

Produktspesifikasjoner – Den mest detaljerte spesifisering av et dataprodukt

Morten Borrebæk

Morten Borrebæk: Data product specification – the most detailed specification of a data product

KART OG PLAN, Vol 66, pp. 238–242. P.O.Box 5003, N-1432 Ås, ISSN 0047-3278

In Norway there are product specifications covering different kinds of geographic data, made for different usage and applied by different user communities. However, these product specifications do not conform to any commonly approved standard and the specifications are different in content and structure, although there are several similarities. Still, much data that is published do not follow any product specification at all.

ISO 19131 Data Product Specification defines how to set up a detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party. It is a precise technical description of the data product in terms of the requirements that it should fulfil.

There is a strong relationship between a data product specification and metadata. The data product specification only defines how the dataset should be. For various reasons, compromises may need to be made during the process. The metadata associated with the product dataset should reflect how the product actually is.

The purpose of this International Standard is to provide practical help in the creation of data product specifications, in conformity with other existing standards for geographic information.

This International Standard is intended for use by producers, providers and potential users of data products, and is an important part of a spatial data infrastructure.

Keywords: Data Product Specification, Spatial data, spatial infrastructure

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

E-mail: morten.borrebæk@statkart.no

Definisjoner og forkortelser

En moderne geografisk infrastruktur krever tilgang til geografisk informasjon i form av utveksling av datasett på filbasis, men også gjennom predefinerte tjenester (f.eks API'er, Application Program Interface). I Norge har vi i lang tid benyttet SOSI formatet som et utvekslingsformat for geografiske data, samt at vi har grensesnitt som NGIS-API og MATRIKKEL-API, etc basert på standarder. Både utveksling av data i form av filer og tilgang til data implementert som tjenester fordrer at dataene følger en felles spesifisering, dvs. at innhold og struktur er i henhold til en systemuavhengig spesifisering som er tilgjengelig for både leverandør og mottaker, og som kan implementeres på tvers av implementasjoner og plattformer.

ISO 19131 Data Product Specification er en detaljert beskrivelse av hvilke krav som er satt

til et datasett eller datasett serier sammen med nødvendig tilleggsinformasjon som gjør det mulig å etablere og forvalte slike data, samt gjøre disse anvendelige for brukerne. Det er den mest detaljerte tekniske beskrivelsen av data (ofte betegnet som dataprodukt) i form av krav som skal eller kan oppfylles.

Standarden ble sendt fra ISO/TC 211 til ISOs sentralsekretariatet i midten av august 2006, for votering som FDIS. FDIS står for Final Draft International Standard og er siste nivå før standarden blir utgitt som en IS, Internasjonal Standard. Denne artikkelen er skrevet med utgangspunkt i FDIS versjon av ISO 19131 Data Product Specification, men det er liten grunn til å forvente større tekniske justeringer.

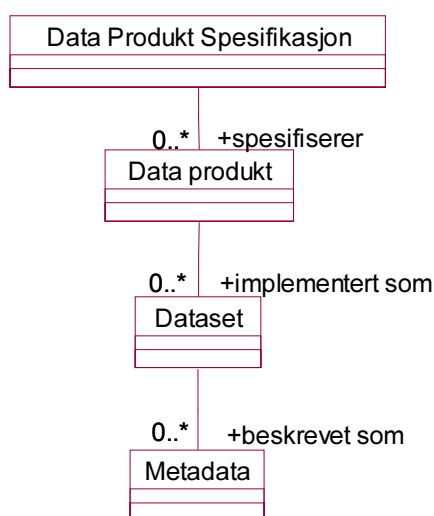
Standarden dekker spesifisering av geodataprodukter for geodata. I artikkelen er begrepet data produkt synonymt med produkt, og 'dataproduktspesifisering' eller

'produktspesifikasjon for geodata' er forenklet til produktspesifikasjon, underforstått at det er produkter og spesifikasjoner innenfor området geografiske data. Standarden omhandler ikke spesifikasjon av tjenester.

Hva er forskjellen på en produktspesifikasjon og metadata

En produktspesifikasjon beskriver hvordan et produkt var ment å være. Av ulike grunner kan det være nødvendig å gjøre diverse kompromisser i selve produksjonsfasen. Eksempel på dette er at en kanskje ikke oppfyller krav til stedfestingsnøyaktigheten eller at kostnadene knyttet til registrering av visse objekttyper viser seg å være uakseptabelt høye. Hvordan produktet egentlig ble er beskrevet i form av metadata.

Figur 1 viser hvordan en produktspesifikasjon spesifiserer et produkt. Dette produktet er igjen implementert som datasett, som igjen skal beskrives i form av metadata (for geografisk informasjon fortrinnsvis NS_EN ISO 19115:2003 standard).



Figur 1

En produktspesifikasjon kan lages ved ulike anledninger, for eksempel for selve produksjonsfasen, men kan også lages for produkter som avledes fra eksisterende data. Den kan lages av produsenten for å spesifisere pro-

duktet, men også av brukere for å beskrive deres krav. Et tidlig fokus på produktspesifikasjoner vil i alle tilfelle forenkle prosessen med å fremskaffe metadata, som det ofte er et krav om å lage og forvalte.

Inndeling av produktspesifikasjonen i delprodukter (delmål)

En viktig egenskap ved standarden er muligheten for inndeles et produkt i delprodukter, og knytte produktspesifikasjonsinformasjon kun til dette delproduktet.

Et delprodukt kan bestå av et nærmere angitt geografisk område, ulike objekttyper eller egenskaper, deres geometriske eller topologiske angivelse eller et spesielt raster/bilde dersom et produkt består av flere. Modellen for dette er mer detaljert beskrevet i standarden, med informasjon om hvilke egenskaper som er obligatoriske for å definere et slikt delprodukt, samt hva som er påkrevet og opsjonelt.

Et eksempel på at angivelse av delprodukt er en nyttig mekanisme er at en i produktspesifikasjonen kan si at kravet til fullstendigheten for en objekttype «bygning» kan settes til 99 % mens kravet til fullstendigheten for en objekttype «bekk» kan settes til 90%. Likeledes kan en spesifisere at kravene til stedfestingsnøyaktighet kan være forskjellig i ulike geografiske områder innenfor et produkt, jfr FKB's områdeinndeling, slik at kravet til stedfestingsnøyaktighet kan spesifiseres å være høyere i FKB-A enn f.eks FKB-C.

Produktspesifikasjonens inndeling

En produktspesifikasjon er inndelt i to hovedseksjoner, en generell oversikt samt klart definerte pakker med informasjon.

Oversikt

Dette er en ren tekstlig fremstilling om hvordan produktspesifikasjonen er blitt til. Den kan inneholde informasjon om tittel, ansvarlig etat, begreper med definisjoner, forkortelser, etc. I tillegg kan den inneholde en enkel beskrivelse av hva produktspesifikasjonen inneholder, hvilket område den dekker, hensikten bak spesifikasjonen, etc.

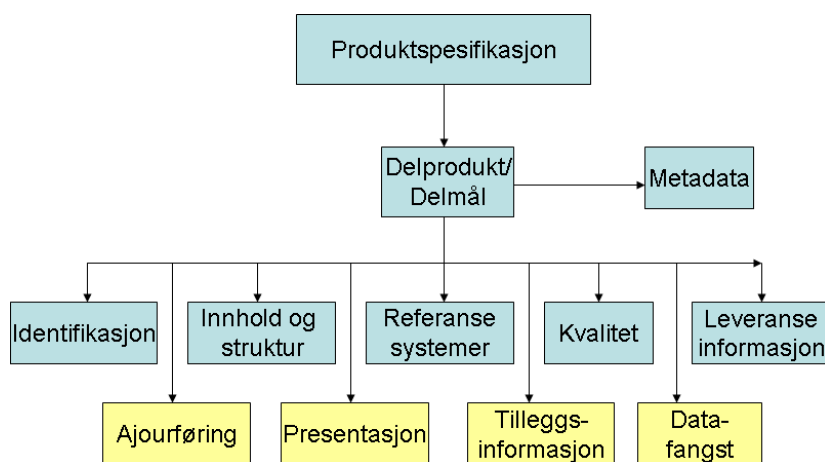
Oversikten kan inneholde en enkel beskrivelse av de informasjonselementer som er

beskrevet mer detaljert under de andre seksjonene.

Klart definerte pakker med informasjon, hver beskrevet i UML

I likhet med de fleste standardene som er utviklet i regi av ISO/TC 211 benyttes UML (Unified Modelling Language) for å beskrive

innholdet av de ulike informasjonspakkene. Disse modellene er mer presise og omfattende enn det som gjengis i denne artikkelen. Det henvises til standarden for den mest presise informasjonen i form av UML-modeller. Den engelske teksten på disse pakkene er tatt med i parentes for å lette koplingen mot den internasjonale standarden.



Figur 2 viser informasjonspakkene som inngår eller kan inngå i en produktspesifikasjon

Informasjon som skal være med i en produktspesifikasjon

Identifikasjonsinformasjon (data product specification)

Dette er informasjon som identifiserer produktet i form av egenskaper som tittel, introduksjon, det geografiske området som produktet dekker eller den tidsperioden som produktet gjelder for. En viktig egenskap her er hovedkategori, en inndeling ut fra produktets hovedtema. Lovlige verdier er beskrevet i metadatastandarden.

I tillegg kan spesifikasjonen inneholde en alternativ tittel og nærmere beskrivelse av målsettingen med spesifikasjonen, angivelse av om det er vektor data eller raster data, eventuelt andre typer data, hvilken oppløsning produktet skal ha (kvalitetsinformasjon) samt mulighet for å angi annen beskrivende informasjon om produktet.

Innhold og struktur (data content and structure)

Produktspesifikasjonen inneholder en detaljert beskrivelse av produktet i form av objekttyper med tilhørende egenskaper og assosiasjoner, eller som et 'coverage' eller bilde.

Innholdet i et objekttype-basert (vektorbasert) produkt beskrives i form av et applikasjonskjema og en objektkatalog. Et eksempel på et slikt produkt er FKB, som er basert på SOSI generell objektkatalog. Applikasjonskjemaet til FKB består av et nærmere spesifisert utdrag av SOSI generell objektkatalog, men er mer detaljert med tanke på hvilke objekttyper med tilhørende egenskaper som inngår, samt hva som er obligatorisk og opsjonelt.

Et applikasjonskjema inneholder den formelle beskrivelsen av strukturen i dataene i form av en implementasjonsuavhengig og plattformuavhengig UML-modell. Reglene for modellering er angitt i NS-EN ISO 19109 Geografisk informasjon – Regler for applika-

sjonsskjema (spesielt kapittel 7 og 8) og standarden 'Retningslinjer for modellering i UML' som er en ytterligere detaljering av hvordan vi gjør dette i Norge. Sistnevnte standard kommer som en del av SOSI 4.0 og er tilgjengelig på kartverkets WEB sider.

En objektkatalog er en katalog som inneholder beskrivelse av alle objekttyper, egenskaper og egenskapsverdier, samt assosiasjoner mellom objekttyper innenfor et fagområde. En objektkatalog skal beskrives i henhold til NS-EN ISO 19110 Geografisk informasjon – Metodikk for objektkatalogisering. [Standardprodukter innenfor Norge Digitalt skal være basert på SOSI generell objektkatalog, som er en generell nasjonal objektkatalog som er et utgangspunkt for mange produktspesifikasjoner].

Innholdet i et 'coverage-' / bildebasert produkt beskrives i form av en identifikator samt en kortfattet beskrivelse, hva slags type bilde / coverage produktet beskriver, samt ytterligere informasjon som er nødvendig for å beskrive produktet .

Et 'coverage' er en subtype av et objekt (feature). Den oppfører seg som en funksjon som returnerer en eller flere egenskapsverdier knyttet til geografisk område. Den mest vanlige formen for 'coverage' er raster, og denne type produkter kalles ofte raster eller bildeprodukt.

Det finnes ingen predefinert kodeliste for 'coverage' typer, dette angis som fri tekst. Men standarden gir eksempler på at et 'coverage' kan være raster bilde (f.eks Landsat , Spot), digitale høydemodeller, digitale temperatur- eller nedbørmodeller, etc.

Referansesystem (reference system)

En produktspesifikasjon skal inneholde informasjon om hvilket referansesystem som skal ligge til grunn for stedfestingen eller tidfestingen i produktet.

Referansesystem for stedfesting er enten et koordinatbasert referansesystem som angir stedfesting i form av lengde og bredde eller kartprojeksjon, eller et system hvor stedfestingen skjer ved geografiske identifikatorer, slik som stedsnavn, kommune, grid-system, eller lignende.

Koordinatbasert stedfesting skal være i henhold til NS-EN ISO 19111 Geografisk in-

formasjon – Modell for stedfesting med koordinater, stedfesting i form av geografiske identifikatorer skal være konform med NS_EN ISO 19112 Geografisk informasjon – Modell for indirekte stedfesting.

Referansesystem for tidfesting skal være konform med NS-EN ISO Geografisk informasjon – Modell for å beskrive tidsaspekter.

Kvalitet (data quality)

Produktspesifikasjonen skal spesifisere de kvalitetskrav som skal oppfylles før produktet kan sies å være konformt med spesifikasjonen. Kvalitetskrav kan stilles i form av krav til stedfestingsnøyaktighet, fullstendighet, logisk konsistens, egenskapsnøyaktighet og tidfestingsnøyaktighet. Disse er nærmere beskrevet i NS EN ISO 19113 Geografisk informasjon – Geografisk informasjon – Prinsipper for spesifisering av kvalitet og NS-EN ISO 19115 Geografisk informasjon – Metadata.

Leveranseinformasjon (data product delivery)

En produktspesifikasjon skal identifisere kravene knyttet til leveransen av et produkt. Følgende egenskaper er obligatoriske :

- angivelse av de formater produktet kan leveres på i form av et navn
- leveranseenheter, f.eks lag, geografisk område, kartblad, etc

Til formatbeskrivelsen kan det også knyttes informasjon om hvilken versjon av formaterne som er benyttet, navn på eventuelle profiler, strukturen i leveransefila, språk som er benyttet innen datasettet, samt angivelse av hvilke karakterkoding som benyttes.

Det kan også knyttes informasjon om størrelsen (Mbyte) på en leveranseenheter, navnet på leveransemediet samt annen informasjon.

Metadata

Kjerne-metadataene (ISO 19115 Core Metadata elements) skal inkluderes i produktet, dvs være tilgjengelig når produktet er utviklet og tilgjengelig for bruk. Metadataegenskaper ut over kjerne-egenskapene som en i spesifikasjonsarbeidet vurderer som nødvendige skal spesifiseres i produktspesifikasjonen, i henhold til NS_EN ISO 19115 Metadata.

Formatet og kodingen av metadata skal beskrives i produktspesifikasjonen.

Informasjon som kan være med i en produktspesifikasjon

Ajourføring (data maintenance)

En produktspesifikasjon kan inneholde informasjon om hvordan dataene er eller skal vedlikeholdes, dvs ajourføres. Dersom ajourføringsinformasjon oppgis skal denne inneholde ajourføringsfrekvensen.

Presentasjon (Portrayal)

En produktspesifikasjon kan inneholde informasjon om hvordan produktet skal presenteres i form av et kart eller et bilde på en data-skjerm. Enkelte produkter har standardisert presentasjon, f.eks ECDIS (elektroniske sjøkart) hvor det er forskjell på hvordan dataene vises på en skjerm om dagen og natten.

En bør også ta høyde for at presentasjonsreglene for et objekt kan endre seg med skalaen (målestokken) objektet presenteres i.

Dersom produktspesifikasjonen skal inneholde presentasjonsinformasjon skal denne utgjøre en referanse til et sett presentasjonsregler og et sett presentasjonsspesifikasjoner jfr ISO 19117 Portrayal.

Datafangst (Data capture)

Et produkt kan fremskaffes på flere måter, ved bruk av ulike typer teknologi og leverandører. En produktspesifikasjon skal i utgangspunktet være så god at ulike typer teknologi kan anvendes for å oppfylle de krav som spesifikasjonen angir, uten at dette er knyttet til en spesiell datafangsmetode. Slik vil en ikke hindre ny teknologi i å bli tatt i bruk.

På den annen side vil det kunne forekomme tilfeller hvor en ønsker at data skal fremskaffes etter gitte prinsipper og metoder, og dette kan beskrives i en produktspesifikasjon i form av en tekstlig beskrivelse av kildene og prosessene som skal benyttes.

Tilleggsinformasjon (additional information)

Informasjon som er vesentlig for produktet og som ikke er beskrevet under de andre informasjonspakkene kan beskrives som til-

leggsinformasjon, i form av en ren tekstlig beskrivelse.

Oppsummering

Standarden beskriver hva slags informasjon som inngår i en produktspesifikasjon for geodata. På enkelte områder er den meget detaljert, på andre områder er det tilstrekkelig å beskrive dette som ren tekst.

Standarden inneholder også mulighet for å definere ulike krav til ulike deler av et produkt. De fleste produkter er ganske homogene, og vil ikke ha behov for en slik oppdeling.

Erfaringene en har med utarbeidelse av produktspesifikasjoner i Norge indikerer at strukturen i denne standarden er dekkende for de behov vi har for både basis geodata, tematiske geodata og avleda data.

Men standarden sier ikke noe om hvordan innholdet skal presenteres i form av et dokument. Spesielt med tanke på rekkefølgen av informasjonspakkene og de mulighetene som ligger i inndeling av delprodukter, kan innholdet presenteres på svært ulike måter. Det er derfor viktig å videreføre dette internasjonale arbeidet i en nasjonal kontekst, slik at produktspesifikasjoner som utvikles og publiseres i Norge (f.eks i forbindelse med Norge Digitalt) ser mest mulig like ut. Dette vil gjøre det lettere å lese og forstå produktspesifikasjoner på tvers av fagområder.

ISO 19131 Data Product Specification er en standard som har referanse til en rekke andre ISO 191xx standarder, slik som 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, ISO 19123 Geographic information – Coverages, etc,etc. Det er følgelig viktig å lage nødvendige veiledere samt applikasjoner som bistår brukerne i å lage produktspesifikasjoner.

Å spesifisere våre data i form av produktspesifikasjoner er et viktig ledd i utviklingen av en nasjonal geografisk infrastruktur. For spesifisering av innhold og struktur har vi har et godt utgangspunkt i SOSI generell objekttallog.