



Semanttinen yhteentoimivuus

Riku Moisio

JHS-seminaari 5.4.2005

"Semanttisen yhteentoimivuuden haaste on siinä, että tietojärjestelmien tulisi ymmärtää automaattisesti teknisellä tasolla ihmisten ja organisaatioiden toiminnasta lähtöisin olevia käsitteitä."

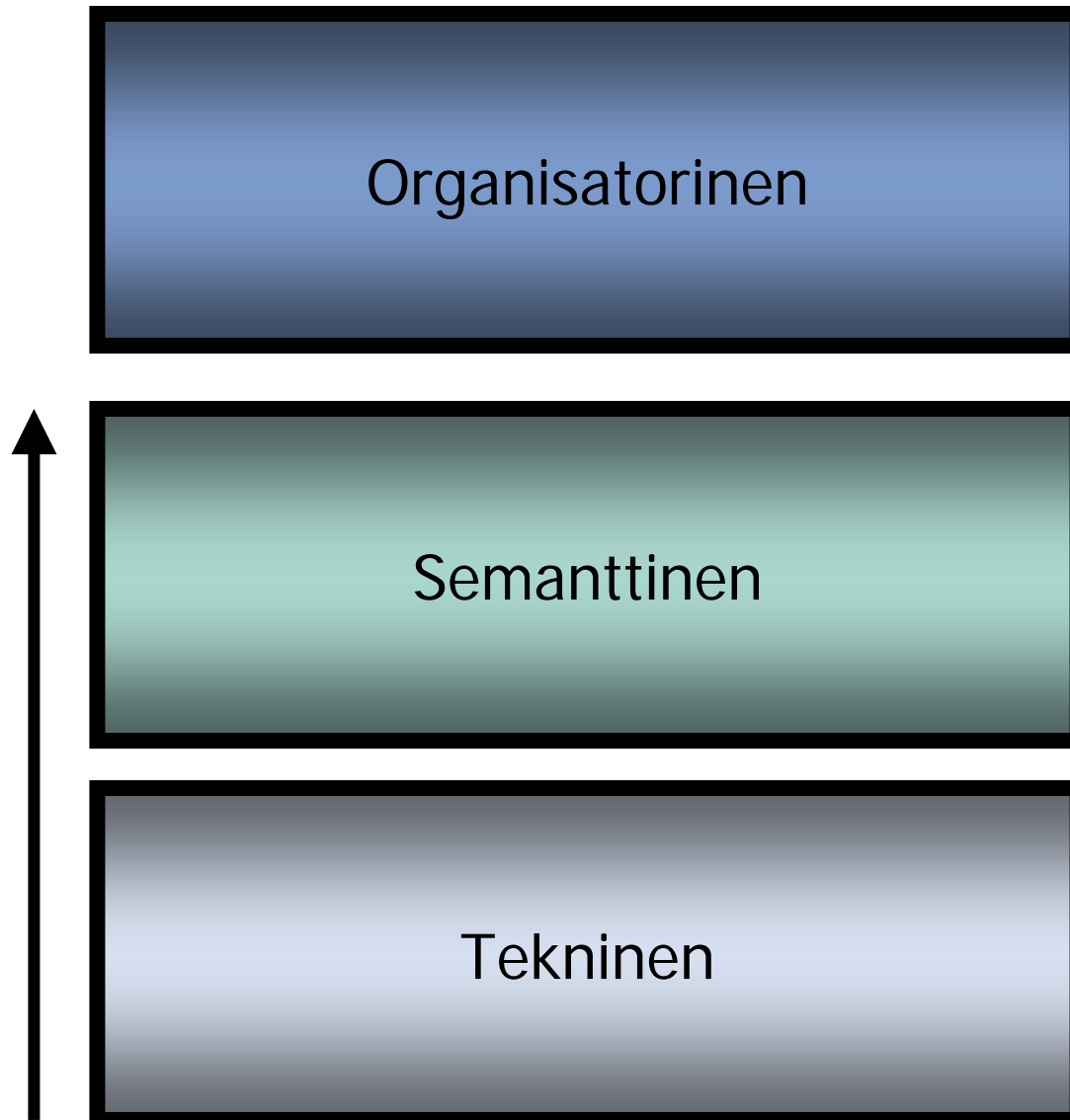


Agenda

- Peruskäsitteitä, motivointia ja kipupisteitä
- Visio ja kehityspolku
- Keskustelu

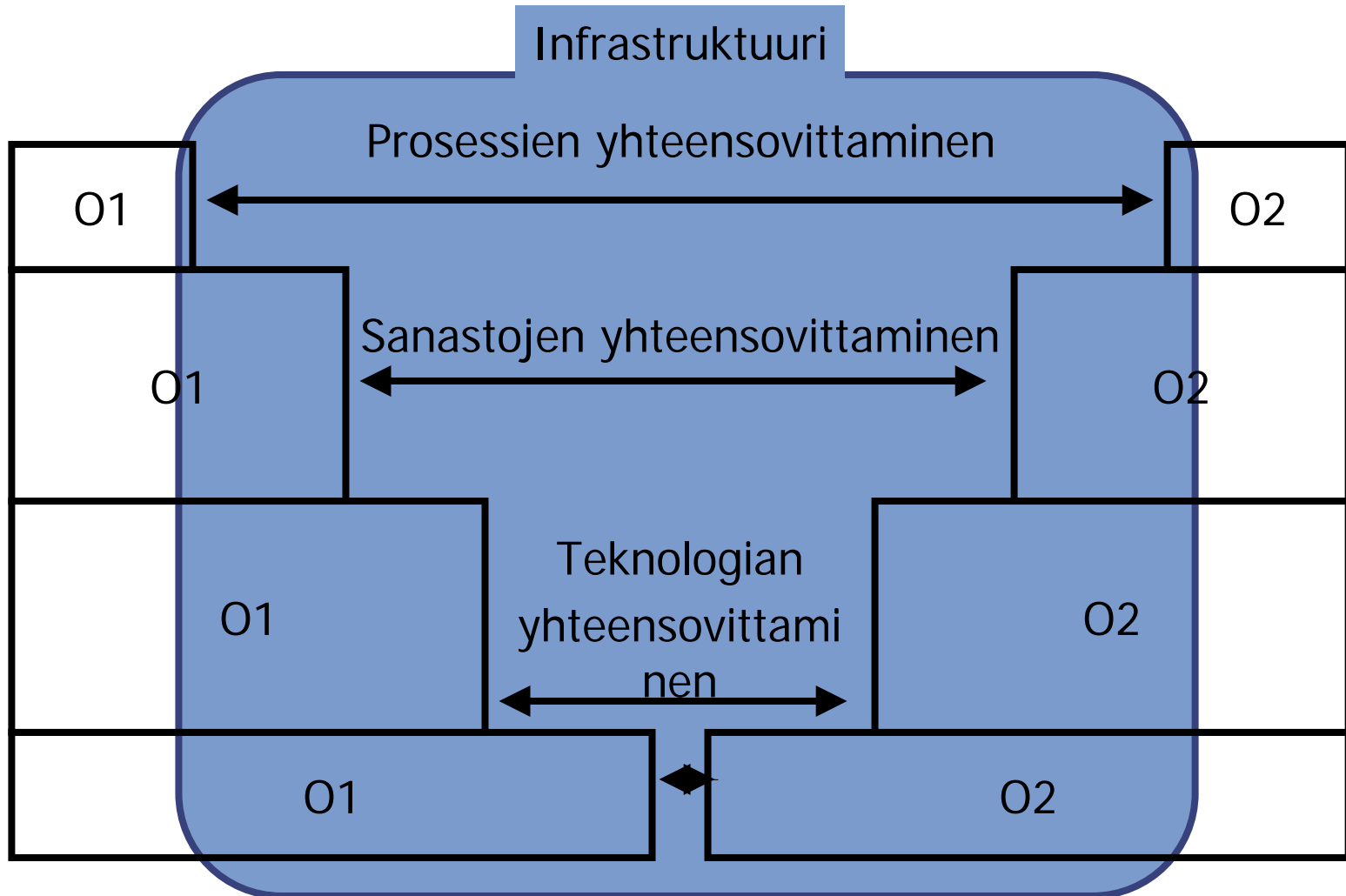


Yhteentoimivuuden jaottelu





Tasot ja niiden välisen konversion luonne (vrt. ISO/OSI NW-pino)





Miksi tekninen ei riitä? (Burkett)

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



Miksi semantiikkaan pitää nyt kiinnittää huomiota?

- Yhä suurempia kokonaisuuksia tarkastellaan yhdessä
 - Sovellusarkkitehtuureista yritysarkkitehtuureihin
 - Tietovarastojen konsolidointi
- Verkostoituneemmat [liike]toimintamallit
 - Tietojärjestelmiltä edellytetään yhä suurempaa prosessiulottuvuutta
 - SOA-pohjaiset liiketoimintamallit ja arkkitehtuuri
- XML-pohjainen integrointi on tehnyt tiedonvaihdosta teknisesti helpompaa
 - Web-teknologiapino
- Julkisella sektorilla tiivistyvä EU-yhteistyö
 - Subsidiariteettiperiaate
 - Jäsenmaan näkyminen "yhtenä" unioniin



Peruskäsitteet: Semanttinen yhteentoimivuus

- *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.
- *Semanttinen yhteentoimivuus on edellytys automaattisesti käsiteltävän tiedon uudelleenkäytettävyyteen*



Peruskäsitteet: Sanastotyö

- *Sanastotyö* on väline, jolla semanttista yhteentoimivuutta pyritään kehittämään
 - Raportissa käsitetty laajasti
- Sanastotyötä ovat mm.
 - Ontologiatyö
 - Viitemallipohjainen sanastotyö
 - Metatietomäärittelyt
 - Terminologia- ja käännöstyö
- Tietosisällön saattamista yleispätevään/yhteismitalliseen koneymmärrettävään muotoon
 - Vrt. uudelleenkäytettävyys



Kipupisteitä

- Organisoimattomuus
 - Teknislähtöinen puuhailu johtaa erittäin merkittäviin kustannuksiin epäyhteensopivien sanastojen konvertoinneissa ja konvertointisääntöjen ylläpitämisessä
 - Muutostenhallinta erittäin työlästä ja riippuvuudet tarpeettoman monimutkaisia ja itseään ruokkivia
 - Johtuu monessa tapauksessa siitä, ettei "asiaa" ole tunnistettu semanttisen tason ongelmaksi



Kipupisteitä

- Peruskäsitteistä ei ole sovittu
 - Edes peruskäsitteitä ei ole “kanonisoitu”, vaikka niitä käytetään joka järjestelmässä
 - Perusrekisterit ja vastaavat tietokannat avainasemassa



Kipupisteitä

- Yhdellä käsitteellä on useita käyttötarkoituksia tai yhteen käyttötarkoitukseen käytetään useita eri käsitteitä
 - Käsitelmäärittelyä ei ylipäänsä ole tehty kunnolla
 - Tietojärjestelmien tietosisällön eksaktiusvaatimus aiheuttaa usein filosofisen tason kysymyksiä, joihin vastaaminen on haasteellista
 - Epätarkka tai väärä käsite vakiintuu
 - Esim. paalutetaan järjestelmän toimintalogiikkaan
 - Käsitteitä ei ole dokumentoitu kunnolla



Kipupisteitä

- Luokitukset ja kooditukset vaihtelevat
 - Sama tietokenttä, eri luokitukset
 - Sama tietokenttä, samat luokitukset, eri tulkinta
- Käsitteet ovat teknisiä ja tietokantalähtöisiä
 - Vrt. Burkett
 - Asiaa ei voi ulkoistaa koodarille!
- Sanastot eivät ole koneellisesti hyödynnettävissä
 - Jo tehtyä sanastotyötä ei hyödynnetä
 - "Vanhan" sanastotyön tekijät eivät mukana "uudessa"
- Inhimillinen tekijä



Visio ja kehityspolku



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**

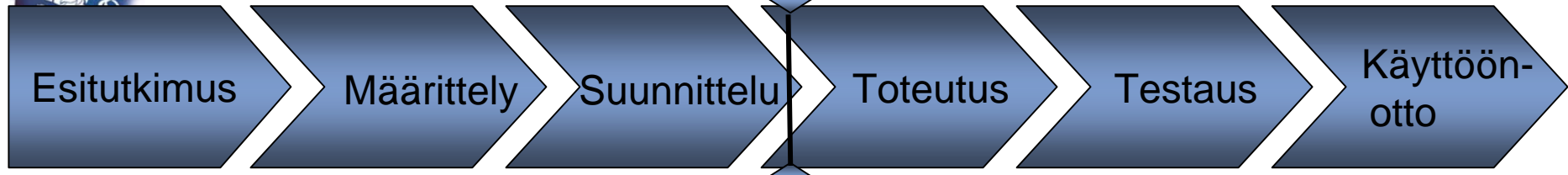


hyödynnetään suunnitelmallisesti...



Käsitteiden analyysi, määrittely ja formalisointi





Tarkastuspiste

- Semanttisen yhteentoimivuuden tarpeiden selvitys
- Minkälaista sanastotyötä hanke edellyttää?
- Mitä sanastoja on jo olemassa?
- Mitä sanastoja täytyy kehittää ja mitkä ovat lähtökohdat?
- Sanastotyötä edellyttävien aihealueiden esimerkkimateriaalin kerääminen

Abstrakti määrittely

- Asiakirjarakenteet
- Viestit
- Luokittelut ja koodistot
- Käsittemallit

Tekninen määrittely

- Esimerkiksi XML-skeemat

Sanastojen implementointi-suunnittelu

- Tietokannat
- Rajapinnat
- Liittymät

Sanastojen implementointi järjestelmään

Semanttisen yhteentoimivuuden toteaminen

• Järjestelmän tietojen ylläpito sanastojen mukaan

• Sanastojen ylläpito ja kehitys



Tietojärjestelmien kehittämisen ja sanastojen laatimisen tasorakenne

Toimialariippumat on			YSA	UN/EDIFACT, eOrdering	UBL, Dublin Core
-----------------------------	--	--	-----	--------------------------	------------------------

...VIISAS PÄÄSEE VÄHEMMÄLLÄ

1) HYÖDYNNÄ VALMIIT SANASTOT

2) JOS EI OLE:

-ÄLÄ YLIMITOITA SANASTOTYÖTÄ

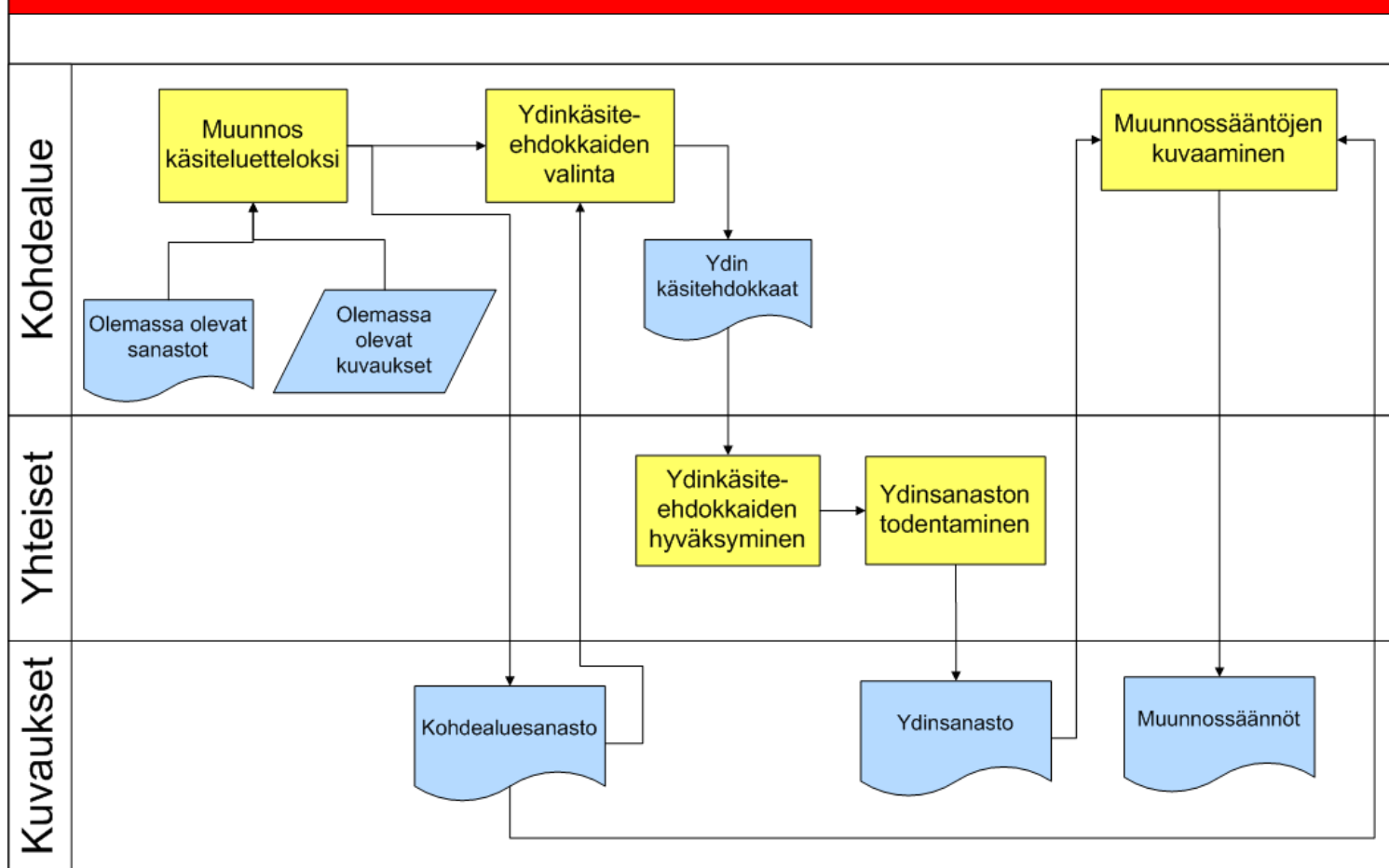
-VERKOSTOIDU

<i>Kattavuus</i>	Tieto- järjestel mä	Organisa atio	Kansallin en	EU:n laajuinen	Maailman - laajuinen
------------------	------------------------------------	--------------------------	-------------------------	---------------------------	-------------------------------------



Sanastojen yhtenäistämisen prosessi

Sanastojen yhtenäistäminen





Esim. asiakirjan metatiedot

Apuvälineet

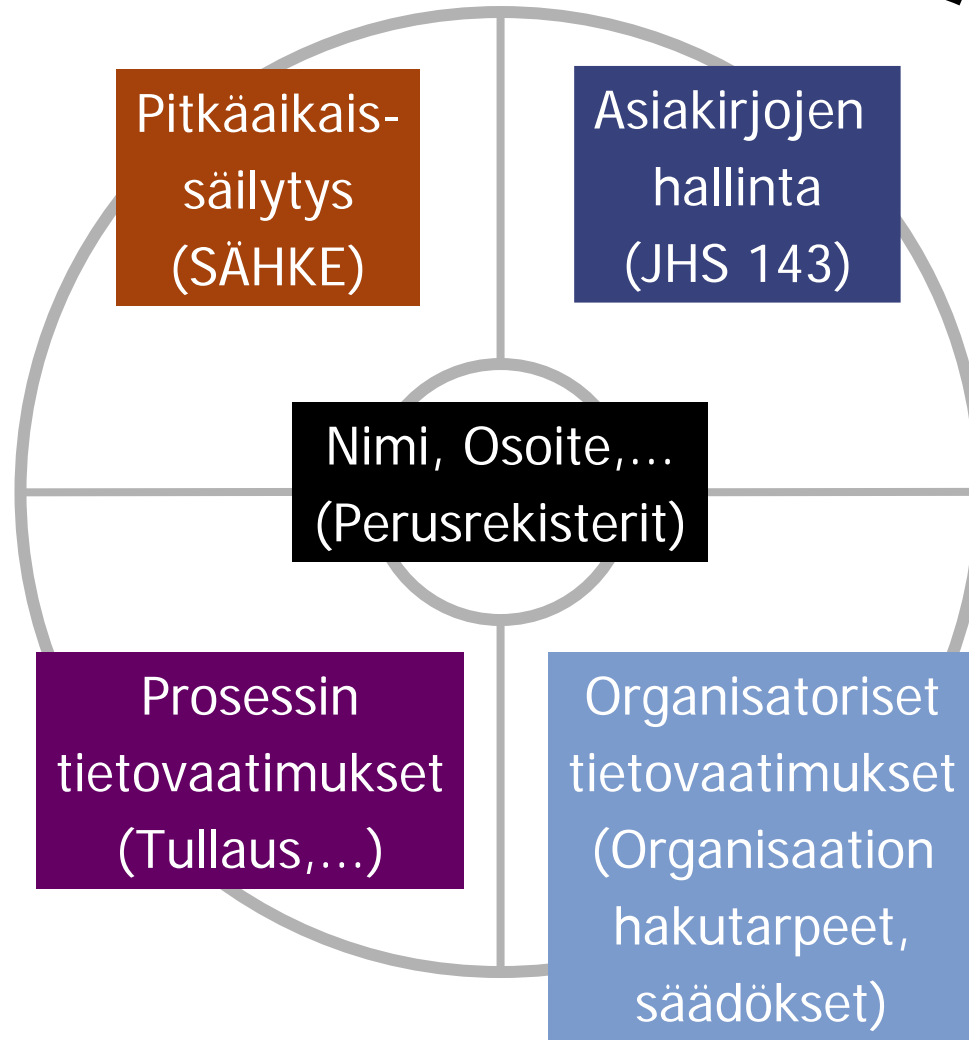
Sanastopalvelin: Sanastot

Termipankki: Termit ja käännökset

Menetelmät: Standardit, parhaat

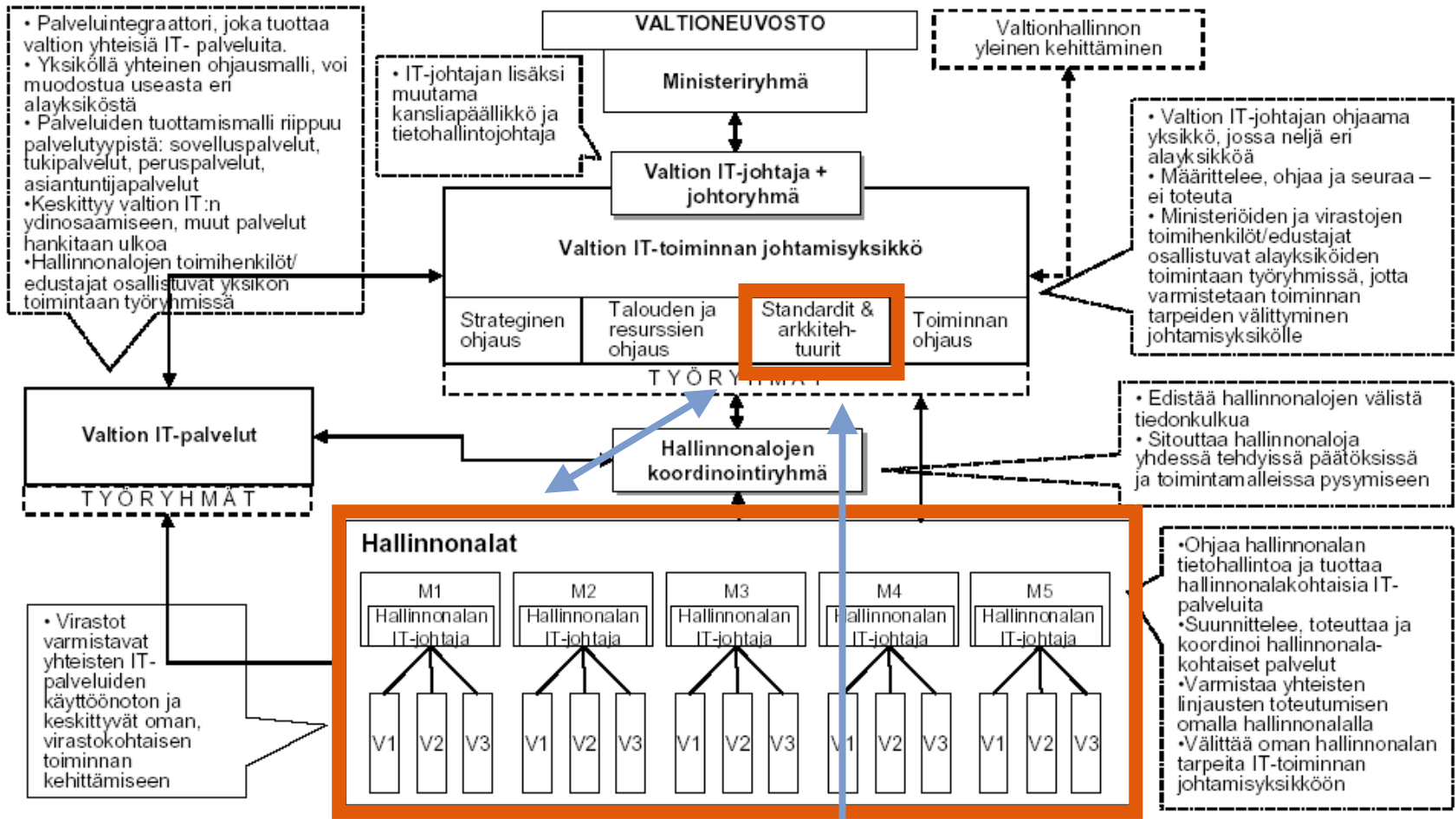
käytännöt, nimeämishojeet,

viitesanastot





koordinoitu ja resurssoitu...



Toimialakohtainen koordinointi

Semanttisen yhteentoimivuuden koordinointi



askellus



Vaihe 0: Nykytila

- Semanttiseen yhteentoimivuuteen liittyvät ongelmat tunnistetaan, mutta niitä ei osata korjata oikealla tasolla
- Sanastotyötä ei ole organisoitu eikä resursoitu
- Semanttisen tason yhteentoimivuusongelmat estävät tietojärjestelmien tehokkaan yhteentoiminnan

Vaihe 1: Tietoisuuden kasvattaminen

- Perustetaan jatkotyöryhmä koordinoimaan semanttisen yhteentoimivuuden kehittämistä ja sanastotyötä
- Tehdään tämän raportin pohjalta JHS-suositus sanastotyön ja semanttisen yhteentoimivuuden edistämiseksi
- Aloitetaan tiedottaminen ja markkinointi hallinnossa yleisesti ja organisaatiokohtaisesti sekä sidosryhmien keskuudessa
- Kehitetään menetelmiä ja työkaluja
- Tuetaan yhteistyön synnyttämistä sanastotyön ympärille
- Ylempien tasojen sanastotyötä aletaan hyödyntää

Vaihe 2: Koordinointi ja resursointi

- Järjestetään julkishallinnon sanastotyön keskitetty koordinointi
- Semanttisen yhteentoimivuuden kehittäminen ja sanastotyö vastuutetaan sekä valtion IT-johdossa että virastoissa
- Asianmukaiset resurssit varataan työn tekemiseen sekä valtion IT-johdossa että virastoissa
- Perustetaan sanastotyötä tukeva keskitetty, julkinen hakemisto
- Yhteistyö organisoidaan (Communities of Practice, CoP)

Vaihe 3: Visio

- Semanttisen yhteentoimivuuden tarve ymmärretään
- Sanastotyö on luonnollinen osa jokapäiväistä järjestelmien kehitystyötä
- Tarpeelliset resurssit ovat käytössä ja vastuut selvillä
- Sanastotyön tarve vähenee, koska päällekkäisen tiedon tallennusta on vähennetty samansisältöisiä tietojärjestelmiä yhdistämällä

2005

2006

2010



Kiitos!

Riku.Moisio@vm.fi

Raportti "Sanastot siltana saumattomalle
tiedonvaihdolle" saatavana osoitteessa:
<http://www.vm.fi>