

Fra ide til utveksling av data i form av WFS/GML

Morten Borrebæk

Morten Borrebæk: From idea to exchange of data as WFS / GML

KART OG PLAN, Vol 66, pp. 265–269. P.O.Box 5003, N-1432 Ås, ISSN 0047-3278

This article describes the process from the idea of a new task (within the geographic information domain) to specification, implementation and making data available that fulfils the requirements of the new task. This process is implemented as a service like Web Feature, and applied using a cyclic methodology based upon international standards. This methodology needs to be repeatable, domain independent and software/platform independent.

In spatial data infrastructure work today there is a focus on making existing data available for the users, both existing users who require better access and quality of the data, and new users who apply geographic information as an important part of their decision-making process.

It is beneficial that new requirements can be solved by applying existing data, since new data captured in most cases will increase the cost of the new project dramatically. But there is little focus on the methodology to acquire and investigate the requirements in a formal way.

One of the main questions regarding the framework design constraints on the content is if a data product should fulfil the requirements of a large and general community or a small specialized community.

To ensure that the user requirements are well defined and understood a use-case methodology should be applied. The requirement of the use-case study should be the basis for a data product specification, which is a specification that contains the product requirements.

The next step in the process is to create an application schema for the feature types, property types and associations required together with a feature catalogue, eventually based upon an existing data dictionary.

Doing an 'as is' analysis and 'gap' analysis will clarify how existing data will conform to this virtual specification. Applying WFS with any form of a query language to filter the information will enable the service to extract data from existing data sources and present the result according to the specification.

This methodology applies several of the ISO 191xx standards.

Keywords: repeatable methodology, use case, data product specification, application schema, feature catalogue, GML application schema, WFS.

Morten Borrebæk, Chief Engineer, Statens kartverk, 3507 Hønefoss.

E-mail: morten.borrebæk@statkart.no

Introduksjon

En moderne geografisk infrastruktur skal dekke mange brukeres behov i form av tilgjengelighet til data og tjenester knyttet til disse. I vår norske infrastruktur legges det stor vekt på å gjøre data fra Norge Digitalt-partene tilgjengelig i et forvaltningssamarbeid. Dette forvaltningssamarbeidet skal igjen sikre tilgang til og tjenester mot partenes data.

Dataene skal dekke behovet til eksisterende brukere og nye brukere i forvaltningssamarbeidet. I tillegg skal de også dekke myndigheters behov for rapportering, der stedfestet

informasjon er en viktig del. Eksempler på dette er rapportering for KOSTRA (KOMMUNE-STAT-RAPPORTERING) eller rapportering i henhold til direktiver og forskrifter slik som EU's vannrammedirektiv. Det er ønskelig i størst mulig grad å dekke disse behovene med utgangspunkt i eksisterende data og tjenester. I mange tilfeller vil dette la seg, gjøre, i noen tilfeller må en gå til innsamling av nye data, hvilket ofte er svært kostnadskreven. Denne artikkelen beskriver en metodikk som i størst mulig grad bygger på nasjonale og internasjonale standarder for geografisk informasjon, med fokus på følgende standarder:

ISO 19103	Geographic Information – Conceptual schema language
NS-EN ISO 19109	Geografisk informasjon – Regler for applikasjonsskjema
NS-EN ISO 19110	Geografisk informasjon – Metodikk for objektkatalogisering
NS-EN ISO 19119	Geografisk informasjon – Tjenester
ISO 19131	Data Product Specification

Det bemerkes at en slik metodikk er under utarbeidelse innenfor RISE (Reference Information specification in Europe), som er et prosjekt innenfor EU's sjettede rammeverksprogram og som tar sikte på å utvikle en metodikk for harmonisering av data innen Europa. Denne metodikken tar spesielt utgangspunkt i behovet for harmonisering mellom ulike land, men metodikken er også anvendelig på nasjonalt eller lokalt nivå. Implementasjon av en slik metodikk er under utprøving, men foreløpig ikke implementert.

Metodikk

Metodikken er bygd opp rundt følgende aktiviteter:

- Identifikasjon av krav (use cases)
- 'As is' analyse (Analyse av hvilke data og tjenester som finnes i dag)
- 'Gap' analyse (Forskjellen mellom eksisterende data og de krav som er identifisert gjennom 'use case' en).

Metodikken krever følgende personressurser eller roller:

Domene ekspert:

Fagekspert innenfor det fagområde som representerer kravene

GI-arkitekt:

Fagekspert innenfor standarder knyttet til geografisk informasjon slik som ISO/TC 211 og OGC , herunder modellering

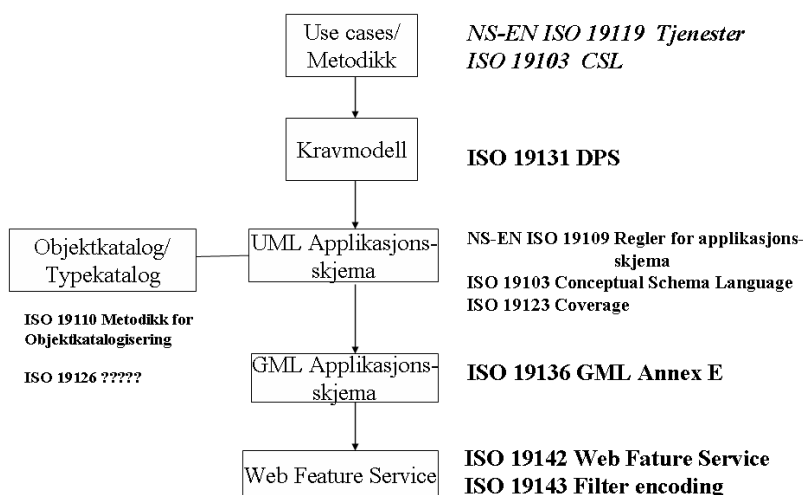
Koordineringsansvarlig:

Den som håndterer selve prosessen og sørger for at alle data harmoniseringskrav er identifisert og adressert. (Facilitator)

Software ingeniør:

Spesialist innenfor implementering og anvendelse av relevante spesifikasjoner

Avhengig av kompleksiteten kan samme person ha flere roller. Aktivitetene går selvsagt ikke upåvirket av hverandre. En vil skaffe seg informasjon om eksisterende data allerede i kravfasen, men en vil også kunne komme i en situasjon hvor det er umulig å skaffe nødvendige data ut fra eksisterende kilder.



Figuren over identifiserer de ulike aktivitetene i metodikken. Det første leddet er en 'use case' metodikk. En 'use case' adresserer blant annet hvordan en skal gjøre funksjonelle krav lesbare og forståelige. Kravene som identifiseres inngår i en kravmodell, denne er utgangspunkt for en produktspesifikasjon i henhold til ISO 19131 Data Product Specification.

Produktspesifikasjon inneholder blant annet en beskrivelse av innholdet, enten i form av vektordata med objekttyper, egenskaper og assosiasjoner, eller som et bilde / coverage. I henhold til standarden er det ikke noe absolutt krav om at en beskriver dette innholdet ved hjelp av et generelt modelleringspråk, f.eks. UML (Unified Modelling Language), men denne metodikken forutsetter at dette gjøres. Innholdet beskrives i form av et UML applikasjonsskjema. Objekttyper, egenskaper og assosiasjoner kan være nærmere beskrevet i en objektkatalog.

UML applikasjonsskjema er utgangspunkt for å generere et GML applikasjonsskjema. Dette skjemaet vil være utgangspunkt for en WFS (Web Feature Service) som er en tjeneste som henter data fra ulike kilder og leverer dette som GML (Geography Markup Language) -data, i henhold til GML-applikasjonsskjema.

Identifikasjon av krav og prosessen mot applikasjonsskjema

Identifisering av krav (use cases)

Det finnes mange ulike 'Use case' metoder innenfor informasjonsteknologien. Ulike deler av IT bransjen har utviklet ulike løsninger. Det ligger ikke innenfor denne metodikken å velge en spesiell av disse. Det er viktig å merke seg at geografisk informasjonsteknologi blir en integrert del av generell informasjonsteknologi, hvor det bare er stedfestingsdelen som er spesiell. Det er viktig at ulike miljøer kan fortsette med sine 'Use Case' metoder men samtidig spesifisere kravene tilstrekkelig detaljert til at disse kan være utgangspunkt for et applikasjonsskjema.

For å bistå denne prosessen er det i regi av RISE benyttet en 'use case' mal utviklet av OGC (Open GIS Consortium), og det er utviklet en mal for sjekklister i form av Excel reg-

neark. Malene skal avdekke områder hvor harmonisering er påkrevet, og gi grunnlag for utviklingen av produktspesifikasjonen.

Selv om denne metodikken ikke beskriver noen standardisert 'use case' metodikk, er dette som metode omtalt flere steder i de internasjonale standardene innenfor geografisk informasjon. NS-EN ISO 19103 Conceptual Schema Language beskriver blant annet at 'use case' er et nyttig hjelpemiddel i blant annet utviklingen av kravmodeller samt analyse- og arkitekturmodeller. NS-EN ISO 19119 Tjenester beskriver at 'use case' sammen med UML Interaction diagram (diagramteknikk som illustrerer hvordan objekter påvirkes gjennom meldinger) er nyttige for å beskrive de dynamiske aspektene av en tjeneste.

Utviklingen av produktspesifikasjon

En annen artikkel i dette fagtidsskrift forklarer hvordan en produktspesifikasjon i henhold til ISO 19131 Data Product Specification er beskrevet. Spesifikasjonene bygges opp på bakgrunn av 'use case'ene som identifiserer kravene samt eventuelle sjekklister som er fylt ut i prosessen. Men det er også av avgjørende betydning at en domeneekspert er med i denne fasen av metodikken, selv om vedkommende ikke er noen ekspert på utviklingen av en produktspesifikasjon. Resultatet av denne prosessen er da en produktspesifikasjon for et geodataprodukt som inneholder de objekttyper, egenskaper og assosiasjoner som er nødvendige for å dekke en brukergruppes behov. Produktspesifikasjonen beskriver et virtuelt produkt, som ikke nødvendigvis er identisk med noe eksisterende produkt.

UML applikasjonsskjema og objektkatalog

Den delen av produktspesifikasjonen som inneholder beskrivelsen av innholdet i form av objekttyper eller bilde/Coverage beskrives i henhold til UML og blir et UML applikasjonsskjema. Dette er en beskrivelse av de data som inngår og som en trenger for å fylle de krav som er identifisert gjennom 'use case' fasen. UML-applikasjonsskjemaet beskrives i henhold til NS/EN ISO 19109 Regler for applikasjonsskjema, for den tekslige beskrivelsen følges NS-EN ISO 19110 metodikk for objektkatalogisering.

GML applikasjonskjema

GML-applikasjonsskjemaet genereres automatisk fra UML applikasjonskjemaet, og er et skjema for beskrivelse av data i XML-syntaks (Extensible Markup Language). GML, inkludert GML-skjema, er standardiserert i ISO 19136 Geography Markup Language. Annex E inneholder regler for hvordan en beskriver et GML-skjema på bakgrunn av et UML applikasjonskjema, dvs 'mapping' mellom konsepter i UML og GML.

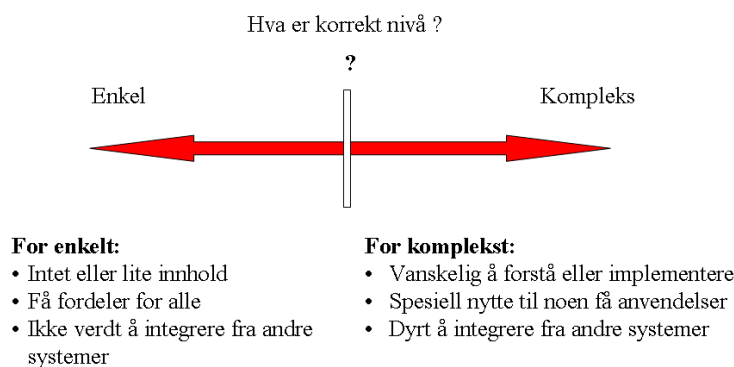
'As is' analyse

En 'As is' analyse beskriver de data som faktisk eksisterer, og fremkommer i forbindelse med 'Use case' prosessen. Dette er data som

stort sett finnes i eksisterende geodatabaser, men mye viktige data kan også finnes som tekstdokumenter i Word eller Pdf-format, som excel regneark, eller kanskje ikke engang som digitale data.

Utfordringene for en dataleverandør er i hvor stor grad en er i stand til å dekke brukernes behov ved utviklingen av et gitt dataset, f.eks fra Statens kartverk.

Figuren under illustrerer at dersom dataene innholdsmessig er for enkle er de også av liten interesse, dersom de er kompliserte vil de være av stor nytte for noe få, og tilsvarende vanskelig å ekstrahere informasjon fra i andre brukersammenhenger. Dette vil fremkomme gjennom 'use case' prosessen.



(fra: Douglas Nebert, FGDC, 2005)

Det kan også være tilfelle at det aktuelle geografiske område går på tvers av de respektive dataleverandørers dekningsområder og krever data fra flere leverandører. Dette er spesielt aktuelt der det geografisk interesseområde dekker flere land, slik som f.eks nedbøfelter som utgangspunkt for rapportering i henhold til vannrammedirektivet, hvor data må hentes fra eksempelvis både Norge og Sverige, og hvor de eksisterende data kan være svært forskjellige. Harmoniseringskomponenter bør identifiseres tidlig i prosessen, og ulike typer sjekklister vil være til god hjelp i dette arbeidet.

'Gap' analyse

Forskjellen mellom resultatet av 'as is' analysen og den virtuelle produktspesifikasjonen fremkommer som et resultat av 'Use case'

se' prosessen, og er igjen utgangspunkt for implementering av WFS (Web Feature Service). WFS er beskrevet i andre artikler i dette fagtidsskriftet.

Oppsummering/konklusjon

Metodikken involverer flere kjente prosesser som nå er standardisert gjennom internasjonalt standardiseringsarbeid i regi av ISO/TC 211 og OGC. I Norge har vi kommet langt i utviklingen av en generell objektkatalog for geografiske objekter. Vi har også kommet langt i utviklingen av produktspesifikasjoner for geodata, og er i ferd med å gjøre disse konforme med internasjonale standarder.

Men den delen som kanskje er den mest spennende delen av metodikken, spesielt i fra et teknologisk synspunkt, er WFS (Web

Feature Service). Denne teknologien muliggjør at data kan hentes fra forskjellige servere og gjennom ulike prosesser (spørring og filtrering) presenteres i henhold til et omforent GML skjema.

Jo mer avanserte spørre- og filtreringsmekanismene blir, jo større er muligheten for å integrere data fra ulike leverandører og administrative områder. Web Feature Service, sammen med 'rene' WEB services (Web baserte tjenester som benytter SOAP, WSDL og XML) vil kunne løse flere av dagens behov basert på eksisterende typer data, i mange tilfeller også uten at brukeren sitter på avanserte GIS systemer. Avanserte GIS analyser kan utvikles som 'WEB services' og integre-

res i metodikken. Dette gjør det også lettere å integrere geografiske data i beslutningsprosesser, også på områder hvor det i dag ikke benyttes stedfestet informasjon.

Det er fortsatt et stykke frem før dette er en realitet, og det gjenstår mye arbeid før denne metodikken er utprøvd og klar til å bli implementert. Uten internasjonale standarder hadde en slik metodikk vært svært vanskelig å realisere. Denne metodikken har også som utgangspunkt i å realisere INSPIRE's ønske om å integrere eksisterende infrastrukturkomponenter (f.eks data) fra ulike Europeiske land uten at det skal være nødvendig å igangsette ny datafangst, eller gjøre endringer på datainnholdet.