

# Sähköisten dokumenttien hallinta: peruskäsitteet ja kuvausmenetelmät

Airi Salminen

Jyväskylän yliopisto

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

VIVA-seminaari

Tampere 3.12.2002

# Sisältö

1. Mikä on dokumentti?
2. Dokumenttien hallinta
3. Rakenteiset dokumentit
4. Dokumenttianalyysi
5. Dokumenttianalyysin kuvausmenetelmiä
6. Yhteenveto

Esitys perustuu pääosin artikkeliin:

Airi Salminen, Methodology for document analysis. Teoksessa A. Kent (Ed.), *Encyclopedia of Library and Information Science*, Vol. 67 (Supplement 30) (s. 299-320). New York: Marcel Dekker, 2000.

# 1. Mikä on dokumentti?

## *dokumentti*

asiakirja, -paperi, todistuskappale

Nykysuomen sanakirja, 1996

## *document*

1. a written or printed paper furnishing information or evidence, as a passport, deed, bill of sale, or bill of lading; a legal or official paper
2. any written item, as a book, article, or letter, esp. of a factual or informative nature.

Random House Webster's Dictionary, 1998

# 1. Mikä on dokumentti?

*dokumentti*

keskeisiä ominaisuuksia dokumenttien hallinnan näkökulmasta

1. Tarkoitettu ihmisten aistittavaksi (nähtäväksi, kuultavaksi, tunnettavaksi) ja ymmärrettäväksi jossain asiayhteydessä.
2. Tallennettu jollekin medialle.
3. Voidaan tunnistaa ja käsitellä kokonaisuutena.

# 1. Mikä on dokumentti?

*dokumentti* keskeisiä ominaisuuksia dokumenttien hallinnan näkökulmasta (jatkoa)

4. Dokumentissa on *sisältö* ja sillä voi olla yksi tai useampia *ulkoisia esitysmuotoja*.
5. Sisältö koostuu *osista*, osat koostuvat *symboleista*.
6. Sisällön rakenteen tarkoitus on auttaa ihmistä ymmärtämään sisältöä.

# 1. Mikä on dokumentti?

*sähköinen dokumentti*

lisäominaisuuksia

7. Tallennettu sähköiselle medialle.
8. Tarkoittaa yleensä samaa kuin *digitaalinen dokumentti*.
9. Dokumentin ihmiselle esittämiseen tarvitaan tekninen laitteisto.
10. Digitaalisen dokumentin esittämiseen ja käsittelyyn tarvitaan myöskin ohjelmisto.

## 2. Dokumenttien hallinta

### *dokumenttien hallinta*

Toimenpiteet, joilla ainakin jossain määrin kontrolloidusti pidetään yllä dokumenttien kokoelmaa suunniteltuun tarkoitukseen.

## 2. Dokumenttien hallinta

### *dokumenttien hallinnan toimenpiteitä*

- käyttö- ja tuotantotapojen suunnittelu
- sisältörakenteiden ja esitystapojen suunnittelu
- kokoelman organisoinnin suunnittelu
- metatietojen suunnittelu
- käytettävän teknologian (laitteiston, ohjelmiston) valinta ja/tai suunnittelu
- käyttöoikeuksien suunnittelu



## 2. Dokumenttien hallinta

### *dokumenttien hallinnan toimenpiteitä*

- dokumenttien luominen ja tallentaminen
- metatietojen luominen ja tallentaminen
- muuttaminen
- jakelu
- yhdisteleminen
- hävittäminen

## 2. Dokumenttien hallinta

### *dokumenttien hallinnan toimenpiteitä*

- dokumenttien hakeminen (ihminen ja/tai ohjelma)
- tietojen hakeminen (ihminen ja/tai ohjelma)
- katselu tai kuuntelu
- metatietojen hakeminen (ihminen ja/tai ohjelma)

## 2. Dokumenttien hallinta

### *metatietoja*

- sisältöä ja sen merkitystä kuvaavat tiedot (esim. avainsanat, indeksit, ontologiat)
- tuottamiseen liittyvät tiedot (esim. tekijä, tekopäivä, prosessi, organisaatio)
- sisällön rakennekuvaus (esim. DTD, sisällysluettelo)

## 2. Dokumenttien hallinta

### *metatietoja*

- sijainti ja esitystapa (esim. URL, tiedostoformaatti)
- käyttöoikeudet (esim. luku- ja kirjoitusoikeudet, tekijänoikeudet)
- versiotiedot

## 2. Dokumenttien hallinta

### *metatietoja*

- luotettavuuteen liittyvät tiedot (esim. allekirjoitus, salakirjoitusavain, annotaatiot)
- metatietoihin liittyvät tiedot (esim. metatietoskeema, metatietojen tuottajat)

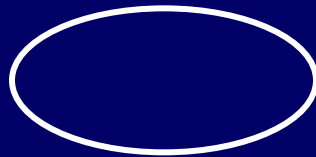
## 2. Dokumenttien hallinta

ICN = Information Control Net (Ellis, 1979)

ICN:n symbolit:



resurssi (resource)



toiminto (activity)



kontrollivuo (control flow)

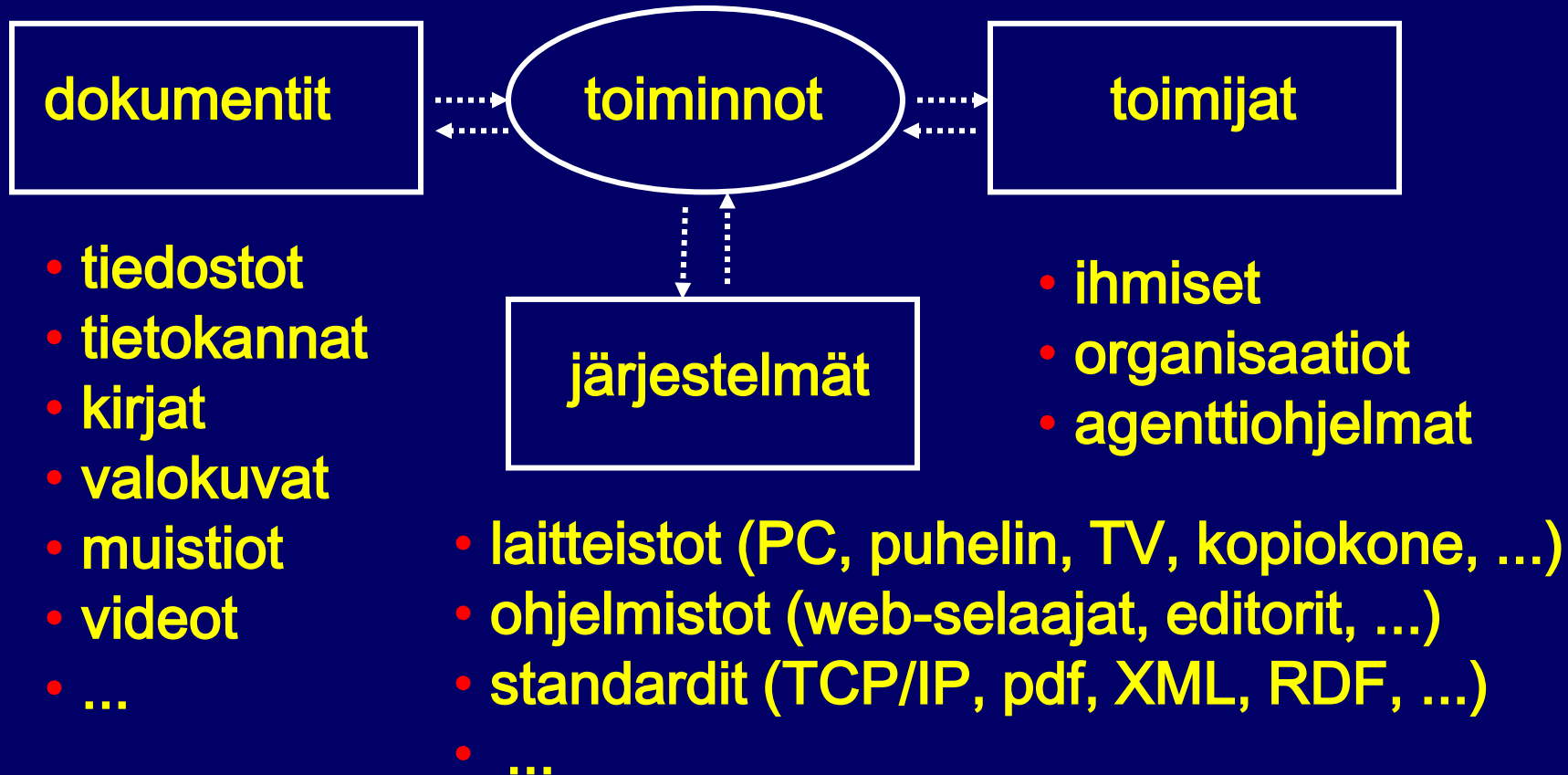


tietovuo (information flow)

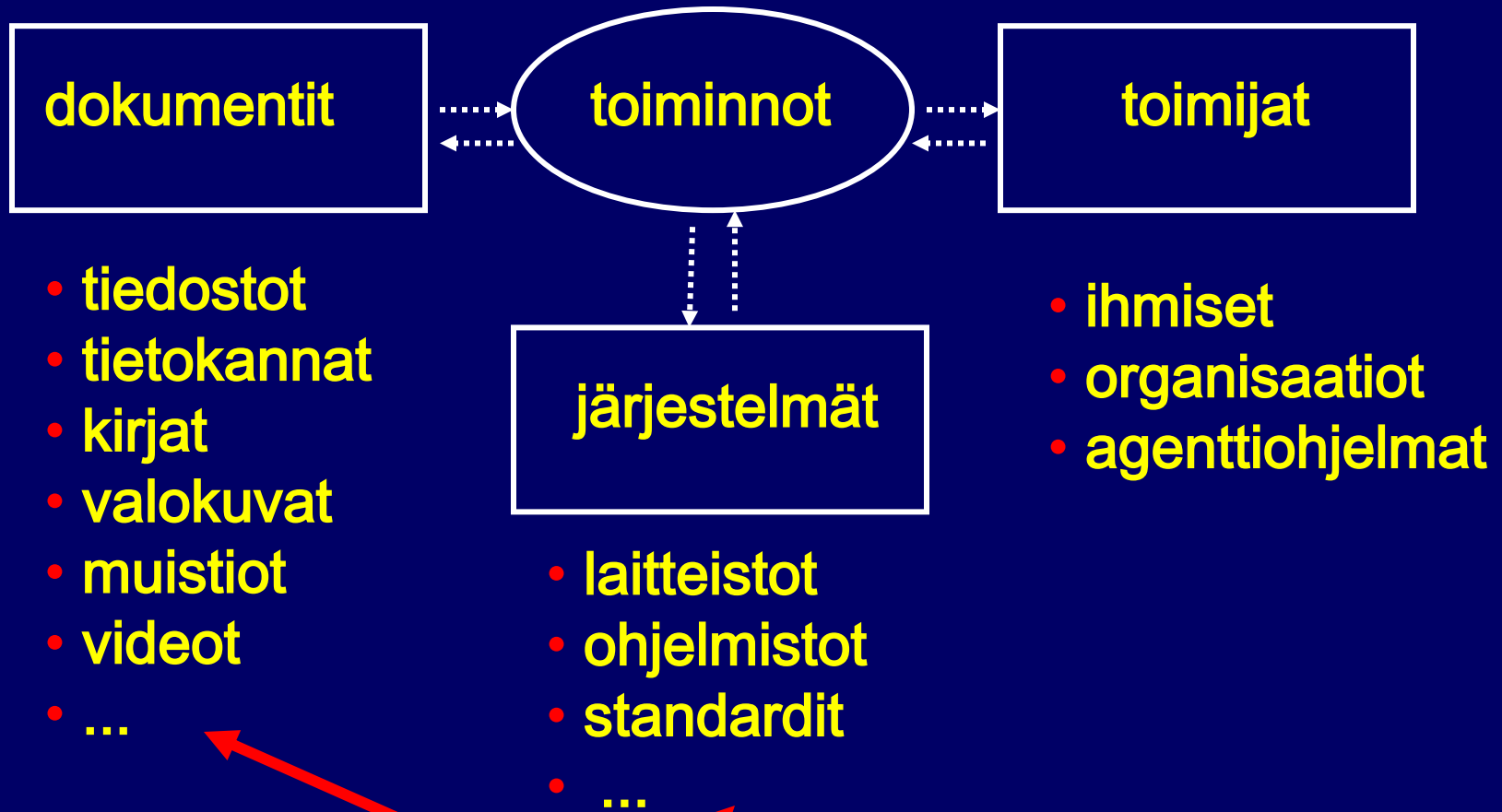
Ellis, C.A. (1979). Information Control Nets: A mathematical model of office information flow. *Proceedings of the Conference on Simulation, Measurement and Modeling of Computer Systems, ACM SIGMETRICS Performance Evaluation Review*, 8 (3), 225-238.

## 2. Dokumenttien hallinta

### Dokumenttien hallintaympäristö



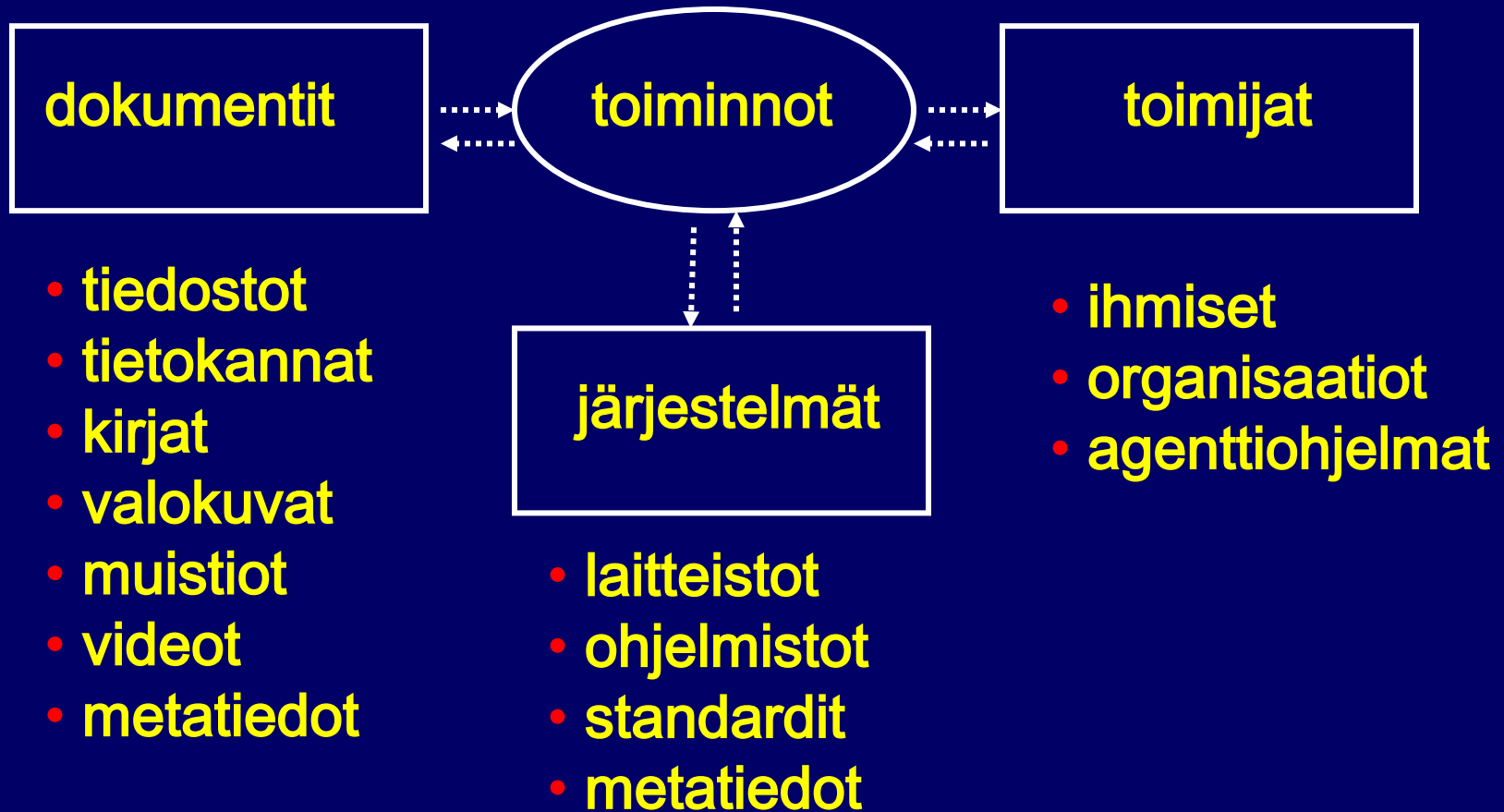
## 2. Dokumenttien hallinta



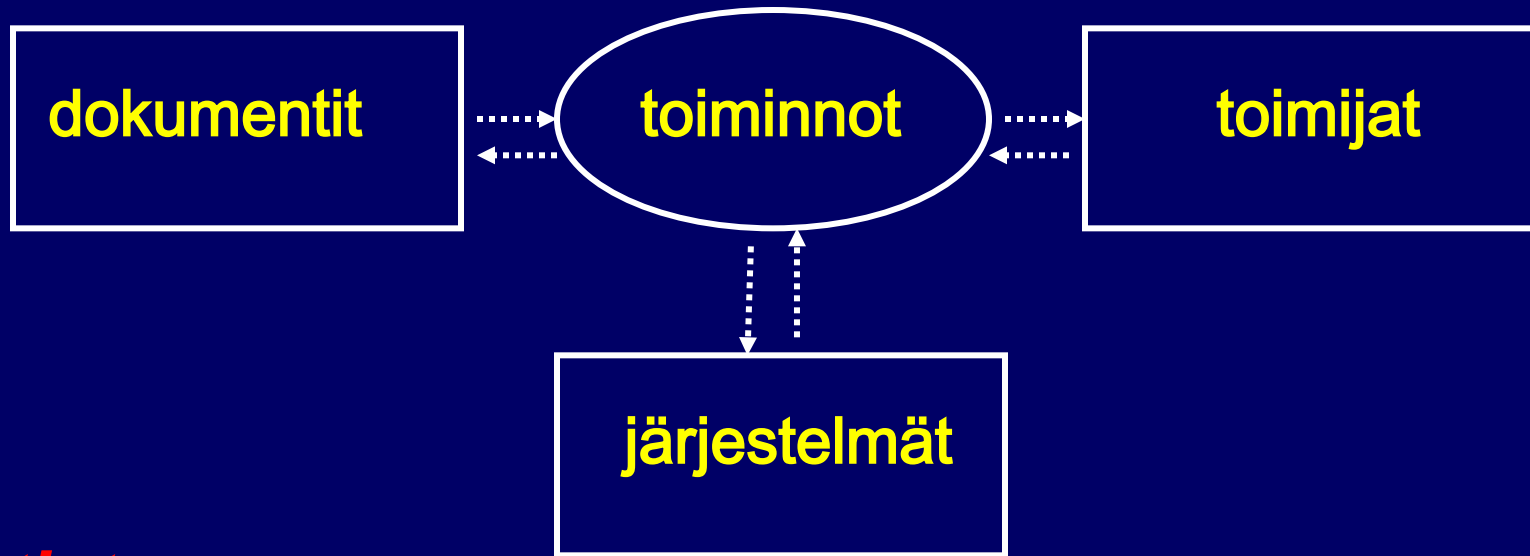
**Missä on metatiedon paikka?**



## 2. Dokumenttien hallinta



## 2. Dokumenttien hallinta



**tieto**  
(*'information'*)

- tuotetaan toiminnoissa
- käytetään toiminnoissa
- tallentuvat resurssien tietovarastoihin (dokumentit, järjestelmät, toimijat)

# 3. Rakenteiset dokumentit

## *rakenteinen dokumentti* (*'structured document'*)

- Dokumenttien hallintaympäristössä on käytettävissä standardoitu tapa sisällön rakenteen esittämiseen ja määrittelemiseen (esim. XML tai SGML).
- On mahdollista suunnitella dokumenttirakenteet tiettyjä toimintoja varten (esim. HTML web-julkaisemista varten).
- Dokumenttien hallintaympäristössä on käytettävissä ohjelmistoja, jotka pystyvät tunnistamaan rakenneosat (esim. web-selaajat).

## 3. Rakenteiset dokumentit

XML = Extensible Markup Language

**Joukko sääntöjä informaation esittämiseksi rakenteisina dokumentteina**

T. Bray, J. Paoli, C. M. Sperberg-McQueen, and E. Maler (Eds.), Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation 6 October 2000, <http://www.w3.org/TR/REC-xml>

# 3. Rakenteiset dokumentit

```
<?xml version = "1.0"?>
<runo tekija = "Murasaki Shikibu" tekija_syntyntyt = "974">
<info_linkki xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xlink:href=
    "http://digital.library.upenn.edu/women/omori/court/murasaki.html">
  Lis&auml;tietoa runoilijasta
</info_linkki>
<sae>
<rivi>This life of ours would not cause you sorrow</rivi>
<rivi>if you thought of it as like </rivi>
<rivi>the mountain cherry blossoms</rivi>
<rivi>which bloom and fade in a day. </rivi>
</sae>
</runo>
```

**Note:** The text of the rivi elements is taken from <http://www.slip.net/~knabb/rexroth/translations/japanese.htm>, containing Kenneth Rexroth's translations of Japanese poetry

### 3. Rakenteiset dokumentit

**This life of ours would not cause you sorrow  
if you thought of it as like  
the mountain cherry blossoms  
which bloom and fade in a day.**

[Lisätietoa  
runoilijasta](#)

# 3. Rakenteiset dokumentit

## Dokumenttityypin määritys (DTD) runolle

```
<!DOCTYPE runo [  
<!ELEMENT runo (info_linkki? nimi?, sae+)>  
<!ATTLIST runo  
    tekija CDATA #REQUIRED  
    tekija_syntynyt CDATA #OMITTED>  
<!ELEMENT otsikko (#PCDATA) >  
<!ELEMENT info_linkki (#PCDATA) >  
<!ATTLIST info_linkki  
    xmlns:xlink CDATA #FIXED "http://www.w3.org/1999/xlink"  
    xlink:type CDATA #FIXED "simple"  
    xlink:href CDATA #REQUIRED >  
<!ELEMENT sae (rivi+) >  
<!ELEMENT rivi (#PCDATA) >]
```

## 4. Dokumenttianalyysi

Dokumenttien hallinnan suunnittelu ja kehittäminen edellyttävät dokumenttianalyysiä. Analyysissä selvitetään ja kuvataan:

- nykyiset toimintatavat
- tarpeet tuleville toimintatavoille
- ehdotukset ja suunnitelmat tuleviksi toimintatavoiksi



# 4. Dokumenttianalyysi

## RASKE-metodologia

- Joukko menetelmiä ja tekniikoita dokumenttianalyysiin.
- Kehittäminen aloitettu RASKE-projektissa (RASKE = Rakenteisten ASiakirjastandardien KEhittäminen).
- RASKE-projektissa menetelmiä käytettiin dokumenttien hallinnan uudistamiseen Eduskunnassa ja ministeriöissä (SGML-standardin käyttöönottoon).
- Erityisesti *rakenteistamisprojekteihin*, joissa tavoitteena rakenteisten dokumenttien käyttöön siirtyminen.

# 4. Dokumenttianalyysi

## Julkaisuja RASKE-metodologiasta ja sen käytöstä:

- Salminen, A., Lehtovaara, M., & Kauppinen, K. (1996). Standardization of digital legislative documents - a case study. In M.S. Lynn (Ed.), *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences* (pp. 72-81). Los Alamitos, CA: IEEE Computer Society Press.
- Salminen, A., Kauppinen, K., & Lehtovaara, M. (1997). Towards a methodology for document analysis. *Journal of the American Society for Information Science, Special Issue on Structured Information/Standards for Document Architectures* 48, 7 (July 1997), 644-655
- Lyytikäinen, V., Päivärinta, T., Salminen, A., & Tiitinen, P. (1997). Valtion talousarvioon liittyvien asiakirjojen rakenteistaminen. RASKE-projektin raportti. Helsinki: Eduskunta, 139 s.
- Tiitinen, P., Päivärinta, T., Salminen, A., & Lyytikäinen, V. (1997). Suomalaisten EU-lainsäädäntöasiakirjojen rakenteistaminen. RASKE-projektin raportti, Tietohallinnon selvityksiä. Helsinki: Ulkoasiainministeriö, Tietohallintolinja, 170 s.

# 4. Dokumenttianalyysi

## Julkaisuja RASKE-metodologiasta ja sen käytöstä:

- Salminen, A., Tiitinen, P., & Lyytikäinen, V. (1999). Usability evaluation of a structured document archive. In R.H. Sprague, Jr. (Ed.), *Proceedings of the Thirty-Second Hawaii International Conference on System Sciences* (file ddhfu06.pdf at CD-ROM). Los Alamitos, CA: IEEE Computer Society
- Salminen, A. (2000). Methodology for document analysis. In A. Kent (Ed.), *Encyclopedia of Library and Information Science*, Vol. 67 (Supplement 30) (pp. 299-320). New York: Marcel Dekker.
- Salminen, A., Lyytikäinen, V., Tiitinen, P., & Mustajärvi, O. (2000). SGML for e-Governance: The case of the Finnish Parliament. In A.M. Tjoa, R.R. Wagner, & A. Al-Zobaidie (Eds.), *Proceedings of the 11th International Workshop on Database and Expert Systems Applications (DEXA 2000)* (pp. 349-353). Los Alamitos, CA: IEEE Computer Society Press.

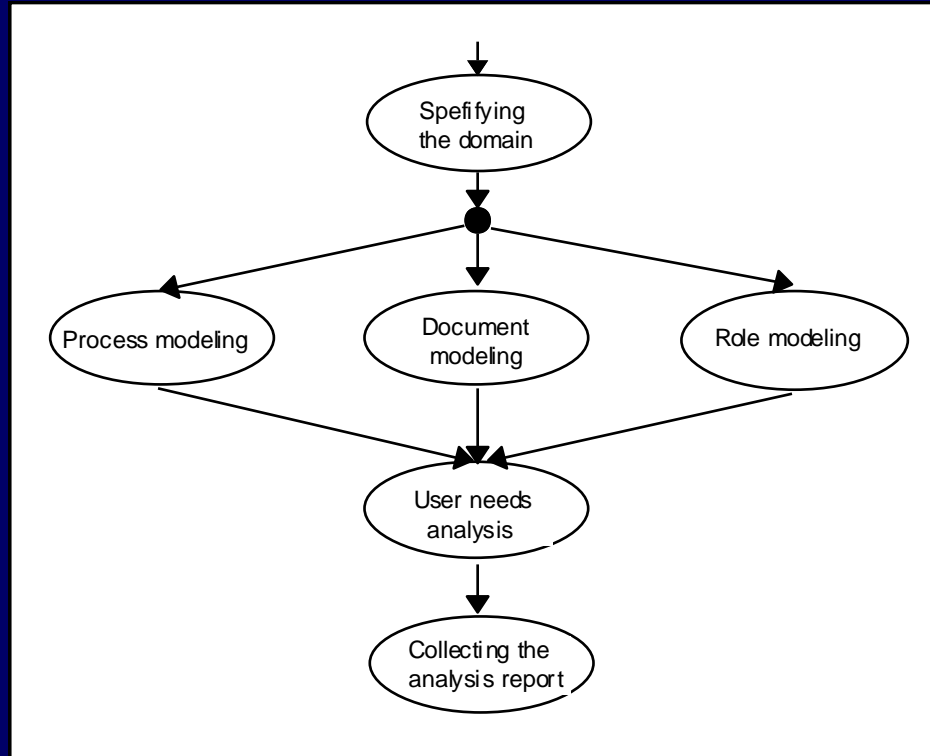
# 4. Dokumenttianalyysi

## Julkaisuja RASKE-metodologiasta ja sen käytöstä:

- **Tiitinen, P., Lyytikäinen, V., Päivärinta, T., & Salminen, A. (2000).** User needs for electronic document management in public administration: a study of two cases. In H.R. Hansen, M. Bichler, & H. Mahrer (Eds.), *Proceedings of ECIS 2000, European Conference on Information Systems* (pp. 1144-1151). Wien: Wirtschaftsuniversität Wien.
- **Salminen, A., Lyytikäinen, V., & Tiitinen, P. (2000).** Putting documents into their work context in document analysis. *Information Processing & Management* 36, 623-641.
- **Salminen, A., Lyytikäinen, V., Tiitinen, P., & Mustajärvi, O. (2001).** Experiences of SGML standardization: The case of the Finnish legislative documents. In R.H. Sprague, Jr. (Ed.), *Proceedings of the Thirty-Fourth Hawaii International Conference on System Sciences* (file etegv01.pdf at CD-ROM). Los Alamitos, CA: IEEE Computer Society.

# 4. Dokumenttianalyysi

## Dokumenttianalyysi RASKE-metodologiassa



# 4. Dokumenttianalyysin kuvausmenetelmiä

## RASKE-metodologiassa kuvauksen kohteina ovat

- organisaation tai organisaatioverkoston toiminnot
  - organisaatiot ja ihmiset toimijoina
  - dokumentit ja niihin liittyvät metatiedot
  - järjestelmät
  - organisaatioiden ja ihmisten tarpeet
- 
- Lyytikäinen, V., Päivärinta, T., Salminen, A., & Tiitinen, P. (1997). Valtion talousarvioon liittyvien asiakirjojen rakenteistaminen. RASKE-projektin raportti. Helsinki: Eduskunta, 139 s.
  - Tiitinen, P., Päivärinta, T., Salminen, A., & Lyytikäinen, V. (1997). Suomalaisten EU-lainsäädäntöasiakirjojen rakenteistaminen. RASKE-projektin raportti, Tietohallinnon selvityksiä. Helsinki: Ulkoasiainministeriö, Tietohallintolinja, 170 s.

# 4. Dokumenttianalyysin kuvausmenetelmiä

## RASKE:n rakenteistamisraportin osia:

- Tiedonkeruutavat
- Sovellusalueen yleiskuvaus
- Sovellusalueeseen liittyvä dokumentaatio ja sen käsittelijät
- Dokumenttien rakenne
- Ongelmat, tarpeet ja kehittämisajatukset

## Esimerkkiraportteja:

•Lyytikäinen, V., Päivärinta, T., Salminen, A., Tiitinen, P. (1997). Valtion talousarvioon liittyvien asiakirjojen rakenteistaminen. RASKE-projektin raportti. Helsinki: Eduskunta, 139 s.

<http://www.cs.jyu.fi/~airi/raportit/RASKE-talousarvioraportti.pdf>

•Tiitinen, P., Päivärinta, T., Salminen, A., Lyytikäinen, V. (1997). Suomalaisten EU-lainsäädäntöasiakirjojen rakenteistaminen. RASKE-projektin raportti, Tietohallinnon selvityksiä. Helsinki: Ulkoasiainministeriö, Tietohallintolinja, 170 s.

<http://www.cs.jyu.fi/~airi/raportit/RASKE-EU-raportti.pdf>

# 4. Dokumenttianalyysin kuvausmenetelmiä

## RASKE:n kuvausmenetelmiä:

- osapuolikuvaus (organizational framework description)
- dokumentaation tuottamisprosessin kuvaus (document output model)
- dokumenttien käyttö tuottamisprosessissa (document input model)



# 4. Dokumenttianalyysin kuvausmenetelmiä

## RASKE:n kuvausmenetelmiä:

- suhdekaavio (document-relationship diagram)
- tilakaavio (state-transition diagram)
- käsittelijätaulukot (document-role relationship table)
- käsittelytilannetaulukot
- rakennekuvaukset (peruskomponentit ja niiden käyttö, elm-kaaviot, DTD:t)

## 4. Dokumenttianalyysin kuvausmenetelmiä

RASKE:ssa tarpeiden analysointi ja raportointi jakaantuu seuraaviin osiin

- Tarpeet liittyen dokumentteihin
- Tarpeet liittyen tietotekniikkaan
- Tarpeet liittyen työhön dokumenttien kanssa (sisältää tarpeet liittyen ihmisiin ja organisaatioihin työssä)

Tiitinen, P., Lyytikäinen, V., Päivärinta, T., & Salminen, A. (2000). User needs for electronic document management in public administration: a study of two cases. In H.R. Hansen, M. Bichler, & H. Mahrer (Eds.), *Proceedings of ECIS 2000, European Conference on Information Systems* (pp. 1144-1151). Wien: Wirtschaftsuniversität Wien.

## 5. Yhteenveto

- **Dokumentti:** keskeiset ominaisuudet dokumenttien hallinnan näkökulmasta.
- **Dokumenttien hallinta:** suunnittelua, tuottamista, käyttöä tiettyyn tarkoitukseen.
- **Metatiedot:** hyvin suunniteltu metatietojen hallinta edellytyksenä dokumenttien hallinnalle.
- **Metatietojen pitäisi olla mahdollisimman pitkälle järjestelmäriippumattomia.**

## 5. Yhteenveto

- **Rakenteiset dokumentit:** standardoitu, ihmisten ja tietokoneen tulkittavissa oleva muoto tietylle sovellusalueelle tai tiettyyn käyttötarkoitukseen.
- **Dokumenttianalyysi:** edellytys dokumenttien hallinnan uudistamiselle organisaatiossa tai organisaatioverkostossa.
- **RASKE-metodologia:** menetelmiä dokumenttianalyysiin.

## 5. Yhteenveto

- RASKE:n kuvauskohteina toimintoprosessit, organisaatiot, ihmiset ja heidän työtehtävänsä, tekniikka, dokumentit, ihmisten tarpeet työtehtävissään, organisaatioiden tarpeet tavoitteidensa suorittamisessa.

## 5. Yhteenveto

Lisätietoja Jyväskylän yliopiston dokumenttien hallinnan tutkimuksesta:

Document Management Research  
<http://www.cs.jyu.fi/~airi/docman.html>

Esityksessä mainituista artikkeleista osa saatavana elektronisessa muodossa mainitulta sivulta.