

Distribuerade geografiska metadata

Problem och möjligheter



Rapport utarbetad av
Anders Östman, Luleå Tekniska Universitet
1999

ULI rapport 1999:2

Distribuerade geografiska metadata



Problem och möjligheter

Rapporten är utarbetad av
Anders Östman, Luleå Tekniska Universitet, 1999

ISSN 1101-8895
ISRN ULI-RS-99/2—SE

Titel: Distribuerade geografiska metadata
Utgivare: Utvecklingsrådet för landskapsinformation (ULI)
© Utvecklingsrådet för landskapsinformation
Utgiven: Juni 1999
Upplaga: 350 exemplar
Tryck: Sandvikens tryckeri AB

Rapporten kan beställas från ULIs kansli.
Pris exkl. moms: 100:-/st för ULIs medlemmar, 200:-/st för övriga.
Adress till kansliet:
ULI, 801 82 Gävle.
Tel: 026-61 10 50. Fax: 026-61 32 77. E-post: uli@uli.se
Rapporten finns också att hämta på ULIs hemsida under adress:
www.uli.se

INNEHÅLLSFÖRTECKNING

	<u>Sida</u>
Förord	8
1. Sammanfattning	9
2. Inledning	10
3. Elektronisk handel	12
4. Försökets genomförande	14
5. Överföring av förfrågningar och metadata	15
6. Upprättande av katalogtjänst	17
7. Synpunkter från slutanvändare	18
8. Framtida arbete	20
9. Slutsatser	22
10. Finansiering	22
11. Referenser	23

FÖRORD

ULI, Utvecklingsrådet för Landskapsinformation, verkar för ett effektivare användande av geografisk information i samhället.

ULI har de senaste åren verkat för etablerandet av en fungerande nationell infrastruktur för geografiska metadata (data om data).

På kort sikt stödjer ULI den nationella databaskatalogen som förvaltas av Lantmäteriet som förutsätter att producenter skickar uppgifter till Lantmäteriet och kan nås via ULIs hemsida (www.uli.se) Detta förfarande kommer att effektiviseras genom att en ny version av den nationella databaskatalogen, MEGI, sjösätts sommaren 1999.

Ett problem med etablerandet av en fungerande metadatastruktur är de svaga drivkrafterna för dataproducenterna att generera aktuella metadata. Det vore förmånligt om metadata mer eller mindre automatiskt genereras ur dataproducenternas egna dataadministrationssystem och om de kan nås hos dataproducenterna.

En struktur med distribuerade geografiska metadata förutsätter att användarna/köparna av data hittar relevanta metadata. Detta skulle kunna underlättas genom någon form av katalogtjänst.

Föreliggande studie av professor Anders Östman, Luleå tekniska universitet, har upprättats på uppdrag av ULI för att belysa förutsättningarna för att söka, ajourhålla och distribuera katalogiserade geografiska metadata.

Gävle i maj 1999

Lars Hansen, Kanslichef, ULI

1. SAMMANFATTNING

Metadata är en central resurs, såväl vid utvecklingen av infrastrukturer för information som vid intern administration av data. Vid utveckling av infrastrukturer behövs metadata bl. a vid upprättande av mäklartjänster där man kan söka efter data. Vid administration av interna data behövs metadata bl. a för att hålla rätt på versioner och varianter av dokument, kartfiler etc.

De flesta försök med att upprätta mäklartjänster har haft en centralistisk grundidé. Det innebär att metadata skickas in till en central metadatabas, där metadata lagras, uppdateras och administreras. Den centrala metadatabasen tillhandahåller också ett generellt gränssnitt mot olika slutanvändare. Av flera olika orsaker har dessa centralistiska angreppssätt sällan varit lyckade. Dels saknas ofta ekonomiska incitament för dataproducenter att leverera sina metadata till en central instans. Hanteringen av exempelvis uppdateringar är också förhållandevis omständlig. Prestigekonflikter i att en enskild organisation ges ett övergripande ansvar har också noterats vid ett flertal tillfällen.

Inom elektronisk handel växer en ny typ av "organisationer" fram, nämligen de "digitala biblioteken". Dessa är egentligen nätverk av intressenter som via Internet samverkar för att kompletteras varandras produkter och nå en större marknad. Denna typ av handel går således mot en allt mindre central kontroll och ett ökande inslag av distribuerat ansvar och nätverksbyggande.

The Open GIS Consortium (OGC) är en organisation som domineras av systemleverantörer. Man arbetar här med att utforma specifikationer för morgondagens GIS-programvaror. Den modell som här har presenterats för hantering av metadata, baseras på de nya strukturerna för elektronisk handel och distribuerade metadata. Tanken här är att varje dataleverantör ansvarar för sina egna produkter (data och metadata). Genom s k katalogtjänster, sammanställs metadata och presenteras för olika kundgrupper. Olika katalogtjänster kan då inrikta sig mot olika marknadssegment.

För att studera problem och möjligheter som hantering av distribuerade metadata medför, har en förstudie genomförts. En katalogtjänst upprättades vid Miljödatacentrum i Kiruna. Förfrågningar om metadata skickades sedan ut till fyra dataleverantörer, nämligen Miljöförvaltningen i Stockholms Stad, Sveriges Geologiska Undersökning, Lantmäteriverket samt Länsstyrelsen i Norrbottens län. Förfrågningarna bearbetades hos respektive leverantör och metadata skickades tillbaka till Miljödatacentrum. Dessa metadata sammanställdes sedan i katalogtjänsten och gjordes åtkomliga via systemet EnviCat.

Denna studie har i huvudsak varit producentorienterad. För att fånga upp synpunkter från konsumenter, har en mindre enkätundersökning genomförts. Syftet med denna enkätundersökning var att ange riktlinjer för eventuellt fortsatt arbete inom området.

Bland de resultat som erhöles, kan följande nämnas:

- Kommunikationen mellan katalogtjänsten och dataleverantörerna baserade på epost-teknik. Denna teknik fungerade tillfredsställande. Dock pågår idag en utveckling av mer avancerade tjänster som erbjuder större flexibilitet för organisationer som vill etablera denna typ av nätverk (digitala bibliotek).
- Överföringen av metadata baserades på ISO's standardförslag, conformance level 1. Detta förslag är dock alltför "kartorienterat", speciellt för icke-traditionella producenter av geografisk information.
- Kvaliteten hos data är ibland otillräcklig och dokumentation om datakvalitet saknas oftast helt och hållet.
- Det saknas idag färdiga system för administration av metadata. Det föreligger således här ett behov av produktutveckling och / eller systemutveckling. Produktutvecklingen kan exempelvis avse innehåll och struktur på metadata, användargränssnitt och koppling till produktionssystem. Systemutvecklingen kan exempelvis avse frågor rörande ajourhållning av metadata, versionshantering etc.
- Leverans av geografisk information behöver effektiviseras. Frågor som behöver lösas är exempelvis leveranstider, priser, nyttjanderätter och beställningsmekanismer.

2. INLEDNING

Med geografiska metadata (hädanefter kallat metadata) menas här data som på ett översiktligt sätt beskriver geografiska data. Sådana beskrivningar skall bl. a innehålla uppgifter som hur datamängden identifieras, dess datakvalitet, hur data representeras (vektor eller raster etc), koordinatsystem samt hur man får tillgång till datamängden (OGC, 1998). En metadatabas är då följdaktligen en databas som innehåller geografiska metadata.

Arbetet med att upprätta metadatabaser tog fart i april 1994. Vid denna tidpunkt gav president Clinton order om att de federala myndigheterna i USA skulle tillhandahålla sina geografiska data på Internet (Executive Office of the President, 1994). Man påbörjade då arbetet med att bygga en nationell infrastruktur för geografisk information (NSDI) i USA.

En NSDI (National Spatial Data Infrastructure) består av flera olika komponenter. I allmänhet ingår åtminstone följande komponenter, nämligen:

- grundläggande data, d v s data som anses vara av allmänt intresse och av stor betydelse. Exempel på detta kan vara data om landets topografi, fastigheter, befolkning, företag etc.
- standarder för överföring av data, så att informationsutbytet mellan sändare och mottagare blir meningsfullt.
- mäklartjänster, med vars hjälp man kan hitta lämplig information
- regler, som anger vilka rättigheter och skyldigheter de olika aktörerna har.

Metadata används bl. a i mäklartjänster och utgör således en viktig komponent i en NSDI. För att metadata skall vara användbara, krävs också någon form av standardisering. Vidare har också de federala myndigheterna i USA ett krav på sig att tillhandahålla metadata på Internet. Frågor rörande metadata berör således samtliga grundläggande komponenter i en infrastruktur för geografisk information.

Frågor rörande metadata, sökmotorer och datatillgänglighet står idag högt på dagordningen, såväl politiskt som inom forskning och utveckling. Det anordnas idag speciella konferenser om infrastruktur för geografiska data och om interoperabilitet. Inom EU har man övervägt att tillsätta en s k "High level working party" som skall arbeta med dessa frågeställningar. För att fånga upp synpunkter, har man gjort en problembeskrivning i ett s.k "green paper" (European Commission, 1999) och välkomnar här synpunkter och debattinlägg.

De mest vanliga orsakerna till att upprätta metadatabaser är enligt OGC (1998) och CEO (1999)

- att katalogisera sin egen information och på så sätt säkerställa de investeringar som är gjorda i geodata
- att tillhandahålla information om den egna organisationens dataprodukter till datakataloger, datamäklare och sökmotorer.
- att tillhandahålla den information som behövs för att bearbeta de externa data som erhålls via transferering etc.

Det finns ett flertal nationella initiativ till att skapa NSDI (Masser 1999). I Europa är kanske initiativen i Holland, England och Portugal de mest kända. Detta arbete går nu vidare och flera länder är nu i färd med att skapa den andra generationens NSDI. Diskussioner pågår också för att skapa en europeisk eller global infrastruktur (EGII eller GSDI)

I Sverige finns också flera initiativ inom området, såväl nationella som regionala / lokala. Lantmäteriverket har exempelvis arbetat med att utforma en nationell databaskatalog. Som resultat av detta arbete kommer det att finnas en applikation utvecklad i Access, för hantering av metadata. Dessutom kommer det att upprättas en metadatabas som innehåller ett urval metadata från ett urval av dataleverantörer.

Andra initiativ i Sverige är exempelvis BD-GIT-projektet i Norrbotten, där en dataväxel upprättades och utgjorde grunden för en regional samordning i länet. Inom projektet Pilot-GIS har provats att tillhandahålla data över Internet. Ett resultat av detta arbete är att behovet av metadata i form av en databaskatalog uttrycks tydligt. Inom Vägverket har arbete med att bygga upp metadatabaser påbörjats. Hos Miljödatacentrum i Kiruna (MDC) har sett förhållandevis intensivt utvecklingsarbete skett, bl.a med mål att tillhandahålla miljödata och annan geografisk information inom olika internationella projekt.

Internationella standarder för metadata för geografisk information har utarbetas av det europeiska standardiseringsorganet CEN samt det internationella organet ISO. Det finns också nationella standarder utvecklade, bl.a FGDC-standarderna i USA. Diskussioner har förts och fortsätter att föras rörande dels vilka uppgifter som metadata skall innehålla, vilka uppgifter skall vara obligatoriska samt huruvida fri text eller definierade kodlistor skall användas. De olika standarderna och standardförslagen skiljer sig härvidlag åt.

The Open GIS Consortium (OGC) är en sammanslutning bestående av ledande programvaruleverantörer, större användargrupper och andra intressenter. OGC arbetar för att skapa öppna system, där data kan fritt

utbytas mellan systemen och där programvaror från olika leverantörer kan samverka. Man har inte som primärt mål att skapa standarder, men eftersom de ledande leverantörerna av programvaror medverkar, kommer de specifikationer som utarbetas troligen att bli industri-standarder inom området.

OGC utarbetar således specifikationer (Open Geodata Interoperability Specifications, OGIS). Dessa specifikationer beskriver bl.a de funktioner som standardiserade komponenter skall kunna utföra. Inom OGIS arbetar två arbetsgrupper med metadata-relaterade frågor, nämligen topic 11 (metadata) och topic 13 (catalog services). Vad avser själva metadata (termer, fältspecifikationer etc) hänvisar OGC till arbetet inom ISO. Uppenbarligen avser man att i kommande programprodukter implementera ISO-standarderna för metadata när den blir färdig.

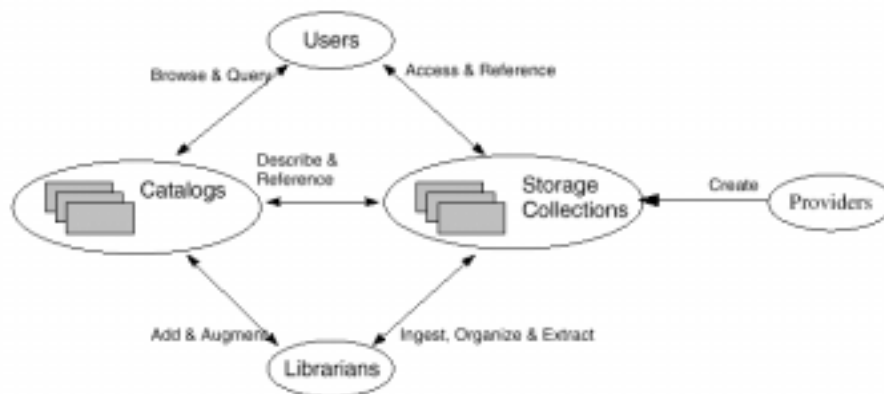
Men att enbart ta fram en metadatastandard är ej tillräckligt för att nå målet med samverkande programvaror och databaser. Trots att FGDC-standarderna har varit i bruk i enbart ett fåtal år, kan man redan nu notera sökproblem. Det finns mycket metadata på Internet idag. Dagens sökmotorer klara inte av att på ett effektivt sätt tillhandahålla metadata som svar på olika frågor. Sökmotorer som exempelvis AltaVista ger ett antal länkar som svar, varefter användaren själv får klicka sig fram och tillbaka mellan länkarna.

3. ELEKTRONISK HANDEL

IT-utvecklingen innebär att nya former av affärsverksamhet kan utvecklas. Det tydligaste exemplet är kanske elektronisk handel, där köpare beställer varor över Internet, ibland också kombinerat med betalningsfunktioner. Man har sedan länge kunnat konstatera att "den teknik, kunskap och övriga resurser som krävs för att tillfredsställa en ökande kundbas, håller på att växa och bli alltför ohanterlig för en enskild organisation" (Lopez och Larsgaard, 1998). Man har därför sedan början på 1980-talet arbetat med nya organisationsformer (Barett och Konsynski, 1982). Deras koncept rörande nätverksorganisationer har rönt stor uppmärksamhet inom den vetenskapliga litteraturen.

På senare år har konceptet "digitala bibliotek (digital libraries) introducerats. Ett digitalt bibliotek är ett exempel på en nätverksorganisation och kan definieras som "en koordinerad mängd heterogena aktörer, som samverkar genom en elektronisk och kommersiell infrastruktur, för att utveckla, producera, förbättra, distribuera och marknadsföra elektroniska produkter, tjänster och kunskap" (Lopez och Larsgaard, 1998). De digitala biblioteken är således exempel på de framtida "företag", där informationstjänster säljs på det elektroniska torget.

OGC har noterat de problem som dagens metadatabaser har att brottas med. Man har också uppmärksammat de möjligheter som den framtida handelsplatsen, t ex de digitala biblioteken, medför. Man föreslår därför att man inrättar speciella katalogtjänster, som var och en är utvecklade speciellt för egna målgrupper. En målgrupp kan exempelvis vara kommunala förvaltningar inom en kommun, medan en annan kan vara skolektorn. En principskiss visas i figur 1.



Figur 1. Princip för metadatasystem enligt OGC, 1998.

Principerna för metadatasystemet är i stort sett följande (se figur 1). En dataproducent (provider) producerar data och metadata som lagras i databaser (storage collections). En producent använder ofta metadata i sin produktionsuppföljning, t. ex. verifikation av enskilda bearbetningssteg etc. Det bör således ej vara förenat med något större merarbete för producenten att tillhandahålla produktionsorienterade metadata i en "storage collection".

De produktionsinriktade metadata är enbart till viss del av intresse för användargrupperna. Katalogtjänsten innebär att en bibliotekarie (librarian) hämtar metadata från producenterna, extraherar för honom/henne intressant information, tillför ev. ytterligare information och lagrar dessa nya metadata i en katalog. Strukturen hos denna katalog kan vara avsedd för en viss grupp användare, t. ex. kommunala förvaltningar, gymnasieskolor, journalister etc. eller vara avsedda för allmän dataförsäljning. Med jämna mellanrum uppdaterar sedan bibliotekarien sin katalog genom att avsöka dataproducenternas metadatabaser.

En användare (user) frågar slutligen den "kundinriktade" katalogen efter information. Användaren får referens till dataleverantören och hämtar/köper information där.

Enligt befintliga tidsplaner beräknas denna del av OGIS vara färdig under 1999. Det innebär att om den tidsplanen håller, så kan vi förvänta oss de första produkterna i början av år 2000.

Även om de svenska insatserna inom metadataområdet hittills varit förhållandevis blygsamma, finns det ett stort intresse för frågor rörande metadata och söksystem. Även små begränsade försök ger ofta viktiga erfarenheter och nationella skillnader blir synliga. Exempelvis kan man mot bakgrund av den av Lantmäteriet genomförda studien konstatera att en centraliserad lösning med en enda stor metadatabas förefaller vara svårt att förverkliga i Sverige. I Portugal har man å andra sidan arbetat med andra förutsättningar och har här mycket positiv erfarenhet av en centraliserad lösning.

Om nu en centraliserad lösning fungerar dåligt i Sverige, hur kommer då en distribuerad lösning enligt OGC's koncept att fungera? Är drivkrafterna hos dataleverantörerna så starka att de tillhandahåller sina produktionsorienterade metadata? Vilka drivkrafter finns för att utveckla katalogtjänster för olika avnämarkategorier? Vilka strategiska åtgärder bör man och kan man vidtaga på ett tidigt skede, så att man står väl rustad när de nya programvarorna kommer i början av år 2000? Vad gör man med existerande metadatabaser? Passar de termer som här används, är lagringsstrukturen och metoderna för åtkomst lämpliga? Skall man migrera dessa till ny struktur och i så fall när och hur? Hur minimerar man effekterna av driftsstörningar? Hur sköter man om ajourhållningen av metadata?

Listan på angelägna frågor kan göras lång. Denna förstudie har ej ambitionen att besvara alla dessa frågor. Dessa är istället planerade att studeras i en eller flera efterföljande större studier. Syftet med denna förstudie är istället att skaffa en överblick över viktiga frågor och andra frågor som hittills har varit förbisedda. Tanken är att vi, med denna överblick som grund, skall utforma en handlingsplan för fortsättningen. Syftet med denna studie är således att identifiera viktiga frågeställningar, kunna bedöma dess komplexitet samt ge förslag på hur vi ämnar arbeta vidare med dessa frågeställningar.

4. FÖRSÖKETS GENOMFÖRANDE

Syftet med detta projekt är att ge fördjupad kunskap om hur metadata för geografisk information kan hanteras i distribuerade system. Målet är att

- utveckla ett demonstrationssystem där metadata från tre olika leverantörer regelbundet avsökts och katalogiseras i en katalogtjänst.
- identifiera viktiga frågeställningar för fortsatta studier samt att ge förslag på hur dessa frågor bör studeras.

Försöket baseras på OGC's modell för katalogtjänster. Med denna modell som grund, har arbetet indelats i tre olika delar, som beskrivs mer ingående i kommande avsnitt. Dessa tre olika delar är

1. Överföring av förfrågningar och metadata. Denna del har genomförts av Miljödatacentrum i Kiruna AB (MDC) som har skickat ut förfrågningar och mottagit metadata. Leverantörer av metadata har varit Miljöförvaltningen i Stockholms Stad, Sveriges Geologiska Undersökning, Lantmäteriverket och Länsstyrelsen i Norrbottens län.
2. Upprättande av katalogtjänst. Denna del har utförts av MDC.
3. Inhämtande av synpunkter från slutanvändare. Denna del har utförts av Luleå tekniska universitet, som också har haft ansvaret för projektledningen.

5. ÖVERFÖRING AV FÖRFRÅGNINGAR OCH META DATA

Tekniken för överföring av meddelanden och metadata baserades i detta försök på elektronisk post. Det innebär att MDC skickade iväg en förfrågan om metadata till de dataproducerande organisationerna. Speciella epost-adresser inrättades för detta ändamål. Innehållet i brevet var formaliserat med hjälp av ett antal ledord med efterföljande värden. Ett exempel på en förfrågan om metadata visas i figur 2.

```
Metadata förfrågan version 0.1:  
    Förfrågan ID: [id-begrepp, t ex. mdc0001]  
Datum: [t ex. 981005]  
Adressat: [t ex SGU]  
Dataprodukt: [t ex Brunnsarkivet ]  
Version: [Samtliga] eller [ändringar efter <datum>]  
Område: < xmin,ymin,xmax,ymax > Koordinater anges i Rikets Nät  
Retur: [ftp (nod, dir, user, passwd)] eller [epost (epostadress)]  
Slut förfrågan
```

Figur 2. Struktur på förfrågan om metadata

När dataproducenten erhöll en förfrågan om metadata, påbörjades arbetet med att generera ett svar. Ledordet "Dataprodukt" avser en databas eller produkt som den är definierad hos dataleverantören. Ledordet "Retur" anger hur svaret skall skickas till katalogtjänsten, i detta fall MDC. Om svaret skall skickas med ftp, anges nod, bibliotek, användarnamn och dess lösenord. Om svaret skall skickas med epost anges epostadressen. I försöket skickades samtliga svar med epost.

Svaret på förfrågan skickades tillbaka till MDC. Figur 3 anger vilken information som önskas för varje dataset. Denna struktur följer den föreslagna kataloginformationen som krävs för att uppfylla Conformance level 1 enligt ISO's förslag till standard för metadata. Dessutom valdes att använda i huvudsak engelska ledord i filen istället för att skapa svenska motsvarigheter.

Själva överföringen av förfrågningar och metadata fungerade utan problem. Däremot skilde sig metoderna åt för hur man genererade svaren. Länsstyrelsen i Norrbotten genererade exempelvis sina svar automatiskt från sin metadatabas. Dessa svar var dock ej så fylliga, främst beroende på att de uppgifter som efterfrågades ej är anpassade till länsstyrelsens databaser och därför ej finns tillgängliga. SGU å sin sida sammanställde sina svar manuellt och sparade dem som textfiler. Dessa textfiler skickades sedan iväg automatiskt när förnyade förfrågningar kom. Svaren var således statiska, men det beror på att själva metadata är förhållandevis statiska för de databaser som ingick i försöket.

Några problem som finns rapporterade är

1. Formaten för förfrågning och metadata kan förenklas
2. Begrepp som "dataprodukt" och "overviewname" är luddiga
3. Vissa uppgifter saknas som kontaktperson samt databasens namn i kortform
4. Nyckelordslistorna enligt ISO är undermåliga. De är bland annat alldeles för kartinriktade.

Epost-tekniken är förhållandevis enkel och fungerade utan större problem under försöket. Yaser Bishr har identifierat 6 olika nivåer av interoperabilitet, dvs hur datorsystem kan samverka med varandra (Bishr, 1998). Denna indelning påminner om OSI-modellen som utvecklades under 1980-talet för att beskriva nivåerna i datorkommunikation. Epost-tekniken som här har använts motsvarar nivå 2 i Bishr's modell, d v s nivån precis ovanför nätverksprotokollen. Dagens systemleverantörer arbetar idag med att utveckla produkter på nivå 4, vilket här innebär att man kan söka efter information på nätet utan att man vet var de finns lagrade. Vi kan således förvänta oss att inom en snar framtid ha mycket kraftfullare hjälpmedel tillgängliga som kan hantera förfrågningar om metadata på ett effektivare sätt.

Svar på metadataförfrågan:

Förfrågan > mdc0001

Antal filer i svaret> 15

Detta filnummer> 3 (möjliggör att ett svar kan bestå av flera epost-brev)

Title> Benämning på dataset som används hos avsändaren

Version>

Responsible party information:

Responsible party role code> Vilken roll som organisationen har enligt tabell

Responsible party organisation:

Responsible party organisation name> Namn på organisation som är ansvarig för datasetet.

Postal address> Postadress

City> Postnummerområde

Administrative Area>

Postal Code> Postnummer

Country> Land

Overview:

Overviewname> Namn på översiktsinformation om denna skall utnyttjas av flera dataset.

Language of dataset code> Kod för språk använt i datasetet enligt ISO-639 kod (Sv för Sverige)

Abstract> Sammanfattande text som beskriver datasetet

Purpose> Sammanfattande text som beskriver syftet med att ta fram datasetet

On-line resource linkage > En http-adress till mer information om datasetet

Progress code> Hur långt har datasetet framskridit enligt tabell

Dataset extent:

Geographic extent name: Anger den geografiska utsträckningen mha av ett fördefinierat utsnittsnamn. Giltiga namn är: Länsnamn, Kommunnamn, topografiska kartblads namn.

Geographic reference system: Anger vilket koordinatsystem som utsnittskoordinater är angivna i. Standard om inget anges är RT90 2,5 gon Väst.

Geographic extent coordinates: Anger geografisk begränsning av datasetet i koordinatsystemet som är angivet ovan

West bounding coordinate> Västligaste koordinaten

East bounding coordinate> Östligaste koordinaten

North bounding coordinate> Nordligaste koordinaten

South bounding coordinate> Sydligaste koordinaten

Reference date> Datum då datasetet publicerade eller gjordes tillgängligt.

Resolution level code> En skalfaktor enligt tabell

Category:

Theme code> Tematisk kod som anger huvudsakliga innehållet för datasetet

Figur 3. Format för metadata

Som sammanfattning av denna del av försöket kan således följande slutsatser dras

- Ett system för utbyte av distribuerade metadata kan i dagsläget gärna baseras på epost.
- Viss oklarhet föreligger vad avser tolkning av termer och begrepp. Detta medför å sin sida också att katalogtjänsterna också behöver hantera olika tolkningar av dessa begrepp.
- ISO's Conformance level 1 förefaller ej vara generellt användbar. Den är i dagsläget alltför kartinriktad och passar ej annan typ av information, t. ex sådan som levereras av SGU och Länsstyrelsen i Norrbotten. ISO's värddlistor måste också anpassas till svenska förhållanden även för den mer kartinriktade informationen.
- Informationen (metadata) ligger ofta lagrad på många ställen inom organisationen. Den interna hanteringen av metadata är således ett problem.

6. UPPRÄTTANDE AV KATALOGTJÄNST

En katalogtjänst har två viktiga funktioner. För det första skall tjänsten regelbundet samla in metadata från olika dataleverantörer. Under detta försök gjordes en sådan insamling vid två olika tillfällen. För det andra skall katalogtjänsten presentera metadata och annan information på ett för slutanvändarna lämpligt sätt.

För att utforma ett söksystem som är direkt anpassat för en specifik användarkategori, krävs förhållandevis omfattande förarbete i form av formulering av kravspecifikation och implementering. Eftersom det primära syftet med denna studie var att prova leveranser av metadata i ett distribuerat system, valdes därför en befintlig mer generell sökmiljö, nämligen EnviCat från Miljödatacentrum.

EnviCat är ett söksystem för geografiska data som har utvecklats av Miljödatacentrum i Kiruna. Det baseras på databashantering i Oracle-miljö med tillägget SDE från ESRI för hantering av rumslig information. För denna studie kompletterades EnviCat med funktioner för automatisk förfrågning efter metadata samt automatisk laddning av inkommande metadata i en metadatabas. Att införa dessa funktioner mötte inga större tekniska problem. För att ladda data utvecklades dock funktioner för att kontroll av värdetyper, domänvärden, fältlängd, nycklar etc.



Figur 4. Exempel på användargränssnitt i EnviCat

Metadatabasen är implementerad i en ren relationsdatabas. Den är anpassad mot såväl FGDC's standard som ISO's standardförslag. Databasen innehåller c:a 80 tabeller. Katalogtjänsten utnyttjar dock enbart c:a 10 av dessa tabeller. Sökbara variabler är exempelvis datamängdens titel, ord i sammanfattningen, klassificering, datum, geografiskt läge och skala. Bland de tjänster som katalogen kan utföra är

- Söka data
- Bläddra runt
- Visa egenskaper för en datamängd
- Visa exempelbild
- Visa geografiskt utsnitt
- Lägg datamängd i kundvagn för ev beställning och leverans.

Vid överföring av metadata från dataproducenterna, fanns det önskemål om att kunna använda uppslagstabeller för ofta återkommande värden. Om denna teknik skall användas, fordras det dock rutiner för ajourhållning av dessa värdelistor och regler för hur värdelistorna skall publiceras.

7. SYNPUNKTER FRÅN SLUTANVÄNDARE

Även om denna studie primärt inte var avsedd att studera användarnas behov och synpunkter, gjordes en enkätundersökning med detta syfte. Orsaken till det är att dessa synpunkter behövs när det gäller att föreslå en hållbar strategi för fortsatt utveckling.

Synpunkter från olika användarkategorier insamlades vid ett seminarium i Uppsala den 11 januari 1999. Problemställningarna presenterades under några inledande presentationer. Därefter blev det fri diskussion och deltagarna fick dokumentera sina synpunkter i en enkätundersökning. Syftet med enkätundersökningen var ej att försöka kvantifiera förekomst av problem eller andra åsikter, utan enbart att dokumentera dem.

Metadata behövs för såväl internt producerade data som externa data. Enkäten tar upp frågor rörande såväl metadata som dataförsörjning. Av enkätsvaren framgår bl. a följande:

1. Internt producerade geografisk information
 - 1.1. Datakvalitén är ofta dåligt dokumenterad i metadata
 - 1.2. Metadata saknas som beskriver var data finns, vad som finns och uppgifter om senaste version
 - 1.3. Det finns problem med leverans av data (lång leveranstid, konverteringar krävs)
 - 1.4. Data har bristande datakvalitet, främst bristande fullständighet
2. Externt producerad geografisk information
 - 2.1. Kvalitetsuppgifter saknas ofta
 - 2.2. Dokumentation (metadata) håller ojämn nivå (standarder efterfrågas)
 - 2.3. Det är svårt att hitta metadata
 - 2.4. Data har ofta bristande datakvalitet, främst i form av bristande aktualitet och fullständighet
 - 2.5. Dataleverantören använder ibland en annan klassindelning än mottagaren
 - 2.6. Konvertering av data krävs ofta
 - 2.7. Samma tema kan vara spritt på flera olika dataset
3. Åtgärder för att öka kännedom om data
 - 3.1. Man bör bygga upp och vidareutveckla metadatabaser på intranet / internet
 - 3.2. Man bör göra produktionsinformation mer tillgänglig inom den egna organisationen
 - 3.3. Standardiserade databasbeskrivningar efterfrågas
 - 3.4. Standarder för att hantera och distribuera data efterfrågas
 - 3.5. Fler sökprotokoll efterfrågas
 - 3.6. Användarvänliga gränssnitt för att söka efter geografisk information efterfrågas
4. Problem med att upprätta metadatabaser
 - 4.1. Osäkerhet råder rörande databasdesign (struktur, relevant innehåll)
 - 4.2. Hur skall metadatabaserna ajourhållas?
 - 4.3. Hur bör gränssnitten mot slutanvändarna se ut?
 - 4.4. Hur erhåller man aggregerade beskrivningar ur exempelvis detaljerad produktionsinformation?
 - 4.5. Hur hanterar man dataleverantörernas bristande kvalitet på metadata?
 - 4.6. Det är idag svårt att prioritera verksamhet inriktad mot hantering av metadata
 - 4.7. Ansvar för uttag och produktion är idag spritt på många avdelningar
5. Skulle en katalogtjänst underlätta?
 - 5.1. Det skulle bli snabbare och enklare att hitta rätt information
 - 5.2. Fler skulle få kännedom om våra data
 - 5.3. En katalogtjänst skulle innebära förenklad administration av data
 - 5.4. En katalogtjänst skulle kunna innebära att missförstånd reduceras
 - 5.5. Ett distribuerat system ger bättre möjligheter till färsk information

Av de synpunkter som lämnas i enkäten, kan bl. a följande slutsatser dras

- Kvalitén hos data är ibland otillräcklig. Dokumentationen om datakvalitén kan dessutom vara dåligt eller saknas helt. Det bör påpekas att denna enkätundersökning ensam inte kan ge besked om hur omfattande dessa problem är. Det var dock en allmän uppfattning bland deltagarna i undersökningen att dessa problem är mycket vanliga. Detta indikerar att insatser inom detta område är mycket angelägna.
- Det saknas idag färdiga system för administration av metadata. Behov av systemutveckling och produktutveckling föreligger således. Produktutveckling kan exempelvis avse innehåll och struktur på metadata,

användargränssnitt och koppling till produktionssystem. Systemutveckling kan å sin sida exempelvis avse frågor rörande ajourhållning av metadata, versionshantering etc.

- Det finns ett behov av standarder inom området. Dessa standarder kan exempelvis avse metadata (termer och innehåll), filformat, sökprotokoll och klassindelningar. Som tidigare nämnts, så föreligger det ett omfattande standardiseringsarbete även utanför GI-området. Standarder för metadata tas exempelvis fram inom World Wide Web Consortium med tillhörande filformat och sökprotokoll (W3C, 1999). Man arbetar här med en modell för metadata (RDF, Resource Description Framework) och tänker implementera den i XML. Detta till skillnad mot standardiseringen inom CEN, som är baserad på EXPRESS. I takt med att standarder inom GI-området dröjer, kan andra mer generella standarder som RDF och XML komma att bli dominerande.
- Leverans av geografisk information behöver effektiviseras. Frågor som behöver lösas är exempelvis leveranstider, priser, nyttjanderätt och beställningsmekanismer. Dessa frågor är dock i huvudsak av administrativ och politisk natur.
- Frågor rörande metadata röner i dagsläget liten uppmärksamhet inom den egna organisationen. Huruvida dessa frågor verkligen förtjänar mer uppmärksamhet eller ej, har ej studerats inom denna studie. Det förefaller dock klart att en effektivare hantering av metadata ger rationaliseringsvinster och en förbättrad användning av såväl intern som extern information. För att erhålla en effektivare hantering av metadata kan vissa organisatoriska åtgärder vara nödvändiga, som exempelvis att integrera metadatahanteringen med produktionssystemen samt genom att förenkla ansvarsfördelningen mellan olika avdelningar.

8. FRAMTIDA ARBETE

För att ett fortsatt arbete med metadata skall bli framgångsrikt, krävs att insatserna inriktas mot åtgärder som ger reella nyttoeffekter. För utvecklingsinsatser krävs att nyttoeffekterna visas på relativt kort sikt, medan forskningsinsatsernas kan motiveras trots att nyttoeffekterna kan komma att visa sig först på längre sikt.

De ekonomiska drivfaktorer som har kunnat identifieras är i korthet

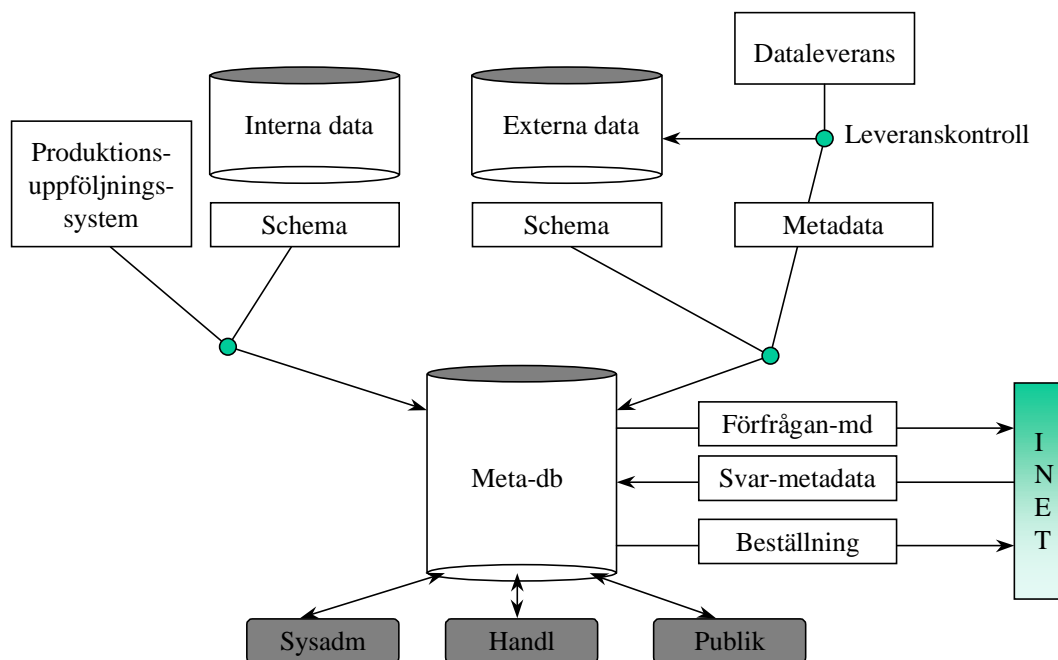
- En god tillgång till metadata ökar användningen av geografisk information. Ökad användning av geografisk information antages i sin tur förbättra verksamheten.
- Ett fungerande system för metadata förenklar systemadministrationen
- Metadata, och då speciellt kvalitetsuppgifter, behövs för att kunna utvärdera slutresultatet.

Antag att en effektiv hantering av metadata ger nyttoeffekter som motiverar investeringar i ett system för metadata. Figur 5 visar ett exempel på hur ett sådant system skulle kunna se ut.

Det system som föreslås i figur 4 består av 5 delar, nämligen en del för hantering av egenproducerade data, en del för hantering av externa data, en del för förfrågningar på nätet, en metadatabas samt en del bestående av ett antal olika användargränssnitt. En viktig uppgift för metadatabasen i detta exempel är att den innehåller accessvägar till data. Det innebär att metadatabasen hanterar uppgifter om var data finns lagrat, senaste version etc.

Hantering av egenproducerade data består av ett system för produktionsuppföljning. Här lagras uppgifter om produktionsprocesser och kvalitetsbedömningar som är gjorda. Det primära syftet för ett sådant delsystem är att övervaka, dokumentera och följa upp själva produktionen. Men dessa data kan också aggregeras till mer översiktliga metadata som lämpar sig för lagring i en mer allmän metadatabas. Exempel på aggregerade metadata kan vara datumuppgifter, sammanfattning av bearbetningssteg, sammanfattning av indata, sammanfattande kvalitetsbedömningar etc.

Externt producerade data kan levereras via nätet, CD-ROM eller annat media. Vid all leverans krävs någon form av leveranskontroll. Vid denna kontroll kan också diverse bearbetningar göras, för att exempelvis passa in data i den interna strukturen. De metadata som levereras av dataleverantören, eller som kan härledas ur själva datamängderna, läggs in i metadatabasen.



Figur 5. Exempel på möjligt system för hantering av geografiska metadata

Viss information kan behövas beställas över nätet. Det kan exempelvis vara information där aktualitetskraven är höga, t ex vissa miljödata. Det kan också vara data som organisationen ej har använt tidigare och som således behöver införskaffas. För denna typ av verksamhet behövs söktjänster och beställningsfunktioner. Själva dataleveranserna förutsätts dock gå igenom den gemensamma funktionen för leveranskontroll.

Den centrala komponenten i hela systemet för metadata är själva metadatabasen. Här lagras övergripande information om olika datamängder som exempelvis accessvägar, rättigheter, nyttjanderätt, datakvalitet, lämpliga användningsområden etc.

För att använda systemet krävs en uppsättning användarmiljöer. Dessa miljöer kan utformas olika, beroende på behov. I figur 4 visas 3 olika användarmiljöer. Dessa kan exempelvis avse en miljö för systemadministrativ personal, en miljö för handläggare och en miljö för publik åtkomst över Internet.

Ett eventuellt fortsatt arbete föreslås att inriktas mot att prova det system som illustreras i figur 4. Av speciellt intresse är då givetvis de kostnads- och nyttoeffekter som ett sådant system medför. Men innan en sådan studie kan genomföras, behöver vissa komponenter studeras ytterligare, nämligen:

1. Metoder för att upprätta katalogtjänster
2. Metoder för att extrahera metadata ur produktionssystem
3. Metoder för leveranskontroll vid leverans av externa data

För genomförande föreslås att de olika utvecklingsinsatserna bedrivs i ett antal separata utvecklingsprojekt. Tänkbara tillämpningsområden kan vara kommunala tillämpningar (översiktsplanering eller detaljplanering) eller räddningstjänst. För en sammanhållen strategi föreslås att ett nätverk bildas, där olika intressenter kan hitta gemensamma lösningar.

9. SLUTSATSER

Baserat på de i denna studie redovisade försök, kan följande slutsatser dras.

OGC's koncept med katalogtjänster separerade från dataproducenter stämmer helt överens med modern syn på elektronisk handel och s k digitala bibliotek.

För kommunikation mellan katalogtjänst och dataleverantör, kan i dagsläget epost användas. Inom en snar framtid kan dock mer moderna tekniker bli att föredra.

De nuvarande standardförslagen inom GI-sektorn vad avser metadata förefaller vara alltför kartorienterade för många "icke-traditionella" producenter av geografisk information.

Kvalitén hos data är ibland otillräcklig och dokumentation om datakvalitet saknas oftast helt och hållet.

Det saknas idag färdiga system för administration av metadata

Leveranser av geografiska data behöver effektiviseras.

10. FINANSIERING

Denna studie har finansierats av Utvecklingsrådet för Landskapsinformation (ULI). Dessutom har Miljödata-centrum i Kiruna, Stockholms Stad, Sveriges geologiska Undersökning, Lantmäteriverket samt Länsstyrelsen i Norrbottens län själva finansierat sina egna insatser.

11. REFERENSER

Barett S, Konsynski B, 1982. Inter-organisational information sharing systems. *MIS Quarterly* (Special Issue), pp 93-105.

Bishr Y, 1998. Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science*, Vol 12, No 4, pp 299-314.

CEO, 1999. Recommendation on Metadata – Describing the data, services and information you have available!. Center for Earth Observation Programme of the European Commission, JRC, Ispra, Italy.

European Commission, 1999. Public Sector Information: A Key Resource for Europe, <http://www2.echo.lu/info2000/en/publicsector/greenpaper.html>

Executive Office of the President, 1994. Coordinating geographic data acquisition and access the National Spatial Data Infrastructure, Executive Order 12906. Federal Register 59, pp 17671-17674.

Masser I, 1999. All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science*, Vol 13, No 1, pp 67-84.

Lopez X.R, Larsgaard M, 1998. Towards a California Geospatial Digital Library: A Strategy for Networked Knowledge. *Cartography and Geographic Information Systems*, Vol 25, No 3, pp 133-142.

OGC, 1998. The OpenGIS Specification Model, Topic 11: Metadata, Version 3.1. OpenGIS Project Document Number 98-111r2. <http://www.opengis.org/public/abstract/98-111r2.pdf>.

W3C, 1999. Resource Description Framework (RDF) Model and Syntax Specification. <http://www.w3.org/TR/1999/PR-rdf-syntax-19990105>.

Tidigare utgivna rapporter

1988:1

Program för GPS-verksamheten i Sverige, LMV-rapport 1988:19

1988:2

Research and Development within the
Field of Geographic Information Systems in Sweden Gävle 1988

1989:3

Standardisering inom området landskapsinformation Stockholm/
Gävle 1989

1989:4

Geografiska informationssystem Gävle 1989

1990:5

GIS i Sverige Gävle 1990

1990:6

The Swedish Standardization Project within in the Field of GIS Gävle 1990

1991:1

Införande av GIS Gävle 1991

1991:2

GIS programvaror Gävle 1991

1991:3

GIS-användning i Sverige Gävle 1991

1992:1

Swedish R&D in GIS 1991 Gävle 1992

1992:2

Handlingsprogram för forskning och utveckling inom
geografiska informationssystem Gävle 1992

1995

Bok - Introduktion till GIS. ISBN: 91-630-3245-7 Gävle 1995

1996:1

GIS i Sverige 1995 Gävle 1996

1997:1

Kostnads/nyttoanalyser av GIS projekt Gävle 1997

1997:2

GIS i Sverige 1997 Gävle 1997

1998:1

GIT i Kalmar Gävle 1998

1998:2

ULIs historia Gävle 1998

1999:1

GIS i Försvarsmakten Gävle 1999

ULI

Utvecklingsrådet för landskapsinformation

ULI är en ideell förening som verkar för effektivare användning av landskapsinformation. ULI leds av en styrelse och har ett kansli som sköter den löpande verksamheten. Föreningen är öppen för svenska organisationer. Verksamheten finansieras främst genom medlemsavgifter.

Föreningens kansli: Lars Hansen, Christina Wallström
ULI, 801 82 Gävle. Tel: 026-61 10 50. Fax: 026-61 32 77
E-post: uli@uli.se. Hemsida: www.uli.se