Avoimet standardit ja arkistointi

Ossi Nykänen
ossi@w3.org

Tampereen teknillinen yliopisto (TTY)
Hypermedialaboratorio
W3C Suomen toimisto

Esitelmä

Hyvin lyhyt versio:

- World Wide Web Consortium (W3C) kehittää ja standardoi keskeisiä Web-teknologioita. W3C-työ luo siten osaltaan puitteet myös Webin arkistointityölle

Pidempi versio, otsikoita:

- W3C ja universaalit Web-standardit
- Web-teknologiat ja arkistointi
- Dokumenttien arkistointi vs. Web-sovellusten arkistointi

Esityksen tavoite on esitellä W3C-työtä ja lyhyesti luonnehtia Web-teknologioiden mahdollisuuksia arkistointityön näkökulmasta

World Wide Web Consortium (W3C)

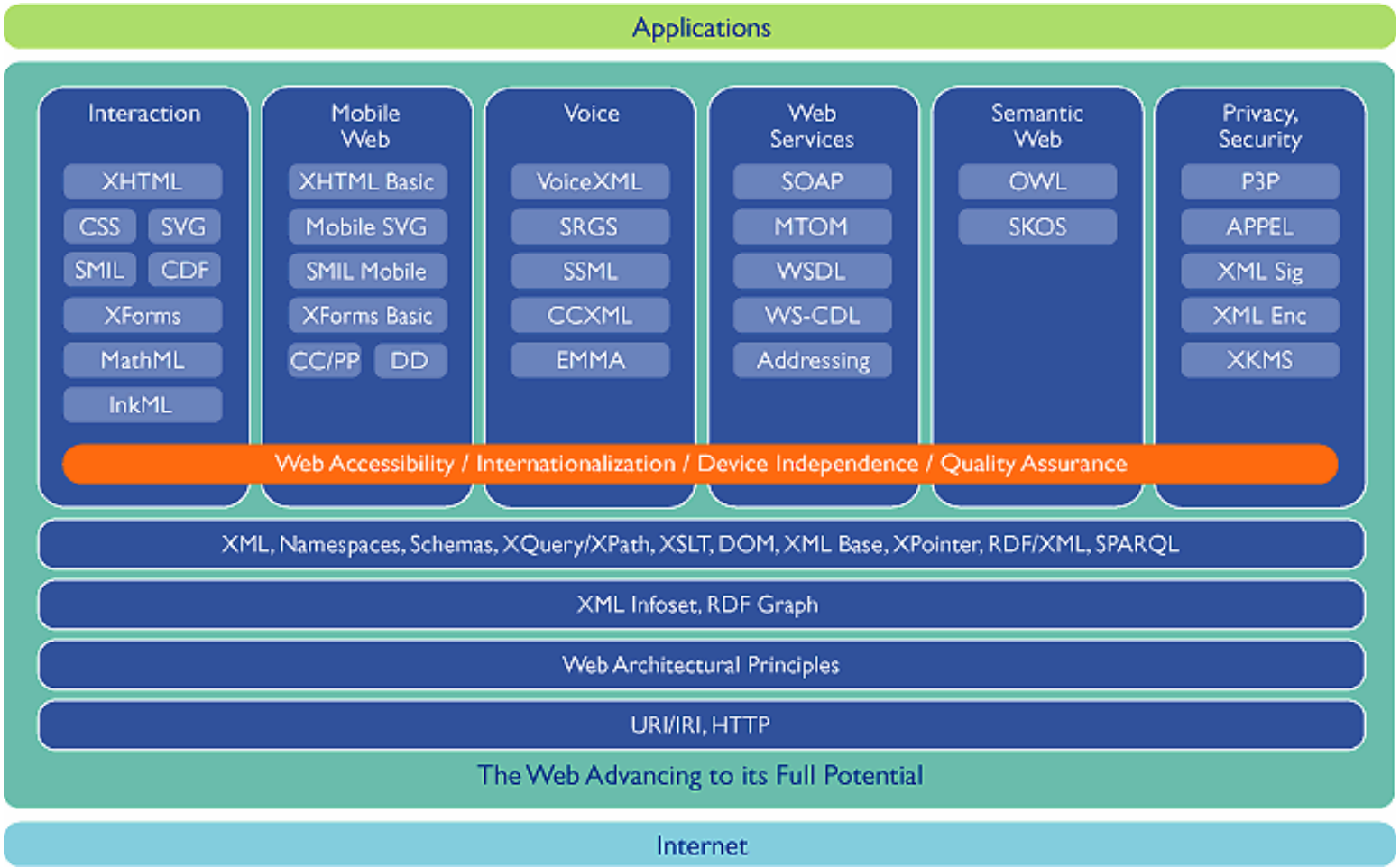
W3C kehittää yhteensopivia teknologioita ja siten ohjaa Webin kehittymistä täyteen mittaansa

- ...asettamalla teknisiä suosituksia (esim. HTML, XML, WAI)

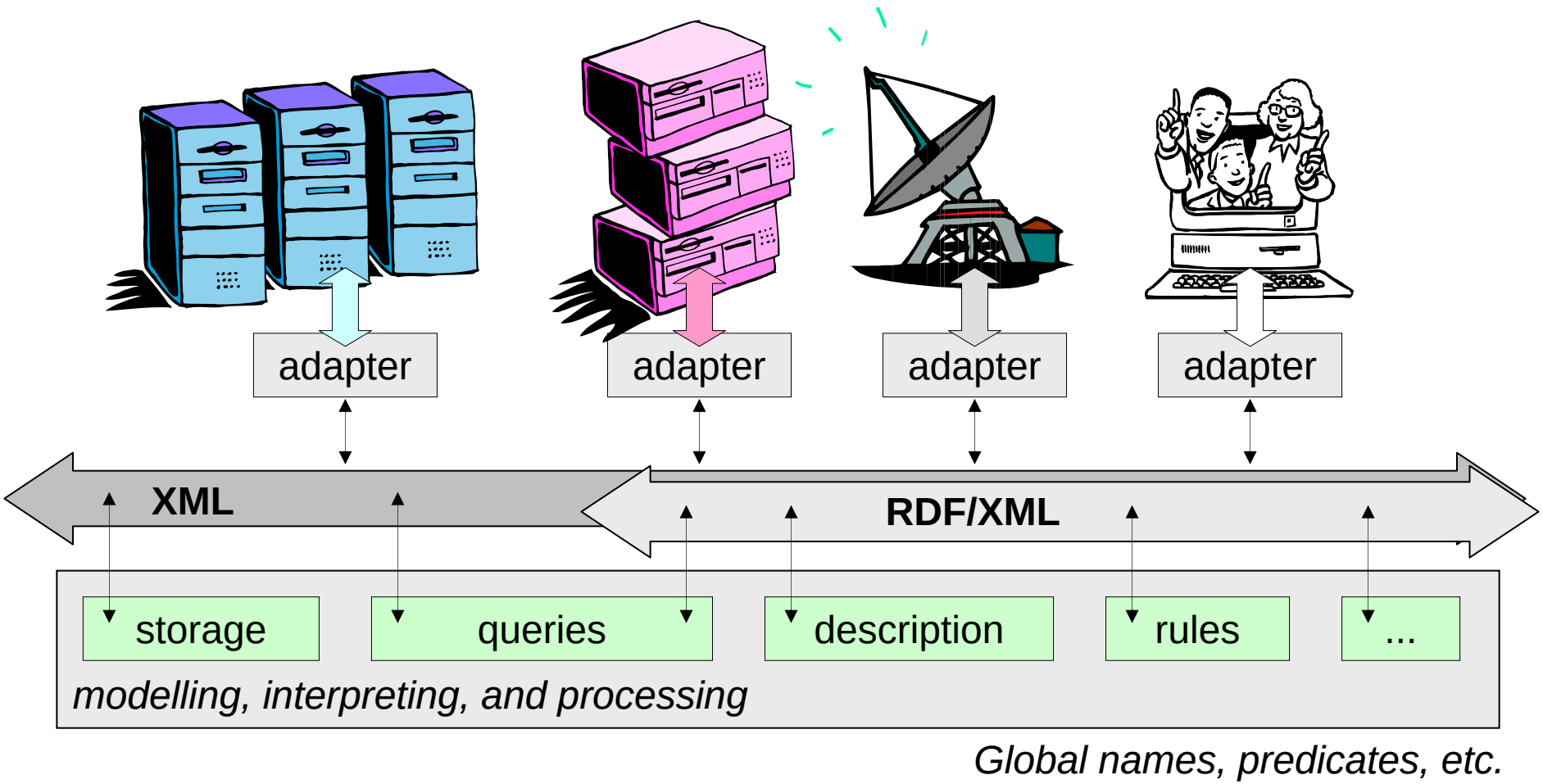
3 päätstoa, 17 aluetstoa, yli 400 jäsenorganisaatiota

Jäseneksi? http://www.w3c.tut.fi/joining.html

Web-infrastruktuuri ja sen standardeja



Web-infrastruktuuri ja tiedon integraatio



W3C ja suositusten kehitysprosessi

W3C kehittää ja standardoi teknologiaa konsensuksen periaatteella siten että…

1. Suositukset edistävät yhden ja yhteisen Webin kehitystä (**”One Web”**)
2. Suosituksille on tunnustettu tarve ja W3C-yhteisön tuki
3. Suositukset ovat teknisessä mielessä päteviä
4. Suositukset ovat (dokumentteina) vapaasti saatavilla verkossa ja niiden soveltaminen onnistuu ilman lisenssimaksuja

W3C kehittää **teknisten suositusten** ohella myös **ohjeita** ja ns. **hyviä käytäntöjä**, mm.

- Web-arkkitehtuuri (TAG)
- Mobiili Web (Mobile Web Initiative, MWI)
- Semanttinen Web (Semantic Web)
- Saavutettavuustyö (Web Accessiility Initiative, WAI)

W3C myös osallistuu aktiivisesti Web-teknologioiden kehitystyöhön muissa standardointi- yms. foorumeissa

- Ks. http://www.w3.org/2001/11/StdLiaison

W3C on sitoutunut ylläpitämään työnsä tuloksia ja varmistamaan niiden vapaan saannin

- Ks. http://www.w3.org/TR/

W3C:n jäseneksi voi hakea mikä tahansa organisaatio; W3C:lla on selkeä toimintaohje joka on vapaasti saatavilla verkossa

- http://www.w3.org/2005/10/Process-20051014/

HTML, XML ja rajapintoja tietoon

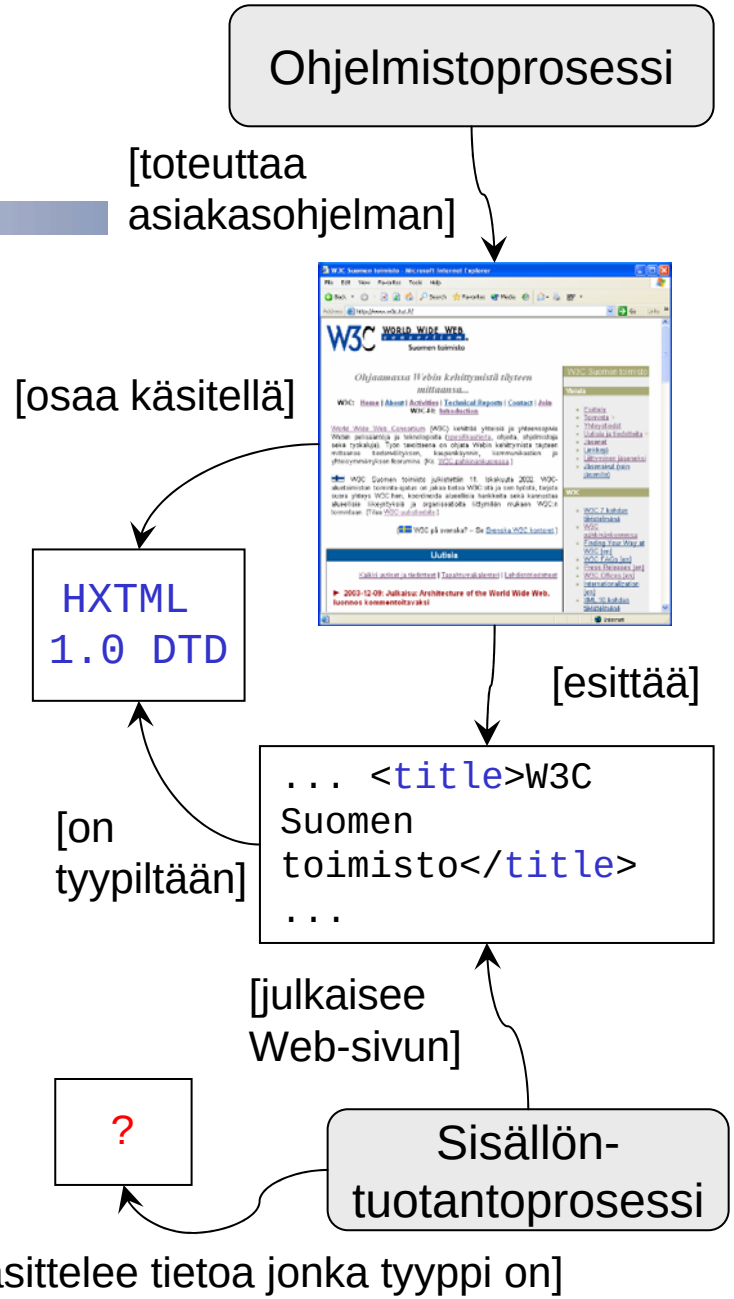
W3C:n XML eli **Extensible Markup Language** tarjoaa yhteensopivan metakielen ja teknologiaperheen jonka avulla voidaan kuvata ja käsitellä mitä tahansa tietoa

XML-teknologiaan perustuvien tekstiformaattien etuina voidaan pitää mm.

1. Tiedonesityksen tekninen perusta on verraten yksinkertainen
2. Tietoa voidaan käsitellä tekstimuodossa (tarvittaessa yhteisen XML-kieliopin tasolla)
3. Erilaisia ja eri toimittajien välineitä on runsaasti saatavilla (vrt. virkistäminen)

”Mutta”

1. Esim. HTML on pohjimmiltaan ”julkaisuformaatti”
2. Tiedon käsittely edellyttää yleensä myös sovelluskohtaista tyyppi- ja skeematietoa (ts. esim. pelkkä HTML-kielioppi ei aina riitä tulkintaan tiedon tasolla)



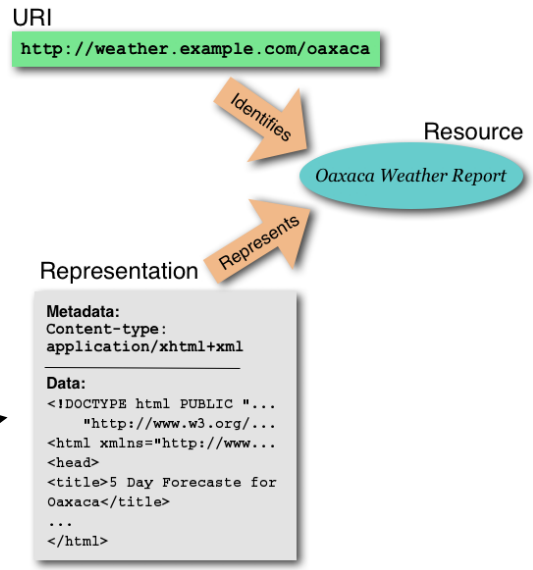
Web ja arkistoinnin haasteita

Laajemmin tarkasteltuna Web ei ole pelkästään joukko julkisia hakemistoja joista löytyviin tiedostoihin voidaan viitata hypertekstilinkillä

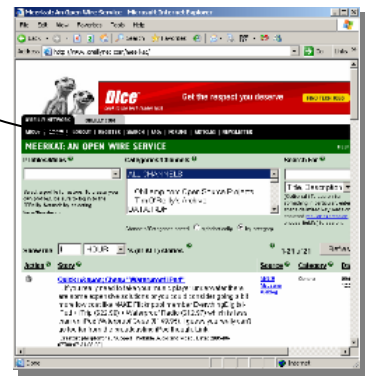
Web-sovellusten näkökulmasta arkistointia voi tapahtua (ainakin) kolmella eri tasolla:

1. **Palvelukokonaisuuksien tasolla** (esim. organisaation uutispalvelu tai taidenäyttely)
2. Palvelun tarjoamien resurssien **esitystavan tasolla** (esim. kämmentietokoneympäristöön tarjottava hypertekstimuotoinen esite)
3. Resursseja vastaavien **dokumenttien (tai datan) tasolla** (esim. uutistiedote tai näkymä näyttelyesineistön tietokantaan)

Sovelluskehitystä (ja arkistointia) **tiedon tasolla** voidaan aktiivisesti tukea esim. Semanttisen Webin tekniikoiden avulla



[sisältää osana palvelua]



Web Architecture Principles, Constraints, and Good Practice Notes

Identification

- Global Identifiers (principle, 2)
- Identify with URIs (practice, 2.1)
- URIs Identify a Single Resource (constraint, 2.2)
- Avoiding URI aliases (practice, 2.3.1)
- Consistent URI usage (practice, 2.3.1)
- Reuse URI schemes (practice, 2.4)
- URI opacity (practice, 2.5)

Interaction

- Reuse representation formats (practice, 3.2)
- Data-metadata inconsistency (constraint, 3.3)
- Metadata association (practice, 3.3)
- Safe retrieval (principle, 3.4)
- Available representation (practice, 3.5)
- Reference does not imply dereference (principle, 3.5)
- Consistent representation (practice, 3.5.1)

Data Formats

- Version information (practice, 4.2.1)
- Namespace policy (practice, 4.2.2)
- Extensibility mechanisms (practice, 4.2.3)
- Extensibility conformance (practice, 4.2.3)
- Unknown extensions (practice, 4.2.3)
- Separation of content, presentation, interaction (practice, 4.3)
- Link identification (practice, 4.4)
- Web linking (practice, 4.4)
- Generic URIs (practice, 4.4)
- Hypertext links (practice, 4.4)
- Namespace adoption (practice, 4.5.3)
- Namespace documents (practice, 4.5.4)
- QNames Indistinguishable from URIs (constraint, 4.5.5)
- QName Mapping (practice, 4.5.5)
- XML and "text/*" (practice, 4.5.7)
- XML and character encodings (practice, 4.5.7)

General Architecture Principles

- Orthogonality (principle, 5.1)
- Error recovery (principle, 5.3)

Hyviä huomioita

Tekniset standardit jättävät liikkumavaraa tiedon sovelluskohtaisen mallinnuksen sekä sovellusten teknisen toteutustavan suhteen

Avoimet standardit tukevat laajamittaista arkistointityötä merkittävästi

Arkistointiin sisältyy sekä ”kokonaisten” Web-sovellusten näkökulma että niiden takaa löytyvän tiedon näkökulma

- Yksinkertaisimmillaan kyse on sivukohtaisesta arkistoinnista
- Monimutkaisimmillaan arkistoitavaa tietoa saattaa olla vaikea erottaa sitä käsittelevästä (palvelinpohjaisesta tai hajautetusta) tietokoneohjelmasta
- Arkistointia *sinänsä* ei (toistaiseksi) ole tarkasteltu yhtenä Web-teknologioiden keskeisenä reunaehtona – vaikuttaminen W3C-työn tavoitteisiin tapahtuu osallistumalla yhteisöön (keskeisesti jäsenyyden kautta)

Lopuksi

World Wide Web Consortium (W3C) kehittää Web-standardeja

Web-tiedon arkistointia voidaan tarkastella eri tasoilla

- Toimiva arkistointi edellyttää tyypillisesti yhteistä määrittelytyötä ja (etukäteistä) suunnittelua; tähän työhön hyviä välineitä tarjoavat mm. XML-perhe ja Semanttinen Web sekä suositus Web-arkkitehtuurista

Web-teknologiat tarjoavat sovelluskehittäjille rikkaan infrastruktuurin jonka tiedostaminen voi helpottaa monimutkaisten sovellusten arkistointiin liittyviä kysymyksiä (ja arkistoinnin rajanvetoa)

Jäikö joku W3C-asia mietityttämään?
Allekirjoittaneen tavoittaa helposti:

http://www.w3c.tut.fi
ossi@w3.org

Liite: Lisätietoja ja osoitteita eteenpäin

W3C

- http://www.w3.org/ (kotisivu)
- http://www.w3c.tut.fi/ (W3C Suomen toimiston kotisivu)

Aiheeseen liittyvää W3C-työtä

- http://www.w3.org/Consortium/Activities (yleiskuva)
- http://www.w3.org/XML/ (XML)
- http://www.w3.org/TR/webarch/ (Architecture of the World Wide Web, Volume One)
- http://www.w3.org/2001/sw/ (Semantic Web)

W3C:n suositukset, raportit ja teknistä tietoa

- http://www.w3c.org/TR/ (kaikki tekniset dokumentit)

Jäseneksi!

- http://www.w3.org/Consortium/Prospectus
- http://www.w3c.tut.fi/joining.html