

Strukturella Metadata - Referensgrupp 2 dec

Närvarande

Benny Lund (Bolagsverket)
Cilla Öhnfeldt (Naturvårdsverket)
Erik Mossing (Bolagsverket)
Fredrik Emanuelsson (Riksarkivet)
Fredrik Klingwall (Kungliga biblioteket)
Fredrik Persäter (Lantmäteriet)
Johan Oelrich (Arbetsförmedlingen)
Marjan Akhavan (E-hälsomyndigheten)
Matthias Palmér (Digg)
Mattias Ekhem (Digg)
Michalis Vassilas (Digg)
Olov Johansson (SGU)
Stefan Jakobsson (Svensk nationell datatjänst)
Tomas Lindberg (SGU)
Ulrika Domellöf Mattson (Digg)

Bakgrund och syfte

Arbetet sker inom byggblock metadata, en del av Ena, Sveriges digitala infrastruktur och drivs av Myndigheten för digital förvaltning. Syftet med referensgruppen är att arbeta fram en profil för att främja interoperabilitet vid datadelning genom återanvändning av klasser, egenskaper och koncept.

Allt material kommer att publiceras till Github: <https://github.com/diggs sweden/interoperable-specifications>

Materialet som tagits fram är på engelska är för att underlätta dialog med bland annat SEMIC (Europeiska kommissionens initiativ för att förbättra den semantiska interoperabiliteten) på EU-nivå. Referensgruppsdeltagarnas namn kommer att synas på Digg:s GitHub. Om ni inte vill att era namn ska synas på GitHub hör av er till Ulrika Domellöf Mattsson.

Tidigare förstudie finns även den på Github: <https://github.com/diggs sweden/information-models-investigation/tree/main>

Repetition om interoperabla specifikationer (se även ppt)

En specifikation är en behållare med resurser. En samling av en mängd olika artefakter såsom informationsmodeller, guidelines diagram och exempeldata. Förr var specifikationer ofta i PDF-format (goda exempel, modell etc.).

En interoperabel specifikation i sin tur möjliggör återanvändning av klasser, egenskaper (attribut och relationer) samt begrepp. Klasser är i sammanhanget en kategorisering av data eller identitet med liknande egenskaper. Egenskaper som i UML betecknas som attribut (i RDFS talar man om properties).

Syftet med interoperabla specifikationer är att stödja återanvändning mellan specifikationer. Förhoppningen är bland annat att förenkla, effektivisera och förbättra kvaliteten i anslutning till att nya informationsmodeller skapas då myndighetsaktörer inte behöver uppfinna hjulet på nytt utan kan återanvända befintliga klasser. Data som följer olika specifikationer kan vara partiellt interoperabla.

Genomgång av den iteroperabla specifikationens (INSPEC) ingående delar i allmänhet och applikationsprofilen i synnerhet (Se även ppt och Github)

Mathias presenterade kort att reglerna för interoperabla specifikationer (med den nya förkortningen INSPEC) för de olika specifikationsdelarna och specifikationen som helhet:

- Regler för interoperabla specifikationer - PROF-INSPEC
- Regler för datavokabulärer - RDFS-INSPEC
- Regler för terminologier - SKOS-INSPEC
- Regler för applikationsprofiler - SHACL-INSPEC
- Regler för diagram - SVG-INSPEC

Dessa delar täcker olika aspekter av interoperabilitet och standarder för specifikationer, datavokabulärer, terminologier, applikationsprofiler och diagram, vilket säkerställer enhetlighet och kompatibilitet i olika sammanhang.

En centralt insikt sedan förra mötet är att vi kan låta den interoperabla specifikationen (prof:Profile) vara samma uttryck som applikationsprofilens centrala resurs, dvs samma URI bara ett rikare uttryck. Ytterligare en insikt sedan sist är att data vokabulär kan använda OWL:Ontology klassen. En tredje insikt är att vi klarar oss utan att introducera egna properties eller egenskaper genom att använda relationer från Dublin Core (dcterms:references, dcterms:hasPart, dcterms:requires, dcterms:isVersionOf) och Prof (prof:isProfileOf, prof:isInheritedFrom). Rätt mycket går att återanvända med andra ord.

För att en specifikation ska anses vara interoperabel måste den och dess delar ha URIs och beskrivas med PROF samt vara typad som prof:Profile. En interoperabel specifikationsdel kan vara en datavokabulär, en terminologi, en applikationsprofil eller ett diagram medan andra delar kan finnas men har ingen föreskriven betydelse enligt den interoperabla specifikationsprofilen.

Syftet med applikationsprofilen är att tydligare specificera hur klasser, egenskaper och koncept ska återanvändas i nya sammanhang. Behovet kan till stor del täckas av SHACL Shapes Constraint Language, ett språk för att validera RDF-grafer mot ett antal villkor. Dock är SHACL för flexibelt för användningsfallet med applikationsprofiler, därför definierar vi SHACL-INSPEC för att fånga de specifika krav/begränsningar som behöver uppfyllas vid användning av SHACL för att uttrycka applikationsprofiler.

Eftersom SHACL är ett komplext språk räcker inte det inte att räkna upp reglerna i punktform som tidigare, utan läsaren hänvisas även till en separat fördjupning kring hur SHACL-INSPEC som finns här: <https://github.com/diggsweden/interoperable-specifications/blob/main/docs/ap.md>. Mathias utgick från fördjupningsmaterialet och visade med detaljerade exempel på hur SHACL går att använda som grund för att beskriva den semantiska specifikationen. Fördjupningen är ett försök att grunda informationsmodellerna i länkade data med hjälp av SHACL.

Synpunkter och funderingar

En fråga lyftes kring PROF-Vokabuläret och om det var tillräckligt stabilt eftersom det är en working group som publicerat den. Vi har dock redan tidigare använt oss av PROF på dataportalen och det är samma working group som tagit fram DCAT. Man kan säga att vi valde väg redan för några år sedan när vi började uttrycka Profiler baserat på PROF på dataportalen. Det kan komma något nytt som är bättre, men vi testar detta i Norden och Europa och så får vi se om fler hakar på eller kommer på något bättre.

En eventuell utmaning i sammanhanget är att den variant av SHACL-uttrycket som presenterats kanske inte alltid stöds av valideringsmotorer. Det här borde vi testa så att valideringen fungerar i exempelvis en JavaScript-implementation. Fråga om Kungliga biblioteket (KB) har använt någon SHACL-motor för validering. Nej, myndigheten har än så länge inte använt SHACL och jobbat med den typen av validering. Det kan dock bli aktuellt för KB att titta på validering och

SHACL. EU-kommunionen har en motor för validering av DCAT-AP-SE, som man skulle kunna titta på den motor som används i det sammanhanget.

Allmän fråga - Borde vi hitta ett sätt att göra det kompatibelt med XML- ur ett valideringsperspektiv?

Förbättringsförslag kopplat till bakgrundsbeskrivningen på Github - Man kanske skulle lyfta ut en del av diskussionsavsnittet på Github som en separat del - Ett stildokument skulle kunna förklara lite mer om vad som är rimliga sätt att använda tekniken på. Någon sorts korsning mellan tutorial och förklaring om varför t.ex. klasshierarkier kan bli problematiska.

Delar av referensgruppen lyfte att arbetet med att uttrycka den interoperabla specifikationen är komplicerat och även om man förstår och sympatiserar med syftet är det svårt att helt sätta sig i detaljerna. Samtidigt är det en fara om vi försöker förenkla för mycket, då vi kanske inte tar klivet mot att konsumera informationsmodeller som länkade data på det sätt som vi vill. Det kom även frågor kring hur den framtagna profilen ska implementeras och vad det innebär för dataproducenterna i praktiken. Gruppen var återigen överens om att konkreta exempel behöver tas fram.

Vi behöver kanske inte alla förstå exakt hur profilen är uppbyggd utan det viktiga är att man förstår och är överens om de grova dragen och att får fram en första version av den interoperabla profilen. En profil som vi kan utgå ifrån i kommande pilot där vi har för avsikt att visa på hur en informationsmodell i EA skulle kunna representeras som länkade data samt i processen tydliggöra vilken annotering som krävs för att en sådan export och konvertering ska vara möjligt.

Inför fortsättningen

Till nästa möte den 9 december siktar vi på att få fram ett konkret exempel över hur den interoperabla specifikationen kan användas i praktiken.

Efter mötet skickas den förslagna specifikationen med alla sina specifiktionsdelar beskrivna ut på remiss så att referensgruppsdeltagarna har tid att gå igenom materialet och komma med återkoppling. Referensgruppen återsamlas sedan i februari för att gå igenom eventuella förbättringsförslag och synpunkter innan publicering. På mötet i februari vill vi få ett kvitto från referensgruppen om att det här är rätt väg att försätta på. Det här är inget som görs i en handvändning utan det är en resa där målställningen är att jobba iterativt.

Målet i februari är komma överens om en första version av den interoperabla profilen som vi sedan kan utgå ifrån i kommande pilot där vi har för avsikt att visa på hur en informationsmodell i EA skulle kunna representeras som länkade data med möjlighet att navigera ett interaktivt diagram kopplat till applikationsprofilen.