

Introduksjon til XML og XML som forvaltningspolitisk virkemiddel.

G. Christian Bull
Rikstrygdeverket EPI/UiO/AFIN

Målsetning med forelesningen

- XML, med alle sine avarter og standarder, er et enormt felt, som ikke kan dekkes på 90 minutter.
 - En rekke sentrale begreper er utelatt, eller omtales kun i forbifarten.
 - Målsetningen er at dere skal få en liten smak på hva XML er, og hva man kan gjøre med det, uten å gjøre dere til eksperter.
-

Disposisjon.

- Introduksjon til markup språk og HTML.
 - En enkel introduksjon til XML.
 - XML som forvaltningspolitisk virkemiddel: EML.
 - Introduksjon til OASIS.
 - EML-historie.
 - EMLs verdensbilde.
 - Praktiske EML-eksempler.
 - Tilpasning av EML til nasjonale forhold.
 - Avslutning.
-

Hva er markup språk?

- Markup språk stammer fra trykkerier.
 - "Markup men" skrev instruksjoner til typografene i margen på manuskripter.
 - Eks: "<ingress>", "<avsnitt>"
 - Forklarer hvordan den teksten som er "marked up" skal behandles.
 - IBM startet utviklingen av ML for datamaskiner på slutten av 1960 tallet.
-

HTML

- På begynnelsen av 90-tallet så ML ut til å lide en stille død, til fordel for WYSIWYG og binære filformater.
 - Binære filformater er *proprietære*, det vil se at de kun kan leses av programmene som har laget dem (eller programmer som er laget for å lese dem). Eksempelvis powerpointpresentasjoner.
 - Dette gir leverandører med stor markedsandel et konkurransefortrinn.
 - "Skal du snakke med trygdekontoret må du ringe fra en Nokia-telefon!"
-

HTML

- Introduksjonen av Hypertext Markup Language (HTML) i 1991 et vannskille.
 - HTML bruker "tagger" (merker) i teksten for å angi for fortolkningsprogrammet hvordan teksten skal se ut.
 - Benytter kjente tagger, som en HTML-fortolker (webleser) må vite hvordan den skal forholde seg til.
 - HTML-fortolkeren vet at når den finner <-tegnet, så er det starten på noe den skal fortolke. >-tegnet avslutter.
 - Eksempel: "Dette er vanlig tekst. Denne teksten blir fet. Dette blir vanlig tekst."
 - Fortolkeren ser først -taggen, som forteller den at teksten som følger skal vises i fet skrift.
 - Så finner den som forteller at den skal slutte med fet skrift.
-

HTML

- HTML er utelukkende til for å fortelle fortolkeren hvordan et dokument skal *se ut*, ikke *hva det inneholder*.
- Det revolusjonerende lå i at ulike fortolkningsprogrammer kunne lese den samme filen og vise det samme på skjermen.
- HTML har relativt løse grammatiske regler.
 - Eks: En `
` (linjeskift) behøver ikke å følges av en `</br>` (avslutt linjeskift!?)
- Generelt syns datamaskiner det er greiest å forholde seg til streng grammatikk.
 - HTML: `` Dette er en tekst med `
` et linjeskift og fet skrift.
 - XHTML: `` Dette er også en tekst med `
` linjeskift og fet skrift. ``

XHTML?

- Den rotete grammatikken i HTML, kombinert med at stadig flere plattformer (mobiltelefoner, PDAer, kjøleskap) skulle aksessere WWW utløste et behov for en opprydding.
- Da XML ble standardisert i 1999 fikk man et språk med streng grammatikk, som HTML lett kunne tilpasses.
- HTML ble erstattet av XHTML som anbefalt standard for dokumenter på WWW i Januar 2000.

Så hva er XML?

- XML er et *selvdefinierende, utvidbart markup-språk*.
- Selvdefinierende: XML er skrevet i XML; det er ikke ordene i seg selv som betyr noe, men hvordan de tolkes.
- Utvidbart: Alle kan finne på nye ord, og gi dem en mening. Et nytt sett med ord, som utgjør et språk, kalles et XML-subset. XHTML er ett slikt subsett, EML et annet. Det finnes hundrevis av offisielle og trolig tusenvis av uoffisielle XML-subset.

XML Syntaks

- Enkel syntaks (grammatiske regler) gjør det lett å lage fortolkningsprogrammer.
 - Alle tagger må være lukket.
 - `"<tag></tag>"` eller `"<tag/>"`
 - Tagger kan ikke overlappe.
 - Ikke `"<tag1><tag2></tag1></tag2>"`
 - men `"<tag1><tag2></tag2></tag1>"`
 - Det som står mellom taggene kalles et *element*.
 - Et XML-dokument har ett roelement.
 - Et element inne i et element kalles et barne-element (child element)
 - Elementnavn (det som står inne i taggene) må overholde visse navngivningskonvensjoner.
 - Det skilles mellom store og små bokstaver i elementnavnene.
 - Luft i element-teksten beholdes.
- Et xml-dokument som overholder disse reglene kalles "velformet" (well formed)

XML Metaspråk

- Metaspråk er språk (basert på XML) for å angi nærmere regler for hvordan et dokument kan og må struktureres, hva tagger kan inneholde et.c.
 - DTD (Document Type Definition) stammer fra 1999.
 - Ikke helt vellykket. For liten kontroll over forekomster av elementer, dårlig håndtering av namespaces et.c.
 - XML Schema introdusert i 2001 for å rette opp svakhetene.
 - Gir meget god kontroll, men er hemmet av at det er svært inflekt.
 - Relax NG relativt nylig tatt inn i W3C-varmen som konkurrerende standard til XML Schema.

Namespaces.

- Det er allment akseptert at det er dumt å finne opp hjulet gjentatte ganger.
- Derfor er det ønskelig å kunne inkludere eksisterende XML-subset i sine egne.
 - Siden man skal gi taggene beskrivende navn er det ikke vanskelig å tenke seg situasjoner hvor man gjerne vil bruke et navn noen andre alt har brukt.

Namespaces.

- Løsningen er å angi hvilket subset en gitt tag henter navnet sitt fra.

```
<h:html xmlns:xdc="http://www.xml.com/books"
  xmlns:h="http://www.w3.org/HTML/1998/html">
  <h:head><h:title>Book Review</h:title</h:head>
  <h:body>
    <xdc:bookreview>
      <xdc:title>XML: A Primer</xdc:title>
      <h:table>
        <h:tr align="center">
          <h:td><xdc:author>Simon St. Laurent</xdc:author</h:td>
          <h:td><xdc:price>31.98</xdc:price</h:td>
          <h:td><xdc:pages>352</xdc:pages</h:td>
          <h:td><xdc:date>1998/01</xdc:date</h:td>
        </h:tr>
      </h:table>
    </xdc:bookreview>
  </h:body>
</h:html>
```

Hvilke muligheter gir XML oss?

- Vi kan lage våre egne subset av XML, hvor vi blir enige med oss selv om hva et element kan inneholde.
- Så kan vi publisere dette i form av skjemaer og dokumentasjon. Andre, som leser dokumentene våre, kan da forstå hva vi mener med taggene.
- Vi kan benytte eksisterende standarder sammen med nye.
- Man kan i tillegg validere dokumenter mot XML syntaks og skjema/DTD. Vi kan da klare oss med færre "sanity checks" i programkoden.
- Dette gjør det lett for ulike programmer å utveksle data.

Et eksempel fra Rikstrykdeveket: Min familie i XML.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<tpsPersonData xmlns="http://www.riv.no/namespaceTPS"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.riv.no/namespaceTPS
  H:\SYSTEM-1\SYSTEM-TVPDSTP-TVPDSTP-2.XSD">
  <tpsServiceRutine >
    <serviceRutinenavn>F903-PDNUMMER-PERSRELA-O</serviceRutinenavn> >
    <fnr>010675nnnn</fnr>
    <aksjonsdato>
    <aksjonskode>A</aksjonskode>
    <aksjonskode2>0</aksjonskode2>
  </tpsServiceRutine>
  <tpsSvar>
    <svarStatus>
      <returStatus>00</returStatus>
      <returMelding> </returMelding>
      <utfyllendeMelding> </utfyllendeMelding>
    </svarStatus>
  </tpsPersonData>
```

```
<personDataS005>
  <fnr>010675nnnn</fnr>
  <spesregType> </spesregType>
  <aktivitetsDIFP>aktivitetsDIFP
  <typeAdresse>OFFA</typeAdresse>
  <adresse>STOCKFLETHS GATE 35 H0302</adresse>
  <endringDato>2003-08-22</endringDato>
  <antallRelasjoner></antallRelasjoner>
  <relasjoner>
    <relasjon>
      <typeRelasjon>BANK</typeRelasjon>
      <fnrRelasjon>14101nnnn</fnrRelasjon>
      <spesregType> </spesregType>
      <kortnavn>BULL CORNELIA</kortnavn>
      <fornavn>CORNELIA</fornavn>
      <mellemnavn> </mellemnavn>
      <etternavn>BULL</etternavn>
    </relasjon>
  </relasjoner>
</personDataS005>
</tpsSvar>
</tpsPersonData>
```

XML som forvaltningspolitisk virkemiddel: EML

- Election Markup Language.
- EML er et XML subset, definert med XML Schema.
- Beregnet på bruk i alle steg av alle slags demokratiske valg.

OASIS

- OASIS er en åpen non-profit organisasjon finansiert av medlemsorganisasjoner. En rekke store selskaper, som Boeing, Intel, AMD, IBM, Microsoft og Oracle, deltar i finansieringen.
- I prinsippet kan hvem som helst bli medlem av OASIS og ha innflytelse i en hvilken som helst teknisk komité (TC).
- Tre medlemmer må til for å starte en ny teknisk komité.
- Imidlertid har ikke individuelle medlemmer stemmerett når standarder skal vedtas – kun organisasjoner.
- Full åpenhet i alle ledd av prosessen er et grunnleggende prinsipp. All skriftlig kommunikasjon innad i komiteen må foregå i åpne fora. Møtereferater er åpent tilgjengelige.
- Målsætningen er å hindre at noen enkeltinteressent får dominere utviklingen av standarden.
- Når en TC har fullført en foreslått standard, må den prøves ut i minst 3 piloter før den eventuelt kan vedtas som OASIS-standard.
- Arbeidet koordineres med bl.a. ISO, NIST og ITU.

EML: Litt historie.

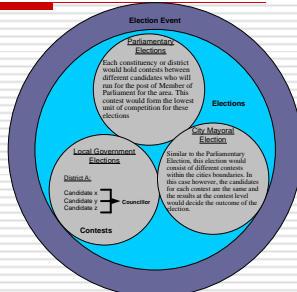
- Arbeidet startet mai 2001.
- Gregg McGilvray fra election.com initiativtager. Microsoft og Accenture var med-initiativtagere.
- Målsetning: å gjøre det mulig for små aktører å hevde seg i konkurransen.
- UK Office of the e-envoy overtok ledelsen i september 2001 fordi det var vanskelig å få aksept fra potensielle bidragsytere for at en leverandør ledet gruppen.
- EML v3.0 ferdig i mars 2003, nesten 1 år etter at arbeidet opprinnelig var planlagt sluttført.

Hva gjør EML?

- EML angir:
 - Hvilke data som skal eller kan være tilstede.
 - Hvorledes data skal struktureres.
- Dette gjør det mulig for applikasjoner levert av ulike leverandører å utveksle data.

EML universet.

- Nesting av "hendelser"
 - ElectionEvent er topp-elementet.
 - ElectionEvent inneholder én eller flere Election-elementer.
 - Hvert Election-element inneholder ett eller flere Contest-elementer.
 - En Election er, eksempelvis, et Stortingsvalg.
 - Hvert Stortingsvalg er 19 fylkesvise Contests.
 - Et kommune og fylkestingsvalg kan tenkes som en eller to Election-elementer i en ElectionEvent.



EML prosessmodell.

- Skjemaene er ordnet i 5 serier.
 - 100-serien definerer Election Events.
 - 200-serien vedrører kandidater og nominasjoner.
 - 300-serien vedrører velgere og manntallslister.
 - 400-serien vedrører stemmesedler.
 - 500-serien vedrører telling.
 - Reflekterer at valg foregår stegvis.

EML i bruk.

- Eksempelvis må folkeregisteret ha programvare, som kan generere EML 310-meldinger for hver enkelt stemmeberettiget. De aktuelle myndighetene (kommunene) må ha programvare som tar i mot melding 310 og legger hver enkelt velger inn i manntallsdatabasen.

EML i bruk: Utdrag av en skjema definisjon (420-Authentication.xsd)

- Dette er et av de enkleste skjemaene i EML. Det angir den informasjonen en velger som ønsker å stemme må oppgi til den/de/det som har myndighet til å gjøre om en velger har stemmerett.

```
<?xml version="1.0" encoding="utf-8"?>
<!-- edited with XML Spy v4.4 U (http://www.xmlspy.com) by Paul Spencer (alphaXML Ltd.) -->
<xsd:schema targetNamespace="urn:oasis:names:tc:evs:schema:eml"
  xmlns="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
  attributeFormDefault="unqualified" version="1.0" id="420">
  <xsd:annotation>
    <xsd:documentation>
      Changes from version 0.3 to 1.0 (2002-10-10):
      None
      Changes from version 0.2 to 0.3 (2002-09-20):
      TransactionId removed (now in emlcore)
      Changes from version 0.1 to 0.2:
      AuditInformation removed
    </xsd:documentation>
  </xsd:annotation>
</xsd:schema>
```

420-Authentication.xsd forts.

```
<xsd:include schemaLocation="emlcore.xsd"/>
<xsd:element name="Authentication">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element ref="VotingChannel" minOccurs="0"/>
      <xsd:choice>
        <xsd:element ref="VToken"/>
        <xsd:element ref="VTokenQualified"/>
      </xsd:choice>
      <xsd:element name="LoginMethod"
        type="xsd:string" minOccurs="0"/>
      <xsd:element name="Language" type="xsd:language" minOccurs="0"/>
      <xsd:any namespace="##any" minOccurs="0"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
</xsd:schema>
```

EML i bruk. 440-CastVote.xsd .

- Et mer interessant eksempel. En enkel, utfyllt stemmeseddel, som et velformet instansdokument:

```
<?xml version="1.0" encoding="UTF-8"?>
<CastVote xmlns="440-castvote.xsd">
  <VToken Type="VID">
    1234567890123456
  </VToken>
  <ElectionEvent>
    <Event>
      <EventName Id="12ab"> /* Id=NMTOKEN */
        Kommune og Fylkestingsvalg 2007
      </EventName>
    </Event>
```

440-CastVote forts.

```
<Election>
  <ElectionName>Oslo Fylkeskommune</ElectionName>
  <Contest>
    <ContestName>Bystyre</ContestName>
    <Selection>
      <Option>
        <OptionName>Det Pragmatiske Parti</OptionName>
        /* Alternativt kan hver option ha sin egen VToken knyttet til
        seg, som viser retten til å avgi stemme i denne contest. */
      </Option>
    </Selection>
    /* Her kan man ha en Seal-struktur, som benytter seg av W3C XML
    digital signatur, samt EML timestamp. "The votes are sealed at
    this level if there is a chance that the message will be divided,
    for example so that different elections can be counted in
    different locations." */
  </Contest>
</Election>
```

440-CastVote forts.

```
<Election>
  <ElectionName>Kommune og Fylkestingsvalg
  2007</ElectionName>
  <Contest>
    <ContestName>Bydelsutvalg</ContestName>
    <Selection>
      <Option>
        <OptionName>Snillistlisten</OptionName>
      </Option>
    </Selection>
  </Contest>
</Election>
</ElectionEvent>
/* Til slutt kommer valgfri "audit information." */
</CastVote>
```

EML i bruk.

- Innenfor en <Election> kan vi her kun ha én <Contest>. Dette skiller seg fra f.eks 110-skjemaet:
 - <xsd:element name="Election" maxOccurs="unbounded"/>
- Elementer med samme navn *kan* med andre ord ha ulikt innhold i ulike skjema.
- Dette er et bevisst valg, for å unngå at elementer blir unødige rike på data.

EML i bruk.

- Det første eksemplet hadde ingen mulighet for kumuleringer og strykninger.
- Hver kandidat på en liste må være en option.
- <WriteinOptionName> kan da fungere som slenger.
- <Value> kan representere personstemme.
- Forhåndskumulering vha. <Value> er ikke mulig uten modifikasjon av skjemaene.
 - 410-ballots.xsd mangler noen plass for <Value>.
 - CandidateStructure (emlcore.xsd) mangler <Value>.
 - Mulig løsning er å oppføre forhåndskumulerte kandidater to ganger?

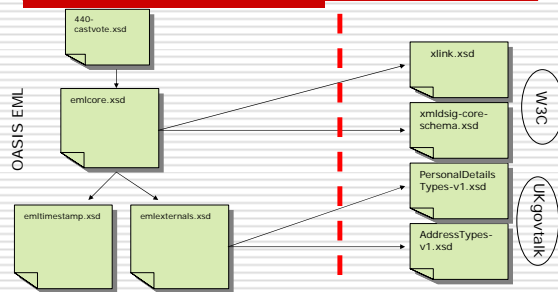
En 440-CastVote med kumulering:

```
<Election>
  <ElectionName>Kommunevalg 2007</ElectionName>
  <Contest>
    <ContestName>Kommunestyre</ContestName>
    <Selection>
      <Option>
        <OptionName>Peder Ås</OptionName>
        <Value>2</Value>
      </Option>
      <OptionName>Marte Kirkerud</OptionName>
      <Value>2</Value>
      <OptionName>Hans Tøstadi</OptionName>
      <Value>1</Value>
      <OptionName>Bodhild Bratt</OptionName>
      <Value>1</Value>
      <WriteInOptionName>Lars Holm</WriteInOptionName>
      <Value>1</Value>
    </Option>
  </Selection>
</Contest>
</Election>
```

EML er et puslespill.

- ❑ Alle skjemaene refererer til emlcore.xsd, som inneholder deklarasjoner for mye brukte elementer, som VToken.
- ❑ emlcore.xsd refererer til ytre standarder, som W3C XML Digital Signature.
- ❑ emlexternals.xsd inneholder referanser til eksterne skjema, som er tenkt endret for nasjonalt bruk.

Puslespillet 440-CastVote.xsd



Nasjonale tilpasninger.

- ❑ EML er designet for å kunne brukes i nær sagt et hvilket som helst demokratisk valg/referendum.
- ❑ Det følger av dette at det er mengder av informasjon i de forskjellige EML-skjemaene, som ikke er av relevans for norske valg.
- ❑ Videre er enkelte begrensninger (constraints) ufullstendige.
 - Eksempelvis er antallet kandidater i skjemaene ubegrenset. (`<xsd:element ref="#Candidate" maxOccurs="unbounded"/>`)
 - Det kan være ønskelig å pålegge en nasjonal, eller til og med lokal, minimums- og maksimums-begrensning, slik at ikke lister med for få/mange kandidater er gyldige.
- ❑ Et nasjonalt tilpasset skjema skal imidlertid gi dokumenter, som kan valideres også med EML.
 - Det er ingenting i veien for nasjonale, regionale eller lokale tilpasninger, så lenge man bruker riktige teknikker.

Nasjonale tilpasninger.

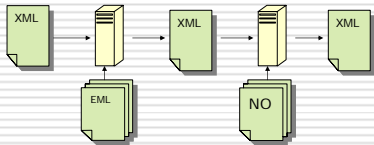
- ❑ Det eneste skjemaet man er ment å endre er imidlertid emlexternals.xsd, som kun vedrører måten navn og adresser behandles.
 - Det skjemaet vi fikk pr. e-post er for øvrig den britiske tilpasningen. Original bruker xNAL.
- ❑ UK Office of the e-envoy utarbeidet en anbefaling, hvor fire ulike metoder utredes:
 - Hensikt: Å kunne validere dokumenter, slik at de er i tråd både med EML og nasjonale krav.
 - Alt 1: Å skrive egne, tilpassede skjema. Altså en egen EML(NO) spesifikkasjon.
 - ❑ Det vil være relativt enkelt å gjennomføre.
 - ❑ Dette frarådes imidlertid, da det vil skape problemer når EML-spesifikasjonen endres.
 - ❑ Man kan dessuten lett komme i skade for å lage skjema, som kommer i konflikt med EML.

Nasjonale tilpasninger.

- Alt 2: Å avlede nasjonale skjema fra grunnspesifikasjonen vha. `xs:redefine`.
 - ❑ Ugreit siden elementer med like navn er definert ulikt i forskjellige skjema innenfor EML. Dermed må det meste redefineres.
 - ❑ Videre er det problemer i forhold til å bruke `xs:redefine` for å begrense `xs:choice`, som brukes mye i EML.
- Alt 3: Å bruke XSLT stylesheet, for å pålegge nye begrensninger.
 - ❑ Ingen større ulemper med dette, bortsett fra at Alt 4 er bedre.

Nasjonale tilpasninger.

- Alt 4: Schematron.
 - En blivende ISO-standard.
 - Schematron er en XML-utvidelse spesielt laget for å pålegge ekstra begrensninger på eksisterende skjema.
- UK benytter Schematron.



Kritikk av EML.

- EML har blitt kritisert for å bite over for mye.
 - EML forsøker å være "all things to all people."
- Komitédesign.
- XML Schema er i seg selv blitt sterkt kritisert for å være overdrevent komplisert.
- Etter pilotene i England påpekte Charbel Aoun, teknisk sjef i election.com og ett av navnene bakte EML v3, at serverne til BT-konsortiet hadde vært svært nær ved å bli overbelastet av de innkommende stemmene, fordi EML skjemaene er for kompliserte.

Alternativer til EML?

- IEEE 1622 Voting Systems Electronic Data Interchange.
 - Fortsatt i startgropen.
 - Vil neppe omfatte alle sider ved e-valg.
- The Open Voting Consortium
 - Har en demoversjon av systemet sitt oppe, med egendefinerte Relax NG-skjema.
 - USA-spesifikt.

Oppsummering.

- EML er det vi har å forholde oss til, og er anbefalt av CoE.
- Har vært brukt med hell i flere ulike scenarier.
- Teknikker for nasjonale tilpasninger foreligger.

Konklusjoner

- EML er drevet frem av kommersielle interesser.
- EML har som primær hensikt å muliggjøre konkurranseutsetting av valg.
- EML passer således godt inn i en britisk modell, hvor lokalvalg "settes ut på anbud", og myndighetene ikke eier systemene.
- Dersom man velger en sveitsisk modell, hvor myndighetene eier alt, inklusive kildekoden, blir EML mindre interessant.

Oppsummering: XML i offentlig sektor.

- XML gjør utveksling av data enklere.
 - Mellom datamaskinprogrammer internt i én etat.
 - Mellom forskjellige etater.
 - Mellom private og det offentlige.