

Strukturella Metadata - Referensgrupp 5

Närvarande

Andrew Mercer (RAA)
Benny Lund (Bolagsverket)
Cilla Öhnfeldt (Naturvårdsverket)
Erik Mossing (Bolagsverket)
Fredrik Persäter (Lantmäteriet)
Johan Oelrich (Arbetsförmedlingen)
Kristofer Gäfvert (Lantmäteriet)
Lars Näslund (Trafikverket)
Marjan Akhavan (E-hälsomyndigheten)
Martin Brandhagen (Vetenskapsrådet)
Matthias Palmér (Digg)
Mattias Ekhem (Digg)
Michalis Vassilas (Digg)
Lars Nässlund (Trafikverket)
Olov Johansson (SGU)
Olof Olsson (SND)
Patrik Wahlgren (SCB)
Peter Israelsson (Digg)
Stefan Jakobsson (Svensk nationell datatjänst)
Tomas Lindberg (SGU)

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: specifications

<https://github.com/diggsverige/interoperable>

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 Michalis Vassilas.

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

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.

Genomgång av den interoperabla specifikationen (INSPEC) v 0.9

Mathias presenterade den senaste beskrivningen av INSPEC och dess olika delar som ni hittar här:

<https://github.com/diggsverige/interoperable-specifications?tab=readme-ov-file>

Avsnittet "Conceptual design" är omstrukturerat så att specifikationsdelarna introduceras redan från början med illustrationen för att ge översikt. Bytt ordning i bilden då det är Applikationsprofilen är den centrala beståndsdel i INSPEC. Kraven i avsnittet är numera uppdelade i följande delar:

- Machine readability requirements (maskinläsbarhetskrav) - En informationsmodell består av klasser, egenskaper och koncept, som bör återanvändas över olika specifikationer. För att hantera detta används applikationsprofiler, som refererar till befintliga element utan att skapa nya genom arv.
- Human readability requirements - För att en interoperabel specifikation ska vara lättillgänglig för människor bör den innehålla diagram som ger en visuell översikt över hur klasser, egenskaper och koncept kombineras. Formellt tillför diagram ingen ny information då all data redan finns i applikationsprofilen, men de hjälper läsaren att snabbt få en överblick och navigera komplexa informationsmodeller.
- Practical requirements - En interoperabel specifikation behöver betraktas som en container med flera delar vilket innebär att specifikationen måste ha en struktur där dess delar kan beskrivas med metadata och kopplas samman på ett tydligt sätt. Dessutom behövs en

mekanism för att hämta specifikationsdelarna. De centrala delarna i en specifikation är terminologier, diagram och applikationsprofiler, vilka bör beskrivas separat och vara åtkomliga.

I avsnittet introduceras även ett nytt begrepp **Foundations Specifikation** som betäckning för specifikationer som introducerar generiska datavokabulärer eller terminologier det vill säga utan "reused"-märkning och utan att tillhandahålla en applikationsprofil. Fundering kring om foundational är ett bra ord eller om gruppen har ett bättre förslag. Reflektion - På svenska fungerar begreppet ganska bra – Grundläggande specifikation.

Även avsnittet Background and motivation har justerat då det fanns en del överlapp mellan avsnitten. Utgångspunkten har varit att försöka skapa tydlighet och hjälpa läsaren. Matthias har förtydligat detta eller gjort ett försök att förtydliga vad återanvändningen handlar om. Huvudfokus är att återanvända så mycket som möjligt.

Det finns numera ett separat avsnitt kring subklasser - Application profiles and subclassing eftersom den tidigare diskussionen blev för lång och inte längre passade in i "design och motivation". Detta var något som också lyftes på senaste referensgruppsmötet som förbättringsförslag kopplat till bakgrundbeskrivningen. Det finns också egen sida som också förklara varför applikationsprofiler behöver en egen lösning, som inte kan hanteras med RDFS eller OWL. Avsnittet avslutas med en jämförelse med hur Semantic Style Guide talar om återanvändning.

I avsnittet - Rules for interoperable specifications (som beskriver den formella specifikationen) finns numera även regler för hur INSPEC entiteter ska representeras i diagram. Modelleringsspråk är valfritt, men UML är att föredra. VG-INSPEC definierar flera regler för att säkerställa att diagram korrekt representerar INSPEC-entiteter:

- **SVG-1:** Hyperlänka INSPEC-entiteter till deras URI:er.
- **SVG-2:** Attribut: typ av INSPEC-entitet, t.ex. data-inspec-type="class".
- **SVG-3:** ID med MD5-summa av entitetens URI (möjliggör CSS).
- **SVG-4:** Attribut: reusable-märkning, t.ex. data-inspec-reusable="true".
- **SVG-5:** Attribut: "main" eller "supportive", t.ex. data-inspec-weight="main"

Det nya avsnittet Bootstrapping specifications har även lagts till om hur den interoperabla specifikationers olika delar ska identifieras. Detta görs genom att ange att resurser överensstämmer med en specifikation. Bootstrapping-specifikationerna behöver inte själva vara interoperabla specifikationer, men det är fördelaktigt om de är det. Än så länge har dessa specifikationer ingen stabil namnrymd, men ett förslag är <http://purl.org/inspec/>. Specifikationer pekas ut via dcterms:conformsTo för att markera att en:

- specifikation är en interoperabel specifikation → inspec:PROF
- specifikationsdel är en data vokabulär → inspec:RDFS
- specifikationsdel är en terminologi → inspec:SKOS
- specifikationsdel är en applikationsprofil → inspec:SHACL
- specifikationsdel är ett diagram → inspec:SVG

Även avsnittet - [Harvesting interoperable specifications](#) har lagts till som beskriver skördningen av interoperabla specifikationer. Skördningen utgår från **PROF-INSPEC** (INSPEC-1 – INSPEC-8) och omfattar hantering av datavokabulärer, terminologier, applikationsprofiler och diagram.

I kommande arbete kommer skördning av UML enligt inspec:UML-OSLO och skördning av CSV enligt inspec:CSV behöva redas ut. När det kommer till UML och INSPEC finns ett antal initiativ som tagit fram användbart material:

- **OSLO-ramverket** – initiativ från den flamländska myndigheten "**Digital Flanders**" (*Open Standards for Linking Organisations*)
 - Tekniskt ramverk för transformation:
GitHub: OSLO-UML-Transformer
- **Arbetet har delvis tagits över av SEMIC** (De som var med i tidigare arbete är **ven aktiva i SEMIC**, framför allt:
 - Vocabularies använder detta
 - Semantic Style Guide på hög nivå
 - Tekniskt ramverk för transformation:
GitHub: SEMIC Toolchain Manual
- **Bygger vidare på ISO-TC211** – Geographic Information/Geomatics (*The Harmonized Model Maintenance Group*). Kanske något för Lantmäteriet som behöver mappa från Inspire?

Inom initiativen ovan har det tagits fram tekniska ramverk för att generera AP från UML, SKOS, och massa andra delar i INSPEC. Det här är något som vårt initiativ kan bygga vidare på och använda i den tillämpningspilot som ska genomföras under året. För de som modellerar i UML måste det vara överkomligt. Man har gjort försök att göra det smidigt, men det är inte helt trivialt.

Matthias gick även igenom utkast kring hur entiteter i UML skulle kunna annoteras där Data vokabulär och Applikationsprofiler känns naturliga, men där terminologier även de kan beskrivas skulle kräva en egen UML-modell, kanske bättre att tillhandahålla i SKOS är arbetshypotesen just nu.

Synpunkter och funderingar

En kort sidoreflekation är att det antagligen gör det lättare med dark mode för specifikationen. Går använda CSS media queries för att lägga till standardfärger om webbläsaren föredrar darkmode: `@media (prefers-color-scheme: dark) { }` så kan vara en del av en standard CSS.

Frågan om vi inte borde titta närmare på DCTAP för att uttrycka Applikationsprofiler och SHAC lyftes återigen och ett ärende finns på Github <https://github.com/diggsverige/interoperable-specifications/issues/16>. Inom DDI-aliansten så har dom tagit fram autogenererade svg-diagram där begränsas grafen till en klass med dess arv, properties och relationer t. ex: https://ddi-cdi.github.io/ddi-cdi_v1.0/field-level-documentation/DDICDILibrary/Classes/Conceptual/InstanceVariable.html. Om man begränsar det till en klass så kan det vara en väg att gå.

Svar - Det är mycket man kan göra, men frågan är vad vi ska fokusera på. Vi har tänkt oss att utgå från det manuellt skapande diagrammet i EA. Om man markerar huvudklasserna på något sätt och det leder till en ett begränsat urval kanske automat-layouten skulle kunna fungera.

Inför fortsättningen

Fram till nästa möte den 24 februari har referensgruppsdeltagarna tid på sig att gå igenom den första versionen av den interoperabla specifikationen INSPEC och återkomma med återkoppling på Github.

Det vore bra om samtliga deltagare åtminstone läst igenom 1 – 3 samt 8, men även inspel på övriga avsnitt välkomnas. Fundera gärna kring om foundational specification är ett bra begrepp eller någon i gruppen har ett bättre förslag.

Alla behöver inte förstå exakt hur profilen är uppbyggd utan det viktiga är att förstå och vara överens varför den interoperabla profilen är viktig. En profil som vi kan utgå ifrån för att konvertera en informationsmodell i EA så att den kan representeras som länkade data samt tydliggöra vilken annotering som krävs för att en sådan export och konvertering ska vara möjligt.

På nästa möte den 24 februari återsamlas gruppen för att gå igenom eventuella förbättringsförslag och fastställa en första version av INSPEC som vi sedan kan utgå ifrån i det fortsatta arbetet. Efter mötet den 24 februari pausas referensgruppens arbete för denna gång och kommer att återupptas när Digg eller någon annan myndighet/aktör ser behov av att uppdatera specifikationen. Digg har även initierat och kommer att föra en dialog med SEMIC för att förankra det arbete som gjorts inom referensgruppen med EU. Om Matthias får förändringsförslag till INSPEC av större karaktär kommer han lägga den i en separat branch på Github för framtida översyn. Mindre ändringar kommer att göras direkt i nuvarande version.

Fokus i vår kommer att ligga på att pröva tillämpningen av INSPEC i praktiken och jobba vidare med en pilot vilket innebär att ett webbgränssnitt kommer tas fram för att utforska strukturella metadata som följer de rekommendationer som tagits fram i förstudien om "systemstöd för informationsmodeller" och ska motsvara scenarion A och B, se detaljer här:

[https://github.com/diggsweden/information-models-](https://github.com/diggsweden/information-models-investigation/blob/main/leverabler/kravställning.md)

[investigation/blob/main/leverabler/kravställning.md](https://github.com/diggsweden/information-models-investigation/blob/main/leverabler/kravställning.md) Även ett koncept för uppackning av specifikationer som följer uppfyller INSPEC profilen kommer tas fram vilket också innebär funktionalitet för uppackning av format 2 (UML) enligt tidigare plan tas fram för piloten. Piloten beräknas vara klar oktober 2025.

Grunddatadomänernas informationsmodeller skulle kunna vara bra kandidater att testa den praktiska tillämpningen av INSPEC, men här behöver en dialog föras inom Digg med ansvarig för ansvarig för konceptet nationella grunddata. Skatteverket ansvarar för grunddatadomänen Person, bolagsverket ansvarar för grunddatadomänen Företag. Trafikverket ansvarar för utveckling av en grunddatadomän för Transportsystem. Lantmäteriet ansvarar för grunddatadomänen Fastighets- och geografisk information. E-hälsomyndigheten ansvarar i sin tur för grunddatadomän för Hälsa, vård och omsorg.