

Interoperabla specifikationer

En utredning inom byggblock metadata

Matthias Palmér

(metadataexpert för Digg)

Ulrika Domellöf Mattsson

(byggblocksansvarig)

Deltagare i referensgruppen presenterar sig kort

Var och en tar 1-2 minuter att säga något om:

- Vem man är och vilken organisation man representerar
 - Tidigare erfarenhet av frågan om beständiga identifierare
 - Förväntan på deltagandet / projektet
 - Snabba förslag / tankar redan nu
- (kom ihåg 1-2 minuter)

Agenda

- Bakgrund- info om byggblock metadata
- Referensgruppsarbete
- Presentation av deltagare
- Konceptuell genomgång av materialet på Github
- Nästa möte

Byggblock metadata

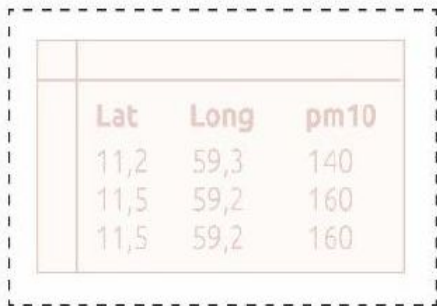
Myndigheten för digital förvaltning, Digg leder arbetet med att etablera en förvaltningsgemensam digital infrastruktur (Ena) för att information ska kunna utbytas på ett säkert och effektivt sätt.

- Samverkan
- Nationella grunddata (ramverk)
- Kompetensområden
- Förutsättningsskapande byggblock

<https://www.digg.se/ledning-och-samordning/ena---sveriges-digitala-infrastruktur>

Beskrivande metadata (Fokus på datamängden)

Titel: Luftkvalitet
Utgivare: Naturvårdsverket
Uppdat: Dagligen

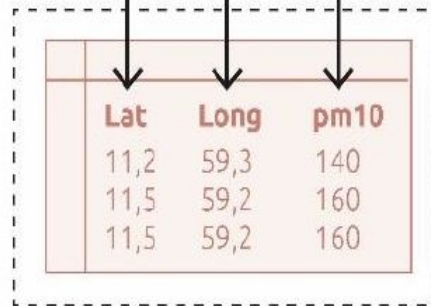


A dashed box containing a table with three columns: Lat, Long, and pm10. The table contains three rows of data. An arrow points from the descriptive metadata box above to the top of this table.

Lat	Long	pm10
11,2	59,3	140
11,5	59,2	160
11,5	59,2	160

Strukturella metadata (Fokus på datauttrycket)

Kolumn: lat
long
pm10
Label: latitud
longitud
partiklar
Datatyp: decimaltal
decimaltal
heltal



A dashed box containing a table with three columns: Lat, Long, and pm10. The table contains three rows of data. Three arrows point from the structural metadata box above to the column headers of this table: 'lat' to 'Lat', 'longitud' to 'Long', and 'heltal' to 'pm10'.

Lat	Long	pm10
11,2	59,3	140
11,5	59,2	160
11,5	59,2	160

Lite förenklat motsvarar beskrivande metadata FAIR principerna Findable och Accessible medan strukturella metadata motsvarar Interoperable och Reusable.

Syfte med referensgruppen

Syftet med referensgruppen är att arbeta fram en profil för att främja interoperabilitet och återanvändning av delar i en specifikation.

En specifikation kan innehålla en blandning av bakgrund, motivering och mer formella beskrivningar. Det är ett paket med olika delar/resurser, vissa riktade mot mänsklig konsumtion, andra maskinläsbara.

Genom att ge användare möjlighet att söka djupare i en datamängd, på den bakomliggande olika delarna möjliggörs interoperabilitet och återanvändning av dem.

Tidigare förstudie: <https://github.com/diggsweden/information-models-investigation>

Referensgruppsarbete

Material:

- Allt läggs up på Github
<https://github.com/diggsweden/interoperable-specifications>
- Bakgrundsmaterial, specifikation, exempel etc.
- Material på engelska då vi förutser Nordisk / Europeisk samverkan

Process:

- Ärendehantering på Github - mallar för styra kommunikationen
- Ändringsförslag (PR = Pull Requests)
- Referensgrupp cirka 4 möten - deltagares namn visas på github
- Första referensgruppsmötet tisdag den 18/11 kl. 10

Uppdelning av innehåll på Github

- Bakgrund och motivation
- Konceptuell design
- Regler om helheten och de fyra delarna
- Vad är en Applikationsprofil
- Hur uttrycks en applikationsprofil
- Kompatibilitet med UML, CSV on the web osv.

(Mer saker planeras)

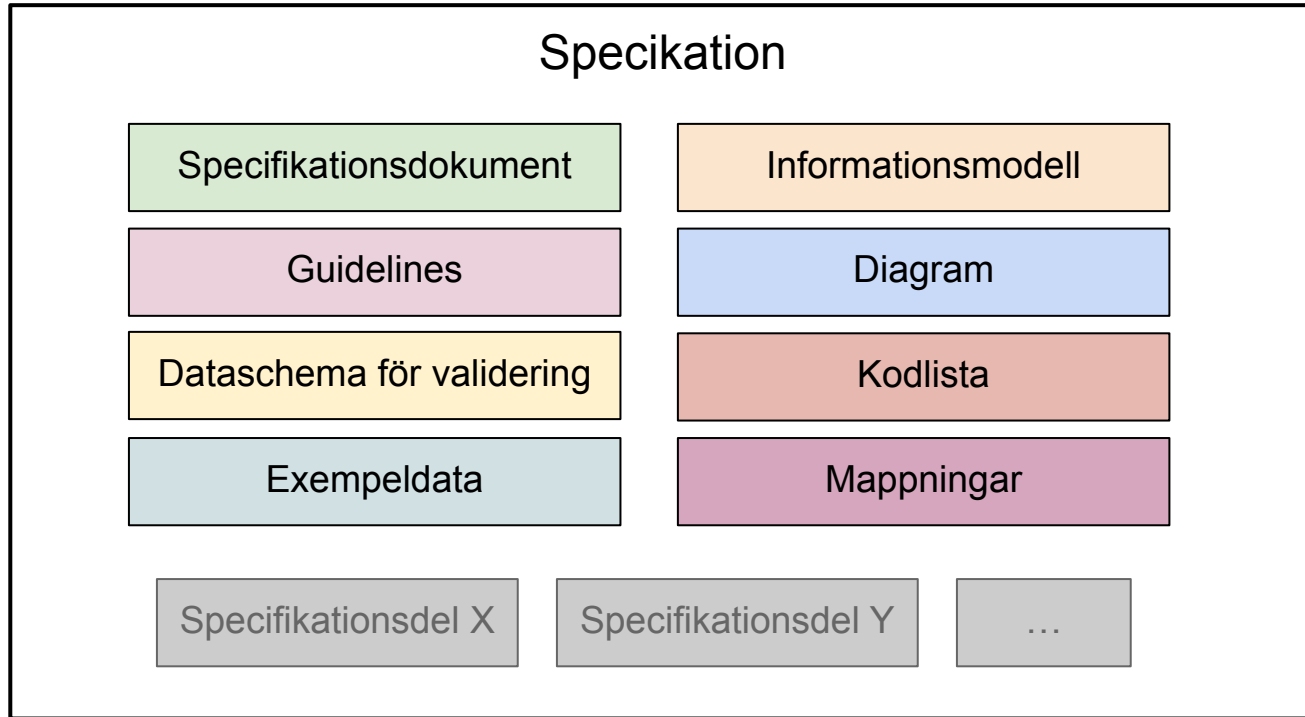
Bakgrund / behov

Syftet med interoperabla specifikationer är att stödja återanvändning mellan specifikationer

Förhoppningen är att:

- **Förenkla / effektivisera** hur man skapar nya informationsmodeller då man inte behöver börja från början
- **Förbättra kvaliteten** då återanvända delar redan är genomtänka och testade
- **Minska tiden för att lära sig** nya specifikationer då det finns mer överlapp
- Data som följer olika specifikationer kan vara **partiellt interoperabla**
- **En community** skapas kring interoperabilitet / återanvändning
- **Erfarenheter och kompetens** mellan organisationer kan delas

En specifikation är en behållare med resurser



Vad kan delas mellan specifikationer

Klasser

- Kategorisering, typning eller indelning av data som har liknande karaktär
- Kan ses som en mängd av instanser, subklasser är delmängder
- Bakgrund från kunskapsrepresentation, programmering (objektorientering)

Properties

- Delas ibland upp i attribut och relationer
- Kan vara självständiga eller hårt knutna till en klass
- Kan ha datatyper

Klasser och properties kallar vi tillsammans för **data vokabulär**

Vad kan delas mellan specifikationer fort.

Begrepp

- Kallas ibland koder, värden, taxoner, termer etc.
- Innehåller ibland:
 - Flerspråkiga lablar, alternativa namn etc.
 - Definition, exempel, användningsanmärkning, historisk anmärkning etc.
 - Relationer samt en hierarkisk struktur

Ett antal begrepp förvaltas i en **terminologi**

En **begreppssamling** är en mängd begrepp som kan vara en delmängd av en terminologi eller till och med samla begrepp mellan terminologier.

Hur uttrycker vi återanvändningen?

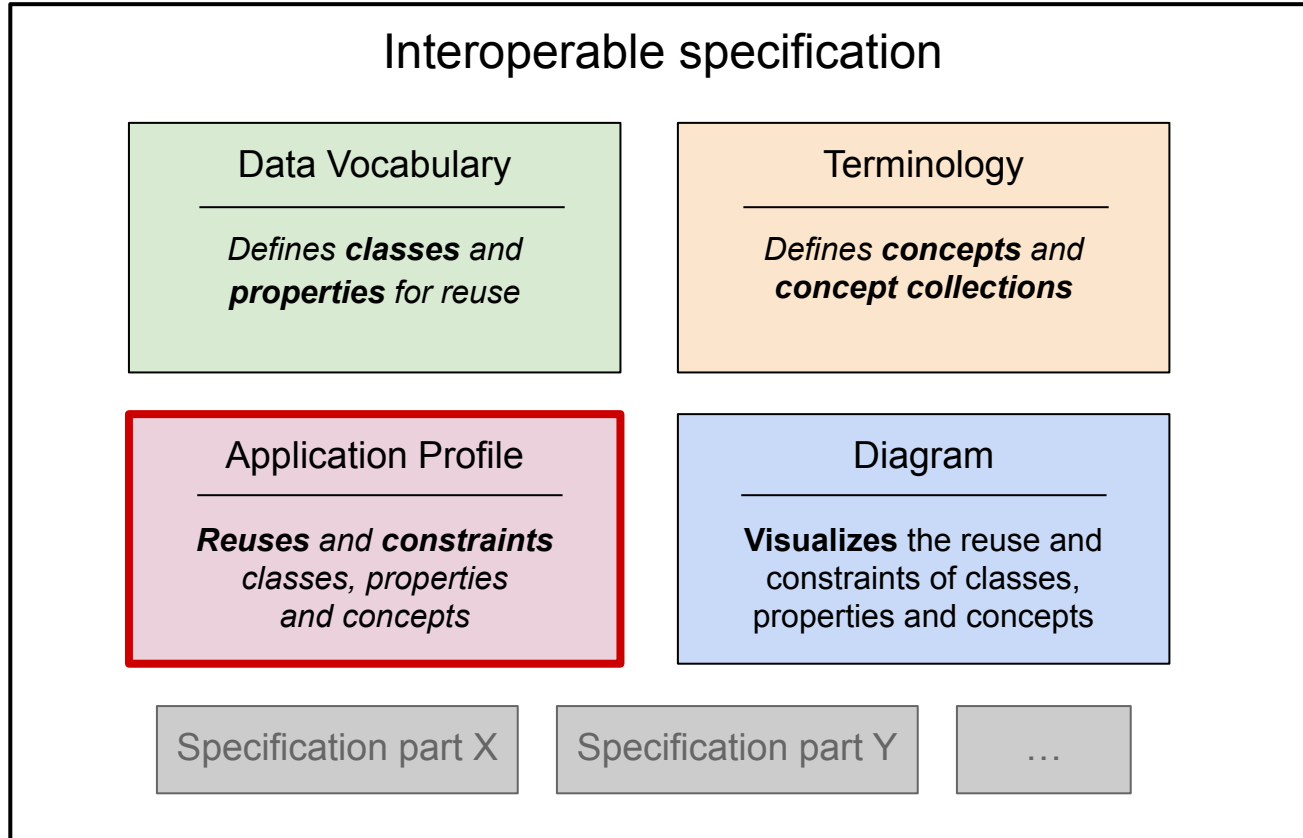
Vi behöver ett speciellt uttryck som klarar att:

1. Uttrycka vilka klasser och properties som ska tas med
2. I vilken konstellation de ska kombineras
3. Vilka begränsningar i värden som får användas (datatyper)
4. Vilka begränsningar i fördefinierade värden som får användas (begrepp)
5. Kardinalitetsbegränsningar

En traditionell informationsmodell klarar inte alltid detta.

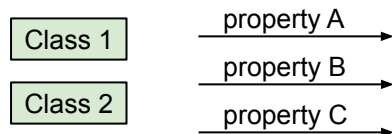
Vi väljer att använda **Applikationsprofiler** för att täcka detta behov.

En interoperabel specifikation har en applikationsprofil

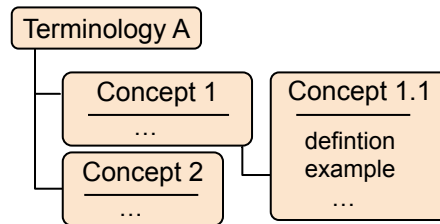


Interoperable specification

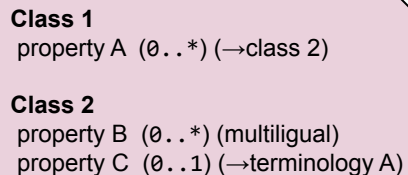
Data Vocabulary



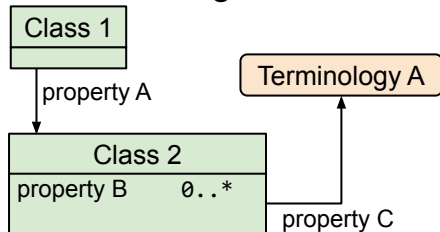
Terminology



Application Profile



Diagram



Specification part X

Specification part Y

...

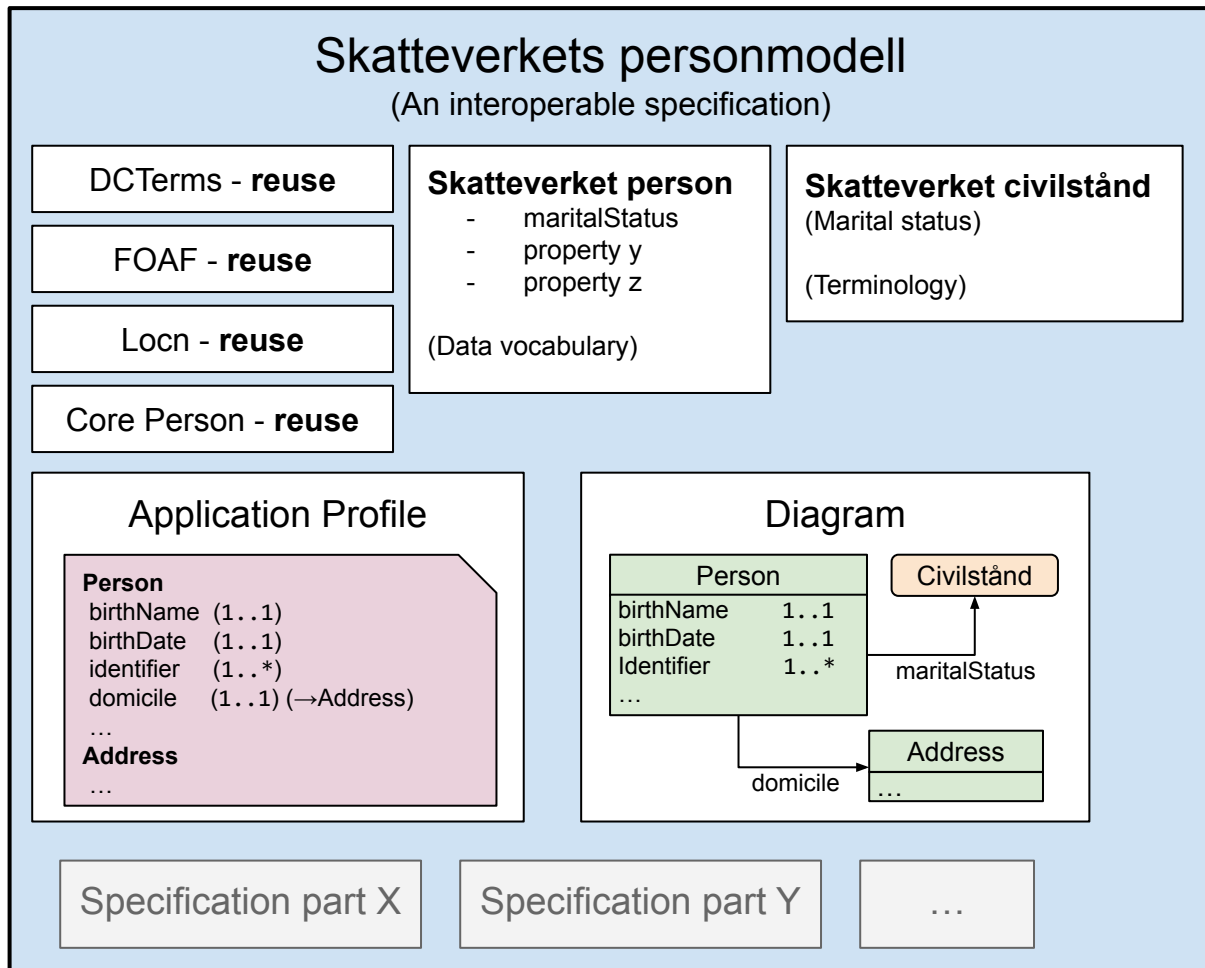
Vad är annorlunda jämfört med informationsmodeller

- Var noga med definition och syfte med klasser och properties (detta är en förutsättning för korrekt återanvändning)
- Större fokus på properties, då de definieras oberoende av klasser (undvik att knyta till specifika klasser om det inte är nödvändigt)
- Klasser, properties och begrepp behöver ha webb-scale stabila identifierare, Dvs de behöver beständiga identifierare i form av URI:er (Se tidigare arbete med beständiga identifierare)
- Var återhållsam med subklasser och subproperties återanvänd direkt när så är möjligt inne i applikationsprofilen

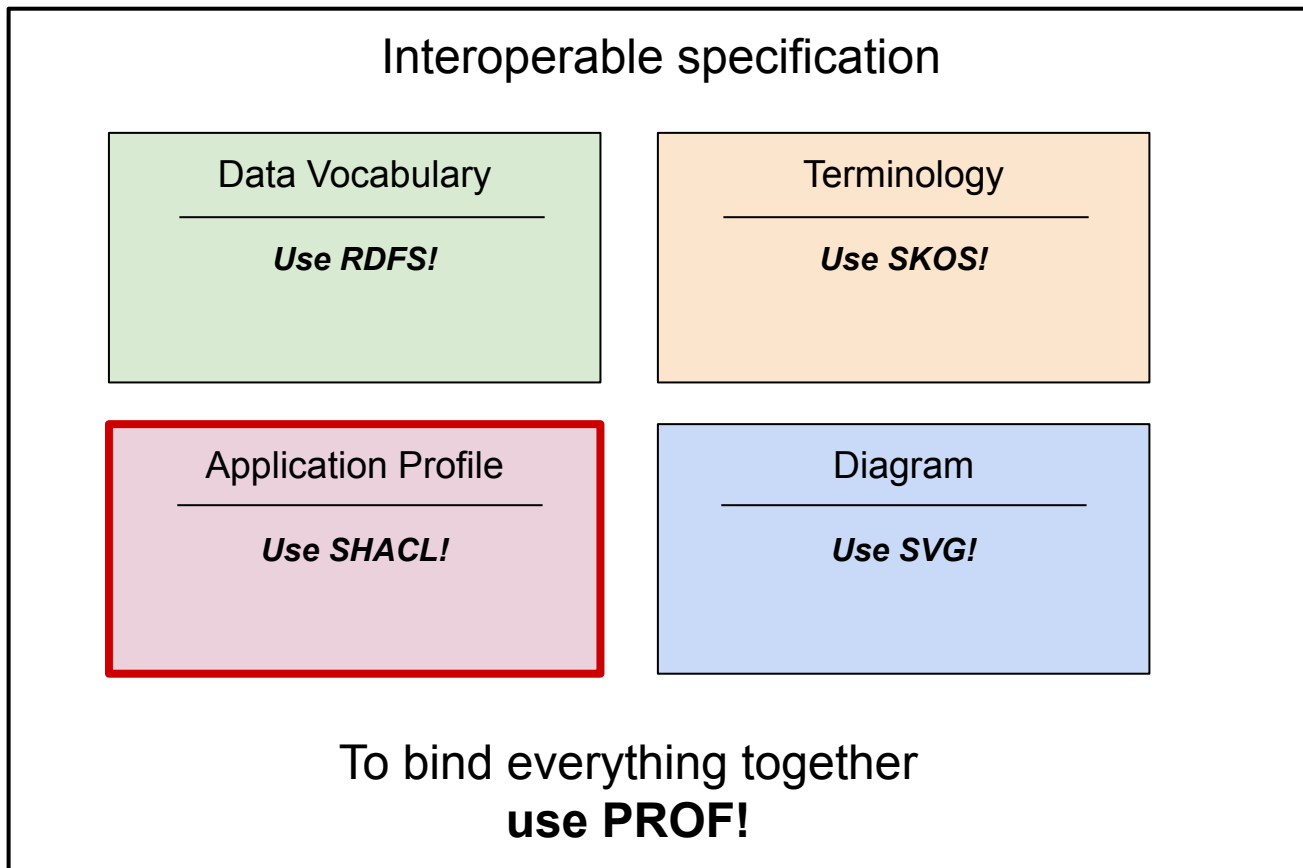
Skatteverkets personmodell

Observera att detta bara är ett exempel, inte en fullständig vy av hur Skatteverkets personmodell skulle kunna se ut.

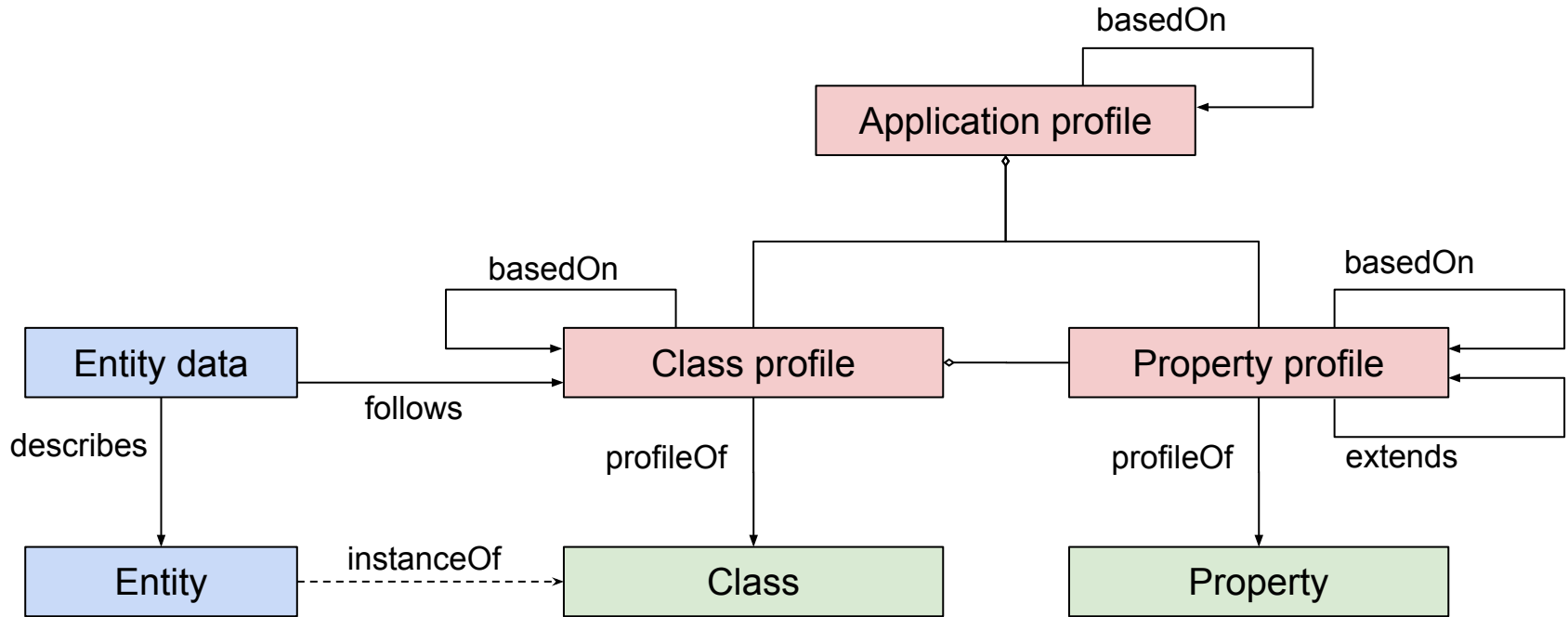
Utgångspunkten är att den isåfall görs baserad på EUs Core Person vocabulary.



Vi försöker att inte återuppfinna hjulet



Modell för Applikationsprofiler



Kompatibilitet med andra uttryck

Försöka tolka UML i skördningssteget:

- Förutsätter att man annoterar klasser, attribut och relationer
- OSLO ramverket kommer användas för detta
- Del av pilot

Tolka andra uttryck som:

- FHIR
- CSV on the webb (tabellannotationer)
(Förutsätter att man specificerat property och datatyp per kolumn)