

<p>Versjon:2.0 Endret dato:20.11.2015</p>	<h1>Retningslinjer for dataformater ved formidling av stedsdata</h1>
<p>Dokumentansvarlig: Gerd Mardal</p>	



Innhold

1	Innledning	3
1.1	Formål.....	3
1.2	Ansvar.....	3
1.3	Revisjon	3
2	Forvaltningsprosesser for stedsrelaterte data	3
2.1	Standardisering av geodata.....	4
2.1.1	Modellering av geodata	5
2.1.2	Koding av data - dataformater.....	5
2.2	Produksjonsløypa for geodata	5
2.2.1	Dokumentasjon (produktspesifikasjon)	5
2.2.2	Validering av stedsdata	7
2.2.3	Transformasjoner mellom formater.....	7
3	Formater.....	8
3.1	Vektorformater	8
3.1.1	Geography Markup Language (GML)	8
3.1.2	SOSI-format – SOSI.....	9
3.1.3	S-57	9
3.2	Rasterformater.....	9
3.2.1	GeoTIFF	9
3.3	Formater for distribusjon av GNSS data (Global Navigation Satellite System)	10
3.4	Andre tilleggsformater.....	10
	Referanser	11
	Vedlegg	12

1 Innledning

Tilfanget av stedsdata er økende, og bruken sprer om seg på internett. Stedfestet data blir brukt i alle deler av offentlig forvaltning og privat næringsliv, og i økende grad av private 'konsumenter'.

EU ønsker fri flyt av varer og tjenester og bruker standardisering som et virkemiddel for dette. For stedsdata gir INSPIRE-direktivet (Europaparlamentets og rådets direktiv 2007) føringer for standardisering. Geodataloven (2010) og Geodataforskriften (2012) oppfyller INSPIRE-direktivet. Stortingsmelding 23 (Kommunal- og moderniseringsdepartementet 2013) viser at også norske politikere har fokus på standardisering og åpenhet. Bruk av proprietære formater hindrer konkurranse og kan føre til merarbeid. Det er et politisk krav, både i EU og Norge, at det offentlige skal levere data på åpne, standardiserte formater. Det kan likevel leveres data på andre formater i tillegg.

Kartverket som geodatakoordinator er i ferd med å organisere en National Spatial Data Infrastructure (NSDI). I den forbindelse ble det opprettet en NSDI-formatgruppe i Kartverket, som har utarbeidet dette dokumentet.

1.1 Formål

Dette dokumentet har som formål å gi retningslinjer for bruk av dataformater for data som er direkte eller indirekte stedfestet. Retningslinjene er en del av en nasjonal infrastruktur for stedfestet informasjon.

1.2 Ansvar

Innholdet skal utformes i samarbeid med divisjonene, formidlingstjenesten og andre relevante enheter.

1.3 Revisjon

Dokumentet revideres ved behov og senest etter to år. Underlagsdokumentene (vedlegg) vil være under kontinuerlig revisjon.

Arbeidsgruppen i Kartverket for denne versjonen har bestått av:

Geir Vidar Mørner, Tinglysningsdivisjonen

Kristian Kihle, Formidlingstjenesten

Tone Kristiansen, Landdivisjonen

Arild Bråthen, Geodesidivisjonen

Per Arvid Jackobsen/Joan P Hansen, Sjødivisjonen

Thomas Hirsh, IT-avdelingen

Nils E Søvde, IT-avdelingen

Kent Jonsrud, IT-avdelingen

Gerd Mardal, IT-avdelingen

2 Forvaltningsprosesser for stedsrelaterte data

Et av hovedmålene til Kartverket er å levere ferske kvalitetsdata. Disse dataene kan være juridisk bindende. Slike autoritative data blir benyttet i offentlig saksbehandling, og det kreves god kvalitet. Andre hovedmål er at våre data skal være lett tilgjengelige og at vi skal være en aktiv pådriver for økt bruk av stedsdata. Det er en forutsetning for alle hovedmålene til Kartverket at dataene og forvaltningsprosessene er godt dokumentert.

Dette dokumentet gjelder også for den nasjonale stedsdatainfrastrukturen som skal bygges opp. Dette arbeidet vil neppe føre til mindre behov for dokumentasjon av stedsdata eller prosessene.

2.1 Standardisering av geodata

Arbeidet med dokumentasjon av stedsdata i Norge har vært basert på prinsipper om harmonisering og standardisering siden begynnelsen av 1980-tallet, da *Samordnet Opplegg for Stedfestet Informasjon (SOSI)* ble påbegynt. Det er mange grunner til å benytte slike prinsipper, for eksempel at kvaliteten på dokumentasjonen blir bedre, at dokumentasjonsarbeidet blir enklere eller at dokumentasjonen blir enklere å forstå.

Med harmonisering menes en samordning av begrepsbruk for alle geodata. Norge digitalt har mer enn 600 medlemmer, og mange av disse deler egne geodata. Arbeidet med harmonisering har neppe blitt mindre viktig siden 1980-tallet.

Samordnet Opplegg for Stedfestet Informasjon innførte standardisering av geodata. I starten var standardiseringen på nasjonalt nivå, men etter opprettelsen av *ISO/TC 211 Geographic information/Geomatics* i 1994 er det blitt skrevet internasjonale standarder for de fleste fagområdene innen geomatikk. Forvaltningsprosessene for stedsrelaterte data er i tråd med utprøvd og allment akseptert praksis.

De aller fleste internasjonale standarder på området er også vedtatt som norske standarder og utgis av Standard Norge. Disse standardene vil ha navn som starter med *NS-EN ISO*, der NS står for norsk standard, EN for europeisk standard (under den europeiske standardiseringsorganisasjonen CEN), ISO for internasjonal standard utgitt av ISO. I dette dokumentet refereres til den norske standarden der denne finnes, ellers til den 'originale' publisert fra ISO.

Formater benyttet for data relatert til sjø er definert av ekspertgrupper innen *the International Hydrographic Organisation (IHO)*. Kartverket ved sjødivisjonen er Norges medlem i IHO. Eksempler på slike data er navigasjonsdata og barymetriske data. Regelverk for data for navigasjonsformål er forvaltet av *the International Maritime Organisation (IMO)*.

Direktoratet for forvaltning og IKT (Difi) har publisert en veileder for publisering av åpne data (Difi-veileder 2013). På våre fagområder er beste praksis likevel langt bedre utviklet enn det Difi beskriver generelt.

2.1.1 Modellering av geodata

Geodata er modeller som beskriver verden, og dokumentasjon av geodata vil i de fleste tilfeller innebære modellering. De fleste standarder for geodata har krav eller føringer om slik modellering.

2.1.2 Koding av data - dataformater

Koding av data betyr å representere dataene på digital form, vanligvis i form av en datafil eller som en strøm av lesbare eller binære tegn fra en tjeneste. Vanligvis innebærer dette også en serialisering av dataene, dvs. en representasjon der data som hører sammen, er lagt ut etter hverandre og beregnet for lesing sekvensielt fra start til ende.

Koding av stedsdata skal være i samsvar med *NS-EN ISO 19118:2011 Geografisk informasjon – Koderegler*.

Dette betyr at det skal finnes dokumenterte regler for hvordan data i samsvar med sitt applikasjonsskjema skal kodes og legges ut i et datasett. De obligatoriske formatene tilfredsstiller dette kravet.

2.2 Produksjonsløypa for geodata

Valg av formater henger sammen med hvilke grensesnitt som brukes og hvor i produksjonsløypa for stedsrelaterte data vi er. Vi vil her forutsette at:

1. Dataene er dokumentert i en produktspesifikasjon med en UML-modell.
2. Produksjon av stedsdata inkluderer en validering mot UML-modeller.
3. Tjenester og leveranser skjer (om nødvendig ved transformasjon) fra et originalformat¹.

2.2.1 Dokumentasjon (produktspesifikasjon)

Digitale stedsdata representerer fysiske objekter eller fenomener i den virkelige verden (eller planlagte slike). Avhengig av tiltenkt bruksområde vil utvalget av objekter eller fenomener variere og likeledes de egenskaper som knyttes til disse og de relasjoner som er mellom dem. Dokumentasjon av data tar sikte på å gi en nødvendig og tilstrekkelig beskrivelse for at data skal forstås og brukes på riktig måte.

Kartverket skal dokumentere sine dataleveranser.

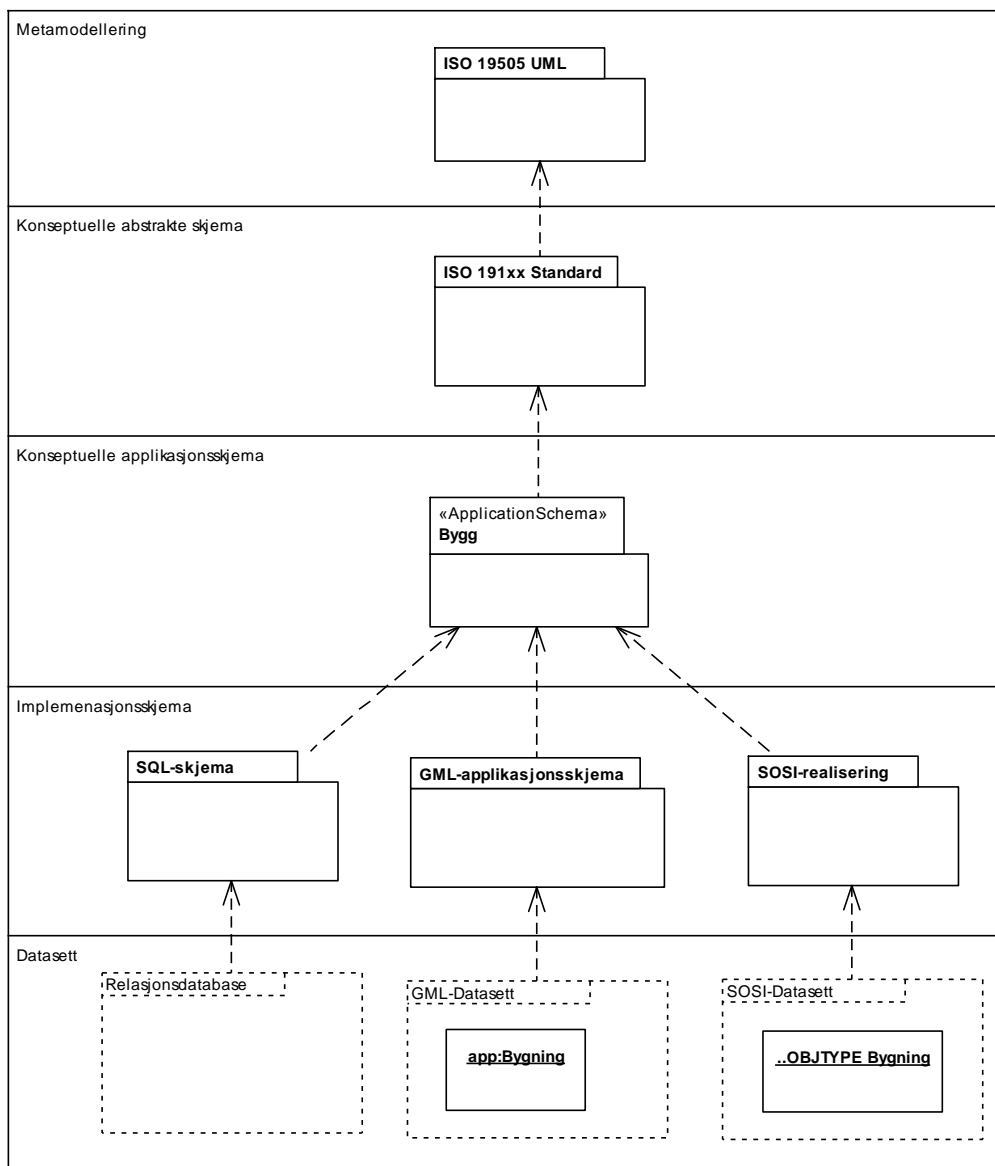
En produktspesifikasjon i samsvar med *NS-EN ISO 19131:2008 Geografisk informasjon - Produktspesifikasjoner* er rammen for den mest utførlige

¹ Begrepet *originalformat* blir i dette dokumentet brukt i forbindelse med transformasjoner og har en videre betydning enn filformater.

datadokumentasjon. Den skal representere den mest omfattende semantiske og strukturelle beskrivelsen av et datasett.

Alle våre offisielle datasett skal ha en produktspesifikasjon. Andre, mer *ad hoc*, datasett kan ha en forenklet dokumentasjon. Dokumentasjonen skal i alle tilfelle gi brukeren en tilstrekkelig forståelse slik at misbruk og feilbruk unngås.

Produktspesifikasjonen skal inneholde en UML-modell for datasettet i samsvar med *ISO/TS 19103:2015 Geographic information – Conceptual schema language* og *NS-EN ISO 19109:2006 Geografisk informasjon - Regler for applikasjonsskjema*. Denne modellen kalles *applikasjonsskjemaet* tilhørende datasettet (nivå 3 i figur 1).



Figur 1. Abstraksjonsnivåer for stedsdata.

2.2.2 Validering av stedsdata

Fra UML-modeller for produktspesifikasjoner kan det genereres SOSI-definisjonsfiler eller GML-skjema. Disse kan brukes for å kvalitetssikre at datasett følger modellene. Det er i dag kun disse formatene som kan brukes til validering, og de har litt ulik status:

1. GML-formatet er et ledende format som stadig utvikles. Datasett på GML-format kan valideres mot GML-skjema med XML-validatorer, men mer komplekse valideringer er ikke implementert ennå.
2. SOSI-formatet blir i liten grad utviklet, men SOSI-kontroll kan også gjøre en del komplekse valideringer.

2.2.3 Transformasjoner mellom formater

Når prosessen med datafangst og dokumentasjon er ferdig, foreligger et datasett i et originalformat². Hvis kvaliteten på dataene skal opprettholdes, er det viktig at det dokumenteres at transformasjoner til andre formater skjer uten tap. Slik dokumentasjon vil være unik for hvert datasett. Et format inneholder et sett med datatyper, og en UML-modell definerer objekttyper og egenskaper som bruker et subsett av disse. I tillegg kan det hende at et datasett ikke benytter alle objekttypene eller egenskapene i UML-modellen.

² Originalformat kan i denne sammenhengen være et filformat, en database eller andre lagringsformer.

3 Formater

Valg av formater gjøres ved å vurdere hvordan formatene oppfyller et sett med kriterier. Et sentralt kriterium er det politiske kravet om åpenhet. Det offentlige skal i prinsippet ikke formidle data på leverandørspesifikke formater. Dette kan virke konkurranseskadelig i retning av bestemte leverandører. Et leverandørspesifikt format bør kunne tillates som et tillegg til obligatoriske leveranseformater hvis brukere er avhengig av formatet og transformasjonen ikke er ressurskrevende.

Politikerenes krav om bruk av standardiserte formater er også viktig.

Valg av formater som krever transformasjoner, må det vurderes om transformasjonen er kostbar eller ressurskrevende. Det må også vurderes om transformasjonen fører til tap av data, og dette må dokumenteres.

I vedlegg finnes også andre kriterier som bør vurderes ved valg av formater for stedsdata. En totalvurdering av kriteriene vil fortelle om det vurderte formatet er egnet.

En nasjonal infrastruktur for stedsdata inneholder vektordata, rasterdata og geodetiske data.

3.1 Vektorformater

3.1.1 Geography Markup Language (GML)

GML er i dag *state-of-the-art* formatet for koding av stedsdata. Dette er vektorformatet med størst uttrykkskraft, dvs. som har størst funksjonalitet til å kode både enkle og kompliserte strukturer. GML vil i løpet av 2016 bli det primære og obligatoriske vektorformatet for nye datasett.

GML er et XML-format og det finnes derfor en rekke generelle XML-verktøy som brukerne kan benytte for å lese og skrive formatet. I tillegg finnes det også spesifikke GML-verktøy, og formatet støttes av all de store leverandørene. GML ble opprinnelig utviklet i OGC (OGC 07-036), men gjennomgikk en stor forandring i prosessen for å bli en internasjonal standard, *NS-EN ISO 19136:2009 Geografisk informasjon - Geografisk markeringsspråk*. Denne standarden beskriver versjon 3.2.1 av formatet, og det er også denne versjonen som hele INSPIRE-arbeidet med dataspesifikasjoner bygger på – dermed også geodataloven.

Det forutsettes at GML brukes i samsvar med et tilhørende GML skjema som er avledet fra applikasjonsskjemaet etter standardiserte regler. Dette inngår som en viktig komponent i produktspesifikasjonene.

Det er kommet en utvidelse av GML, *ISO 19136-2:2015 Geographic information – Geography Markup Language (GML 3.3) – Part 2: Extended schemas and encoding rules*, som gir koding og koderegler for enda flere datatyper (for eksempel *coverages*).

GML er uten tvil det mest presise og omfattende kodingspråket vi i dag har for geografisk informasjon.

3.1.2 SOSI-format – SOSI

SOSI – Samordnet Opplegg for Stedfestet Informasjon – er en norsk utvikling som har røtter i Miljøverndepartementet og Kartverket tilbake til midten av 80-tallet. Utviklet den gang det ikke fantes noen tilsvarende internasjonale standarder. SOSI er i dag en bransjestandard under Kartverkets myndighetsområde.

SOSI og dets utvekslingsformat har i mer enn 20 år vært det mest brukte og anerkjente formatet i Norge for å kode stedsdata. Det har også vært Kartverkets primære format for koding av alle våre datasett. Den profesjonelle bransjen har i alle år vært en støttespiller for SOSI og innen dette miljøet er det stor kompetanse og mange verktøy for bruk av SOSI.

Det skal benyttes SOSI versjon 4.0 og nyere. Riksarkivaren krever SOSI versjon 2.2 eller nyere for arkivering av kartdata (Arkivforskrift 1999).

Se <http://www.kartverket.no/Standarder/SOSI/>

3.1.3 S-57

For Electronic Navigational Chart (ENC) som benyttet i ECDIS-systemer er S-57 det obligatoriske formatet.

S-57 er videreutviklet og inkludert i standarden S-100. S-100 er tilpasset ISO 19100. I dette formatet vil produktspesifikasjonene for ENC være S-101, for batymetri S-102, undervannsnavigasjon S-103, osv. Det forventes at IHO og IMO vil vedta overgang til S-101 i 2018.

3.2 Rasterformater

Rasterdata er i utgangspunktet matriser med tre dimensjoner der den tredje dimensjonen tilsvarer ett eller flere bånd. Det er mange formater som kan lagre matriseverdier, og tidligere ble metadata lagret i egne filer.

3.2.1 GeoTIFF

GeoTIFF er basert på Tagged Image File Format (TIFF). Matriseverdiene lagres som TIFF, men i tillegg inneholder formatet stedsrelaterte metadata.

GeoTIFF støtter store datasett, tiles, 'overviews' og komprimering med og uten tap. TIFF og GeoTIFF framstår i dag som åpne og gratis formater, og bør være det obligatoriske formatet for rasterdata for nye datasett.

3.3 Formater for distribusjon av GNSS data (Global Navigation Satellite System)

Distribusjon av GNSS data er en viktig del av Kartverkets tjenester og forskningsaktiviteter. Når det gjelder GNSS data, så er det generelt åpne standarder og formater fra organisasjoner som IGS (International GNSS Service) og RTCM (Radio Technical Commission for Maritime Services) som bør benyttes til å distribuere data til eksterne kunder.

3.4 Andre tilleggsformater

Data kan leveres på andre tilleggsformater dersom et av de påkrevde formatene er tilgjengelig.

En forutsetning for bruk av tilleggsformater må være at det finnes en dokumentasjon på hva en finner i datasettet og hvordan dette er strukturert. I tillegg må det dokumenteres at transformasjonen ikke har ført til tap av informasjon.

Referanser

Europaparlamentets og rådets direktiv, 2007, *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*, <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002>.

Geodataloven, 2010, *Lov av 3. september 2010 nr. 56 om infrastruktur for geografisk informasjon*. <https://lovdata.no/dokument/NL/lov/2010-09-03-56>.

Geodataforskriften, 2012, *Forskrift av 8. august 2012 nr. 797 om infrastruktur for geografisk informasjon*. <https://lovdata.no/dokument/SF/forskrift/2012-08-08-797>.

IT-standardforskriften, 2013, *Forskrift av 5. april 2013 nr. 959 om IT-standarder i offentlig forvaltning*. <https://lovdata.no/dokument/SF/forskrift/2013-04-05-959>.

Referanse katalogen, 2014, *Liste over anbefalte og obligatoriske IT-standarder i offentlig sektor*, versjon 4.1, 28. mars 2014. <http://standard.difi.no/filearchive/referanse katalog-v4-1.pdf>.

Arkivforskrift, 1999, *Forskrift av 1. desember 1999 nr. 1566 om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver*. <https://lovdata.no/dokument/SF/forskrift/1999-12-01-1566>.

Difi-veileder, 2013, *Veileder i tilgjengeliggjøring av offentlige data*, versjon 2, januar 2013. <http://data.norge.no/veileder>.

Kommunal- og moderniseringsdepartementet, 2013, *Digital agenda for Norge – IKT for vekst og verdiskaping*, Meld. St. 23 (2012–2013). <https://www.regjeringen.no/no/dokumenter/meld-st-23-20122013/id718084/>.

Østensen Olaf, 2014, *Retningslinjer for dataformater ved formidling av stedsdata*, versjon 1.0, 2. november 2014, Kartverket.

Vedlegg

Det er eget excel-regneark.