa riittäviä metatietoja (esim. tutkijat, Tilastokeskus). Muun muassa verohallinnon sähköistämishankkeessa (Taxit) on jo kehitystyön varhaisessa vaiheessa huomattu, että osahankkeisiin syntyy rinnakkaisia käsitteistöjä. Tämä osoittaa yhtenäisen sanastotyön olevan tarpeellista ja kiireellistä.

Täysin manuaalinen ja osittain vapaamuotoinen tapa hallita käsitteistöä voi aiheuttaa ongelmia järjestelmäkehityksen edetessä. Tällaisia ovat:

- Käsitteiden hallinta on hyvin riippuvaista avainhenkilöistä.
- Tietoja on kuvattu eri näkökulmista (Musta kirja kuvaa virkailijan ja tietojärjestelmän kannalta, täyttöohje kuvaa asiakkaan kannalta), mutta yhdistävää kokonaiskuvaa ei ole olemassa.
- Kaikki kuvaukset ovat sellaisessa muodossa, että niitä ei voida hyödyntää koneellisesti (sisällönhallinta, sähköiset asiointipalvelut).
- Järjestelmät tallettavat sanastotietoa eri paikoissa, mikä saattaa johtaa erilaisten termien käyttöön.
- Käsitteiden hallinta on hyvin tietokantalähtöistä, mikä voi vaikeuttaa niiden käyttöä esimerkiksi prosessitasolla.
- Kuvausten hajanaisuus ja vaihtelevat käytännöt vaikeuttavat järjestelmiin perehtymistä.
- Verohallinnon ulkopuolelle ei voida julkaista yhtenäistä tietoa saatavissa olevista tiedoista.

Verohallinto on huomattu, että yhtenäinen ja keskitetty sanastotyö on välttämätöntä näiden ongelmien ratkaisemiseksi.

4 YMPÄRISTÖHALLINTO

Alla on kuvattu joitakin Ympäristöhallinnossa havaittuja semanttisen yhteentoimivuuden ongelmia sekä pyrkimyksiä ratkaisuihin.

1 Tietokantojen tietojen yhdenmukainen nimeäminen ja esittäminen

Eri tietokantoihin tallennettujen tietojen sujuva yhdistäminen on mahdollista vain, jos samat asiat on nimetty ja esitetty niissä yhdenmukaisella tavalla. Ympäristöhallinnossa on ollut käytössä jo useita vuosia Hakemistot - apuväline tietokantojen yhdenmukaistamiseksi -ohje. Ohje koostuu seuraavista osioista:

- Tunnisteet-osa sisältää ohjeet tietokantojen taulujen ja sarakkeiden nimeämiseksi.
- Koodit-osa sisältää ohjeet tietojärjestelmissä käytössä olevien koodilistojen luonnista, käytöstä ja ylläpidosta.

Ympäristöhallinnossa on kaksi eri sovelluskehitysyksikköä - toisessa näistä ohjetta on noudatettu aika hyvin, toisessa puolestaan ei. Todennäköisesti samankaltaista ohjeiden noudattamatta jättämistä esiintyy yleisesti lähes kaikissa organisaatioissa. Tiukempi johdon ohjaus ja käskytykset saattaisi parantaa tilannetta.

2 Käsitteiden määrittely

Ympäristöhallinnossa on asianhallintajärjestelmän yhteyteen rakennettu hanke-osio, johon tallennetaan erityyppisiin hankkeisiin liittyviä tietoja, erityisesti rahoitustietoja. Osion rakentaminen aloitettiin vesihuoltoavustusvetoisesti. Tämän vaiheen aikana hankkeessa määriteltiin näytöillä esiintyvät käsitteet. Myöhemmin osiota täydennettiin muilla avus-

tustyypeillä. Tuolloin ilmeni ongelmia erityisesti näytöillä olevien kenttien nimien osalta. Samasta tarkoitteesta käytetään eri tilanteissa ja eri käyttäjäjoukoissa eri ilmaisuja. Käyttäjät kokevat sen varsin ongelmallisena. Asia on ainakin väliaikaisesti tarkoitus ratkaista ohjeistuksen avulla. Näiltä ongelmilta olisi välttytty, mikäli määrittelyvaihe olisi kattanut koko tulevan kentän.

3 Luokitukset ja kooditukset

Ympäristöhallinnon tietojärjestelmien kannalta ongelmana ovat lajitietoihin liittyvät lajilistat ja niiden jatkuva ylläpito. Tätä tarkoitusta varten ei ole olemassa käyviä kansallisia tai kansainvälisiä luetteloita. Näin ollen eri tietojärjestelmissä ylläpidetään omia lajilistoja, mikä puolestaan tekee eri lähteistä peräisin olevien tietojen yhdistämisen varsin hankalaksi. Käytännön tilanteessa ei auta, vaikka käsite laji olisi kaikkialla yksikäsitteinen, mutta käsitteen ”taakse” tallennetut yksittäisiä lajeja yksilöivät koodit kuitenkin poikkeavat toisistaan. Sama ongelma pätee myös fysikaalis-kemiallisiin parametreihin.

5 MAA- JA METSÄTALOUSMINISTERIÖ

Sanastotyön koordinointiin on koottu Tikessä (Tietopalvelukeskus) asiantuntijatyöryhmä. Työryhmälle on asetettu tavoitteeksi yhdenmukaistaa hallinnonalan käsitteistöä järjestelmien välisen tiedonsiirron ja järjestelmäkehityksen tueksi.

Elintarviketalouden laatutietojärjestelmässä (ELATI) sanastotyö on ydinhankkeena. Hankkeen yhteydessä on aloitettu sanastopilotin suunnittelu. Tavoitteena on sanastotyön mallintaminen ja eri osapuolten tehtävien selkeyttäminen. Sanastotyön tarve on osoittautunut kiireelliseksi esimerkiksi Eurostatille tapahtuvan elintarviketurvallisuustietojen raportoinnissa.

Sanastotyössä joudutaan ratkomaan asiantuntijoiden, tietojärjestelmien rakentajien ja tilastojien yhteistyönä mm. seuraavan ”sikatiedon” ongelmia: Maatilayrityksiltä kerätään lainsäädännön edellyttämiä sikojen lukumäärätietoja kolmeen eri aikoina ja erilaisiin tarpeisiin kehitettyyn järjestelmään ministeriön hallinnonalalla. Tietoja kerätään sikarekisteriä, kansallisia eläinpalkkioita sekä tilastointia varten. Sikojen tiedot eivät ole yhteensopivia eri järjestelmien välillä. Liittymäjärjestelmissä joudutaan ”tilastosikaa” ja ”tukisikaa” käsittelemään toisistaan poikkeavilla tavoilla, vaikka semanttisesti käsitellään samaa sikaa ja samoja tietoja. Vertailut eivät ole suoraan mahdollisia. Sikatietoja välitetään XML-sanomina Kansallisen eläinterveydenhuollon yhteistyötahoille. Tätä varten sikojen tiedot on jouduttu kuvaamaan jälleen uudella tavalla.

6 VÄESTÖREKISTERIKESKUS

Valtakunnallisessa väestötietojärjestelmässä (VTJ) ylläpidetään tietoja henkilöistä, rakennuksista, huoneistoista ja kiinteistöistä. Henkilötietoja ovat mm. henkilötunnus, syntymäaika, sukupuoli, nimi, äidinkieli ja kansalaisuus, vakituista ja tilapäistä asuinpaikkaa koskevat tiedot, holhous- ja huostaanottotiedot, tietojenluovutuskiellot sekä perhesuhdetiedot (avioliitot, lapset).

Väestötietojärjestelmän tietopalvelun asiakkaita ovat valtion ja kunnan viranomaiset, seurakunnat, tieteellinen tutkimus sekä erilaiset yritykset. Väestörekisterikeskuksen tavoitteena on edistää VTJ:n tietojen käyttöä ja käytettävyyttä kaikissa yhteiskunnan toiminnoissa niiden tehokkuuden parantamiseksi, kuitenkin niin, ettei henkilötietojen tietosuoja ja yleinen tietoturvallisuus vaarannu.

Tyypillinen käyttötilanne Väestötietojärjestelmän tietopalvelussa on, että asiakas pyytää henkilöiden perustietoja sekä henkilöiden välisiä suhteita, kuten syntymäkotikunta- ja asuinpaikkatiedot sekä sukulaisuussuhteet. Asiakas tallentaa tiedot tietojärjestelmäänsä tai niitä voidaan käyttää suoraan asiakkaan toimintaprosesseissa. Tavallisesti Väestötietojärjestelmän ja asiakkaan tietojärjestelmien tietomallit ja tietorakenteet poikkeavat toisistaan. Ennen lähettämistä tietoa on usein koostettu ja muokattu poimintaohjelmilla asiakkaiden tarpeiden mukaiseksi.

Poikkeavien tietorakenteiden lisäksi Väestötietojärjestelmän tietomallin sekä käsitteiden ja niiden välisten suhteiden dokumentointiin ulkoisia asiakkaita varten ei ole kiinnitetty aikaisemmin riittävästi huomioita. Tämä on johtanut siihen, että käsitteiden merkitys ja niiden väliset suhteet saatetaan kadottaa asiakkaiden vastaanottamassa tiedossa. Toisin sanoen järjestelmät eivät ole semanttisesti yhteentoimivia.

Väestörekisterikeskus näkee, että tämän tyyppisiä tietojärjestelmien välisen semanttisen yhteentoimivuuden ongelmia voitaisiin vähentää seuraavasti:

- VRK:n ja sen yhteistyökumppanien käytettävissä olisi jatkuvasti ajantasainen dokumentaatio (yhdessä paikassa ylläpidettynä) VTJ:n tietomallista, eri tietojen nimistä ja niiden merkityksestä.

- Tietojen nykyisten ja potentiaalisten tulevien käyttäjien vapaasti tutkittavana olisi vastaavaa informaatiota etukäteistä tutustumista varten (omien prosessien ja järjestelmien määrittelyä silmällä pitäen esim), siinä määrin kuin on mahdollista vaarantamatta VTJ-kokonaisuuden tietoturvaluutta.
- Keskeisiä asiakaskohtaisia tiedonsiirtoratkaisuja koskevaa informaatiota olisi samalla tavalla kootusti dokumentoitu VRK:n ja yhteistyökumppanien käyttöön (myös ao. asiakkaan käytettävissä).
- Väestötietojen tietopalvelussa / tiedon välityksessä yleisesti suositellut tietomallit ja tiedonsiirtoformaatit olisi määritelty ja dokumentoitu ja ne olisivat yleisesti saatavilla.

7 PERUSREKISTERIT

Tietojen ajantasaisuuden ja oikeellisuuden vaatimukset ajavat etsimään keinoja perusrekistereiden tietojen nykyistä laajemmalle yhteiskäytölle. Teknisesti yhteiskäyttö edellyttää sitä, että rekisterien tietojen tulee olla saatavilla palvelulähtöisesti avoimien rajapintojen kautta. Yhteiskäyttö ei voi perustua pelkästään suurten tietomäärien kopiointiin rekisterien ja asiakasjärjestelmien välillä.

Perusrekistereiden yhteenliittymä Rekisteripooli on kahdessa pilottiprojektissa määritellyt yhteispalveluja, jotka yhdistävät useita peruspalveluja. Määrittelyt ovat sisältäneet rekisterien kohdealuekohtaisten sanastojen sekä yhteentoiminnallisuuden mahdollistavan yhteisen ydinsanaston tuottamisen, rakennettavien palvelujen edellyttämässä laajuudessa.

Perusongelma semanttisen yhteentoimivuuden kannalta on, että eri rekistereissä on eri tarkkuustasolla ja eri tavalla nimettyjä käsitteitä. Myös käsitteiden arvoalueet poikkeavat eri rekistereissä toisistaan. Esimerkiksi yhteiseksi sovittua käsitettä ”jakeluosoite” vastaa kohdealueilla Väestötietojärjestelmän ”vakinainen kotimainen osoite” ja ”tilapäinen kotimainen osoite”, Kiinteistötietojärjestelmän ”lainhuudon saajan vakinainen kotimainen lähiosoite” ja kaupparekisterin ”yrityksen käyntiosoitteen katutieto” ja ”yrityksen posti-osoitteen lokero tai katutieto”.

Rekisteripoolin sanastotyössä on otettu periaatteeksi, että työtä ei tehdä varastoon tai oletetun tarpeen mukaan, vaan se kohdistuu konkreettisen ongelman ja kehittämistyön, kuten yhteisen verkkopalvelun rakentamisen, edistämiseen. Tarvelähtöisesti on valittu myös taso, jolla yhteisiä käsitteitä on pyritty muodostamaan. Yhteisten käsitteiden tarkoituksena on mahdollistaa verkkopalveluiden löytäminen ja ketjuttaminen, johon riittää avain- eli yhdistävistä tiedoista sopiminen. Muut käsitteet ovat kohdealuekohtaisia, eli verkkopalvelujen toiminta ei edellytä laajempaa yhtenäistämistä.

Rekisteripoolin sanastotyössä on sovellettu sanastojen yhtenäistämismenettelyä, joka on systemaattinen menetelmällinen tapa kohdealuekohtaisten käsitteiden ja yhteisten ydinkäsitteiden määrittelyyn. Prosessi tarjoaa kevyen ja järjestelmällisen, pieniin etenemisas-

keliin perustuvan työskentelytavan, joka huomioi julkishallinnon organisaatorakenteen ja muut erityispiirteet. Prosessi on osoittautunut joustavaksi ja nopeaksi tavaksi tehdä sanastotyötä. Sen avulla on pystytty minimoimaan byrokratiaa ja hitautta, mitä tyypillisesti liittyy laajamittaisiin useiden osapuolien yhdessä tekemiin sanastohankkeisiin.

8 MUSEOT

Museoiden tehtävänä on tallentaa, järjestää, luokitella ja tarjota yleisön ja tutkijoiden käyttöön kansallista kulttuuriperintöä, kuten esineitä ja valokuvia. Kulttuuriperinnön digitointiryhmä (KULDI) on esittänyt vision, että vuoteen 2010 mennessä keskeinen muistiorganisaatioiden eli kirjastojen, arkistojen ja kirjastojen kulttuuriperintö on digitoitu. Aineistojen tulee olla tietoverkkojen kautta kaikkien ulottuvilla.

Monet museot ovatkin digitoineet aineistoaan ja tuoneet sitä esille omassa Internet-palvelussaan. Tällaisessa tapauksessa käyttäjän eli verkkomuseossa vierailijan ongelmana on, että kussakin verkkomuseossa on käytävä erikseen ja jokaisella niistä on omanlaisensa rakenne ja jäsenitys. Tämän ongelman ratkaisemiseksi ehdotettiin jo 1990-luvun loppupuolella museoiden yhteisen verkkopohjaisen hakujärjestelmän toteuttamista.

Museovirasto käynnisti vuonna 2003 Suomen Museot online -hankkeen (SMOL), jossa rakennettiin eri museoiden kokoelmat virtuaalisesti yhdistävä hakujärjestelmä. Sen avulla museoiden kokoelmia voidaan hakea, selata ja katsella samasta verkko-osoitteesta (<http://www.suomenmuseotonline.fi>).

Projektin ensimmäinen vaihe oli laatia yhteinen ydinsanasto, koska museoilla ei ollut olemassa yhteistä sanastoa. Museot tietenkin käyttävät samanlaista terminologiaa, mutta niiden täsmällinen käyttö ja tulkinta vaihtelee eri museoissa:

- Eri museoissa käytetään erilaisia kokoelmanhallintajärjestelmiä (luettelointijärjestelmiä).
- Eri museoissa on erilaiset luettelointikäytännöt (ja siten myös erilaiset metatiedot ja näillä erilaiset arvoalueet).
- Eri museoissa käytetään erilaisia luokitusjärjestelmiä ja asiasanastoja, muun muassa museon erikoisalan mukaisesti.

Jotta yhtenäinen haku olisi ollut mahdollista, oli eri lähteistä tuleva aineisto ensin yhdenmukaistettava. Yksi mahdollisuus olisi ollut ottaa mukaan vain ne metatiedot, jotka olivat yhteisiä kaikille museoilla. Tällaisia tietoja kuitenkin oli vain vähän – käytännössä

vain tunnistetiedot löytyivät varmasti jokaisen museokohteen luettelointitiedoista. Tällaisen ”pienimmän yhteisen nimittäjän” käyttämisessä on riskinä, että tiedon merkitys ja kuvauksen rikkaus häviää.

Yhteisten ydinkäsitteiden löytämiseksi hankkeessa sovellettiin sanastojen yhtenäistämisprosessia. Tarkastelun kohteeksi otettiin Musketti-, Siiri- ja Antikvaria-kokoelmanhallintajärjestelmien sanastot ja osin Dublin Core (Culture & Simple). Prosessin tuloksena tuotettiin ydinsanasto, jossa oli 20 käsitettä kuvauksineen ja arvoaluemäärittelyineen. Käsittemäärittelyjen perusteella laadittiin myös RDF-skeema ja XML-sanoma (rajapinta).

Ydinsanaston määrittelyn perusteella museot voivat rakentaa sovelluksen, joka poimii tarvittavat tiedot museon luettelointijärjestelmästä ja tuottaa niistä SMOL-rajapintamäärittysten mukaisen XML-tiedoston. Käytännössä tämä sovellus usein on osa museoiden jo käyttämää kokoelmanhallintajärjestelmää – tällainen poimintaohjelma on toteutettu jo Musketti- ja Siiri-järjestelmiin ja on tätä kirjoitettaessa rakenteilla Antikvaria-järjestelmään.

9 YLIOPISTOT

Esimerkkinä yliopistoissa ja muissa oppilaitoksissa esiintyvistä sanastotyön tarpeellisuudesta esitellään lyhyesti opintojaksojen sisällön kuvaamiseen liittyviä ongelmia. Esimerkki perustuu Teknillisen korkeakoulun Viestintätekniikan laboratorion sanastoprojektiin ”Opintojakson sisällön kuvauksen määrittely”, jossa määritettiin yliopistojen yhteistä OODI-järjestelmää ajatellen sanastoehdotelma (http://www.media.hut.fi/~th/OODI/OODI_complete.pdf).

Opintorakenteen määrittely tapahtuu nykyisellään tietyn rakenteen mukaisesti. Rakenne on kuitenkin niin väljästi määritelty, että siihen liitetty sisältökuvaus ei välttämättä anna riittävästi tietoa opintojaksoista eri käyttötarkoituksiin.

Opetusohjelman tietovarantoa käyttävät eri tahot ja eri tarkoituksiin: opintojaksojen vastuuhenkilöt suunnitteluun, opintokokonaisuuksista vastaavat rakenteiden suunnitteluun ja määrittelyyn, hallintohenkilöt esimerkiksi korvaavuuksien määrittelyyn ja opiskelijat tehdäkseen valintoja.

Sanaston käyttämisen tärkein tehtävä on huolehtia siitä, että käytetyt ilmaukset ovat yksiselitteisiä. Jos esimerkiksi jostain opintojaksosta sanotaan, että siellä käsitellään Internetiä, lukijalle on vielä epäselvää, käsitelläänkö opintojaksossa sisältötekniikkaa, protokollatekniikkaa, juridisia asioita vai jotain muuta Internetiin liittyvää.

Varsin usein eri lähteissä oleva tieto on ristiriitaista keskenään ja tällöin on epäselvää, mihin tietoon voi luottaa. Yleisin tapa ratkaista tämä ongelma oli henkilökohtainen yhteydenotto opintojakson mainittuun vastuuhenkilöön.

Opintojaksojen kehittäjät tulevat harvoin ajatelleeksi sitä, että opintojaksoja tulisi mainostaa tai että opintojaksoa etsittäisiin sisällön perusteella tietämättä, onko sellaista opintojaksoa edes tarjolla. Tällaista etsimistä tapahtuu paljon ja uusien hakumahdollisuuksien myötä tapahtuisi varmasti vielä enemmän. Jokaisen opiskelijan tulee sisällyttää tutkintoonsa vähintään 20 opintoviikkoa vapaasti valittavia opintoja. Jos opiskelijalla on pääaineenaan esimerkiksi sisällöntuotanto, hän voisi olla halukas etsimään vapaasti

valittaviin opintoihin internetiin liittyviä kursseja. Vaihtoehtoisia sisältöjä ovat esimerkiksi protokollatekniikat, tietoturva, juridiikka, ohjelmointi, liiketalous... Itse asiassa on hyvin hankalaa suoraan luetella kaikkia mahdollisia internetiin liittyviä opintojaksotyyppejä. Samaan ongelmaan törmäävät myös opintokokonaisuuksien suunnittelijat. Moneen opintokokonaisuuteen sopisi muitakin alaa käsitteleviä opintojaksoja, joiden olemassaolosta suunnittelija ei tiedä.

Ratkaisuna mainittuihin ongelmiin on tietojen sijoittaminen semanttisesti mahdollisimman täydellisenä yhteen järjestelmään. Tällöin opintojaksojen tietoja päivitettäisiin vain yhteen paikkaan ja siten ne olisivat oikeelliset ja tietojen hakija voisi luottaa löytämänsä tietoon. Opintojaksojen etsiminen tietyin hakukriteerein helpottuisi huomattavasti. Ontologian ja siihen liittyvien sanastojen rooli on opintojaksohakujen ja sisältökuvausten ymmärtämisen kannalta merkittävä asia. Mikäli samasta asiasta puhutaan kahdella erilaisella termistöllä, törmätään semanttiseen epätäydellisyyteen, mikä saattaa aiheuttaa väärinymmärryksiä.

Taloudellisesta näkökulmasta ontologia ja sanastot saattavat tuoda hyötyä siten, että opintojaksojen sisällöt eivät olisi helposti päällekkäisiä. Kullakin opintojaksolla voidaan paremmin keskittyä oleelliseen asiaan, kun on selkeästi tiedossa se, mitä muilla opintojaksoilla käsitellään.

VUODEN 2005 AIKANA ILMESTYNEET JULKAISUT

- 1/2005 Valtionhallinnon ylimpien johtamisvirkojen uudistaminen
- 2/2005 Kuuleminen ja osallistuminen tietoverkoissa

TYÖRYHMÄ- MUISTIOITA



VALTIOVARAINMINISTERIÖ
Snellmaninkatu 1 A
PL 28, 00023 VALTIONEUVOSTO
Puhelin: (09) 160 01
Telefaksi: (09) 160 33123
www.vm.fi

3/2005
SANASTOT SILTANA SAUMATTOMALLE
TIEDONVAIHDOLLE

ISSN 0788-6322
ISBN 951-804-489-9 (nid)
ISBN 951-804-490-2 (pdf)