

# KomInn – Felles språk i eNorge

*Bruk av XML skjemaer for standardisering i offentlig sektor*

*Versjon 1.0 - desember 2003*

*av Plogen*

Åpne standarder for datautveksling i offentlig sektor realiseres med XML skjemaer. Bruk av XML skjemaer alene gir imidlertid liten gevinst dersom de ikke inngår i et felles konsept. Det må etableres strenge regler for struktur, navngiving og gjenbruk av XML skjemaer og skjemakomponenter. Gevinstene er:

- Økt lesbarhet og enklere tolkning av data som utveksles.
- Enklere identifisering og bruk av eksisterende standardiserte skjemakomponenter.
- Enklere utvikling av nye standardiserte skjemaer.

Hensikten med dette dokumentet er å definere regler for utvikling av XML skjemaer. Innholdet baserer seg på resultater fra tilsvarende arbeid i regi av det Danske Ministeriet for Videnskap, Teknologi og Udvikling<sup>1</sup>.

## 1. Skjemavarianter

Grensen mellom skjemaer, som kun er fragmenter i en datamodell, og skjemaer som beskriver en konkret grenseflate, kan være flytende. Hovedprinsippet er at utviklingen av et skjema i størst mulig grad skal baseres på gjenbruk av eksisterende typer og elementer. XML skjemaene grupperes i følgende varianter:

### **Sentrale typer**

Er skjemaer med basis datatyper og verdilister som anvendes overalt. Eksempler på dette er "FirstName", "LastName" og "CivilRegistrationNumber". Typene samles i logiske pakker som f.eks. en personpakke.

### **Sentrale elementer**

Er skjemaer med element- og attributtdefinisjoner og benyttes til å ensrette elementnavngivingen på hyppig forekommende elementer og attributter. En elementdefinisjon skal som hovedregel ta utgangspunkt i en definisjon av tilhørende datatype. F.eks. vil definisjonen av elementet "CivilRegistrationNumber" være basert på typen "CivilRegistrationNumberType".

### **Pakker**

Inneholder referanser til sentrale elementer og typer. Pakker organiseres logisk etter bestemte områder som personer, adresser osv. Fordelen er at skjema utvikleren enkelt kan importere en hel samling relevante typer og elementer ved kun å referere til en enkelt pakke.

### **Sammensatte skjemaer**

Bygger på de sentrale typer og elementer. I motsetning til pakker vil sammensatte skjemaer typisk beskrive struktur, semantikk og kardinalitet på elementer for en større struktur, som igjen kan inngå i andre sammenhenger. Et sammensatt skjema viser et gitt områdes forståelse av f.eks. en adresse. Den samme basispakke for Adresse kan brukes i forskjellige områder.

---

<sup>1</sup> Regelsamling for utvikling av XML Schemaer, den fælles offentlige XML komité, Danmark

### **Grensesnittskjemaer**

Beskriver et grensesnitt mellom 2 systemer og består typisk av en eller flere sammensatte skjemaer. Rotelementet i et grensesnittskjema gjenspeiler konteksten hvor skjemaet inngår.

## **2. Generelt**

### **Nasjonale og internasjonale skjemadatabaser skal kontrolleres før definering av nye elementer, typer eller komplekse skjemaer**

Nye elementer, typer eller komplekse skjemaer skal kun defineres når det i forveien ikke finnes en anerkjent standard på området. Følgende skjemadatabaser bør kontrolleres:

- ebXML Core Component Dictionary v1.04: [http://www.ebxml.org/specs/ccDICT\\_print.pdf](http://www.ebxml.org/specs/ccDICT_print.pdf)
- UK GovTalk: <http://www.govtalk.gov.uk>
- XML.org Registry: <http://www.xml.org/xml/registry.jsp>
- DOD XML Registry: <http://diides.ncr.disa.mil/xmlreg/user/index.cfm>

### **Skjemaer skal defineres i XML Schema Definition Language**

Et XML skjema skal uttrykkes i *XML Schema Definition Language* og overholde W3C's anbefaling av XML skjema av 2. mai 2001. (<http://www.w3.org/XML/Schema>).

### **XML skjema instanser skal opprettes som UTF-8 filer med UTF-8 koding**

UTF-8 kan representere hele ISO 10646 og foretrekkes internasjonalt.

### **elementFormDefault skal alltid være qualified**

Av hensyn til etterfølgende transformasjon av XML-instanser er det viktig at alle typer og elementer er kvalifisert med et prefiks.

### **attributeFormDefault kan være qualified**

### **Datatyper skal defineres sterkest mulig**

Alle XML skjemastandardens muligheter må utnyttes for å typesjekke data innen de oversendes til en applikasjon. Dette gjøres f.eks. ved bruk av attributter for maks, min, regulære uttrykk og verdilister.

### **Norske spesialtegn må ikke benyttes i regulære uttrykk**

I stedet kan man benytte "\p{L}\*" som angir at man kan benytte alle store og små bokstaver i gjeldende tegnsett.

### **Standarden må ikke influeres av begrensninger i et bakenforliggende system**

De sentrale typer og elementer skal være anvendelige på tvers av offentlig og privat sektor og må utformes uavhengig av eksisterende systemer og begrensninger.

### **Null-verdier skal håndteres i XML skjemaet**

Det skal eksplisitt defineres hvorvidt et element tillates å ikke å være utfyllt.

*For binært innhold skal det benyttes en referanse eller base64Binary-typen*

*Include skal benyttes der det er mulig*

### **3. Definerings av sentrale typer og elementer**

#### ***Sentrale typer og elementer skal defineres i selvstendige skjemaer***

For å sikre gjenbruk må man kunne søke frem det enkelte skjema i en database. Enhver type er sentral dersom andre skjemadesignere kan tenkes å ville benytte den.

#### ***Sentrale typer og elementer skal dokumenteres***

Bruk av documentation-elementet eller felles verktøy for utfylling av separate metadata-elementer.

#### ***Sentrale typer og elementer skal ha eksplisitte referanser til dokumentasjonen***

Dokumentasjon i form av metadata skal refereres med source-attributtet på documentation-elementet.

### **4. Strukturering av sammensatte skjemaer og grensesnittskjemaer**

#### ***Skjemaer skal struktureres med "Venetian Blind"-modellen***

For å sikre gjenbruk av skjemafragmenter, skal skjemaet struktureres modulært med anvendelse av navngitte typer og med referanse til elementer og typer.

#### ***Alle skjemaer i et namespace skal være tilgjengelige i en pakke***

Skjemaer skal samles i en pakke vha. include-definisjoner.

#### ***Typer skal defineres globalt***

Datatyper skal defineres i skjemaets rot uansett om de har relevans utenfor det aktuelle skjemaet.

#### ***Elementer som skal kunne gjenbrukes skal defineres globalt***

Dersom elementet skal gjenbrukes innenfor, eller på tvers av skjemaer.

#### ***Avgrens bruken av attributter til metadata***

#### ***Innholdsmodellene skal lages med tanke på transformasjon***

#### ***Dokumentorienterte skjemaer skal kun baseres på blandet innholdsmodell (mixed-content)***

Dokumentorientert markup krever at elementer kan inneholde både data og andre elementer.

#### ***Bruk av substitution groups er kun tillatt dersom disse har foreldre***

Elementet som substitueres må være innlemmet i et statisk element.

## 5. Gjenbruk av sentrale typer, elementer og skjemaer

### **Lag referanser til eksisterende definisjoner av elementer og typer**

Kopiering av eksisterende definisjoner fra et skjema til et annet må ikke forekomme.

### **Eksisterende typer som ikke tilfredsstillter spesifikke krav må utvides eller begrenses**

Med XML skjemaets arvemekanisme kan en ny datatype defineres basert på en allerede eksisterende.

## 6. Navngiving av typer, elementer, attributter mm.

Navngivingskonvensjonen er basert på ISO 11179, "Specification and Standardization of Data Elements" med noen avvik ift. endringer vedtatt i ebXML core components, standardisert av OASIS.

### **Navn skal bygges opp etter modellen: ObjektEgenskapKlassifikasjon**

Det benyttes entallsform. Dersom elementet eller typen opptrer i kontekst av et objekt, kan objekt-prefikset utelates.

### **Typer skal navngis med Type som suffiks**

Unngå denne endelsen på elementnavn.

### **Unngå forkortelser og akronymer**

Kun ord på mer enn 8 tegn bør forkortes. Bruk eksisterende forkortelser. Skriv forkortelsen med store bokstaver.

### **Sentrale typer og elementer skal ha unike navn**

Navngivingen skal sikre at samme definisjon ikke forekommer mer enn en gang.

### **Globale definisjoner i skjemaer under samme targetNamespace skal ha unike navn**

### **Alle navn skal være på engelsk**

Suppleres med en norsk forklaring. Dersom det ikke finnes et dekkende engelsk ord, benyttes norsk med engelsk forklaring.

### **Typer og elementer skal navngis med UpperCamelCase**

Stor bokstav først og stor forbokstav i alle påfølgende ord.

### **Attributter skal navngis med lowerCamelCase**

Liten bokstav først men stor forbokstav i alle påfølgende ord.

### **Verdier i verdilister skal defineres med små bokstaver**

Såfremt det ikke eksplisitt benyttes stor forbokstav som f.eks. i egennavn.

### **Filnavn for sentrale typer og elementer skal være type-/element eller attributtnavn**

For skjemaer benyttes <NamespacePrefixMedVersaler>\_<Element-/typenavn>.xsd.

**Grensesnittskjemaer skal navngis med Interface-klassifikasjonen**

Endelsen ”Interface” legges til.

**Namespace skal navngis etter modellen:**

**`http://<RepositoryDomene>/<InternetDomene>/xml/schemas/<Årstall>/<Måned>/<Dag>`**

**Namespace-prefikser skal navngis etter modellen: <InternetDomene> uten punktum og landkode**

Det må tilstrebes å velge prefikser som ikke overstiger 4 tegn.

**XML-metadata dokumenter skal ha suffiks .meta.xml**

## 7. Regler for versjonering

**Ved mindre endringer skal versjonering baseres på definering av nye element- og typenavn**

Ved å kopiere den gamle definisjonen og tilføye et fortløpende siffer til det nye navnet.

**Ved større releaser skal versjonering baseres på navngiving av namespace**

Samtidig skal alle tidligere versjoner av definerte elementer og typer slettes slik at det nye namespace blir ”rent”.

**Bruk version-attributtet etter modellen: Release.RevisionDraft**

Releasenummeret skal endres når tidligere versjoner av skjemaet vil forhindre validering av eksisterende dokumenter.

## 8. Oppsummering

Åpne standarder for datautveksling i offentlig sektor beskrevet som XML-skjemaer i henhold til dette dokumentet vil bidra til en kosteffektiv realisering av eNorge-visjonen både i sentral- og lokalforvaltningen.

*Plogen* er en norsk rådgivingsorganisasjon som tilbyr et bredt spekter av profesjonelle tjenester innen IKT utvikling og anvendelse.

<http://www.plogen.no>