

Forskningsdatabasen – en tids maskine

Et forskningshistorisk forsøg på en tidsrejse i relationelle databaser

af Jette Rostock

At foretage tidsrejser er som de fleste ved, ikke muligt – i 1998 var arkæologi-studerende Anne Sophie Hansen dog alligevel på en sådan i forbindelse med et kursus i relationelle databaser på Institut for Humanistisk Informatik! På et eksemplar af arkæologen Nina Nikolaevna Gurinas publikation foretog hun en rejse i tid og rum – tilbage til 1950'ernes Leningrad, USSR, hvor hun dumpede ned på Gurinas kontor. Nutidens teknologiske muligheder fulgte med på rejsen, og de besluttede sig derfor for at lave en relationel database, Microsoft Access 1997, til den videnskabelige bearbejdning af det mesolitiske gravfeltet Olenii Ostrov. Men Anne Sophie og Gurina blev noget uenige om databasens udformning.



Om denne fortælling

Her vil jeg fortælle historien om Anne Sophies besværligheder med at lave en relationel database, og hendes diskussioner med Gurina – et forskningshistorisk forsøg på en tidsrejse i relationelle databaser. Anne Sophies og Gurinas uenigheder hvad angår forskellige centrale elementer i databasens design anskueliggør nemlig på glimrende vis de problemer vi arkæologer i dag støder på når vi skal skabe og benytte fælles databaser – Gurina ville designe databasen på sin måde og Anne Sophie på sin. Så med udgangspunkt i denne historie vil jeg belyse hvorledes et databasedesign afspejler sin skabers syn på materialet og dernæst drage nogle konklusioner om relationelle databasedesigns anvendelighed og overlevelsessevne i tid og rum. Er det med andre ord muligt at designe universaldatabaser?

Før afrejse (København år 1998)

"Den hersens eksamensopgave skal bare fyres af " tænkte Anne Sophie højt. Opgaven lød på at designe en database. "En hvilken som helst ... Vælg selv" havde læreren sagt. Nærmest ville det nok have været at lave en over sin CD samling, men det var trods alt for kedeligt. Så i stedet havde hun valgt at lave en til et arkæologisk materiale – til Olenii Ostrov. En lokalitet opkaldt efter øen (Hjorte øen) hvor den blev fundet. Øen ligger i den nordlige del af Onega søen i Karelien.

Olenii Ostrov er et af Europas største mesolitiske gravfelter og er at regne for en Locus Classicus. Den blev udgravet i 1930'erne af en af arkæologiens kvinder, Nina Nikolaevna Gurina. Anne Sophie var en stor beundrer af hende. Så at sidde med Gurinas publikation fra 1956 foran sig, førte Anne Sophies tanker tilbage til dengang hun først var støt på publikationen i Institut for Arkæologis allermest støvede afdeling – den østeuropæiske. Tegningerne i bogen havde været tiltrækkende. Hver eneste af de over hundrede jordfæstegrave af forskellig form og konstruktion var omhyggeligt gengivet med et, to eller tre skeletter. Og hver enkelt af de over syv tusinde genstande var indtegnet. Her var spyd og knive. Her var udskårne figurer og perler af dyretænder. Her var også en eller flere ubearbejdede dyreknogler. Dette havde været så spændende at hun med sit gymnasierussiske og en ordbog havde stavet sig igennem først overskrifter, siden billedtekster og de omfattende tabeller. Anne Sophie havde på denne måde til sidst fået læst det meste af publikationen.

"Det bliver ikke så svært at lave en database til det materiale", tænkte Anne Sophie mens hun vandrede frem og tilbage over gulvet. "Jeg kender det ud og ind ... alle facts ... det er bare at hente facts .. data .. i Gurinas publikation". I Gurinas publikation foreligger Olenii Ostrov materialet naturligvis i en allerede bearbejdet, klassificeret og tolket form, men det bekymrede ikke Anne Sophie. Og hun satsede på at databasen når den var færdig, ikke kun skulle rumme

data fra lokaliteten Olenii Ostrov, men også fra andre mesolitiske lokaliteter og de nedgravede hunde på disse - en ting der havde Anne Sophies særlige interesse.

Anne Sophie satte sig foran computeren og begyndte glad at skrive en liste over hvilke former for oplysninger, data, hun ville putte ind i databasen - data der til sidst skulle udgøre databasens input, som det kaldes i databaseterminologi. I modsætning til output som er de oplysninger databasen kan spytte ud igen. Det svirrede i Anne Sophies hoved: "Input...output...input...output...". Listen over input blev stadig længere. Og det gik efterhånden op for hende, at det også hang sammen med hendes forestillinger om hvilket output hun ønskede at få - hvilke forespørgelser, som det hedder i databaseterminologi, hun senere ville fortage i databasen.

Anne Sophie kiggede på sin liste over de forskellige former for data hun ville putte ind i databasen – den var lang. Var det overhovedet muligt at få orden på den liste? En stor tabel med rækker og kolonner sat op som tabellerne i Gurinas publikation, med de enkelte grave og skeletter i hver sin række, var en løsning hun ville have valgt hvis tingene skulle foregå manuelt. Men her var der tale om en database, et edb program, og for en relationel database er der regler der skal følges hvis den skal fungere. En relationel database er ikke opbygget som én stor tabel – den er ikke som en enkelt kasse proppet hulter til bulter med al verdens facts, hvorfra data kan proppes ned og hives op som man nu har lyst. Nej, relationelle databaser hedder netop sådan fordi de er opbygget ved hjælp af relationer - relationer mellem forskellige grupper af data, entitetstyper som det hedder i databaseterminologi. En database er således opbygget af en masse entitetstyper som er sat sammen med, relateret til, andre entitetstyper i et stort sammenhængende system.

Anne Sophie printede listen ud - de forskellige former for data stod lidt tilfældig imellem hinanden. Hun kiggede på papiret og pludselig bevægede ordene sig ud-og-ind-og-op-og-ned for til sidst at have samlet sig i nogle større grupper. "Ja, sådan .. selvfølgelig!" Der var en gruppering med alle oplysningerne om selve graven, hvor den lå på lokaliteten og hvordan den så ud. Der var alt om skelettet, dets bevaringsgrad og alders- og køns- bestemmelse. Der var en gruppe med alle oplysninger om oldsagen, beskrivelse af den og dens typebestemmelse. Og der var en gruppe med alle oplysninger om den enkelte lokalitet, hvor og hvornår den var fundet og hvem der har gravet den. " Fedt nok! .. Det er jo entitetstyper! .. Det ser rimeligt ud!" Sagde hun til sig selv. Men havde hun nu taget højde for alle de oplysninger hun ville have med? Og hvordan hang det hele sammen? - entitetstyperne skulle jo relateres til hinanden. Desuden var der også lige lidt mere Anne Sophie skulle huske - databasen ville kun kunne fungere korrekt hvis den blev konstrueret efter visse regler. Regler der skal overholdes - nøglekrav, entitetsintegritet og normalformer kaldes nogle af disse regler i databaseterminologi. Reglernes overholdelse får ofte betydning for entitetstypernes indhold og relationernes karakter - kort sagt,

reglerne får også betydning for databasens design. Men Anne Sophie tænkte endnu ikke på reglerne. I stedet åbnede hun Gurinas publikation for at finde ud af hvilke relationer hun skulle knytte mellem entiteterne: Grav, skelet og oldsager. "Gurina må vide det!" Og ganske rigtigt i publikationen havde Gurina foregrebet begivenhedernes gang - i sin måde at stille tabellerne op på havde hun angivet en sammenhæng. Anne Sophie rev sig i håret: "Det kan da ikke passe? .. hvorfor sådan? .. Underligt! .. det ville da være bedre hvis ..."

Det svimlede for Anne Sophie og Gurinas publikation gled ud af hænderne på hende og ned på gulvet hvor den landede med et bump!

På Gurinas Kontor (Leningrad før år 1956)

"Isvinitje!". Det gav et sæt i Anne Sophie. Hun befandt sig i en bulet lænestol med tvivlsom fjedereffekt. Foran hende stod et stort skrivebord hvorpå der lå en del papirer - udgravningsplaner var det vist. Kvinden på den anden side rakte hånden frem og sagde: "De må meget undskylde, men jeg håber ikke De kom til at vente for længe .. Dejligt at møde Dem igen! .. De har taget Deres avancerede skrivemaskine med, ser jeg". Det gik op for Anne Sophie at kvinden var den russiske arkæolog Gurina og forsøgte sig med et gebrokkent: "Dobrie djen! .. God dag! Ja, jeg har taget computeren med". Samtidigt undrede hun sig over hvad både computeren og hun selv gjorde her, for over døren hang et gulnet portræt af Stalin. Og den Gurina som sad på den anden side af skrivebordet var bestemt ikke tæt på de 90 år. Gurina blev født allerede i 1909 - så det her kunne ikke være år 1998! En tanke der blev bekræftet da Gurina sagde: "De må endelig medens De er her benytte dem af vores nye metro" Leningrads metro blev åbnet i 1955. "Jeg er meget taknemmelig for at De vil hjælpe mig med at få bearbejdet min neolitiske gravplads Olenii Ostrov ... Deres forslag om en database, det var hvad De kaldte det, ikke?, lyder spændende .. jeg har tænkt på sammenhænge ... relationerne ...".

Havde Anne Sophie hørt rigtigt? Havde Gurina sagt at Olenii Ostrov var en neolitisk gravplads? Og mumlede en smule tøvende Gurinas tale: "Olenii Ostrov er da en mesolitisk gravplads", men kom i det samme i tanke om at i 1956 regnede man stadig denne lokalitet for neolitisk ...

Da Anne Sophie igen tav fortsatte Gurina med at tale om hvilke relationer der skulle være mellem de forskellige entitetstyper: "Genstandene er jo fulgt med de gravlagte .. så der er en relation mellem skeletterne og oldsager .. og de andre genstande". Anne Sophie rynkede panden og sagde: "Mener du at genstandene i graven er gravlagtes personligt udstyr?". Og Gurina svarede: "Ja, i hvert fald er det en markering af individernes status og rolle i samfundet ... og med hele 177 individer på gravpladsen får vi derfor en fantastisk mulighed for at studere

samfundsstrukturen .. jeg vil mene at samfundet har været ret egalitært, men individernes udstyr og den måde de er gravlagt på, må alligevel vise hvilken status de har haft i levende live - måske bestemt af individets køn eller den arbejdsdeling der har været i samfundet ... Nogle af de gravlagte har flere hundrede hængesmykker af elg og bæver tænder, mens andre slet ingen har ... se bare her!". Gurina hev en større kasse ud der havde stået under bordet.

I kassen lå knoglerne fra gravlagte nummer 100 og adskillige æsker med forskellige former for genstande. I en af æskerne lå der mange hundrede hængesmykker lavet af elgtænder, i en anden mere end hundrede hængesmykker af bævertænder og i en lille æske lå der to hjørnetænder fra bjørn. Til gravlagte nummer 100 hørte også flere spydspidser af ben, pilespidser af flint, to harpuner af ben, en flintkniv og tre store flade skifferknive m.m.. Den mest iøjefaldende genstand var dog en specielt flot ornamenteret bendolk med flintæg. Og bemærkelsesværdig var ligeledes seks bæverkæber. "Vi har fået individ nummer 100 kønsbestemt til mand ... Med sit rige udstyr skilte han sig ud allerede under udgravningen ... han var i øvrigt også en af de kun fire individer der var gravlagt i nærmest stående position .. han var måske shaman" Gurina lukkede æskerne og skubbede kassen ind under bordet igen. Rttede sig op og sagde: "Jeg ser frem til at vi med databasen kan få klarhed over om der blandt de gravlagte er andre specielle personer ... måske viser det sig at de gravlagte falder i nogle grupperinger med samme type af udstyr".

For Gurina var målet med at lave databasen - altså hendes ønsker til output - at kunne gøre forespørgelser der viste noget om forskelle og lighed mellem de forskellige individer. Hun var interesseret i samfundsstruktur. Derfor var hendes omdrejningspunkt individet. Allerede i udgravningssituationen havde man fokuseret på individerne og det øvrige materiale fra Olenii Ostrov var blevet knyttet, relateret, hertil. Gurina mente derfor at relationerne i databasen skulle struktureres i en kædefølge: På lokaliteten er der grav(e), i graven er der skelet(ter), og til skelettet knytter sig enten oldsager eller dyreknogler.

Anne Sophie var ikke enig. For en så direkte relation mellem individet - i form af et skelet - og genstandene harmonerede ikke helt med Anne Sophies syn på gravlæggelser og ønsker til output. På udgravningsplanerne kunne Anne Sophie se at både skeletter og genstande på ordentlig vis var indmålt og indtegnet. Og på plantegningerne havde man ligeledes med en streg markeret at de lå i nogle grave - anlæg. Anne Sophie ville hellere have at omdrejningspunktet for databasen var graven - eller anlægget. Så hun syntes at relationen skulle være sådan: På lokaliteten er der anlæg, i anlægget er der skelet(ter), oldsager, dyreknogler m.v..

Anne Sophies forslag til databasedesign afspejler et syn på den indre sammenhæng i materialet som adskiller sig en helt del fra Gurinas. En meget væsentlig forskel er at genstande i Anne

Sophies forslag til design ikke relateres til skelettet, men til anlægget. Anne Sophie opfatter anlægget (som kan være andet end en grav) som det centrale knudepunkt. I anlægget er der simpelthen forskellige muligheder for indhold. Indholdet kan være oldsager, det kan være andre genstande, det kan være okker eller det kan være et skelet - fra et menneske eller måske fra en hund?

Anne Sophie forsøgte derfor at overbevise Gurina om at et databasedesign med anlægget som det centrale knudepunkt ville være bedst: "Naturligvis vil det være vældigt interessant hvis vi kan finde frem til om der er sociale forskelle i den gravlagte befolkning ... Men er du sikker på at genstandene i en grav "tilhører" den gravlagte og afspejler den gravlagtes personlige status? ... hvad hvis genstandene eller nogle af genstandene rent faktisk (i dengangs virkelighed) ikke viser noget som helst om personen? ... Man kunne jo f.eks. godt tænke sig at nogle af genstandene i graven markerer individets sociale status/rang, mens andre er offeringer gjort i forbindelse med gravlæggelsen. Ofringer eller nedlæggelser som ikke er bestemt af den gravlagtes sociale status, men af en helt anden grund... Hvis vi ved hjælp af databasen ønsker at undersøge hvordan det forholder sig, må designet være åbent overfor andre muligheder". Gurina virkede ikke synderligt interesseret i Anne Sophies ideer, så Anne Sophie prøvede et nyt argument: "Nu er der naturligvis på Olenii Ostrov fundet velbevarede og hele menneskeletter i alle gravene, men på andre mesolitiske lokaliteter er der tilfælde hvor skeletterne ikke er bevaret eller fundet - og det føles altså lidt dumt at knytte genstande til et imaginært skelet. Det er af samme grund at jeg mener at entitetstypen ikke skal hedde grav men anlæg. For hvem kan i øvrigt være sikker på at sådanne anlæg uden skeletter er grave! ... I Databasedesignet skal der være mulighed for at undersøge alle anlæg for deres indhold og heraf slutte, hvilke der skal tolkes som grave ... hvilke der skal tolkes som offergruber og hvilke der skal tolkes som affaldsgruber fra en bosættelse ... Og... og .. hvad hvis skelettet (den gravlagte) ikke er et menneske men en hund? Hvordan ser jeg forskel på en hund i en grube og et menneske i en grube – hvem er gravlagt og hvem er ofret?".

Gurina var sunket helt ned bag bunkerne af papir på skrivebordet. Hendes briller sad yderst på næsen mens hun skiftevis kiggede ned på et håndskrevet ark og på tangenterne på sin skrivemaskine hvori der sad flere stykker papir med gennemslagsark imellem. Gurina var travlt beskæftiget og hørte ikke Anne Sophies udredninger.

Anne Sophie måtte erkende at den database hun egentlig helst ville designe ikke alene skulle være til lokaliteten Olenii Ostrov, men en database der kunne håndtere også andre mesolitiske lokaliteter. For hvor Anne Sophie kom fra var det eneste nye ikke kun at Olenii Ostrov var blevet C-14 dateret og placeret i en sen mesolitisk kultur. Nyt var også fund af mange nye lokaliteter med mesolitiske gravlæggelser og andre former for anlæg. Og hvor Anne Sophie

kom fra var det heller ikke længere socialstruktur der var hovedspørgsmålet for de mesolitiske forskere, men ideologi og religion.

Ude er godt, men hjemme er bedst!

Anne Sophie bøjede sig ned, samlede Gurinas publikation op fra gulvet og stillede den tilbage imellem de andre østeuropæiske bøger. Det var ikke i Gurinas publikation hun skulle finde svaret på hvordan databasen skulle designes.

At lave en database der rummer oplysningerne fra og kan anvendes på flere lokaliteter er ikke helt lige til. Da oplysninger fra forskellige lokaliteter kan være meget forskellige. Dels fordi der rent faktisk er forskelle i de levn der findes på forskellige lokaliteter. Dels fordi - heri ligger nok det største problem - den enkelte udgraver og hans valg har betydning for hvad og hvordan levn på pladsen bliver registreret og kategoriseret.

Dårligt udgravede, dårligt dokumenterede og dårligt publicerede lokaliteter vil kun give få og måske endda tvivlsomme oplysninger. Mens godt udgravede, godt dokumenterede og godt publicerede lokaliteter vil give mange og detaljerede oplysninger, hvis troværdighed kan vurderes. Således vil to lokaliteter der ligger forskelligt indenfor denne karakterskala kun kunne sammenlignes på et niveau svarende til den på karakterskalaen dårligst beliggende.

En database beregnet på input fra flere lokaliteter vil bedst kunne laves ved at forskeren gennemgår materialet fra alle lokaliteter. Denne gennemgang skal sigte mod at skabe en fælles kategorisering som basis for et fælles databasedesign, der opfylder forskerens behov - hans ønsker til output som er afhængig af den problemstilling han har valgt.

Og hvad kan vi så lære af denne fortælling?

I denne fortælling var Anne Sophies problemer med at skabe en database og hendes diskussioner med Gurina måske nok så banale og simple, men alligevel illustrere de hvor vanskeligt det er i et databasedesign at tage højde for andres krav til en databasen. Gurina var ikke i publikationen fra 1956 i stand til at se hverken hvilke nye syn man i fremtiden ville få på et mesolitisk gravmateriale eller hvilke nye spørgsmål der ville blive stillet til det. Og hun kunne derfor ikke forudse hverken hvilke former for sammenhænge, relationer, senere forskere ville se i materialet eller hvilke oplysninger de ville ønske. Hendes omhyggelige og grundige udgravning og dokumentation af Olenii Ostrov giver dog gode muligheder for at bruge materialet på nye måder - der ligger så mange og temmeligt detaljerede oplysninger i materialet at der kan udtrages nye oplysninger af det. Men havde hun for 45 år siden skabt en database, ville den ikke kunne bruges i dag. Vi ville for det første mangle oplysninger. Disse ville dog

ofte kunne tilføjes som nye felter eller nye entitetstyper. Men vi ville for det andet se helt andre relationer mellem entitetstyperne, og det er især på dette punkt at det ville blive svært at bruge hendes databasedesign. Andre relationer mellem entitetstyperne vil sandsynligvis samtidigt medføre et andet indhold i entitetstyperne - og så er vi allerede i fuld gang med at skabe en ny database med et helt andet design.

Denne historie viser at dét at designe en database - i hvert fald hvis det er med forskning for øje - i sig selv er en del af den videnskabelige bearbejdning af materialet. For bearbejdning og tolkning er nødvendig for at lave databasen. Og databasen er faktisk en tolkning af materialet, den repræsenterer forskerens opfattelse af hvad "sandheden" er. Desuden er målet med bearbejdning og tolkning af materialet at skabe et databasedesign, der tilfredsstillter forskerens ønsker til output. For at sige det kort er der tre ting der betinger udseendet af databasedesignet:

- Databaseprogrammets betingelser
- Materialets indre "sande" sammenhæng (input) - og hvad der er "sandt" er jo et spørgsmål om tolkning
- Forskerens spørgsmål til materialet (output)

En forskningdatabase er et redskab udviklet til et bestemt formål. Og da forskeren til alle tider vil ændre sit udgangspunkt og sit formål, må også redskabet ændres. Forskningsdatabasen er en maskine der er rette ind til én bestemt tid - ikke en tidsmaskine, men en ganske bestemt tids maskine. Måske endda en ganske bestemt forskers.

Men hvad så med databaser der ikke specielt er beregnet til forskning, men til andre formål - fx til den nationale administration og registrering af arkæologiske levn? Selv om man til sådanne formål ikke benytter programmet Microsoft Access 97, mener jeg at det er problemer i stil med Anne Sophies og Gurinas som DKC m.fl. har når de designer store altomfattende databaser. Sådanne skal rumme langt flere forskellige oplysninger og skal desuden kunne anvendes af mange hundrede arkæologer - for ikke at tale om forvaltninger i amter og kommuner. Derfor må det være nærmest helt umuligt at designe sådanne databaser så de tilfredsstillte alle de forskellige behov de kommende brugere kan have. Og det må være endnu mere vanskeligt samtidigt at skulle designe en database som set i et længere tidsperspektiv kan opfylde også fremtidens behov. Mon ikke sådanne problemer er grunden til at man endnu venter på DKC's ¹ løsning ude omkring på museerne? Dette ikke sagt for at bremse nogle der vil forsøge, for databaser er nødvendige både i nutiden og fremtiden.

Litteratur

- Gurina, N.N. 1956
Oleneostrovski Mogilnik (Olenii Ostrov gravpladsen),
Materialij i issledovanija po arkeologii SSSR 47,
Moskva - 1956
- Rostock, J. K. 1992
Mesolitiske gravpladser i Nordøsteuropa:
Popovo, Olenii Ostrov og Zvejnieki
Opgave v. Institut for forhistorisk og klassisk arkæologi,
Københavns Universitet, eget tryk 1992
- Rostock, J. K. 1998
Forskningsdatabasen - en tids maskine. To
databasedesign til et mesolitisk gravfelt: Olenii Ostrov
Rapport v. Institut for Humanistisk Informatik,
Københavns Universitet, eget tryk 1998
- Dalby, J. 1994
Database-Håndbogen. Fra design til brug. Den
relationelle database 2.udgave,
Samfundslitteratur 1994
- Kent, W. 1983
A simple Guide to five normal forms in relational
database theory,
Communications of the ACM Volume 26 Number 2.
- Viescas, J.L. 1997
Running Microsoft Access 97
Microsoft Press 1997
- Hasle, P og Braüner, T 1998
Systemudvikling – logos eller mythos ?
Rhetorica Scandinavica, nr.6, 1998,
ISSN 1397-0534. Side 33-45.
- Madsen, Torsten 1989
Databaser og forskning
KARK Nyhedsbrev, nr.4, 1. årgang, 1989.
Moesgård 1990, ISSN 0904-5570. Side 17-20.

Note

1. DKC = Dansk Kulturhistorisk Centralregister