

Metatietojen merkitys tiedonhallinnassa

Airi Salminen

Jyväskylän yliopisto

<http://www.cs.jyu.fi/~airi/>

Suuntana lainsäädäntötyön semanttinen web

Kohti lainsäädäntötyön tiedonhallinnan tehostamista

Seminaari Eduskunnassa 26.5.2004

Sisältö

1. Tietoa tiedosta
2. Mihin metatietoa tarvitaan?
3. Sisällönhallinta tiedonhallintana
4. Metatiedot sisällönhallinnassa
5. Metatiedot semanttisessa webissä
6. Yhteenveto

metatieto = tietoa tiedosta

esimerkiksi

- **dokumentista**
- **dokumenttikokoelmasta**
- **dokumentin osasta**
- **HTML-sivusta**
- **tietokannasta**
- **ohjelmasta**
- **museoesineestä**
- **henkilöstä**

Metatieto voi kuvata esimerkiksi dokumentista

- sisällön rakennetta
- sisällön merkitystä
- sisällön kieltä
- tuottamisympäristöä (tekijä, organisaatio, prosessi, ohjelma)
- versiota
- tallennusmuotoa
- tallennusmediaa
- säilytysaikaa
- käyttöoikeuksia

1. Tietoa tiedosta

Dublin Core -metatiedot webin tietoresursseille:

- **title**
- **creator**
- **subject**
- **description**
- **publisher**
- **contributor**
- **date**
- **type**
- **format**
- **identifier**
- **source**
- **language**
- **relation**
- **coverage**
- **rights**

1. Tietoa tiedosta

Esimerkki DC-metatiedosta HTML-dokumentissa

```
<head>
  <title>Airi Salminen</title>
  <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
  <meta name="DC.Title" content="Airi Salminen">
  <meta name="DC.Creator" content="Airi Salminen">
  <meta name="DC.Subject" content="Airi Salminen">
  <meta name="DC.Description" content="Home page of Airi Salminen, showing
    her representative publications, research interests, projects, courses, and
    contact information">
  <meta name="DC.Date.Modified" content="2004-04-10">
  <meta name="DC.Type" content="Text.Home page">
  <meta name="DC.Format" content="html">
  <meta name="DC.Identifier" content="http://www.cs.jyu.fi/~airi/index.html">
  <meta name="DC.Language" content="en">
  <meta name="DC.Relation" content="http://www.cs.jyu.fi/~airi/docman.html">
  <meta name="DC.Relation"
    content="http://www.cs.jyu.fi/~airi/publications.html">
  <meta name="DC.Rights" content="Copyright 2002 Airi Salminen">
</head>
```

1. Tietoa tiedosta

JHS 143 Asiakirjojen kuvailuformaatti:

- nimeke
- tekijä
- aihe
- kuvaus
- julkaisija
- muu tekijä
- aikamääre
- laji
- formaatti
- identifiointitunnus
- lähde
- kieli
- suhde
- kattavuus
- oikeudet
- asiakirjan laji
- julkisuus
- versio
- ympäristö
- saatavuus
- vastaanottaja
- toimeksiantaja
- kohdeyleisö

2. Mihin metatietoa tarvitaan?

Kirjaston kokoelmien organisoimiseen, julkaisujen hakuun, saatavuuden varmistamiseen

tieto

Kirjaston
julkaisukokoelmat

metatieto

- kirjastokortit (tekijä, julkaisuvuosi, kustantaja, avainsanat, hyllysijainti, ...)
- luokitukset
- sanastot, hakemistot
- lainaajatiedot
- lainaussäännöt

2. Mihin metatietoa tarvitaan?

Internetin tietojen ja palvelujen etsimiseen ja tarjoamiseen

tieto

Internetin
tietovarannot

metatieto

- hakukoneiden indeksit
- linkkisivut
- Dublin Core -kuvaukset
joko upotettuina
dokumenteissa tai
ulkoisina kuvauksina

2. Mihin metatietoa tarvitaan?

Varmistamaan organisaation asiakirjojen löydettävyys, ylläpidettävyys, todistusvoimaisuus, autenttisuus ja ymmärrettävyys

tieto

Organisaation
asiakirjat

metatieto

- kortistot, hakemistot
- ohjelmistokohtaiset metatiedot (diaari-, dokumenttienhallinta-, päätöksenteko-, arkistointijärjestelmät)

3. Sisällönhallinta tiedonhallintana

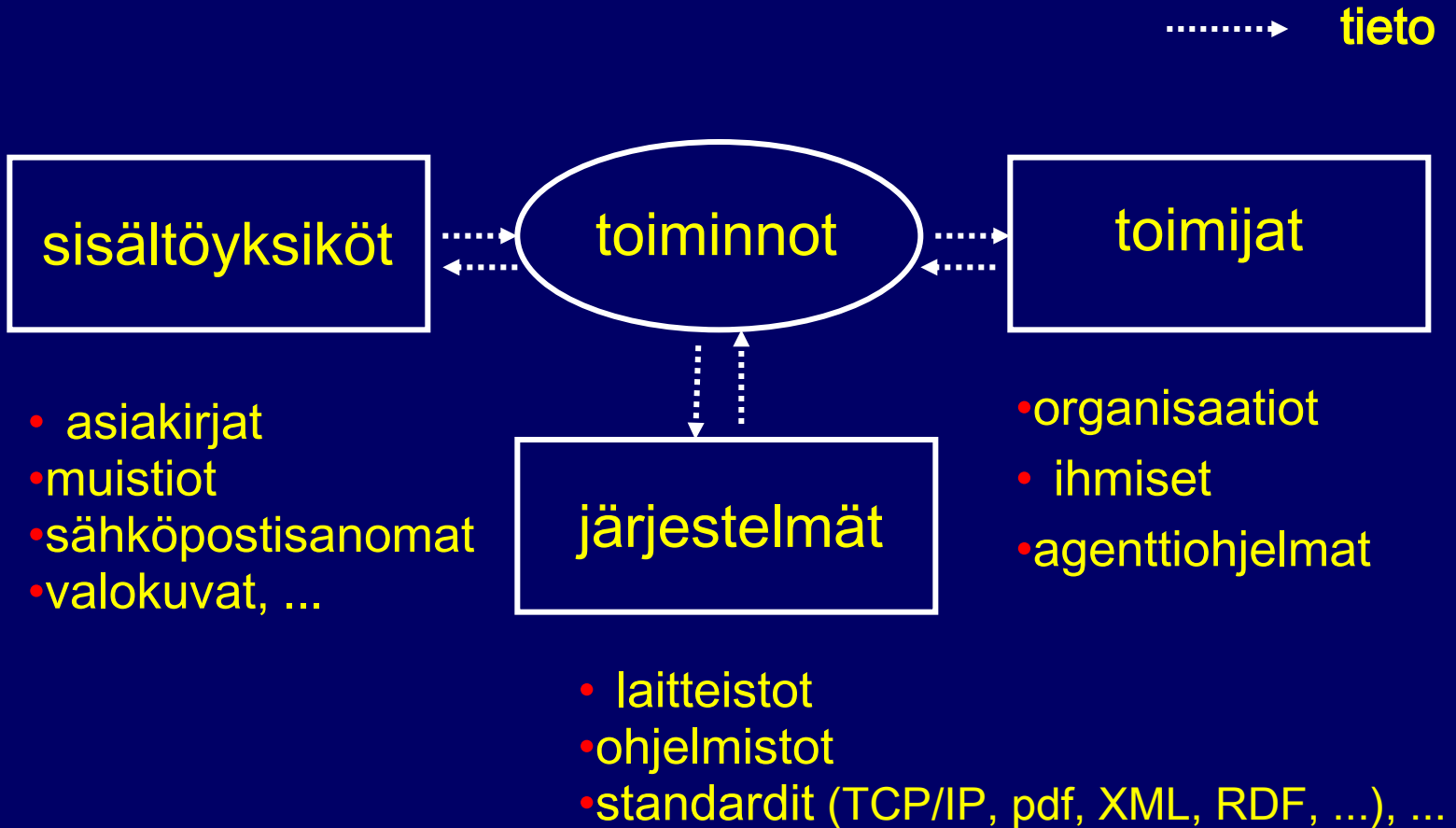
RASKE-metodologia kokonaisvaltaiseen dokumenttien- ja sisällönhallinnan kehittämiseen

- **RASKE1 (1994-1998):** menetelmiä dokumenttien rakenteistamiseen; sovellusalueena valtiopäiväasiakirjojen rakenteistaminen
- **RASKE2 (2003-2006):** menetelmiä sisällönhallinnan kehittämiseen, metatietojen standardoimiseen ja tietoresurssien integroimiseen; sovellusalueena lainsäädäntöprosessin sisällönhallinta

<http://www.it.jyu.fi/raske/>

3. Sisällönhallinta tiedonhallintana

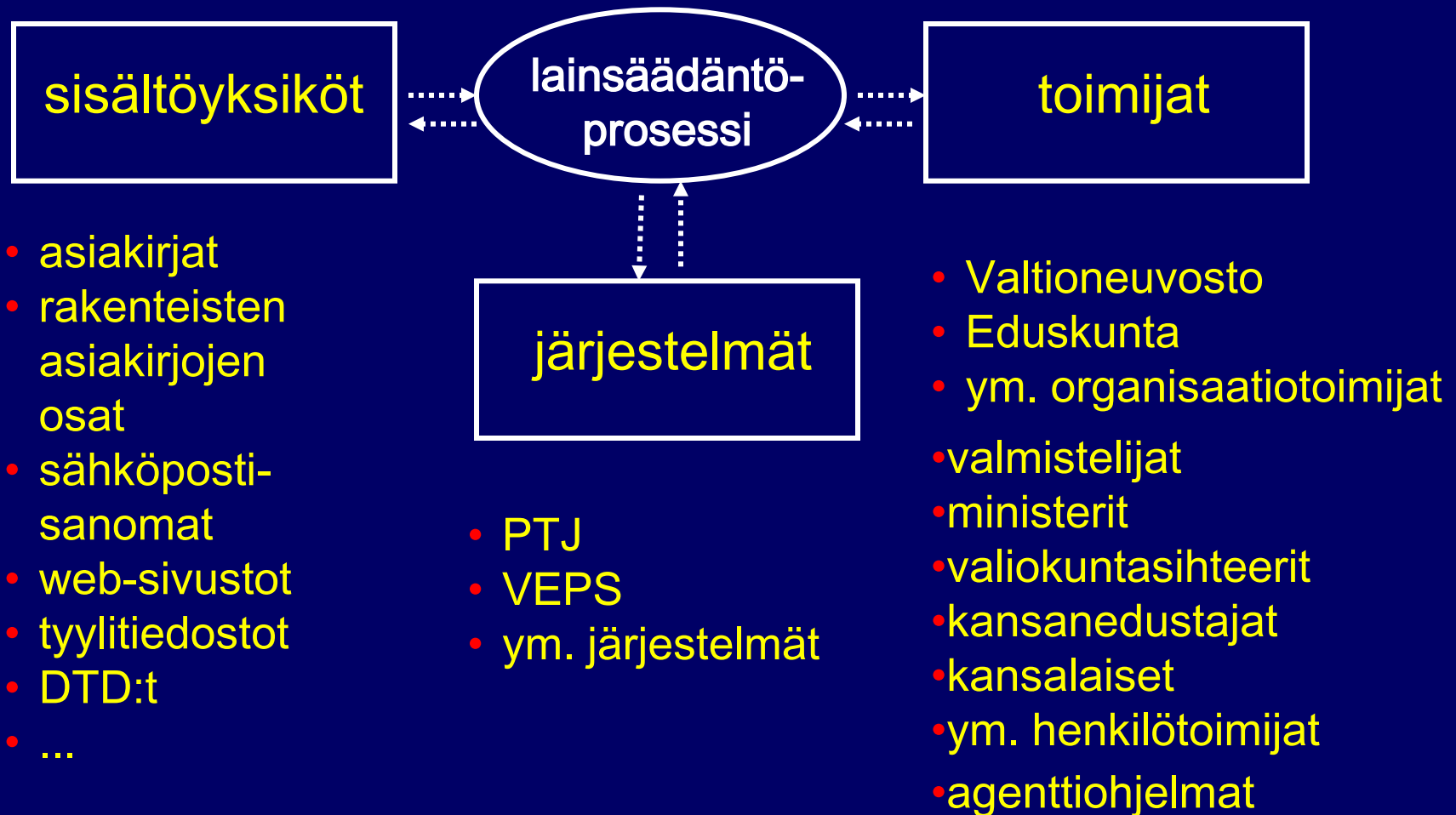
Sisällönhallinnan osa-alueet



3. Sisällönhallinta tiedonhallintana



3. Sisällönhallinta tiedonhallintana



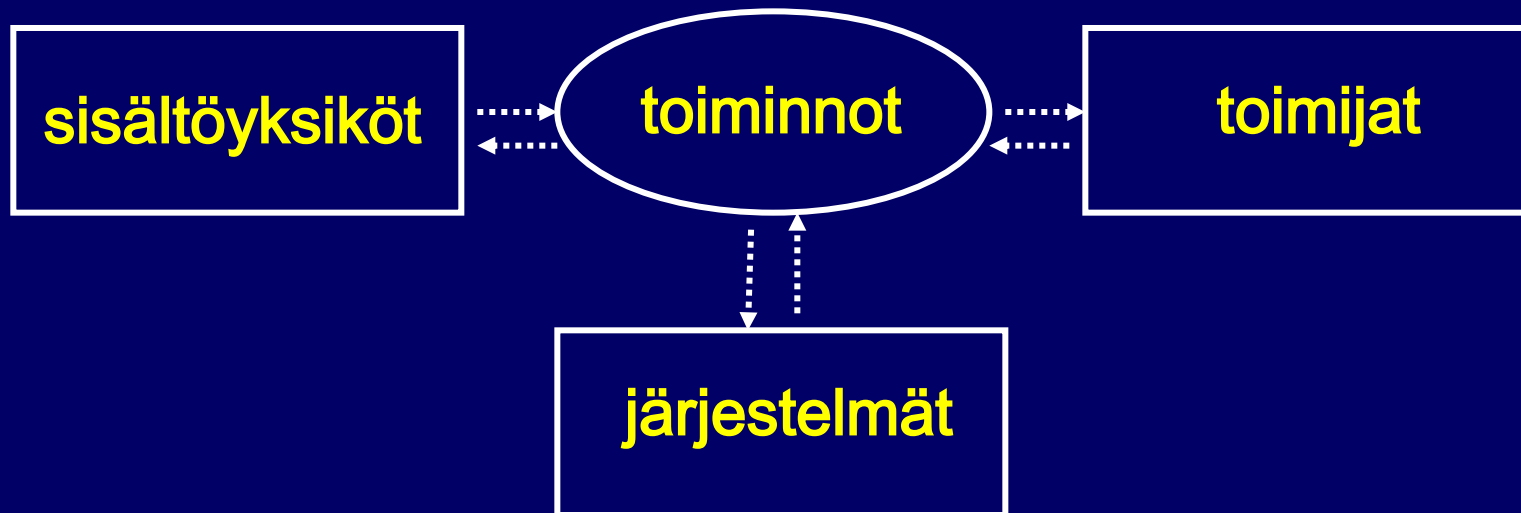
Ongelmia organisaatioiden sisällönhallinnassa

- ▶ sisältöjen määrä kasvanut vaikeasti hallittavaksi
- ▶ sisältöyksiköiden hämärtyminen (asiakirja, dokumentti, html-sivu, tiedosto, sähköpostisanoma, tyylitiedosto)
- ▶ sisältöjen epäyhtenäisyys
- ▶ vastuut metatietojen tuottamisesta ja ylläpitämisestä hämää

Ongelmia organisaatioiden sisällönhallinnassa

- ▶ järjestelmien kirjavuus
- ▶ järjestelmät eivät kommunikoi keskenään
- ▶ tuki työprosesseille puuttuu

4. Metatiedot sisällönhallinnassa

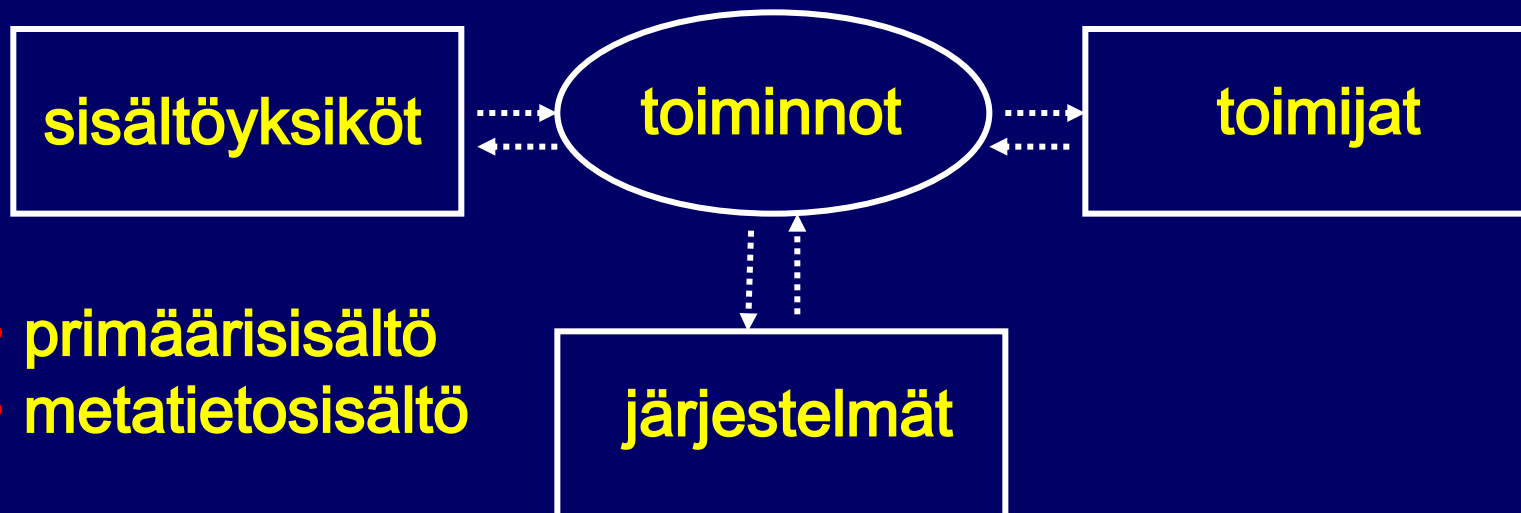


Metatiedot kuvaavat sisältöyksiköitä, toimintoja, toimijoita, järjestelmiä

Tavoitteita metatiedoille

- ▶ formaali, standardoitu muoto niin, että sekä ihmiset että erilaiset ohjelmistot pystyvät niitä hyödyntämään
- ▶ sovellusriippumaton esitysmuoto
- ▶ mahdollisimman paljon automaattista tuottamista
- ▶ täsmällisesti määritellyt tuottamis- ja ylläpitovastuut

4. Metatiedot sisällönhallinnassa



Metatietosisältöä voivat olla esim. rakennekuvaukset, tyylitiedostot, JHS 143 -kuvaukset, prosessikuvaukset, toimijakuvaukset, järjestelmäkuvaukset, standardien kuvaukset

4. Metatiedot sisällönhallinnassa

Vaihtoehtoja metatiedoille

liittyy yhteen sisältöyksikköön	liittyy sisältöyksikköjen joukkoon
upotettu	ulkoinen
keskitetty	hajautettu
ihmisen tuottama	automaattisesti tuotettu
ihmisen käyttöön	tietokoneen käyttöön

5. Metatiedot semanttisessa webissä

Semanttisessa webissä metatiedoilla pyritään varmistamaan yhteisön tietoresurssien käytettävyys niin ihmisille kuin ohjelmillekin (löydettävyys, ymmärrettävyys, luotettavuus, yhdisteltävyys)

tieto

Semanttisen webin
tietovarannot

metatieto

- käsitteistöt ja käsitteiden väliset suhteet, päättelysääntöjä (ontologiat)
- käyttöoikeuksiin ja käyttötarkoituksiin liittyvät kuvaukset
- tietoturvaan liittyvät tiedot (digitaaliset vesileimat, allekirjoitukset, ...)
- tiedot metatietostandardeista

5. Metatiedot semanttisessa webissä

Monipuolista metatietoa

- Nykyistä huomattavasti enemmän webin tietovarantojen merkitykseen, käyttöön, käyttötapoihin ja luotettavuuteen liittyvää metatietoa.
- Metatietojen kieli huomattavasti homogeenisempaa ja kontrolloidumpaa kuin dokumenttisisältöjen.

5. Metatiedot semanttisessa webissä

Metatiedot koneiden ymmärtämässä muotoon

- **Metatiedot formaalissa standardoidussa muodossa niin, että sekä ihmiset että erilaiset ohjelmistot pystyvät niitä hyödyntämään.**
- **Automaattisia päättelysääntöjä.**

5. Metatiedot semanttisessa webissä

Semanttisen webin pelkistetty rakenne



5. Metatiedot semanttisessa webissä

sovellusohjelmat	
primääritietovarannot	DTD:t, XML Schema - määrittelyt, RDF Schema -määrittelyt, RDF-määrittelyt, ontologiat, aihekartat, annotoinnit, ...
URI, Unicode, XML, XML Namespaces, XML Schema, RDF, RDF Schema, XTM, XML-Signature, OWL, Annotea, ...	

6. Yhteenveto

- **Metatiedot organisaatioissa pitkälle järjestelmiin sidottuja.**
- **Ratkaisut hajanaisia, ei tukea työprosesseille**
- **Sisältöjen rakenteisuutta ei pystytä vielä hyödyntämään kovin tehokkaasti.**

6. Yhteenveto

- **Järjestelmäriippumaton standardointi tärkeää.**
- **Metatason määrittelyiden täytyy olla niin formaaleja, että ohjelmat pystyvät käyttämään metatietoja.**
- **Semanttisen webin metatietostandardien käyttö tarjoaa mahdollisuuden yhdistellä ratkaisuja, joita tehdään eri yhteisöissä.**

6. Yhteenveto

- RASKE2-projektissa kehitetään menetelmiä metatietojen standardointiin ja sisällönhallinnan kokonaisvaltaiseen kehittämiseen.
- Lisätietoja RASKE2-projektista ja sitä edeltävästä aihepiiriin liittyvästä tutkimustyöstä:

<http://www.it.jyu.fi/raske/>