

Dokumenttien hallinta & Tietojohtaminen
3. rinnakkaisteema:

Dokumenttien tietosisällön hallinta

Ossi Nykänen

Tampereen teknillinen yliopisto (TTY),
Digitaalisen median instituutti (DMI),
W3C Suomen toimisto



- Dokumenttien tietosisällön hallinta perustuu kykyyn ennakoida ja hallinta tietosisältöjä tuottavia ja niitä hyödyntäviä (osin tietoteknisiä) prosesseja
 - *Hyvä kysymys on tietenkin se, miten tämä tapahtuu*
- Konferenssi: keskiviikko
 - Rakenteiset dokumentit ja sisällönhallinta
 - Mediasisällön hallinta Alma Mediassa
 - Dokumenttienhallintajärjestelmän tekninen toteutus
- Konferenssi: torstai
 - Metatieto asiakirjojen hallinnan välineenä
 - Arkistolaitos ja case SÄHKE
 - Väestörekisterikeskus ja case henkilökortti ja allekirjoitus
 - Dokumenttien hallinnan juridiikkaa

Rakenteiset dokumentit ja sisällönhallinta

- Hyvin lyhyt versio:
 - Rakenteinen dokumentaatio on yleispätevä menetelmä tietosisällön hallintaan tietoteknisiä välineitä käyttäen. World Wide Web Consortium kehittää universaaleja teknologioita joille löytyy runsaasti sovelluksia myös rakenteisessa sisällönhallinnassa
- Pidempi versio, otsikoita:
 - Konteksti: World Wide Web Consortium
 - Lähtökohta: rakenteinen dokumentaatio osana sisällönhallintaa
 - Miten: rakenteiset dokumentit, peruskäsitteet
 - ”Isommat ideat:” tuotantoprosessit ja metatieto
 - Lopuksi: W3C-suositukset dokumenttienhallinnassa
- Esityksen tavoite:
 - Esittää rakenteisen dokumentaation perusidea, luonnehtia dokumentaatioprosessia ja sijoittaa aihepiirin tekniset W3C-suositukset kontekstiinsa

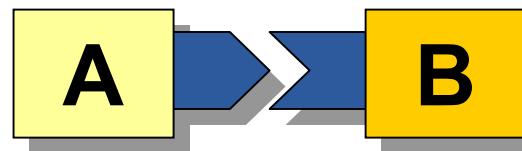
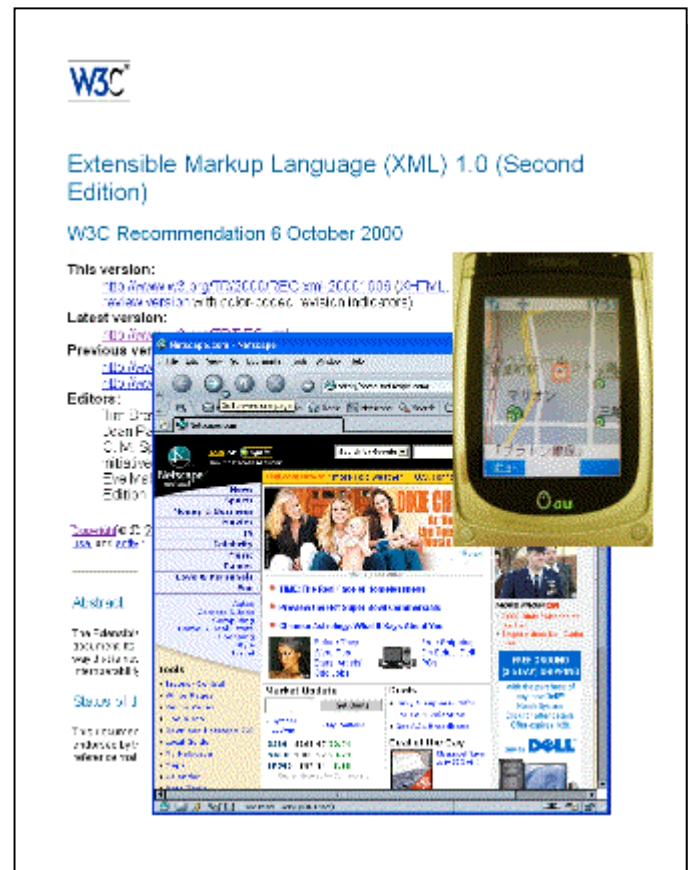
- World Wide Web Consortium
- ...ja sen toiminta
- W3C-suositukset = yleisesti käyttökelpoisia (Web-) teknologioita

- W3C kehittää yhteensopivia (Web-)teknologioita (spesifikaatioita, ohjeita ja ohjelmistoja) ja siten *ohjaa Webin kehittymistä täyteen mittaansa*
- 3 päätstoa, 14 aluetstoa, n. 400 jäsenorganisaatiota



- **Jäseneksi?** – ks. <http://www.w3c.tut.fi/joining.html>

- Vaikutusvaltainen **foorumi**
- **Suosituks**
 - XML-perhe, XHTML-perhe, SVG, RDF, Semanttinen Web, Web-palvelut ...
- **Ohjelmistot**
 - Validaattorit, testiaineistot, open source, ...
- **Ohjeet ja yhteydet**
 - WAI, aluetoimistot, ...
- **Jäsenorganisaatiot**
 - "5 Is": Information, Image, Involvement, Influence, and Impact



W3C:n (nykyiset) toimialueet ja aktiviteetit



<i>DOM</i>	<i>Internationalization</i>	<i>URI</i>	<i>Web Services</i>	<i>XML</i>
<i>Device Independence</i>	<i>Graphics</i>	<i>HTML</i>	<i>Math</i>	<i>Multimodal Interaction</i>
<i>Style</i>	<i>Synchronized Multimedia</i>		<i>Voice Browser</i>	<i>XForms</i>
<i>Patent Policy</i>	<i>Privacy</i>	<i>Semantic Web</i>	<i>XML Key Management</i>	
<i>WAI International Program Office</i>			<i>WAI Technical Activity</i>	
<i>QA</i>				

(Ks. <http://www.w3.org/Consortium/Activities.>)

- W3C:n työllä ja dokumenttien hallinnalla on selviä yhtymäkohtia
 - **rakenteiset dokumenttien käsittely** (XML-perhe, sis. erit. XML 1.0, CSS, XSL/FO, XSLT, XLink, XHTML, ...),
 - **grafiikka** (PNG, SVG, SMIL),
 - **infrastrukturi** (ohjelmointi, jakelu, julkaiseminen, metatiedot, autentikointi, ...)
 - ...
- W3C:llä ei kuitenkaan ole **Content Management** -aktiviteettiä; monet W3C-suositukset kuitenkin käytännössä pohjustavat tähän liittyvää työtä
 - **W3C luo yhteistä perustaa jonka varassa yhteensopivia ja (teknisesti) järkeviä sovelluksia voidaan toteuttaa**



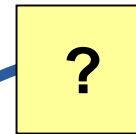
XML 5v 10.2.2003

- Miksi (tavoitteet)
- Teknologia vai prosessi (lue: prosessi)

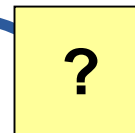
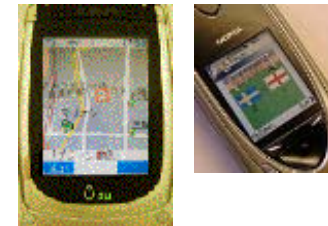
- Dokumenttien hallinnan **tehostaminen**
 - automatisointi ja integrointi
- **Monta sovellusta** samalle sisällölle:
 - prosessorit, tyyli, alustat, ...



Sovellus #1



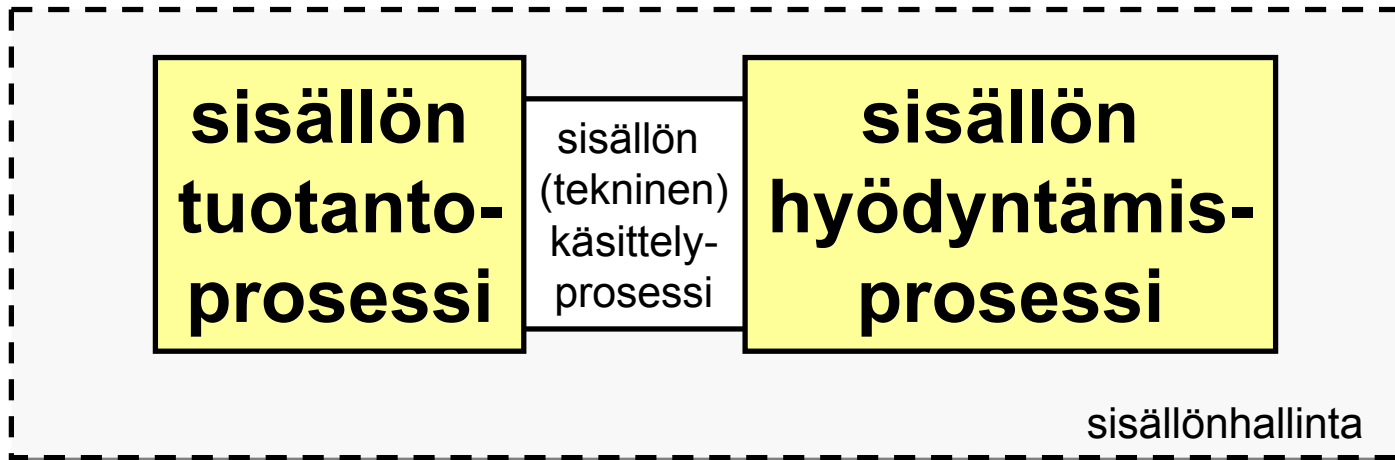
Sovellus #2



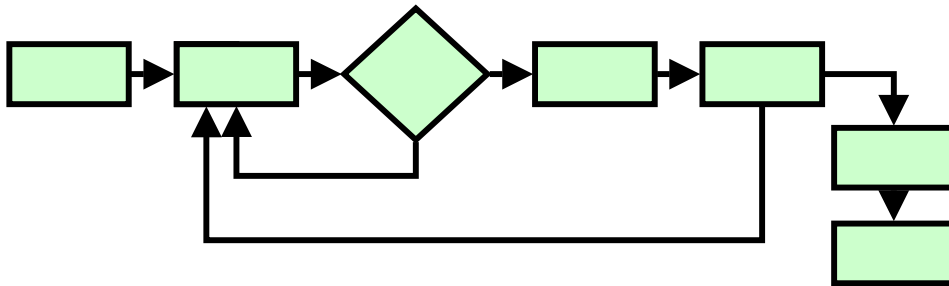
Sovellus #3



Näkökulma: sisällönhallinta = prosessien hallintaa



- Dokumenttituotannosta sisällönhallintaan



- Sisällöntuotanto on hidasta ja kallista
- Ohjelmistot vaihtuvat nopeammin kuin niillä tuotetut sisällöt
- Tekijät vaihtuvat nopeammin kuin sisällöt uudistuvat

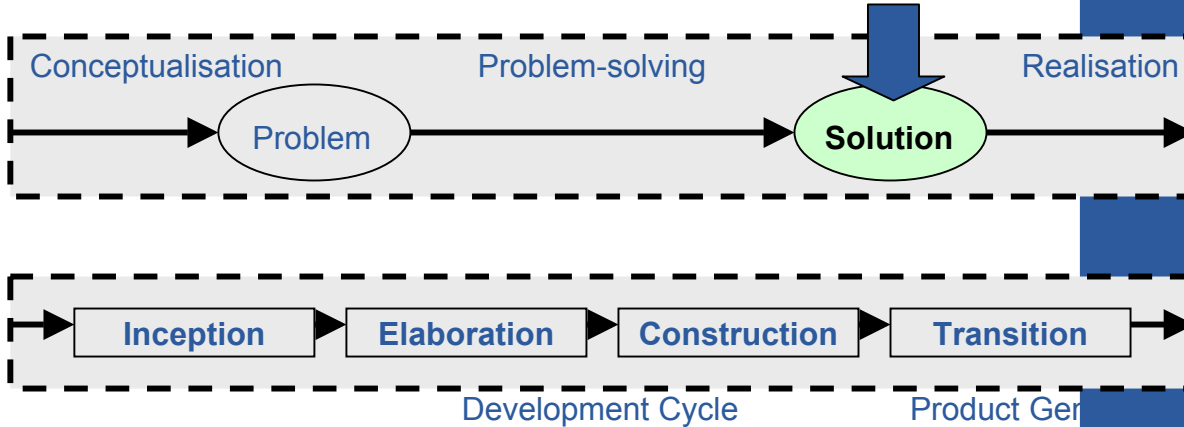
- **Sisällöntuotanto on hidasta ja kallista**
 - tuotantotyö pitää suunnitella huolella
 - kerran tuotettu sisältö tulee tallettaa helposti hallittavassa ja ylläpidettävässä muodossa (→*yhteensopivuus, rakenteisuus, muokattavuus*)
- **Ohjelmistot vaihtuvat nopeammin kuin niillä tuotetut sisällöt**
 - tiedon talletusmuoto ei saa vanhentua
 - sitoutuminen yhden valmistajan tuotteisiin/tiedostotyyppiin on riskialtista
- **Tekijät vaihtuvat nopeammin kuin sisällöt uudistuvat**
 - tekijänoikeuksista yms. pelisäännöistä on sovittava alusta asti
 - työn dokumentoinnista on huolehdittava ajoissa
 - ...koskee myös prosessien dokumentointia

- **Sisällönhallinnan** suurimmat ongelmat syntyvät **tuotannon** vääristä odotuksista ("lyhytnäköisyys")
 - *"kunhan vain saadaan tietokoneelle" -ajattelu ei toimi!*
- Hyvä tuotantoprosessi on huolellisen suunnittelun tulos
 - **suunnittelun lähtökohta on työn lopputuloksena syntyvien tuotteiden määrittely**
- ...suunnittelun perusedellytys on tuotantoprosessin eri vaiheiden tuntemus (ja suhteuttaminen organisaatioon)
 - tekninen kirjoittaminen (menetelmät ja välineet)
 - formatointi, taitto, print yms. (menetelmät ja välineet)
 - jakelu, ylläpito ja päivitykset (menetelmät ja välineet)
- Reunaehdot, rajoitteet ja rajapinnat
 - profilointi? on-line koostaminen? hajautus?
(tuotanto/hyödyntäminen) arkistointi? numeerinen laskenta?

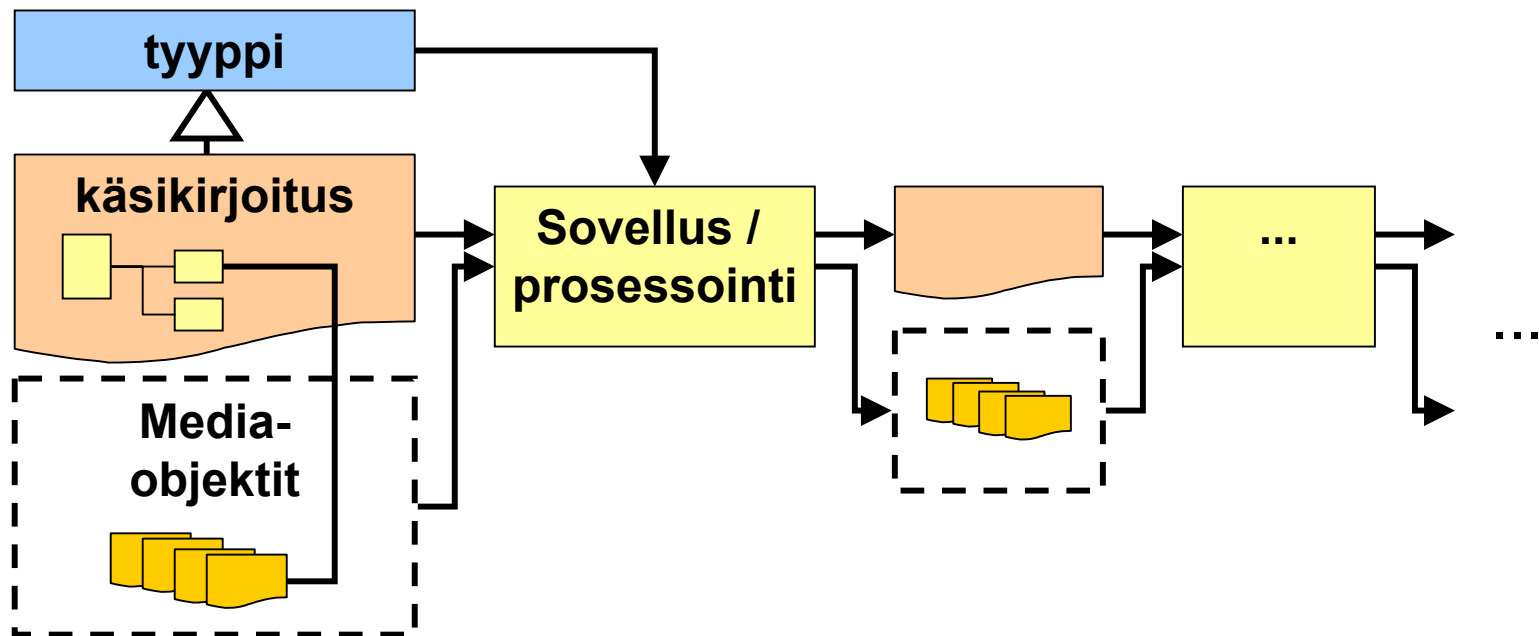
<_XML /> to the Rescue?

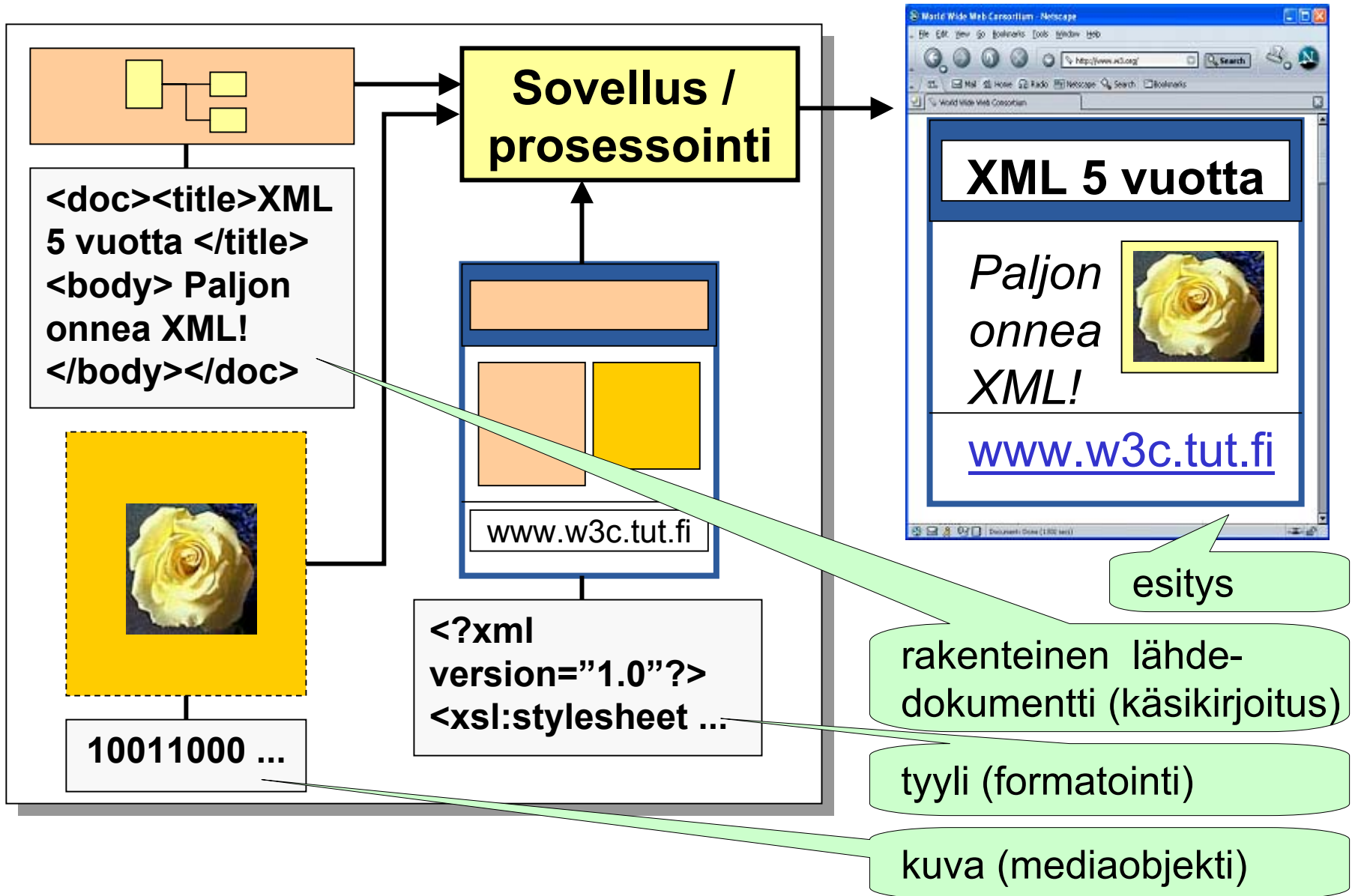
(...mutta miksi? ...ja miten?)

- Rakenteiset dokumentit

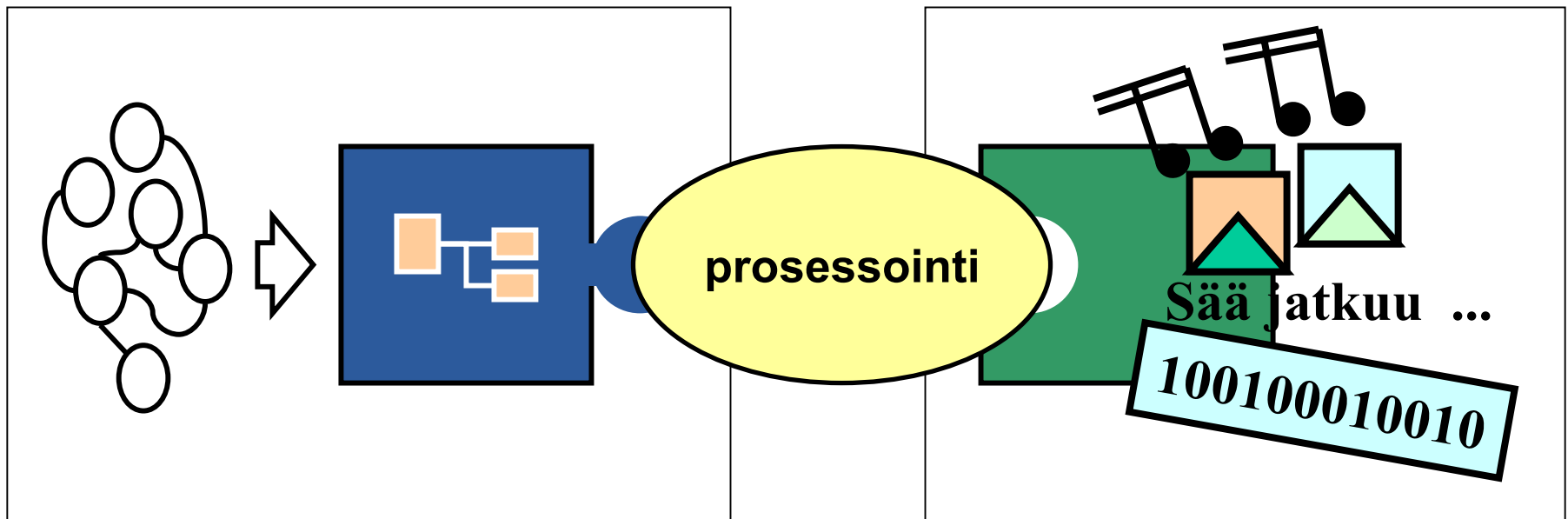


- Lähtökohta **tiedon ja sen sovellusten erottelu**
 - merkkkaus, sanasto, sanaston kielioppi ja sovellus
- Hyödytön sellaisenaan -- **prosessointi** sovelluksessa
- Kaikki data ei rakenteellista -- mukana **mediaobjekteja**



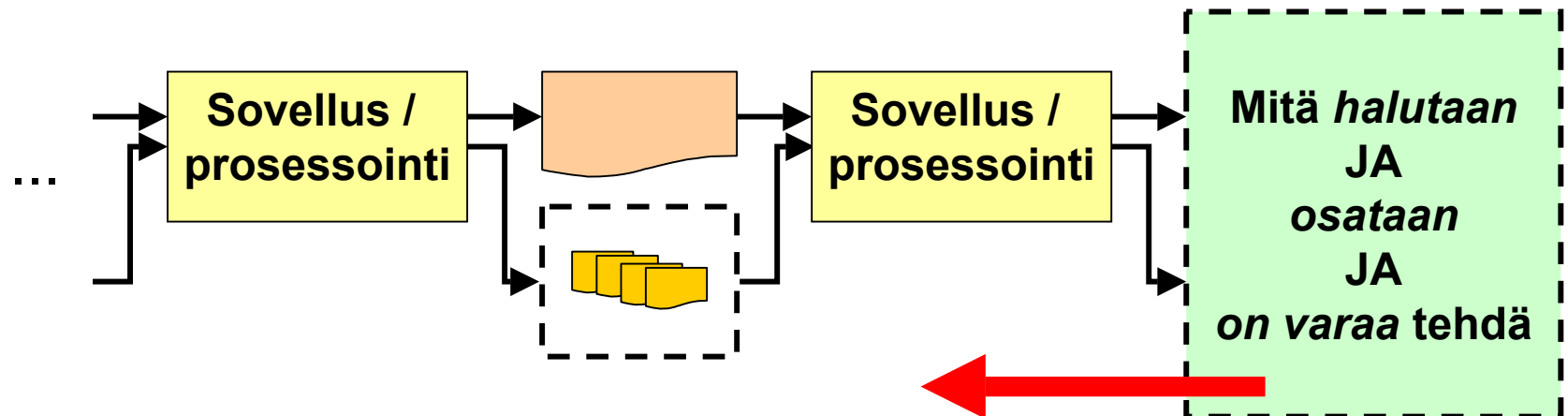


- Rakenteisten dokumenttien (tiedon) käsittelyyn liittyy aina kaksi toisiaan täydentävää näkökulmaa
 - tiedon **mallintaminen** ja
 - sen **esittäminen sovelluksessa** (yleisemmin: tulkinta)



- **Tiedon mallintamisen** perustehtävä on tiedon kuvailurakenteen (skeeman) määrittäminen sen **prosessoinnin näkökulmasta**
 - haaste: mallin rikkaus vs. käytön taloudellisuus
- **Tiedon esittämisen** perustehtävä on sovelluksen **sisällöllisen tavoitteen** toteuttaminen
 - haaste: ylläpito ja versiointi
- Rakenteinen dokumentaatio erottelee näkökulmat sekä teknisesti että työvaiheiden tasolla: molemmat on otettava suunnittelussa huomioon

- Ensimmäinen tehtävä: hyvän **lopputuotteen määrittely**
 - sisällöllisesti tyydyttävä (**tehtävä ja tavoitteet**)
 - teknisesti toteuttamiskelpoinen (**reunaehdot**)
- Puhtaasti teknologiavetoinen lähestymistapa johtaa väärään ratkaisuun ("tehdäänpäs XHTML 1.0:lla")
 - mitä spesifimpiä sisällölliset tavoitteet ovat, sen parempi



- **Yleistäminen**

- kuvaa tieto luonnollisten käsitteiden avulla (**mallinnus**)
- irrota sovelluskohtainen tieto käsikirjoituksista (vrt. **tyylit**)
- tunnista eri prosessien väliset rajapinnat (**abstrahointi**)

- **Jako komponentteihin**

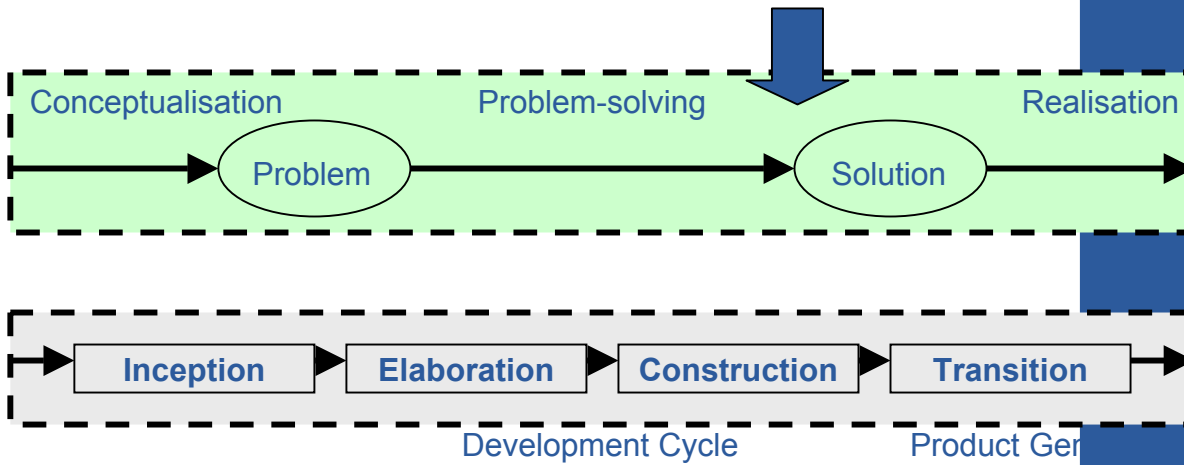
- tunnista dokumenttihierarkia ja nimeä sen osat (**organisointi**)
- kuvaa kaikki mediaobjektit käsikirjoituksen osaksi (**kääreet**)

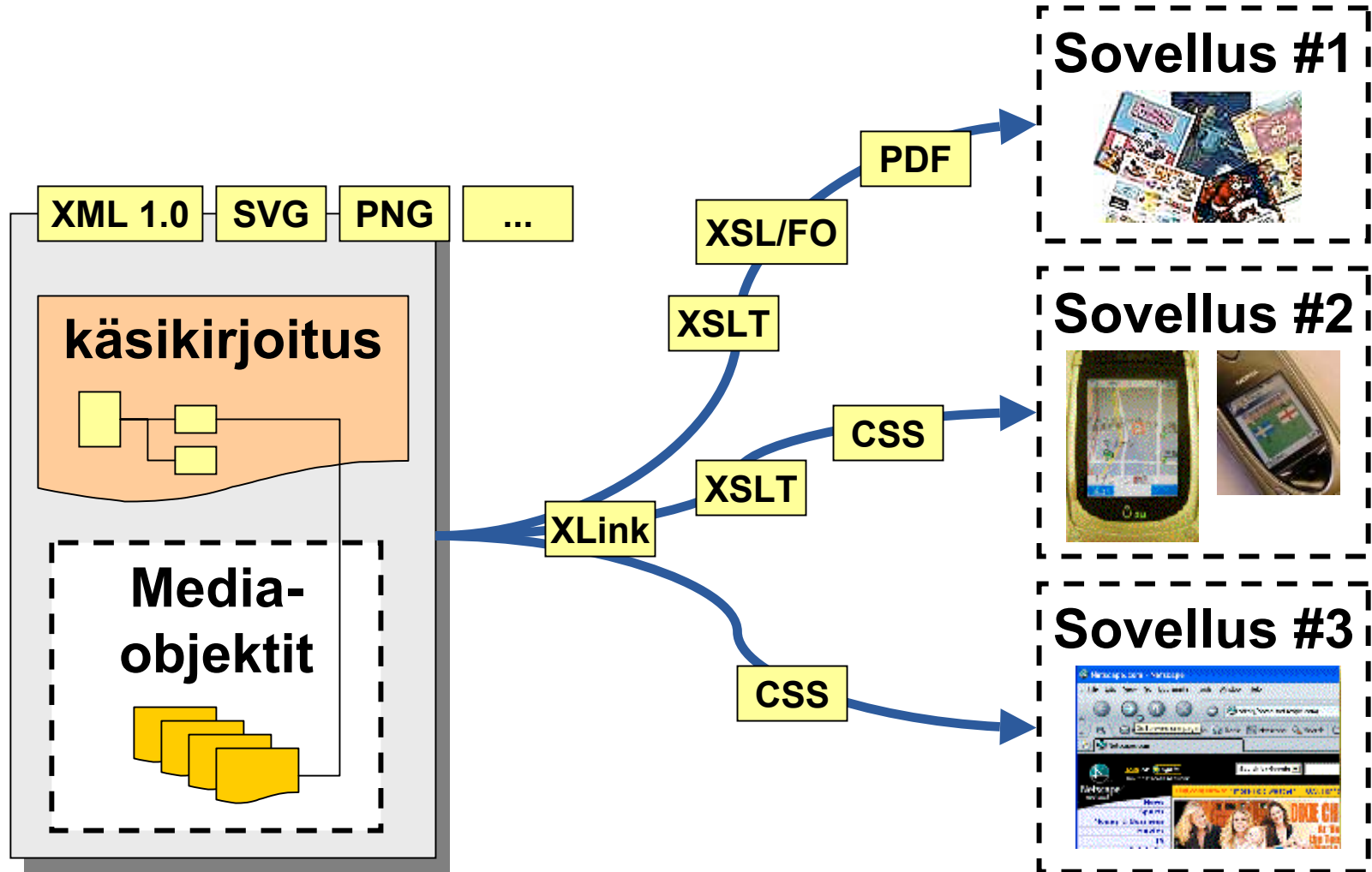
- **Kartoitukset ja kokeilut**

- määrittele lopputuote tarkasti (**tuotannon tavoite**)
- testaa kaikki prosessien rajapinnat (**tuotantojärjestelmä**)
- kerää kokemuksia ja kehitä (**kehitys & sitouttaminen**)
- mukauta organisaatiolle sopivaan muotoon (**käyttöönotto**)

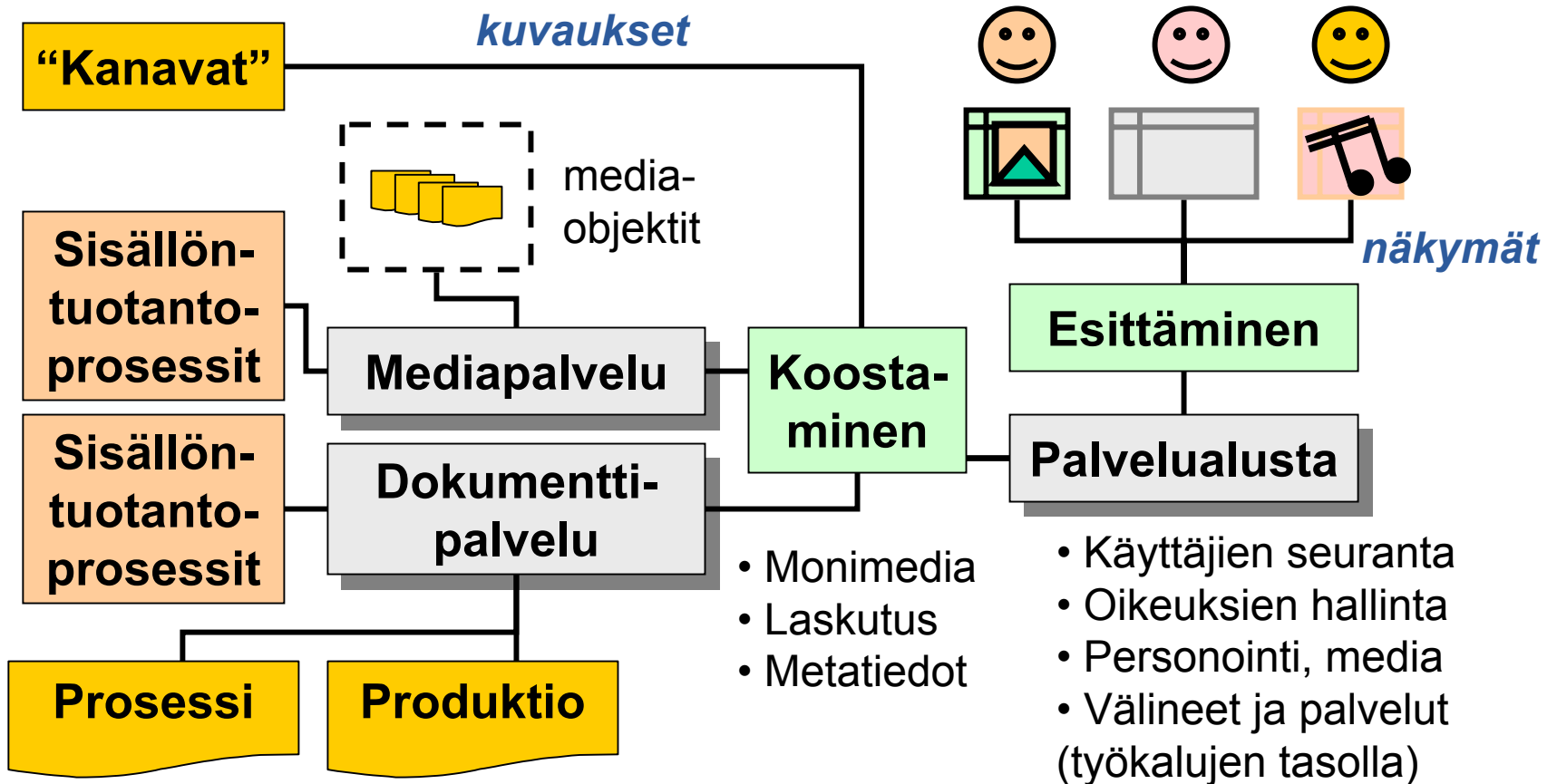
- **Määrittelytekniset** riskit
 - kuultiinko kaikkien prosessien edustajia?
 - unohtuiko oleellisia tavoitteita (*ylläpito, arkistointi, oikeudet, muutokset, versiointi, jatkokehitys, yhteensopivuus(!), ...*)?
- **Poliittiset** riskit
 - ovatko esitetyt tavoitteet ja resurssit todellisia?
- **Tiedolliset ja taidolliset** riskit
 - onko tietotaitoa, entä kun Erkki Ekspertti vaihtaa työpaikkaa?
 - kuka kouluttaa soveltajat – muuttuuko työkulttuuri?
- **Teknologiaan** liittyvät riskit
 - onko tekniikka kypsää?
 - toteuttavat työvälineet halutut kaikki aiotut skenaariot?
 - sopivatko ratkaisut tuotantokäyttöön (*ilmaisohjelmat...*)?

- Tuotantoprosessit

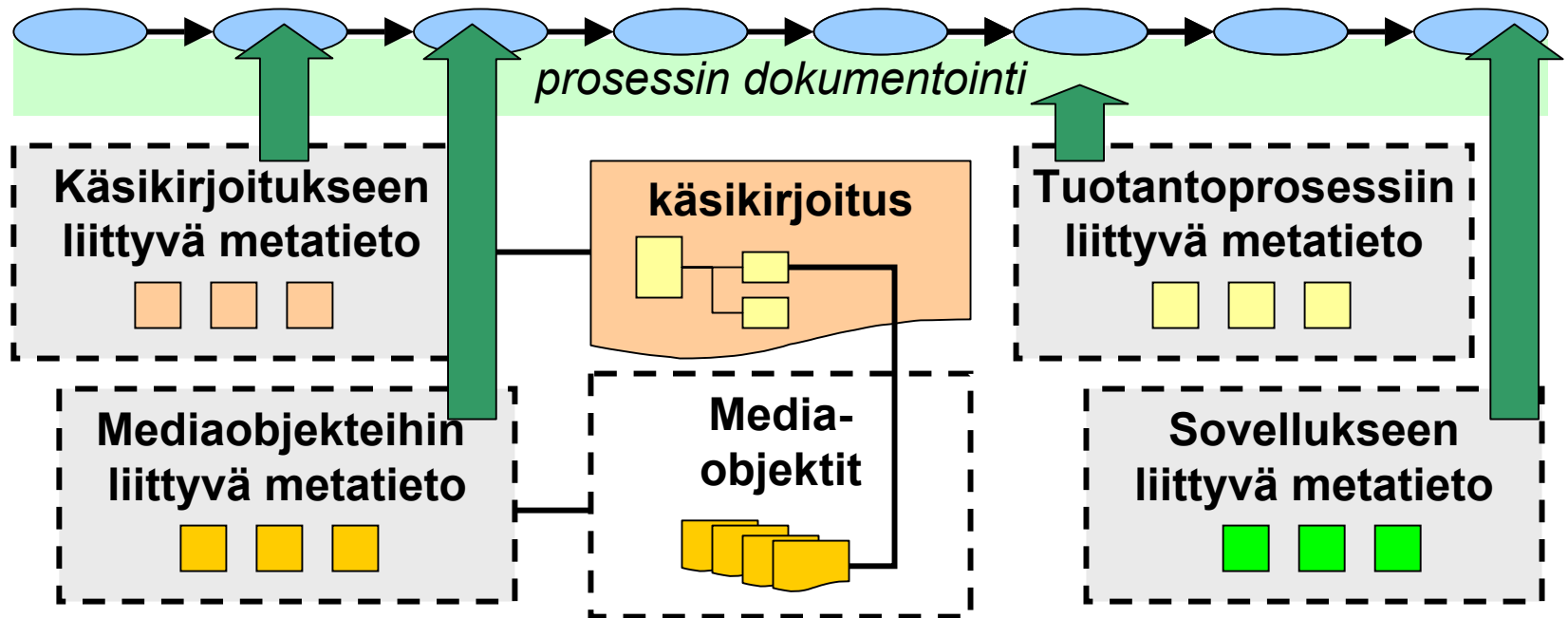




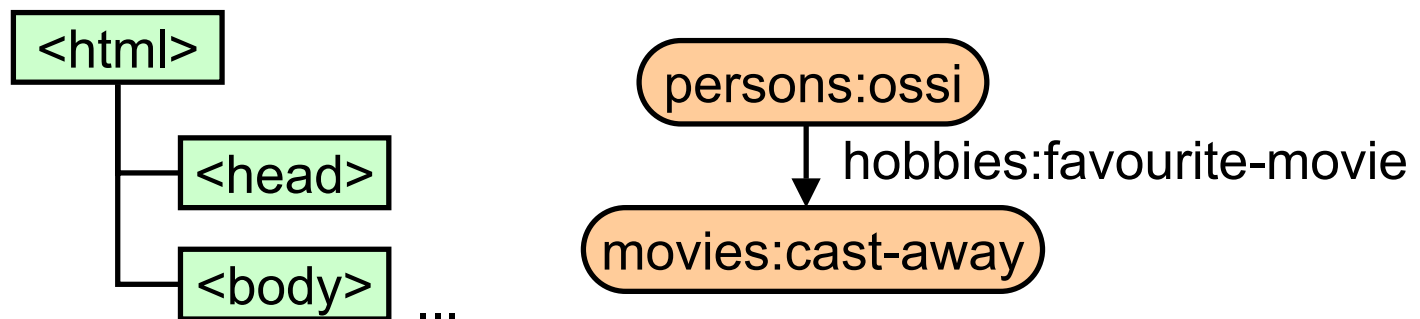
- Sama perusidea, **prosessi** ja **käyttäjät** mukaan kuvaan



- Dokumenttistandardit eivät automaattisesti koodaa/tue niiden tuotamisen prosesseja -- tarvitaan lisätietoja
 - prosessien ja työvaiheiden **dokumentointi**
 - tuotantoprosessin, sovelluksen ja lähdeaineiston **metatiedot**

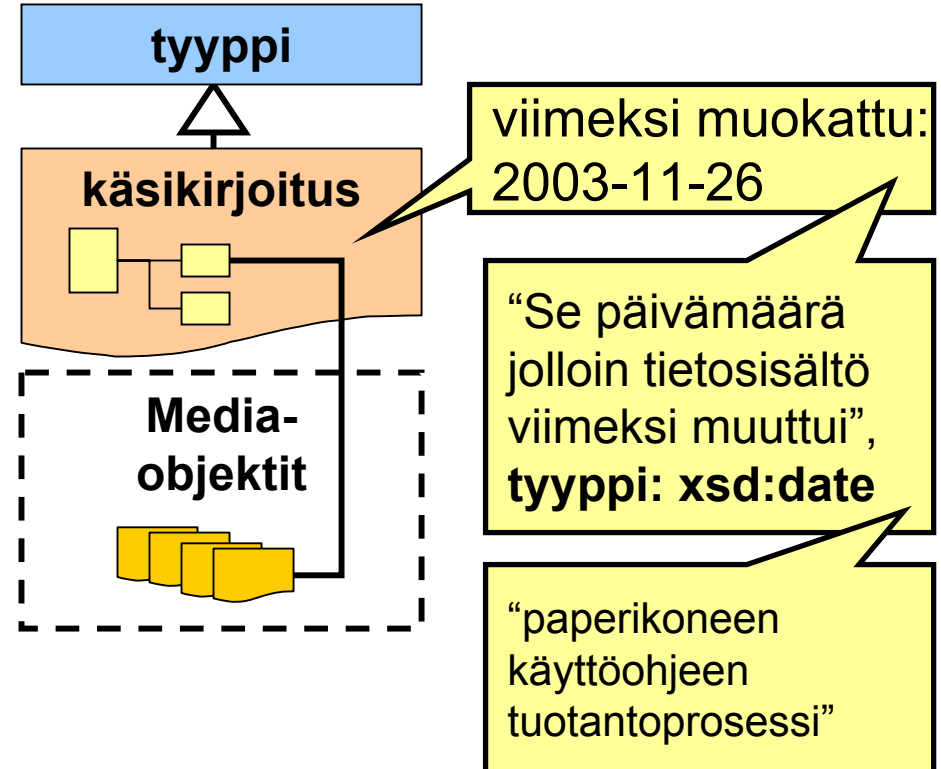


- ”Yksinkertaiset” sovellukset
 - haut, näkymät (profilointi) ja ”tiedon ulkoistaminen”
- ”Edistyneemmät” sovellukset
 - integrointi, päättely, analysointi, business-äly yms.
- Kaksi lähestymistapaa
 - ”attribuuttipohjainen metatieto”
 - tiettyihin sovelluksiin kohdennetut ontologiat
- **Tavoite: lisätä tiedonhallinnan automatisointia**
- Mallinnuksen rooli – tietorakenne vs. tieto:

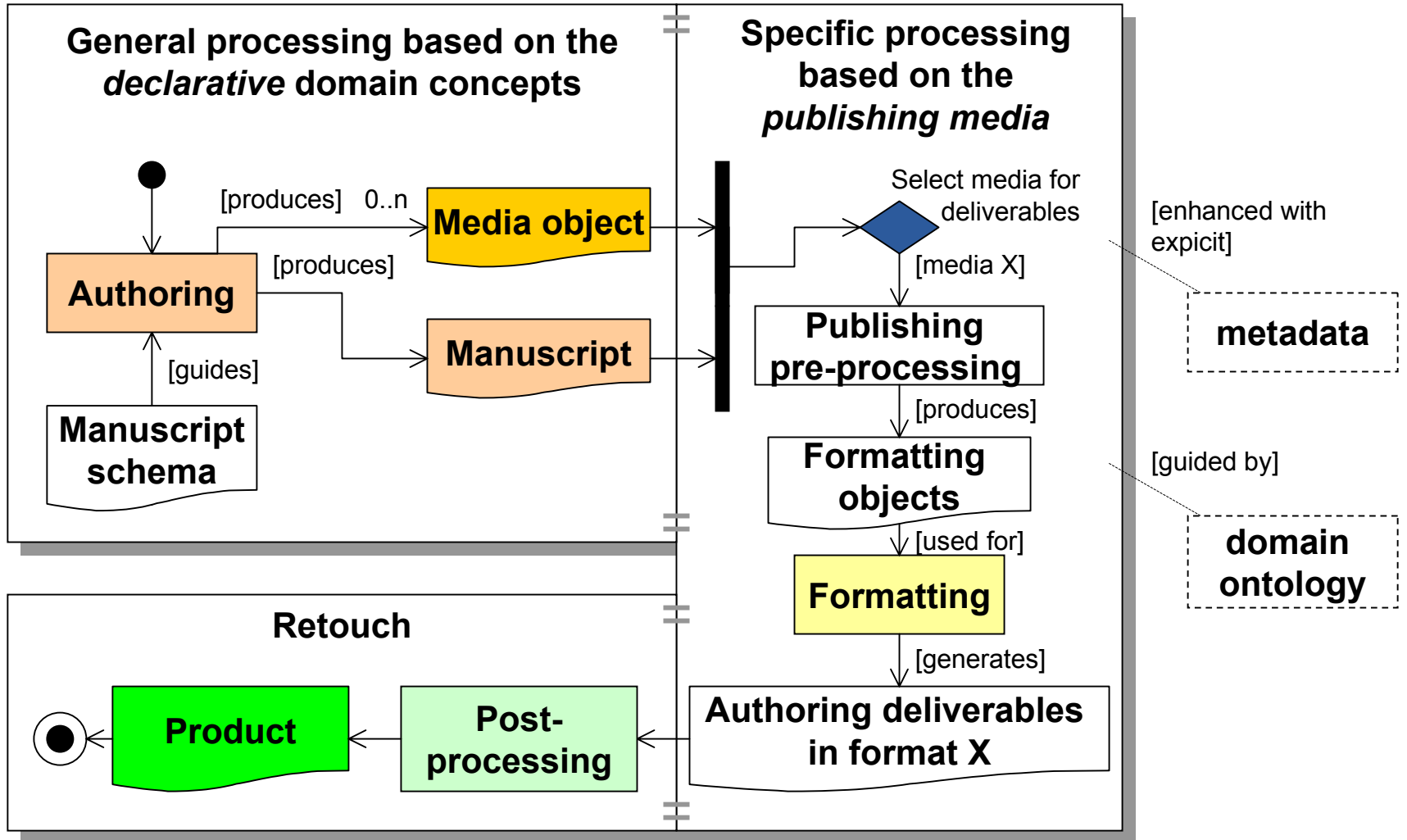


Mikä on metatietoa?

- Metatieto on tietoa jota ei löydy tai jota ei voida helposti poimia käsikirjoituksista tms.
- Metatieto ja tieto -erottelua ei ole pakko lainkaan tehdä
 - dokumenttien tyyppitiedot
 - tekstiformaatti ja ko. formaatin kontrolloitu sanasto
 - ”lisätty, kuvaileva (meta)tieto”
 - metatiedon käyttöä ohjaava skeema (tms.)
 - sovellusalueen ontologia
 - jne.



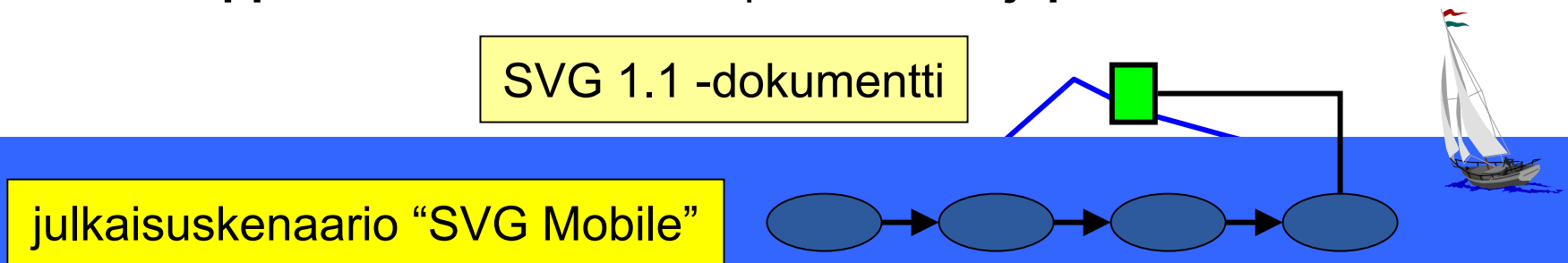
Yksinkertaistus: sisällöntuotantoprosessi



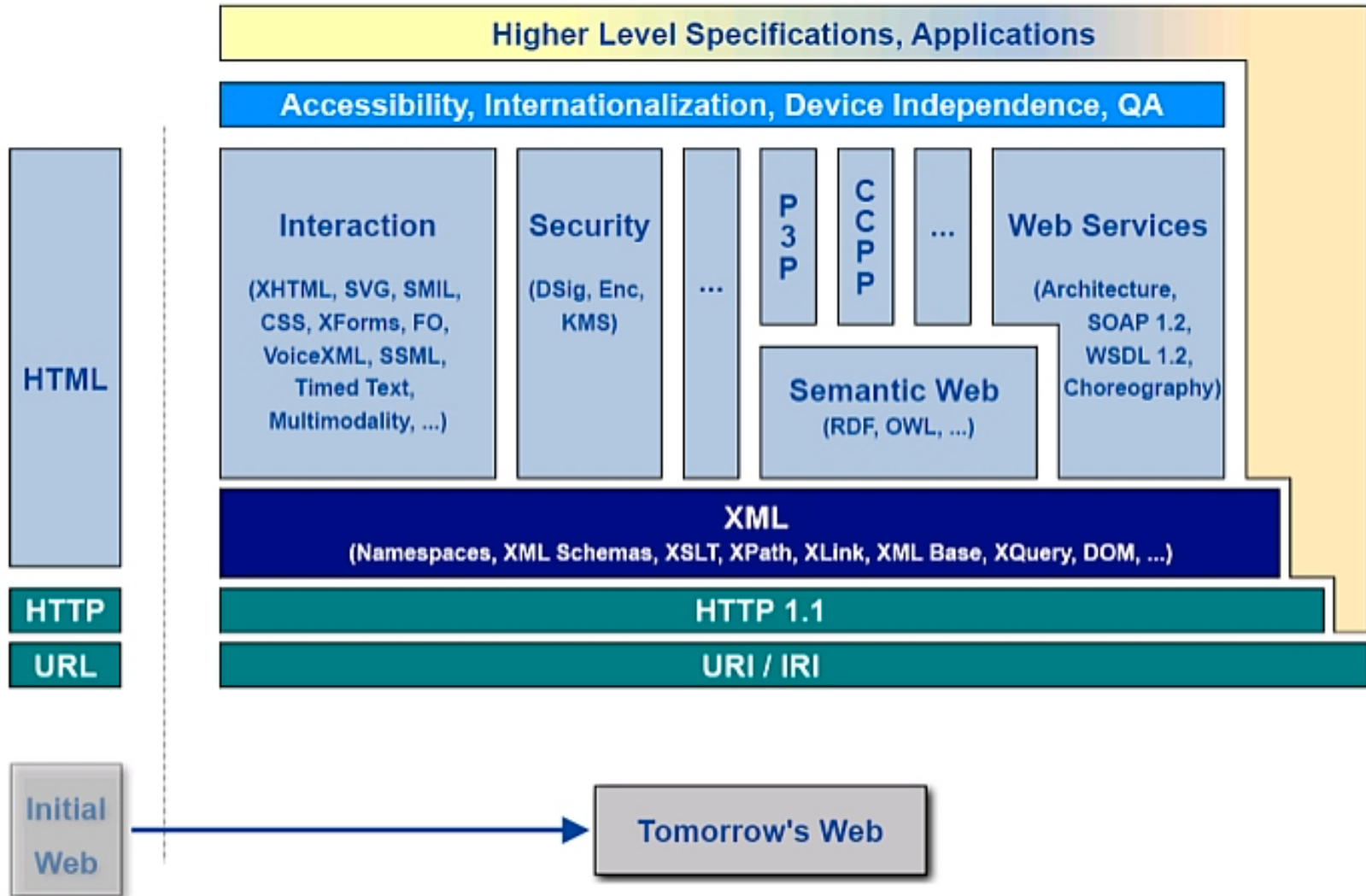
- Tuote löytää käyttäjänsä osuvan mielikuvan perusteella
 - **sähköposti** toimii käsitteellisesti ”posti”-metaforan varassa, mutta käyttöliittymän takana toteutus on hyvin erilainen
- Sama pätee **tuotannon** organisointiin ja ohjeistukseen
 - esim. teknisen kirjoittamisen ja ylläpidon työtehtävät
- **Tuotantoprosessin ohjeistuksen käsitteiden ei tarvitse noudattaa prosessin teknistä suunnittelua!**
 - ~~XML, XSLT, ...~~
 - optimiratkaisu on peilata käsitteet ja työvaiheet oman organisaation työkuultuuriin kautta
- *Ekstraspektio inkrementoi adekvaattia ekspertiisiä +([:-)*

- W3C-suositukset
- Standardien hyötynäkökohtia
- Liitteet

- Sisällönhallinta & W3C-aktiviteetit, yhtymäkohtia:
 - I18N, URI, **XML**, XLink, Device Independence, **Graphics**, XHTML, Math, Multimodal Interaction, **Style**, Synchronized multimedia, Voice Browser, XForms, **Semantic Web**, Web Services, XML Encryption, XML Signature, **WAI** Technical Activity, ...
- Suuntaa-antava piirros: **standardinomaisuus**
 - W3C-suositukset eivät (suoraan) osoita **prosesseja** jotka esim. tuottavat ”hyviä sovelluksia”, vaan määrittelevät prosessien **lopputuotteen** tai kahden prosessin **rajapinnan**



Yhteensopivuus syntyy kokonaisuudesta

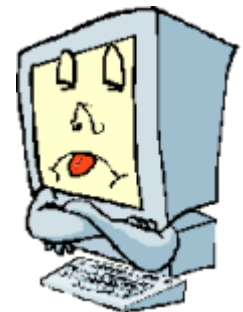


- Valmiiksi mietittyjä **rajapintoja** ja **geneerisiä ratkaisuja**
- **Standardeihin** nojautuminen
 - yhteensopivuus, pysyvyys ja hallittavuus, käyttöehdot
 - laajempia markkinoita, vaihtoehtoisia toimittajia
- **Rakenteisuuden etuja**
 - tuotantoprosessin hallinta (sisältö-esitystapa-sovellus)
 - monikanavajulkaiseminen, profiloitavuus, haut, yms.
- XML-rajapinnan kautta **uusia sovelluksia**
 - esim. metatiedot, salaus ja allekirjoitukset
- **Välineitä** ja 3. osapuolten **caseja** (joista ottaa oppia)
- ...

- Jäikö joku W3C-asia mietityttämään? Allekirjoittaneen tavoittaa helposti:
 - <http://www.w3c.tut.fi>
 - Ossi@w3.org

“The bane of my existence is doing things that I know the computer could do for me.”

– Dan Connolly, The XML Revolution



- W3C
 - <http://www.w3.org/> (*kotisivu*)
 - <http://www.w3c.tut.fi/> (*W3C Suomen toimiston kotisivu*)
- W3C-aktiviteetit
 - <http://www.w3.org/Consortium/Activities> (*yleiskuva*)
- W3C:n suositukset ja raportit sekä teknistä tietoa
 - <http://www.w3c.org/TR/> (*specs*)
 - <http://www.w3.org/QA/TheMatrix> (*spec-tietoa jäsennettynä*)
- Jäseneksi!
 - <http://www.w3c.tut.fi/joining.html>
 - <http://www.w3c.org/Consortium/Prospectus/Joining>

Liite 2: uusia ja tulevia(?) suosituksia

- Recent RECS
 - MathML Version 2.0 (2nd Edition)
 - XML Events
 - XForms 1.0
 - Soap 1.2
 - XPointer
 - SVG 1.1
 - SVG Tiny, SVG Basic
 - XML Encryption/Decryption
 - XML Signature
 - ...
- Proposed RECS:
 - CC/PP Structs & Vocab 1.0
 - PNG (2nd Edition)
- Candidate RECS
 - XForms 1.0 Basic Profile
 - OWL (various)
 - DOM Level 3 Validation, XPath
 - CSS3: TV 1.0, Color, Ruby, Text, Mobile 1.0, Media Queries, Selectors
 - VoiceXML 2.0
 - XML 1.1, XML Namespaces 1.1
 - XInclude 1.0
 - Speech Recognition Grammar 1.0
 - XML Fragment Interchange
- Muuta mielenkiintoista
 - WSDL 1.2, XQuery 1.0, XKMS, DOM Level 3, RDF (various), ...
- Lisää, ks. <http://www.w3.org/TR/>

- XML-pohjaiset lomakkeet: XForms (REC 14.10.2003)
 - HTML:ää monipuolisempi
 - logiikan, esitystavan & datan erottelu
 - XML Schema -tietotyypit & tarkistukset
 - jo paljon toteutuksia, Basic-profiili tulossa
 - täysin XML-pohjainen
 - laiteriippumattomuus, saavutettavuus
 - otettavissa heti käyttöön
- Ks.
 - <http://www.w3.org/MarkUp/Forms/>
 - <http://www.w3.org/MarkUp/Forms/2003/xforms-faq.html>
 - <http://www.w3.org/MarkUp/Forms/2003/xforms-for-html-authors.html>

