

Sammenlægning af geodata - sikring af kvalitet og historik

Søren Tollund, *Informi GIS*

Ideen med denne artikel er at præsentere nogle af de problematikker, overvejelser og arbejdsgange, det er væsentligt at fokusere på, når geografiske data fra forskellige kilder skal lægges sammen – sådan som det netop er tilfældet i forbindelse med kommunalreformen.

Centralt under en sammenlægning af data fra mange forskellige databaser, filservere m.m. er den fælles database, der er i stand til at lagre såvel geometri som egenskabsdata – vi betegner denne løsning Geodatabasen. Hvordan indlæser og samler vi data fra forskellige geografiske områder, evt. lagret i forskellige koordinatsystemer og i forskellige formater og kvaliteter, i den samme database? Geodatabasens struktur kan bruges til at styre indlæsningen, validere datas topologiske kvalitet og sikre fremtidig arbejdsgange.

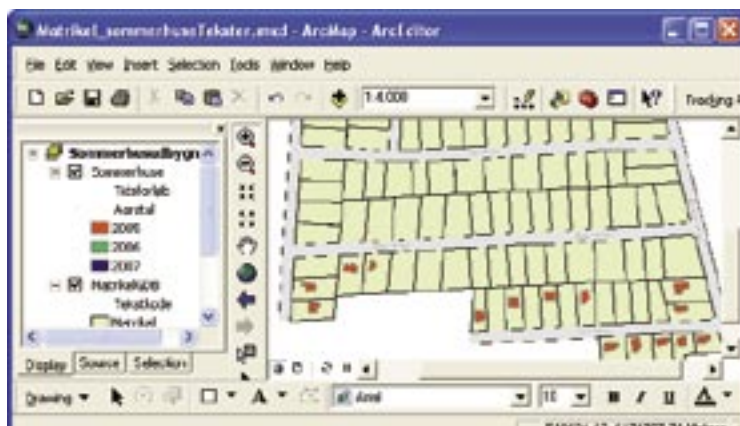
Kommunalreform eller ej, så vil der for langt de fleste geografiske datasæt også være en generel problematik knyttet til opdateringen af data: Hvordan ændrer data sig over tid? – og hvordan sikrer vi, at ændringerne kan følges, når data opdateres? Artiklen præsenterer bl.a. gennem et eksempel, tankegangen bag et historisk arkiv for geodata.

Geodata og historiske ændringer

Verden forandrer sig. Vi mennesker er i høj grad bestemte for forandringerne i kulturlandskabet: Vi lokaliserer råstoffer – og graver efter grus eksempelvis. Hvad stiller man op, når der ikke er mere grus at komme efter? Fylder hullet op igen? Eller omdanner stedet til rekreativt område? Uanset hvad valget bliver, vil vi gerne kunne følge disse forandringer senere.

En eller anden form for tidsstempel er naturligvis afgørende, når der skal arbejdes med historik i geografiske data. Lad os se på nogle forskellige eksempler på registreringsmetoder og muligheder for anvendelse. Først en tænkt nyudstykning – et sommerhusområde.

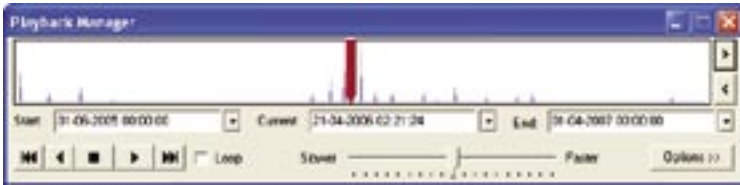
Hvert enkelt hus i det nye sommerhusområde lagres med en indflytningsdato i egenskabstabelen. Dermed er det nemt, at



Figur 1: Et (fiktivt) sommerhusområde under udbygning. I midten af juni 2005 er situationen som vist på kortet med 15 huse klar til indflytning. De populære grunde nærmest vandet (mod syd) er åbenbart blevet solgt først.

OBJECTID	Indflytningsdato	Shape	ShapeArea	Shape*	DBR	FBR
323	01-06-2005	70 Polygon		-Null-	-Null-	
325	01-06-2005	104 Polygon		-Null-	-Null-	
14	01-07-2005	95 Polygon		-Null-	-Null-	
384	01-07-2005	66 Polygon		-Null-	-Null-	

Figur 2: Egenskabstabelen for sommerhus-temaet.



Figur 3: I en 'Playback Manager' vises den tidslige fordeling af indflytningstidspunkterne. Ved at indtaste en ønsket dato eller skubbe på 'skyderen', kan man nemt vælge hvilket tidspunkt man ønsker at få vist i kortet.



Figur 4: Den 21. april 2006 er indflytningen i området som vist i kortet.

følge udbygningen af området. Et andet eksempel på tidsstemplede geografiske data kan ses på hjemmesiden "Det aktive Aalborgkort". Her kan borgeren gå ind og se den historiske byudvikling i et interaktivt kort med graderet farvelægning.

Modellen med at lagre alle objekter i et datasæt med et tidsstempel kan således også anvendes på byområderne. Og den fungerer såvel for byområderne som for det føromtalt sommerhusområde, så længe der blot er tale om at der føjes nye objekter til tabellen. Men hvad gør vi, den dag en af sommerhusejerne får lyst

til at udvide sommerhuset? Så er der jo pludselig tale om en ændring af et eksisterende objekt. De fleste GIS-programmer har værktøjerne til at redigere geometrien, så den afspejler husets grundplan efter ændringen. Hvad så med historikken – muligheden for om et par år at gå tilbage og se på bygningstemaet anno 2005? Måske har jeg en backup af data fra tidligere år; en kopi af alle bygninger i kommunen lagret som vektordata for hvert eneste år. Det ville selvfølgelig give mig mulighed for at grave i fortiden – hvis jeg overhovedet er i stand til at finde de CD'er,

der rummer det ønskede årstal. I praksis kan det være den løsning, man tyer til; men det er tydeligvis ikke nogen ideel løsning at lagre en kopi af den samme bygning år efter år, hvis den aldrig har været udsat for forandringer.

Et andet konkret eksempel på digitale datasæt under stadig forandring er matrikelkortet. Her kan vi fx løbe ind i problemer, fordi en servitut i praksis kan vise sig at være knyttet til et matrikelnummer, der ikke længere er eksisterende pga. udstykninger.

I den til enhver tid gældende version af matrikelkortet er der derfor brug for at kunne følge historiske ændringer – såvel i geometrien som i egenskaberne.

Det er muligt at sammenstille information imellem matrikelkort fra forskellige år som det ses i figuren. Dette kan fx gøres gennem overlayanalyser; men vi ville være betydeligt bedre stillet med en datamodel, der direkte understøtter de tidslige forandringer.

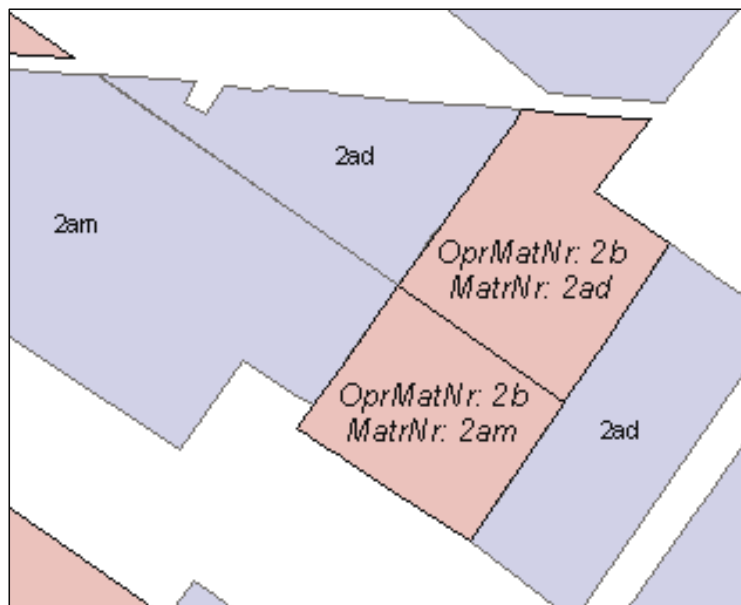
Det historiske arkiv

Vi har med andre ord brug for en datamodel og en arbejdsgang, der tager hånd om tidslige forandringer på det enkelte objekt. Vores IT-administrator vil givet være begejstret for ideen, da perspektivet er, at der kan spares masser af backup-plads...

I de fleste tilfælde er den optimale løsning at arbejde med et egentligt historisk arkiv.



Figur 5: Udviklingen af byområder i Aalborg. Farvelægningen indikerer i hvilken periode, det enkelte område er udbygget. Kilde: <http://www.aalborg.dk/vejviser/>

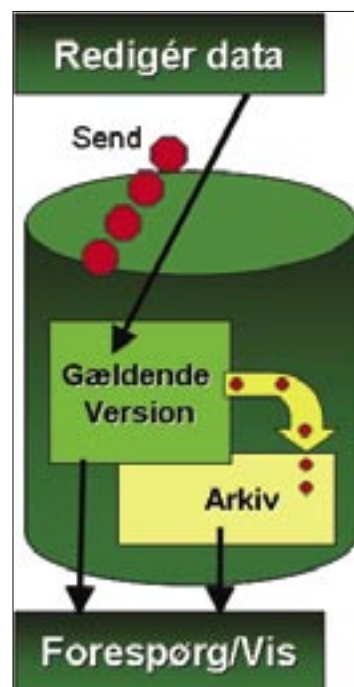


Figur 6: Matrikler i forandring. Ved at sammenstille matrikelkortet med en tidligere udgave er det lykkedes at overføre det tidligere matrikelnummer til de enkelte flader. Områder vist med rød farve har været udsat for forandringer. Eksemplet er venligst udlånt af Aalborg kommunes tekniske forvaltning.

Den grundlæggende ide er, at kun objekter der er blevet ændret, lagres i det historiske arkiv.

Såvel ændringer i geometri som i egenskaber betyder at objektet registreres i arkivet. Denne metodik for lagring af historik giver en række fordele:

- Man undgår at have kopier af hele datasæt fra forskellige år, idet kun objekter der har undergået ændringer lagres i arkivets tabeller.
- Arkivtabeller har kolonner med 'fra' og 'til' gyldighedsdato. Der kan dermed forespørges i arkivtabeller på dato, sådan at man kan se hvordan 'verden' så ud fx 20. oktober 1999. Det kort,



Figur 7: Principperne i en database-model med et historisk arkiv.

man får vist, vil netop vise kombinationen af 'gældende version' og det historiske arkiv.

Det følgende eksempel viser en konkret løsning, hvor der arbejdes med historisk arkivering af geografiske data; nemlig i registreringen af naturtyper i Ribe Amt. Der arbejdes altid i den gældende version, men når et objekt ændres, registreres ændringerne i arkivtabeller med oplysninger om oprindelses-ID og -dato; samt datoer for, hvor længe det pågældende objekt har 'eksisteret'.

for en bestemt kopi af matrikelkortet, mens en transaktion er en ændring i en geografisk database i form af forandring af et objekt eller evt. en afsluttet sekvens af opdateringer.

Eksemplet fra Ribe Amt for understøttelse af historik, stammer faktisk ikke fra data lagret i en geodatabase. Tværtimod er der tale om, at historikken understøttes af en specialudviklet applikation, som kombineret med arbejdsgangen sikrer, at data lagres korrekt i det historiske arkiv. I den næste version af geoda-

Før data fra forskellige datakilder skal indlæses og samles i geodatabasen, er det derfor vigtigt at være bevidst om at få opbygget en hensigtsmæssig databasestruktur, som kan lette det fremtidige arbejde. Det vil her blive for omfattende at beskrive alle overvejelserne og de tilhørende muligheder, det giver. De næste afsnit vil give nogen få eksempler - herunder topologisk rensning af data.

Datakvalitet – Topologisk rensning og validering af data

Geodatabasen er det centrale lager for geografiske data. De samme data kan tilgås af alle organisationens brugere og datamodellen sikrer kvaliteten, når der redigeres i data. Jo mere arbejde, der er lagt i at have en god datamodel – en struktur for databasen – jo flere fordele vil brugerne opleve, når de først begynder at arbejde med data, der er lagret i geodatabasen.

En af de store fordele er muligheden for at validere data. Validering kan fx betyde, at man kun har lov at indtaste bestemte værdier, fx skov/eng/mose – eller værdier indenfor et bestemt interval, fx 4-12 meter - i en kolonne. Dette betegnes normalt attributvalidering (=egenskabsvalidering).

Den anden side af geodatabase-validering er det område, der betegnes spatial (=geometrisk) validering. Det

AREA	ID	DATO	DELKOD	RET	OPDATO
583731 147	3993	23-04-2003	EN	IRIA	95-01-1995

AREA	ID	DELKOD	OPDATO	DATO	HISDATO
583731 147	3993	EN	95-01-1995	95-01-1995	05-10-1995
583731 4281	3993	EN	95-01-1995	95-15-1995	05-10-1995
583731 431	3993	EN	95-01-1995	23-04-2003	29-04-2003

Figur 8: Det samme naturområde - ID 3993 - har været udsat for forandringer i dets registreringer 3 gange i løbet af den periode registreringerne har fundet sted. Eksemplet er venligst udlånt af Ribe Amt.

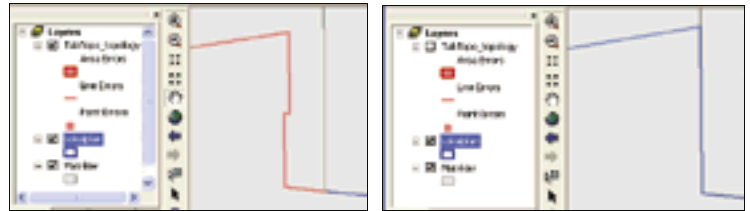
Datamodellen og arbejdsgangen sikrer, at disse informationer overføres automatisk til arkivet.

Opsummerende om historik i geodata kan vi skelne mellem "Snapshots" og "Transaktioner". Et snapshot er et øjebliksbillede af et helt datasæt - f.eks. ortofotos fra et bestemt årstal eller et tidspunkt

tabasen er det imidlertid selve datamodellen, der understøtter historikken på samme måde, som det er illustreret her. Ud over de sædvanlige styrker ved at lagre i en database, giver det den store fordel, at man ikke skal til at programmere og specialudvikle applikationer for at arbejde med historik i data.

giver os eksempelvis mulighed for at sikre, at adressepunkter ligger indenfor bygningpolygoner, at amtsgrænser (og regionsgrænser for den sags skyld) falder sammen med kommunegrænser osv

Lad os tage et eksempel: De fleste kommuner har et digitalt matrikelkort, som de har anskaffet fra anden side – et matrikelkort, kommunen ikke selv skal opdatere. Til gengæld arbejder man i kommunens 'lokalplankontor' med at udfærdige egne lokalplaner. Som regel er det hensigten, at lokalplansgrænser skal følge matrikelgrænser; men der kan selvfølgelig være undtagelser. På kontoret har man under geografiske søgninger været udsat for eksempler på, at nogle matrikler falder indenfor forkerte lokalplansområder pga. små geometriske fejl. Det er imidlertid en stor opgave at skulle undersøge datasættene minutvist for sådanne topologiske fejl,



Figur 10: Lokalplansgrænser der ikke følger matrikelgrænser strider mod de opstillede regler i geodatabasen og markeres derfor med rødt. Figureerne viser lokalplansgrænsen før og efter redigering. Efter redigering er den nordlige del af lokalplansgrænsen markeret som en undtagelse, mens den nord-sydgående del nu følger matrikelskellet.

hvis man ikke har systematiske værktøjer til det.

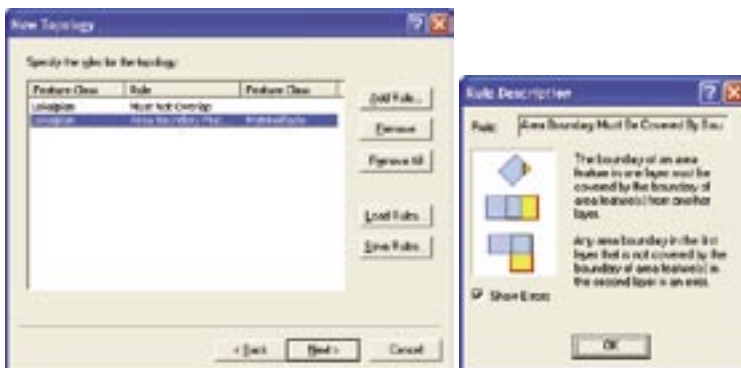
For at kunne foretage en topologisk validering vælger man derfor at indlæse datasættene med lokalplaner og matrikelflader i en geodatabase. I geodatabasen kan man oprette en topologi med regler, hvorefter man kan lade GIS-programmet kontrollere, at reglerne overholdes. I de tilfælde, hvor der er tale om bevidste undtagelser fra reglerne, kan man markere det som 'undtagelser' i databasen.

I forbindelse med sammenlægning af datasamlinger fra flere kommuner vil der med stor sandsynlighed optræde såvel topologiske fejl som fejl i egenskabstabellerne i flere forskellige datasæt. Sammenlægningen er dermed en god lejlighed til at gå data efter i sømmene.

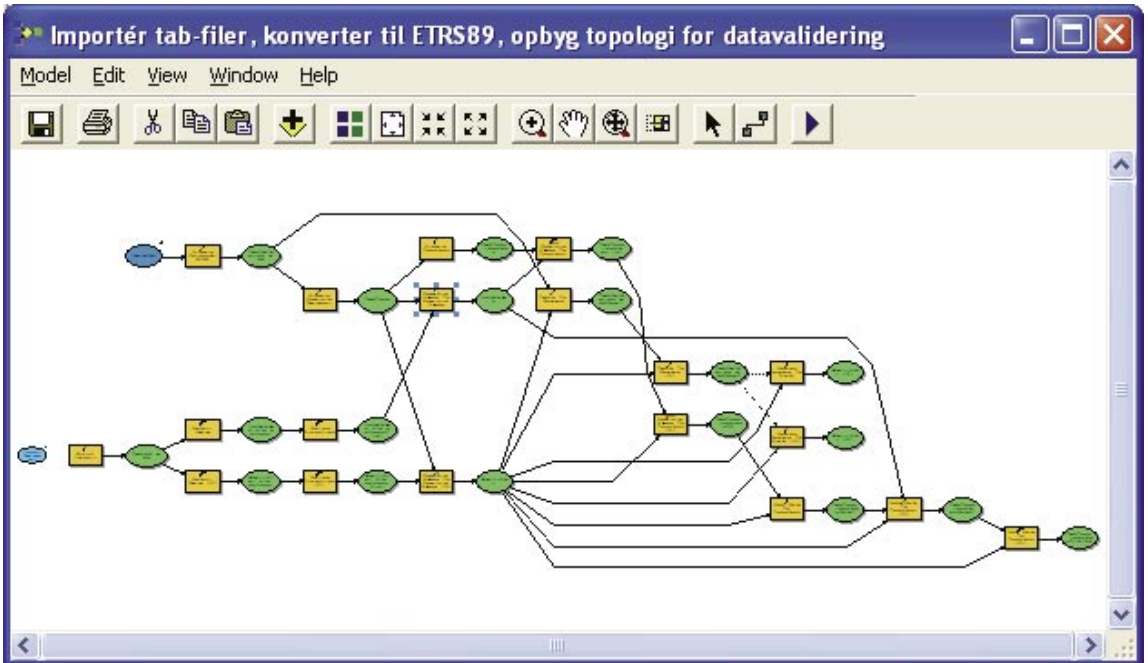
Andre overvejelser i forbindelse med sammenlægningen

Mange kommuner arbejder stadig med geodata registreret i det danske koordinatsystem System 34. I forbindelse med sammenlægningen vil det være oplagt at foretage konvertering til System2000 med referencesystemerne ETRS89 (også kaldet UTM/EUREF89) og DVR90 (højdesystem).

Mange af de arbejdsgange, der skal gennemføres i forbindelse med en sammenlægning, kan modelleres i en interaktiv model. Modellen gør det nemt at ændre på parametre og dokumenterer samtidig arbejdet. Den følgende figur viser et eksempel, hvor tab-filer (her fra GIS-programmet MapInfo) med



Figur 9: Et eksempel på oprettelse af topologiske regler i en geodatabase. Målet er dels at gennemføre en topologisk rensning af lokalplanområder der indlæses - dels at sikre at reglerne automatisk anvendes under fremtidig redigering.



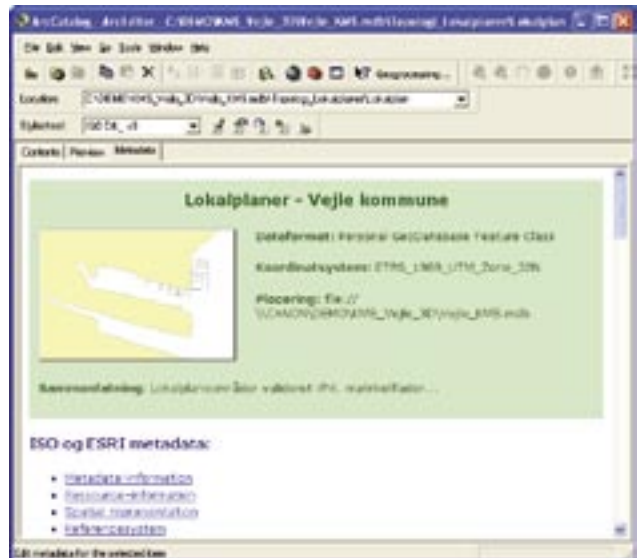
Figur 11: I en model styres hele arbejdsgangen fra import af datasæt til konvertering mellem Sys34 og ETRS89 til oprettelse af geodatabasetopologi.

matrikler og lokalplaner konverteres til ETRS89 og indlæses i en geodatabase for derefter at indgå i en topologi. I topologien specificeres et regelsæt som omtalt i forrige afsnit. Modellen kan nemt udvides, så andre datasæt evt. i andre formater indlæses samtidig.

På KMS' hjemmeside kan man læse mere om System2000 og opgaven med at konvertere fra System34.

Metadata – datas varedeklaration

Det forestående arbejde med sammenlægningen af data er også den oplagte lejlighed til at få indføjet metadata i organisationen, hvis man ikke allerede har gjort



Figur 12: Eksempel på metadata knyttet til en featureklasse - her i en personlig geodatabase. I en organisation som en kommune vil man vælge en 'enterprise' relationsdatabase - SQL Server, Oracle, Informix eller IBM DB2.

det. Metadata er en varedokumentation, der knytter sig til det enkelte datasæt: Hvem har ansvaret for datasættet, hvor tit opdateres det, hvilken kategori tilhører det, hvilke restriktioner knytter der sig til det, osv.?

Udfyldelse af metadata for et nyt datasæt bør være en naturlig del af arbejdsgangen med geodata. Dermed kommer metadata til at optræde som en integreret del af geodata til glæde for alle brugere i organisationen.

Afrunding

Geodatabasen, som det centrale sted for lagring af data, giver os altså gode muligheder for at samle data fra mange kilder, sikre datakvaliteten og dokumentere data gennem metadata.

Derudover giver geodatabasen en række andre fordele, som ligger udenfor denne artikels tema. Der kan bl.a. nævnes flerbruger-samtidig-redigering, distribuerede databaser (replikering), samt funktioner til at understøtte arbejdsan-

ge, eksempelvis versionering. Og endelig den sikkerhed det giver at have data lagret i en database frem for på diverse filservere.

Hjemmesider til inspiration:

Omlægning til System2000: <http://www.kms.dk> - Søg derefter på 'System 2000' Natur i Ribe Amt: <http://www.ribeamt.dk/sw699.asp>

Historik i 'Det aktive Aalborg-kort': <http://www.aalborg.dk/vejviser/> - Her findes også ældre kort.

Om forfatteren:

Søren Tollund Christensen, salgschef for kurser og ArcGIS-udvidelser. Uddannet cand.scient. i geografi og fysik; MTM i geoinformatik, ansat hos Informi GIS A/S siden 1999. Arbejder til daglig med ESRI's ArcGIS Desktop-programmer bl.a. i forbindelse med teknisk support, kurser og pre-sale-opgaver.