

Datalogi – læren om data

.....
Den anden af fem forelæsninger af **Peter Naur**, professor emeritus fra DIKU, der opstod som Danmarks Radios Rosenkjærfrelæsninger i 1966. Forelæsningerne blev holdt i april og maj 1967 og herefter udgivet under titlen Datamaskinerne og samfundet af Munksgaard, København 1967
.....

I den første af disse forelæsninger stillede jeg Dem i udsigt at besvare spørgsmålet: Hvorfor er de elektroniske datamaskiner, eller datamaterne, som jeg foretrækker at kalde dem, af så væsentlig samfundsmæssig betydning? Jeg gennemgik kort regnemaskinernes historie og skitse-rede den vældige udvikling de har undergået i de sidste årtier, den udvikling der blandt andet har medført at vi i dag betragter disse apparater som datamanipulatorer, og ikke i første række som regnemaskiner.

Data, datarepræsentationer og dataprocesser

I dag skal vi beskæftige os med nogle helt andre grunde til datamaternes enestående betydning. Jeg vil skitsere det der ligger i begreberne *data* og de dertil nøje knyttede *datarepræsentationer* og *dataprocesser*, og forsøge at vise at disse begreber i sig rummer et helt nyt syn på en mangfoldighed af menneskets ytringsformer, både matematikken og mere dagligdags foreteelser som sproget. Gennem deres intime forbindelse med disse almenmenneskelige ting bliver datamaternes samfundsmæssige betydning langt mere forståelig.

For nu med det samme at få det værste overstået, her er den internationale definition af ordet ”data”:

DATA: enhver repræsentation af fakta eller ideer på en formaliseret måde, som kan kommunikeres eller manipuleres ved en eller anden proces.

Denne definition lyder formentlig ret så tam, eller endda tom, men just deri ligger det uhyre omfattende i begrebet. For at give Dem en mere håndgribelig forståelse vil jeg nu gennemgå nogle eksempler og sammenholde dem med definitionen.

Lad os begynde med sædvanlig sprogbrug. Som eksempel har jeg her på papiret foran mig skrevet en sætning. Sætningen læser:

”På væggen hænger en sollysende høstscene”.

Disse tegn på papiret foran mig er ifølge definitionen data. Lad os for at forvise os herom først overveje den første del af definitionen, der siger: ”Enhver repræsentation af fakta eller ideer”. Her er der tale om forholdet mellem sætningen på papiret og visse andre omstændigheder i verden, maleriet, væggen, det at maleriet hænger der, og tanken om maleriets motiv. Det er dette forhold mellem sætningen og virkeligheden, der udtrykkes ved at disse data repræsenterer fakta og ideer. Den anden del af definitionen siger at repræsentationen må være formaliseret på en sådan måde



*Jeg vil ...
forsøge at
vise, at disse
begreber i sig
rummer et
helt nyt syn
på en mang-
foldighed af
menneskets
ytringsformer*

at den kan kommunikeres eller manipuleres ved en eller anden proces. Hvad angår den nævnte proces da kan jeg uden videre nævne en sådan, nemlig den der består i at jeg lader mit blik følge teksten og læser den op for Dem som jeg netop gjorde før. Derigennem har jeg kommunikeret disse data til Dem. Tilbage står kravet om at repræsentationen skal være formaliseret. Her tænkes der på den omstændighed at en tekst som den omtalte består af enkelte tegn, bogstaverne, der holder sig inden for et meget begrænset sæt af former, nemlig de henved tredive bogstaver i alfabetet, og ydermere at små uregelmæssigheder i det enkelte tegns fremtræden på papiret er uden betydning. Denne omstændighed er betingelsen for at man kan tale om en perfekt kommunikation af teksten. Hvis for eksempel en af mine tilhørere har nedskrevet sætningen så kan vi gå ud fra at blot han er fortrolig med dansk retskrivning så vil hans version af sætningen være aldeles den samme som min, bogstav for bogstav.

Definitionen af data gemmer dog på endnu mere, gennem det den *ikke* siger. Den udtaler sig aldeles ikke om nogen bestemt måde data kan eksistere og lader derved alle muligheder stå åbne. Allerede den anførte illustration giver en række eksempler herpå. De primære data findes som tryksværte på papir her foran mig. Alene derigennem at jeg har læst sætningen højt og at den er blevet overført til lytterne gennem radioen er disse samme fakta og ideer blevet repræsenteret på en række andre måder, først som en serie lyde jeg udtaler, derefter som elektriske signaler af flere forskellige former, før den omsider igen omsættes til lyd i Deres højttaler. Hver af disse former kan opfattes som data der repræsenterer de samme fakta og ideer.

Endnu en væsentlig ting skal bemærkes ved den givne definition af data, nemlig den udtrykkelige tale om *processer*. En ting eller begivenhed er ikke i sig selv data, men bliver det først når den indgår i en proces hvori dens repræsentation af fakta og ideer er det afgørende. Hvis jeg brænder et beskrevet stykke papir så underkastes skrifttegnene en forbrændingsproces, men dette gør dem ikke i sig selv til data. Dette er derimod tilfældet hvis en person læser dem. Vi ser heraf at databegrebet beskæftiger sig med en måde som mennesker forholder sig på til visse fænomener.

Kunst og modeller som data

Databegrebet ligger nær op ad symbol- og modelbegreberne. Ordet symbol bruges både om matematiske symboler, hvor det knap kan skelnes fra data, og i kunsten. Brugen af symboler i kunsten er dog en kompliceret sag. Nok er der tale om at én ting på en ejendommelig måde peger på, eller henviser til, en anden, men der gives ikke nødvendigvis et krav om en formalisering som ved databegrebet. Således er der forskel mellem et maleri og data. Maleriet er ikke formaliseret og ingen kopi af det vil regnes for lige så god som originalen.

At formelle processer ikke er kunsten uvedkommende fremgår på den anden side af forbindelsen mellem litteraturen og bogtrykkunsten, som må betragtes som en tidlig triumf for mekanisk databehandling. Ved at fremdrage dette eksempel håber jeg også at berolige dem der føler sig uhyggeligt til mode ved de moderne datamater. Vi har alle for længst vænnet os til at omgås særdeles personligt med data som er frembragt gennem mekaniske processer. Der er ingen der føler sig stødt af at kærlighedsdigte er reproduceret ved hjælp af ganske følelsesløse trykkemaskiner. Af og til kan man dog iagttage at der måske er en grænse for hvor vidt vi er parat til at acceptere

dette upersonlige mellemed. Jeg selv føler mig snydt når jeg modtager en julehilsen der er trykt og hvor afsenderen ikke engang har umaget sig til at tilføje sin egenhændige underskrift. Fornemmelsen kan dog også vendes om. Nutildags regnes et festtelegram jo for en fuldt ud passende form for hyldest, næsten bedre end et brev, til trods for at telegramteksten som bekendt har været udsat for en langt mere omfattende håndtering af uvedkommende personer og maskiner.

Databegrebets forbindelse til modelbegrebet er måske ikke ved første øjekast indlysende, da man ved en model normalt tænker på en ting som i sin fremtoning har stærk lighed med det den er en model af, som for eksempel ved modelfly. Modelbegrebet er imidlertid for længst overtaget i de empiriske videnskaber, fysikken, astronomien, biologien, og mange andre, og bruges der om langt mere abstrakte konstruktioner. I astronomien taler man for eksempel om stjernemodeller, der består af kurver eller tabeller, altså data.

Data som værktøj

Men hvis databegrebet på alle disse områder er foregrebet eller indeholdt i disse kendte begreber, hvad bringer det da af nyt, må vi spørge. Svaret herpå ligger allerede i det sagte, på den ene side i den ubegrænsede frihed til som data at vælge hvilken art fænomener som helst, lyde, tryksværte på papir, elektriske strømme, magnetfelter, osv., dels i interessen for processerne. I denne sidste interesse ligger noget aktivt. Data er ikke noget der eksisterer eller som man har, det er noget man bruger, et værktøj for mennesker i deres virksomhed.

Denne opfattelse kan umiddelbart bekræftes ved en overvejelse af sproget. For så vidt sproget anvendes til kommunikation af fakta og ideer mellem mennesker, og ikke som medium for fri fabuleringen, da tjener sproget åbenbart som hjælpemiddel, værktøj, - og er vel endda det vigtigste menneskelige værktøj overhovedet.

Men med en helt fri opfattelse af data åbner vi vejen for en langt mere effektiv udnyttelse af dem. Sagen er at den lethed hvormed vi kan gennemføre en given dataproces i høj grad afhænger af præcis hvilke data vi har valgt til at repræsentere virkeligheden, eller som jeg kort vil sige, af *datarepræsentationen*. Dette forhold er velkendt fra sædvanlig brug af de sproglige repræsentationer, tale og skrift. Blandt de processer der her er tale om kan nævnes adskillige, for eksempel: at udveksle ideer blandt personer; give en fuldstændig beskrivelse af et forhold; at registrere nu tilgængelige fakta for at undgå at de fortabes; at udbrede en enkelt persons ideer til mange. Som enhver ved vælger man sin sproglige form, tale eller skrift, under hensyn til hvorledes processen vil forløbe mest effektivt. Til at udveksle ideer bruger man helst samtale, fordi den skriftlige form er for tung og langsom. Fuldstændige beskrivelser udformes bedst skriftligt, fordi dette er mere fordelagtigt for den der skal bruge dem. Til at registrere tilgængelige fakta bruger mange af os skriftlige notater, men det afhænger af vor indbyggede procesformåen og nogle af os opnår en sikrere registrering ved at bruge hukommelsen. Til at udbrede en enkelt persons ideer til mange benytter man i ikke helt simple tilfælde en kombination af tale og skrift, som vi for eksempel stadig kan iagttage det i arbejdet for at udbrede politiske ideer.

Den helt fri opfattelse af data har i praksis for længst slået sig igennem hvad angår den dataproces der hedder kommunikation. Vi benytter frit en blanding af de gammelkendte repræsen-

tationer, tale og skrift, og en mangfoldighed af andre, telegrafi, telefoni, radio- og fjernsyns-overføring, idet vi i hver livssituation vurderer den indbyrdes fordel af disse i forhold til den kommunikationsproces vi har brug for at få udført.

På området datalagring, som kan opfattes som den proces at kommunikere med en forsinkelse af ubekendt varighed, har en voksende frigørelse fra sædvanlig nedskrivning ligeledes fundet sted med fremkomsten af båndoptagere.

Fra datalogiens synspunkt må denne udvikling opfattes blot som en begyndelse, da den kun berører de allersimpleste dataprocesser. Fremkomsten af datamaterne betyder at langt mere komplicerede dataprocesser meget effektivt kan udføres uden at mennesker behøver at deltage. Vejen er derved blevet åbnet for at arbejdet med datamodeller bliver i høj grad automatiseret og effektiviseret.

Datamodeller i samfundslivet

Betydningen af datamodeller for samfundslivet kan næppe overvurderes. De er grundlaget for enhver planlægning, fra forretningsmandens tabeller og kurver over den erhvervsmæssige udvikling, over ingeniørens arbejde med beregninger og tegninger der beskriver den bro han er ved at konstruere, til fysikerens matematiske teori for atomkernen. På hvert af disse områder tillader arbejdet med data at man skaffer sig viden om hvordan den virkelige verden vil opføre sig under forskellige omstændigheder. Datamodellerne tillader at vi eksperimenterer med virkeligheden, men uden at vi behøver opføre becostelige eksperimentopstillinger og uden at tabe ret meget hvis eksperimentet mislykkes. Værdien heraf er især slående ved planlægningen af ting som skibe, flyvemaskiner, broer, og atomreaktorer, hvor et fejlslagent virkeligt eksperiment øjensynlig kan være katastrofalt. Derimod betyder det ikke alverden om det ved en beregning viser sig at en påtænkt reaktorkonstruktion vil eksplodere.

Hvad er nu betydningen af det datalogiske synspunkt for denne brug af modeller? Her er der to ting at nævne. For det første befrier datalogien os for fordomme om at arbejdet med en bestemt, given problemstilling er knyttet til en bestemt datarepræsentation. Sådanne fordomme er uhyre udbredte. For eksempel arbejder ingeniører og arkitekter traditionelt med tegninger af deres konstruktioner som primære data. Det vil umiddelbart vække modstand hvis man foreslår at denne form måske slet ikke bør være enerådende. Ikke desto mindre er det en kendsgerning at der nu til dags i betydeligt omfang foregår udviklingen af komplicerede konstruktioner, som vejanlæg og skibe, uden nævneværdig brug af tegninger.

Hermed er ikke sagt at den hidtidige brug af tegninger har været en datalogisk fejltagelse. Hvad datalogien siger, er, at datarepræsentationen må afpasses efter den transformation der skal udføres og det procesværktøj der er til rådighed. Indtil de moderne datamaters fremkomst omkring 1945 var arkitekter og ingeniører i det væsentlige henvist til at arbejde kun med mennesker som procesværktøj. I relation til den måde vi mennesker arbejder med hovedet og til vor synsmekanik er tegninger en særdeles velvalgt datarepræsentation.



Datamodellerne tillader at vi eksperimenterer med virkeligheden

Den anden værdi af datalogien for modelbyggeren ligger i selve opbygningen af en datamodel og i beskrivelsen af alle detaljer i de processer der indgår i dens brug. Den nøje formulering af en datamodel virker i høj grad stimulerende for forståelsen af enhver art problem, stimulerende på en måde som må opleves for at forstås. Denne dybere forståelse af problemerne er slet ikke knyttet kun til mere komplicerede problemer. En betydelig del af gevinsten kan også høstes uden at man kommer i berøring med datamater. Det afgørende er oplevelsen af hvordan dataprocesser kan udvikles, beskrives og bearbejdes.

Det fremgår af disse betragtninger at datalogien ikke har nogen nødvendig forbindelse med datamaterne og at den datalogiske betragtningsmåde indeholder meget gammelkendt, som blot ikke hidtil har været samlet under samme synspunkt. Datamaterne har dog det med sagen at gøre at de ved deres fremkomst fremtvinger en ny og velfunderet datalogisk stillingtagen til problemernes løsning. Den hurtighed, sikkerhed og prisbillighed hvormed nutidens datamater gennemfører komplicerede dataprocesser betyder et så afgørende spring sammenlignet med hvad der hidtil har været kendt, at enhver menneskelig aktivitet der i nævneværdig grad har brug for dataprocesser må blive påvirket deraf. Vanskeligheden ved overgangen beror i vid udstrækning på at dyrkerne af de mange specialfag slet ikke er vant til at se deres fag på den led som situationen kræver. Den ingeniør der bygger broer er vant til at vurdere konsekvenserne af nye materialer og nye konstruktive ideer, men ikke i nær så høj grad af nye beskrivelsesformer, nye datapræsentationer. Samtidig griber en ændring i den datarepræsentation der benyttes under projekteringen så dybt ind i arbejdet at den ikke kan overlades til fremmede eksperter alene, men kræver meget aktiv medvirken af det enkelte fags dyrkere.

Dertil kommer at datamaterne selv undergår en udvikling, således at tilpasningen af de enkelte fags problemer til dette værktøj ingenlunde er noget der kan klares én gang for alle.

Datalogi i uddannelsen

Har man således indset at datalogien på den ene side sammenfatter en lang række centrale menneskelige aktiviteter og begrebsdannelser under ét samlende synspunkt, og på den anden side formår at befrugte og forny tankegangen i en lige så lang række fag, da kan man ikke være i tvivl om at datalogien må have en plads i almenuddannelsen. For at nå til en rimelig forestilling om hvordan denne placering bør være er det naturligt at sammenligne med fag af lignende karakter. Man vil da nå frem til sproglære og matematik, som er de nærmeste analoge. Både datalogien og disse to fagområder beskæftiger sig med tegn og symboler der er opfundet af mennesker som hjælpemidler. Fælles for de tre emner er også deres karakter af redskaber for mange andre fag. I uddannelsen må de derfor indgå på to måder, dels som hjælpefag ved studier af mange andre fag, dels som hovedfag ved uddannelsen af specialister i selve disse emner. Vi har jo alle måttet gennemgå meget betydelige mængder sprog, regning og matematik i skolen, uanset at kun ganske få af os er blevet lingvister eller matematikere. På lignende måde må datalogien bringes ind i skoleundervisningen og forberede os alle på i tilværelsen i datamaternes tidsalder, ganske som læsning og skrivning anses som en nødvendig forudsætning for tilværelsen i et samfund der er præget af tryksager.

Datalogien kan tænkes placeret i skoleundervisningen, enten som et selvstændigt fag, eller som en del af matematikken. Det afgørende er indholdet af undervisningen. Hovedtemaerne må være

data, datarepræsentationer, og dataprocesser. Disse fundamentale begreber må belyses gennem en række konkrete eksempler som bekvemt kan tages fra områder inden for skoleelevernes erfaringsområde, talregning, stavning, opslag i registre, og lignende. Der bør også indgå simple eksempler på brugen af mekaniske og elektriske fænomener til datarepræsentation. Disse eksempler bør belyses ved simple forsøg. Datamaterne bør også omtales, men ikke som det centrale i faget, snarere som en afsluttende orientering.

Når datalogien er blevet vel etableret i almenuddannelsen vil den mystik der omgiver datamaterne i manges forestillinger opløse sig i intet. Dette må betragtes som den måske allervigtigste begrundelse for at fremme forståelsen af datalogien. Det vil nemlig være betingelsen for at herredømmet over datamaterne og deres anvendelse ikke bliver en sag for en lille gruppe af eksperter, men bliver en sædvanlig politisk sag, og således gennem det politiske system kommer til at ligge hvor det bør, hos os alle. ❖

”
Når datalogien er blevet vel etableret i almenuddannelsen, vil den mystik der omgiver datamaterne i manges forestillinger opløse sig i intet

PETER NAUR



Peter Naur har været professor emeritus fra DIKU siden 1998. Han er uddannet astronom og erhvervede doktorgraden fra Københavns Universitet i 1957. Snart efter sin ansættelse ved Regnecentralen i 1959 slog han sit navn fast i den internationalt gryende aktivitet omkring datamaskinerne ved sine bidrag til

det internationale programmeringssprog ALGOL 60. I de følgende år førte hans overvejelser over Regnecentralens ALGOL-kompilatorer til, at han i 1966 formulerede en akademisk disciplin under betegnelsen datalogi, overvejelser som i 1972 blev publiceret under titlen "Concise Survey of Computer Methods".

Dette førte til, at Københavns Universitet i 1970 oprettede Datalogisk Institut, hvortil han blev kaldt som professor. Som sådan fortsatte han i de følgende år sine studier over programmeringens problemer med stigende opmærksomhed på at forstå den menneskelige tankevirksomhed i alle dens manifestationer. En lang række af

hans arbejder på dette område blev i 1992 udgivet samlet under titlen "Computing – A Human Activity". Hans fortsatte studier førte derefter til udgivelsen af en række bøger: "Knowing and the Mystique of Logic and Rules" (1995), "Antifilosofisk leksikon: Tænkning – sproglighed – videnskabelighed" (1999), "Antiphilosophical Dictionary: Thinking – Speech – Science/Scholarship" (2001), "Psykologi i videnskabelig rekonstruktion" (2002), "An Anatomy of Human Mental Life – Psychology in Unideological Reconstruction – Incorporating the Synapse-State Theory of Mental Life" (2005) og "The Neural Embodiment of Mental Life by the Synapse-State Theory" (2008). I disse bøger fremlægges en lang række originale indsigter om menneskelig tænkning og den måde, tænkning foregår på i menneskers hjerner og nervesystemer.

Peter Naur blev tildelt G.A. Hagemann Medaillen i 1963, Jens Rosenkjær Prisen i 1966 og Computer Pioneer Award of the IEEE Computer Society i 1986. I 2005 blev han tildelt computerverdens mest prestigefyldte hæderspris, Association for Computing Machinerys A. M. Turing Award.

Den digitale revolution – fortællinger fra datalogiens verden

Bogen er udgivet af Datalogisk Institut, Københavns Universitet (DIKU) i anledning af instituttets 40 års jubilæum med bidrag fra forskere tilknyttet instituttet.

Redaktion:

Tariq Andersen, phd-studerende, Jørgen Bansler, professor, Hasse Clausen, lektor, Inge Hviid Jensen, kommunikationsmedarbejder og Martin Zachariassen, institutleder.

Forsidemotiv: Foto af skulptur af Alan Turing, © basegreen lokaliseret på flickr.com/photos/basegreen

Oplag: 1000 eks.

Grafisk design og produktion: Westring + Welling A/S

ISBN: 978-87-981270-5-5

© Datalogisk Institut 2010. Citater er tilladt under creative commons.

