

Metadata – nøkkelen til geografisk informasjon

Per Ryghaug

Per Ryghaug: Metadata – the key to geospatial data

KART OG PLAN, Vol 66, pp. 232–237. P.O.Box 5003, N-1432 Ås, ISSN 0047-3278

Metadata is a key component in the life cycle of geospatial data and services. The ISO 19115:2003 Metadata standard is an international standard defining the schema required for describing geographic information and services to ensure global interoperability when information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference and distribution of digital geographic data are provided. Why is metadata so important, and what are the key issues when we want to develop them and make catalogue service implementations?

Key words: ISO standard, Metadata, Geospatial data, Infrastructure

Per Ryghaug, Chief Engineer and Team Leader, Geodatamanagement, Geological Survey of Norway. N7491 Trondheim, Norway. E-mail:Per.Ryghaug@ngu.no

Innledning

Denne artikkelen har ikke som mål å forklare selve innholdet i metadatastandarden(e) eller å veilede i bruken av dem. Hensikten har mer vært å forklare betydningen av standardiserte metadata, og gi noen tips for å bidra til hurtigere implementeringer av standardene.

Hvorfor er metadata viktig?

Geografiske data vil alltid være et forsøk på å modellere og beskrive deler av den virkelige verden. Enhver slik beskrivelse av virkeligheten vil imidlertid måtte være én av mange mulige avbildninger, og aldri helt fullstendig. Noe vil være forenklet og andre deler ignorert. Geografiske data er oftest produsert av en leverandør, men blir benyttet av mange ulike brukere. Det er derfor en nødvendighet å beskrive denne avbildningen på en så fullstendig måte som mulig for blant annet å unngå at data brukes til formål de ikke var ment for. Metadata (data om dataene) er et forsøk på en slik beskrivelse.

Digital geografisk informasjon er de siste årene blitt etterspurt og tatt i bruk i langt større grad enn tidligere, og stadig flere brukergrupper ser geografisk informasjon som en nødvendig del av sin beslutningsstøtte. Etter hvert som mengden av og variasjonen i

den digitale geografiske informasjonen øker, vokser også behovet for samtidig å få tilgang til dataenes metadata. Det er gjennom metadata brukerne vil kunne søke etter data, skaffe seg overblikk over datainnholdet, vurdere om kvaliteten er egnet til ønsket bruk, bli informert om hvor man kan få tak i dem og til hvilken pris m.m.

Søkemulighet basert på søk i fri tekst finnes i de fleste moderne applikasjoner og webløsninger. Problemet er bare at slike søk ofte gir vanvittig mange treff, og det vil derfor ta uakseptabel lang tid å finne frem til de geografiske data man egentlig søker. Om man derimot kan utføre mer direkte søk mot spesielle metadatakataloger, hvor dataene er beskrevet og organisert etter en og samme struktur, vil man raskere finne frem til det man søker, og samtidig få tilgang til en strukturert beskrivelse av dataene. Selv om enkelte organisasjoner eller fagmiljøer alt har opprettet interne «standarder» og rutiner for slik informasjonsutveksling, er det lite som vil fungere mellom ulike organisasjoner og mot eksterne brukere dersom struktur og innhold av metadata ikke følger en felles standard. Dette har vært noe av bakgrunnen for å få utviklet en felles internasjonal metadatastandard for geografisk informasjon.

Utviklingen av en ISO metadatastandard for geografisk informasjon

Delvis basert på eldre metadatainitiativ i USA [1] og i Europa [2] ble arbeidet med metadata i ISO/TC 211 startet i 1995. Det var deltakelse fra en rekke land, deriblant Norge. ISO publiserte sin metadatastandard [3] 8 år senere, noe som illustrerer at det tar tid å utvikle globalt aksepterte standarder innenfor et felt der alle ser ut til å ha egne meninger og behov. Standarden ble i 2005 akseptert som en Europeisk Standard (EN), og ble dermed automatisk en Norsk Standard (NS) [4]. I 2006 ble ISO-standarden endret på noen punkter, noe som er beskrevet i et eget ISO-dokument [5]. Bakgrunnen for dette var at mindre feil måtte rettes opp, og et behov for å tilfredsstillere nye krav som var fremkommet under arbeidet med en ny teknisk spesifisering som inneholder en XML skjemaimplementasjon av ISO 19115 [6]. Et nytt tillegg til standarden er også under utarbeidelse for å utvide beskrivelsen av rasterdata [7]. Man må regne med at standarden også i fremtiden vil gjennomgå justeringer og utvidelser etter hvert som nye behov dukker opp.

Metadatastandarden er omfattende og svært modellorientert. Den definerer sammenhengen mellom 404 objektklasser, roller og egenskaper for metadata som sammen definerer de skjema som kreves for å beskrive den geografiske informasjonen og tjenestene (informasjon om identifikasjon, utbredelse, kvalitet, referansesystem, ansvarlig organisasjon, distribusjon med mer). Noe benyttes til katalogisering av datasett, andre elementer brukes for å gi en fullstendig beskrivelse av datasett, datasettsierier samt individuelle geografiske objekter. Standarden er på 140 sider og fås kjøpt gjennom Pronorms nettbuikk på www.standard.no for kr. 1258.

Standarden definerer:

- påkrevde og betingede metadataseksjoner, -roller, -klasser og -egenskaper
- det minimum av metadata som er nødvendig for å opprette en metadatateneste (obligatoriske kjernemetadata)
- valgfrie metadata, som muliggjør en utvidet, standardisert beskrivelse av geografisk informasjon.

- en metode for å lage egne metadataklasser/egenskaper og profiler

I forbindelse med en revitalisering av CEN/TC 287 ble det opprettet en arbeidsgruppe (WG5), hvor en undergruppe fikk i oppgave å utarbeide en profil av EN ISO 19115 [8]. Denne skulle tilfredsstillere de krav til metadata som er fremkommet i et forslag til EU-direktiv (INSPIRE) som fortsatt er inne i en godkjennelsesprosess. Norge har vært sterkt med også i dette arbeidet. Det forventes en revisjon av dette dokumentet når alle krav til implementering av metadata i INSPIRE er utarbeidet.

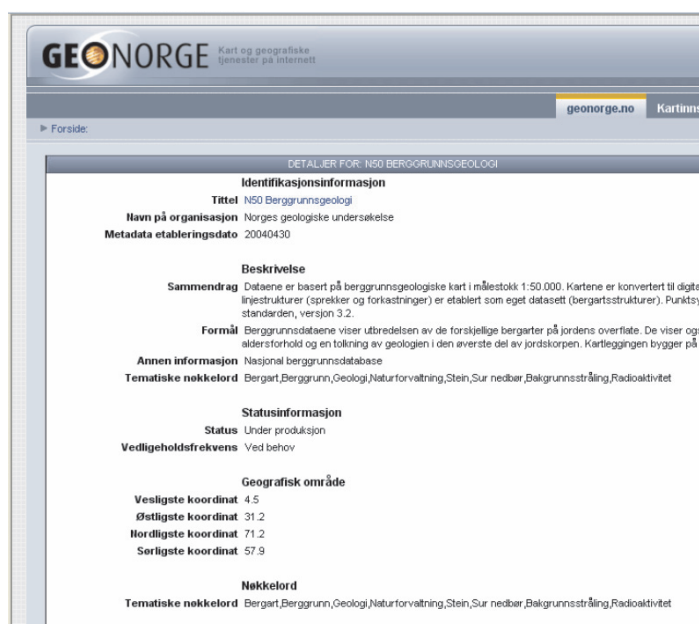
Norsk versjon av ISO-standard

Det fremkom tidlig et behov for å få utarbeidet en norskspråklig versjon av ISO 19115-standard. Dette var nødvendig for å få en bedre forståelse av innholdet, en felles norskspråklig metadatastruktur på ulike datakatalogtjenester og lette sammenkoblingen av disse i et nasjonalt nettverk av tjenester. Den foreliggende versjonen (versjon 1.2) er ikke en fullstendig oversettelse. Engelske betegnelser er beholdt på flere områder for å forankre denne bedre til den engelskspråklige originalen. Den inneholder samtlige UML-modeller og alle tabellariske oversikter. Dette var samtidig et forsøk på å samordne tidligere norske metadataarbeider og fremskaffe en felles plattform å bygge slike tjenester på, tjenester som samtidig kan spille sammen med internasjonale katalogtjenester.

ISO-standard for metadata er en del av en hel familie av standarder tilknyttet geografisk informasjon (ISO 19100-serien), hvor flere av dem henger nært sammen. Den norske versjonen inneholder også klasser som er definert i de andre ISO-standardene og som ikke er oversatt til norsk. Man må derfor slå opp i de internasjonale standardene direkte dersom det er nødvendig å få begreper beskrevet i større detalj enn det som står skrevet i den norske metadataversjonen. Dette gjelder for eksempel en ISO-standard som beskriver tidsaspekter [9], stedfesting med koordinater [10], prinsipper for spesifisering av kvalitet [11] og tjenester [12].

Den norskspråklige versjonen kan lastes ned gratis fra Statens kartverk sitt nettsted <http://www.statkart.no/standard/sosi/html/sosi.htm> under SOSI- DEL1_5 . Det er også utarbeidet en profil (subset) av den norske versjonen tilpasset tematiske vektordata. Denne reduserer antallet metadataelemen-

ter fra opprinnelig 271 til 104 elementer. Som et vedlegg til denne er det utarbeidet en tabell med «folkelige» navn på klasser og elementer til bruk i utformingen av skjermbilder i metadataeditorer og formularer (style sheets).



Figur 1. Eksempel på visning av metadata i metadatakatalogen geoNorge.no etter at det er gjort et søk.

Selv om standarden først og fremst er beregnet på digital geografisk informasjon, vil prinsippene og deler av innholdet også kunne anvendes for andre former for data og tjenester med elementer av stedfesting i seg. Dette kan være for eksempel analoge kart, bilder, diagrammer og tekstdokumenter.

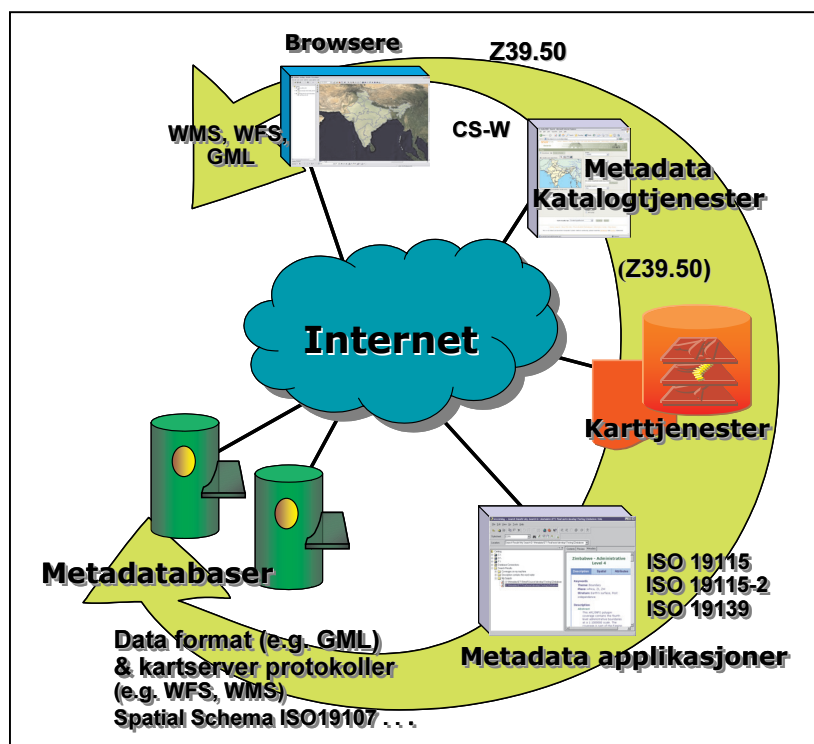
Metadataimplementering

Metadata bør følge med i hele livssyklusen til dataene den beskriver, fra førstegangs dataproduksjon, publisering, tilgang/bruk og eventuelt inn i en ny dataproduksjon (se figur 2.). Alle implementeringer som ønsker å være i overensstemmelse med den norske ISO-profilen, må sørge for å ha metadata spesifisert i henhold til denne standarden. En konform

implementering av en katalogtjeneste skal inneholde minimum det som i standarden er definert som metadatakjernen.

Etableringen av metadata vil være ressurskrevende for dataproduzenten, men kostnadene kan reduseres betydelig dersom metadataene etableres mens datainnholdet produseres, og mye av metadataene bør kunne bli generert automatisk fra dataene selv. Det er en forutsetning at en eller flere personer i organisasjonen har satt seg godt inn i standarden(e). Videre har det vist seg å være viktig å utpeke en person som metadataansvarlig i organisasjonen for å sikre at metadata blir produsert, kvalitetssikret, vedlikeholdt, og at løsninger blir utviklet.

Metadata bør ikke være noe man etablerer i en enkeltstående applikasjon isolert fra selve

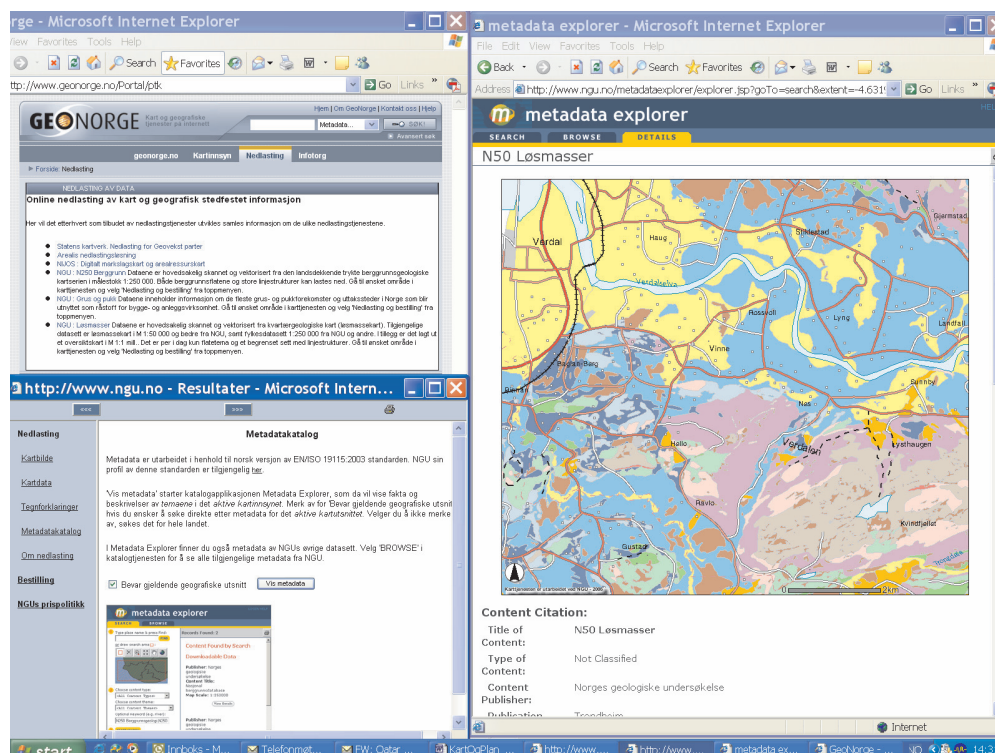


Figur 2. Metadata bør være en nøkkelkomponent i all dataproduksjon og datatjeneste (modifisert etter David Danko, ESRI).

geodataene og forvaltningssystemene. En er derfor avhengig av at den GIT-leverandør man har valgt har skjønnet betydningen av metadata og har integrert en metadataeditor, som er konform med NS-EN ISO 19115, i sin programvare. Metadataene bør være tilgjengelig i forbindelse med funksjoner for nedlasting og bestilling, kun være et tastetrykk unna når man inspiserer selve kartbildene på skjermen (figur 3), og kunne genereres og distribueres sammen med dataene i et XML-skjema. Dersom leverandører på det norske markedet tilpasser slik funksjonalitet basert på den norske versjonen av standarden og bruker de «folkelige» navnene, vil brukerne kunne møte og kjenne seg igjen i et brukergrensesnitt med betegnelser som er felles for det norske geodatamiljøet. ESRI har som eksempel introdusert en slik løsning med sin metadataeditor i Arc-Catalog og i Metadata Explorer. Selv om det fortsatt er en del feil og mangler er det fullt mulig å publisere metadata. Forhåpentligvis

vil XML-skjemaet [6] for metadata gjøre slik applikasjonsutvikling lettere slik at flere leverandører kommer på banen, og dermed sikre at metadatabaser etableres i henhold til EN ISO 19115-standard. Selv om det med tiden helt sikkert vil komme nye måter å etablere katalogtjenester på, vil man trolig ikke behøve å endre selve metadatainnholdet.

Behovet for en samordning av arbeidet rundt norske metadatakataloger og -aktiviteter er etter hvert blitt tydeligere. I forbindelse med etableringen av Norge Digitalt (ND) er kravene til etablering av metadata ytterligere forsterket. En målsetting har vært at geoportalen www.geoNorge.no, med sin katalogtjeneste, skal bli en nøytral og felles inngang til det meste av offentlig geografisk informasjon i Norge. Dette forutsetter at alle involverte geodataprodusenter og tjenesteytere gjør sine metadata og karttjenester tilgjengelig i katalogtjenesten selv om de måtte ha egne publiseringssteder i tillegg.



Figur 3. Metadata tilgjengeliggjort fra geoNorge.no, hvor en nedlastingsfunksjon bringer brukeren inn i datatilbyderens metadatatilgjengelighet. I dette tilfellet løsmassedata på ngu.no.

Den geografiske infrastrukturen for Europa (ESDI), basert på det forestående EU-direktiv kalt INSPIRE, stiller tilsvarende krav til metadata. Implementeringsregler for metadata i INSPIRE, som er under utarbeiding, baserer seg på ISO-standardene og består av metadatakjernen pluss noen tilleggselementer for å oppfylle kravene til databeskrivelse i direktivet. Dette understreker betydningen denne standarden har i forbindelse med den pågående oppbyggingen av nasjonale og internasjonale geografiske infrastrukturer.

Sluttbemerkninger

Mange har lenge snakket og skrevet om hvor viktig metadata er. Selv om det nå er gått 3 år siden ISO offentliggjorde første versjon av metadatatilgjengelighetsstandardene, er det fortsatt alt for få geodataproducenter som har utarbeidet/publisert metadata for sine produkter. Mangelen på applikasjoner for etablering og

publisering av metadata hos flere av de mest brukte GIS-leverandørene må ta noe av skylden, mens systemer som har hatt slik funksjonalitet en tid har hatt mange feil og begrensninger. Systemene følger ikke alltid ISO-standardene helt, og leverandørene har vært trege med å rette på identifiserte feil og mangler. Katalogtjenesten på geoNorge.no har heller ikke levd helt opp til forventningene når det gjelder brukergrensesnitt, funksjonalitet og stabilitet, noe som har fått dataprodusentene til å vente med å involvere seg. Det er nå ønskelig med et sterkere engasjement knyttet til denne portalen i tiden fremover. Portalen bør bli mer inkluderende bl.a. ved at flere av samarbeidspartene i ND er med i utformingen av brukergrensesnittet. Den bør fremheve metadatatilgjengelighetsstandardene sterkere, forbedre søkefunksjonaliteten og sørge for tilgang til alle publiserte karttjenester med offentlig informasjon. Først da kan man forvente at metadata blir prioritert,

og at portalen får en sentral plass i det norske geodatamiljøet.

Refererte dokumenter

- [1] Base Standard for Digital Geospatial Metadata – FGDC-STD-001-1998
- [2] CEN prENV 12657 Geographic information – Data description – Metadata 1998
- [3] ISO 19115:2003 Geographic information – Metadata
- [4] NS-EN ISO 19115 Geografisk informasjon – Metadata
- [5] ISO 19115:2003/Cor. 1:2006 (Technical Corrigendum 1)
- [6] ISO/TS 19139 Geographic information – Metadata XML schema implementation
- [7] ISO 19115-part 2 – Extensions for imagery and gridded data
- [8] prEN N1000 Geographic information – European core metadata for discovery
- [9] NS-EN ISO 19108 Geografisk informasjon – Modell for å beskrive tidsaspekter
- [10] NS-EN ISO 19111 Geografisk informasjon – Modell for stedfesting med koordinater
- [11] NS-EN ISO 19114 Geografisk informasjon – Prosedyrer for kvalitetsvurdering
- [12] NS-EN ISO 19119 Geografisk informasjon – Tjenester