

Perspektiver for data warehouse i organisationen

Af Karsten Boye Rasmussen

Resumé

Puff! Ideen om en tryllestav er tiltalende. Kan vi med et slag gøre virksomheden mere profitabel og opnå mere tilfredshed blandt medarbejdere, kunder og leverandører? Det kan forekomme naivt at tro, at tingene skulle kunne gøres så enkelt, men omvendt er der vel ingen grund til unødigt at komplicere simple løsninger; hvis de findes! Kan vi ændre virksomheden ved anvendelsen af nye tiltag for ledelse eller implementering af ny software? "Data warehouse" er et sådant tryllestavsbegreb og markedsføres af flere forskellige virksomheder som software, hardware og konsulenttjenester, der leverer løsninger på virksomhedens problemer.

Denne artikel beskriver den teoretiske baggrund for en undersøgelse af data warehouses indførelse i virksomheder. Artiklen betragter gennem litteraturstudie forskellige teorier med relevans for data warehouse. Udgangspunktet

er teorien om den informationsbehandlende organisation, hvor nogle centrale parametre findes og dernæst genfindes inden for såvel organisationsteori som teori om systemudvikling. Diskussion omkring disse parametre fører til en beskrivelse af centrale dimensioner for anskuelser af data warehouse. Det er tanken, at disse dimensioner efter sammenstilling til nogle hypoteser om succesfuld indførelse af data warehouse vil blive afprøvet i forbindelse med større virksomheder.

Succes for data warehouse opnås gennem dygtig håndtering af væsentlige organisatoriske dimensioner: 1) eksploration/eksploitation 2) centralisering/decentralisering og 3) usikkerhed/kompleksitet. Men data warehouse er ingen tryllestav. I stedet må succesfuld udvikling af data warehouse opnås gennem succesfuld systemudvikling.

Indledning¹

Indførelse af informationssystemer i virksomheder er sket med forskellig grad af succes. Specielt har offentlige IT-fiaskoer været spektakulære ved deres intense omtale i pressen, mens de samme ulykker nok i nogen udstrækning har været bedre skjult i privatejede firmaer. Men hvad er succes? Opmærksomhed overfor de menneskelige aspekter ved indførelse af teknologi har siden Tavistock-forskningen i 1950'erne ført til, at de (informations-) teknologiske

systemer betragtes som havende både en teknisk og en organisational dimension, som må inddrages for at vurdere informationssystemet.

Systemerne er blevet stadig mere omfattende. Med data warehouse og de tilgrundliggende informationssystemer er der nået det hidtidige kvantitativt højeste niveau af informationssystemer i virksomheden; og yderligere forventes omfanget af informationsopsamling, opbevaring og udbredel-

se fortsat at stige. Desuden er virksomhederne blevet stadig mere sårbare overfor, om informationssystemerne løbende faktisk leverer de ønskede fordele. Uden fordelingen - ingen virksomhed!

Da data warehouse design i sin tekniske udformning allerede anvender termen dimension som en vigtig komponenttype ved design, kan der i denne kontekst optræde en dobbelthed i anvendelsen af termen dimension. For at undgå den tekniske biklang er artiklens titel ikke "Dimensioner i data warehouse i organisationen". I stedet udspejler de begrebsmæssige dimensioner billedligt talt et rum, hvorfra punkter giver "Perspektiver for data warehouse i organisationen". Rummet defineres af de organisatoriske og systemudviklingsmæssige dimensioner: 1) eksploration/eksploitation 2) centralisering/decentralisering og 3) usikkerhed/kompleksitet. De tre dimensioner kan alle ses som balancer. Snarere end et valg mellem ekstremer er det for virksomheden et spørgsmål om at balancere og relatere dimensionerne.

Det organisationsteoretiske fundament

I behandlingen af informationssystemer i virksomheden er det ligetil at vælge teorien om organisationen som informationsbehandlerende enhed som udgangspunkt. Galbraith (1977, s. 3) foretager en introducerende definition af en organisation som "(1) composed of people and groups of people (2) in order to achieve some shared purpose (3) through a division of labor (4) integrated by information-based decision processes (5) continuously through time." Senere i teksten argumenterer Galbraith (1977, s. 30) for en stjernemodel med relationer mellem elementerne: Opgave, struktur, informationsteknologi, personer og belønningssystem. Hermed ligner modellen Leavitts diamant-model fra 1965 med elementerne: Opgave, struktur, personer og teknologi. Herudfra argumenterer Leavitt for, at succesfuld teknologisk forandring ikke kan opnås uden simultan forandring af de

andre elementer af diamant-modellen (se fx Laudon og Laudon, 2000, s. 90). Når virksomhedens informationsbehandling forandres, opnås også en organisationsforandring.

Målet for den oprindelige rationelle beslutningsproces bestod i at udvælge den bedste af alle mulige løsninger. Herbert Simon og andre fra "the Carnegie-Mellon School" eksemplificerede, at overfor det økonomiske "optimerende" menneske stod det administrative menneske, som pga. sine begrænsninger blev "satisficerende". Som bedste løsning vælges løsningen, der er "god nok" (Simon, 1997, s. 199). Godt nok er bedst! På trods af denne fremsættelse og beskrivelse af menneskets begrænsede muligheder for at bearbejde information, vendes der ofte tilbage til den fuldt rationelle model. Og vi tror ofte, at teknik vil kunne hjælpe os dertil. Således var de tidlige epoker af informationsteknologiens fremfærd også mættet med en entusiasme om at kunne omgå individets begrænsninger gennem anvendelse af computere og matematik (Sproull og Kiesler, 1991, s. 143). Hvad enten den rationelle drøm kan virkeliggøres eller ej, bygger beslutningsprocessen på information, og i en tør argumentation anfører Galbraith, at implementeringen af beslutningsstøtte afhænger af indsamlingen af data, og at denne proces vil føre til en formalisering af data.

Galbraith benytter begreberne data og information som synonyme. I opposition til traditionel systemudvikling har Peter Checkland i mere end 20 år arbejdet med en blødere tilgang til systemstemudvikling. Således kaldes hans metode "Soft Systems Development" (SSD). I samarbejde med Sue Holwell tager Checkland udgangspunkt i teorier om systemudvikling og graduerer og raffinerer definitionerne omkring data og information. Således er information i deres teori defineret som "meaningful facts", dvs. data, der gives mening gennem kontekst. I denne begrebsgørelse betragtes data som et univers, hvorfra kun en brøkdel er udvalgt (eller skabt). Den udvalgte del af data kal-

des "capta". Disse "grebne" data bringes til informationssystemet, og kan sættes i sammenhæng og blive til information. "Capta" og processen omkring udvælgelse kan relateres til Simons første trin i beslutningsfølgen², hvor hans begreb "intelligence" samtidigt omfatter indhentning af oplysninger og identifikation af problemer. Sluttelig kan information blive til viden gennem akkumulation og strukturering; viden er da "larger, longer living structures of meaningful facts" (Checkland og Holwell, 1997, s. 90). Er viden i data warehouse også sandhed? Data warehouse kan ses omtalt som "a single version of truth" (Agosta, 2000, s. 15). Negativt betragtet lyder det som en fastlåsning (måske som et produkt fra Orwells Ministry of Truth i romanen "1984"), men i positiv forstand betyder det, at informationen i data warehouse giver et konsistent syn. Denne konsistens er skabt gennem "vask, rensning og skrubning" af data, og denne sammenhæng opnås ud fra metadata, som leverer beskrivelse af "extraction, transformation, loading, meaning and scheduling of the information supply chain" (Agosta, 2000, s. 316). I vidensproduktionen leverer metadata således kontekst for, at capta kan løftes til information, og at videre strukturering kan løfte informationen til viden.

Selektionen af data og struktureringen har afgørende indflydelse på hvorvidt informationssystemet vil blive en succes. Denne kontekst er i Galbraiths model indbefattet i hans begreb om "database scope", og Galbraith er yderligere beskæftiget med informationssystemets friskhed og tilgængelighed, hvad der kaldes systemet "frequency". Galbraith har således introduceret nogle centrale dimensioner til beskrivelsen af informationssystemer: Formalisering, rækkevidde og frekvens. Heroverfor understreger Checkland og Holwell udvælgelsen og dannelse af mening i forbindelse med informationssystemerne. Udvalget er central. Der er ikke længere knaphed på information. Informationsrevolutionen har mangedoblet den mængde information, en person kan sprede i en organisation, men

antallet af timer personen kan fordøje information er ikke blevet forøget (Simon, 1997, s. 23).

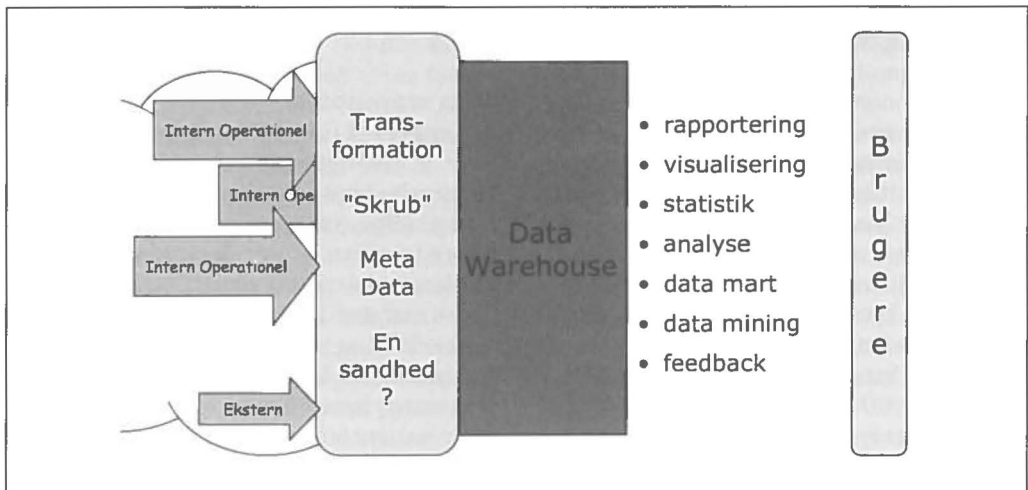
Data warehouse

Der kræves et udvælgelsesarbejde for at sikre, at den rigtige information fordøjes. Virksomhedens operative databaser er mangfoldige, men for data warehouse sker der en udvælgelse (evt. flere udvælgelser for flere problemstillinger). I data warehouse sammenhæng kan begreberne om rækkevidde og udvælgelse kombineres til et spørgsmål om det grundlæggende design af data warehouse. Således efterlades fra organisationsteorien bevarelsen af spørgsmål omkring: Formalisering af data, dannelse af mening og design af data warehouse.

Et data warehouse tjener opgaven at være lager for data. Virksomheder omgives af stadig større strømme af data spændende fra ordreinformation ved kontakt med leverandører til opsamling af kundernes brug af virksomhedens web-sted³. I en input-output anskuelse må der således både være overvejelser omkring produktion (udvælgelse og indhentning) af data, lagring og sluttelig distribution af disse data. Centrale processer og funktioner i forbindelse med data warehouse illustreres på næste side (figur 1).

Figuren viser en datafødsel-side hovedsageligt bestående af input fra virksomhedens operationelle systemer. Langt hovedparten af indholdet i data warehouse findes altså i forvejen i organisationen. I praksis er disse data utilgængelige for direkte udnyttelse, idet data findes dybt begravet i et virvar af database-tabeller i transaktionssystemer. Transaktionssystemerne er optimeret med henblik på operationel hastighed med hurtige svartider og er rettet mod her-og-nu anvendelser og ser bort fra flere forhold, eksempelvis: Historik (som er central for gennem analyse at opnå viden om virksomhedens bevægelser) og data kvalitet (som enkeltstående i transaktionssystemet kan være tilstrækkelig, men som vil udløse

Fig. 1. Data warehouse



konsistensproblemer ved integration fra flere kilder i data warehouse).

Data warehouse understøtter output angivet på højre side af figuren som en lang række applikationer. Data warehouse er en infrastruktur og stiller data til rådighed for applikationer. Et data warehouse adskilles fra transaktionssystemet, idet den uregelmæssige brug og det tunge CPU-forbrug kendetegner et data warehouse, og disse topbelastninger ville ødelægge optimeringen af transaktionssystemet.

Design af data warehouse hviler som udgangspunkt på en begrebsmæssig transition gennem omformning af virksomhedskrav over informationsbehov til overvejelser om, hvorledes disse behov vil kunne opfyldes (Welbrock, 1998, s. 23). Som ved andre indførelser af forandringer er der en dobbelt gevinst: 1) at nå målet og 2) at gennemløbe processen. Informationssystemet ses således som et middel, når Welbrock bemærker, at afklaringen af beslutninger omkring design vil forstærke forståelsen af virksomhedens forretningsstrategi, og at den begrebsmæssige fase således vil vitalisere adskillige dele indenfor virksomheden.

Den følgende definition af data warehouse findes hos en af grundlæggerne af data warehouse:

"a data warehouse is a subject oriented, integrated, non-volatile, and time variant collection of data in support of management's decisions" (Inmon, 1996, s. 33).

Udover Bill Inmon er Ralph Kimball en anden af grundlæggerne ("guruerne") indenfor data warehouse. Kimball bemærker: "The correct job description for a data warehouse manager as publisher of the right data" (Kimball, 2002, s. 5). Så også hos Kimball er der stort fokus på data warehouse som leverandør af støtte for beslutninger. Data warehouse er dog ikke udelukkende leverandør til ledelsesbeslutninger. Der kan også være tale om informationer, der efter frembringelse tilbageføres til det operationelle system. Et ofte anvendt eksempel er net-boghandlen Amazon.com, hvor kunden selv kan træffe beslutninger på grundlag af information fra tidligere kunder (en sammenstilling af "collaboration" og "personalization" (Keen, 2001, s. 178)). Kunden udvælger en bog, og Amazon viser, at andre kunder, foruden den angivne bog, har købt de bøger, der yderligere fremvises. Kunderne danner således et C2C-netværk, der forbinder bøgerne. Det er bl.a. denne type anvendelse, der dækkes af "feedback" i den grafiske model ovenfor. Desuden kan avan-

cerede analyser gennem data mining udpege de overraskende og få situationer i forhold til de mange og velkendte mønstre (Rasmussen, 2001). Den menneskelige informationsforøgelse kan da reserveres til en intensiv betragtning af det langt mindre antal undtagelser.

Udvælgelse betyder, at et data warehouse er langt fra et passivt informationslager.

Kimball skriver om data warehouse:

"The data warehouse provides access to consistent organizational data that can be combined for query, analysis, and presentation of published data with a quality that will act as a driver of business reengineering.

(Kimball, 1996, samlet fra s. xxiii-xxv).

Tætheden af de positivt farvede "buzz-words" (her kursiverede) er overordentlig høj i denne sætning. Nogle af problemerne i forbindelse med etablering af data warehouse kan skyldes intense og meget høje aspirationskrav, der medfører overspændte forventninger til data warehouse som trykkestav.

Begreberne konsistens, integrerede, ikke-flygtigt i citatet ovenfor kan opfattes som

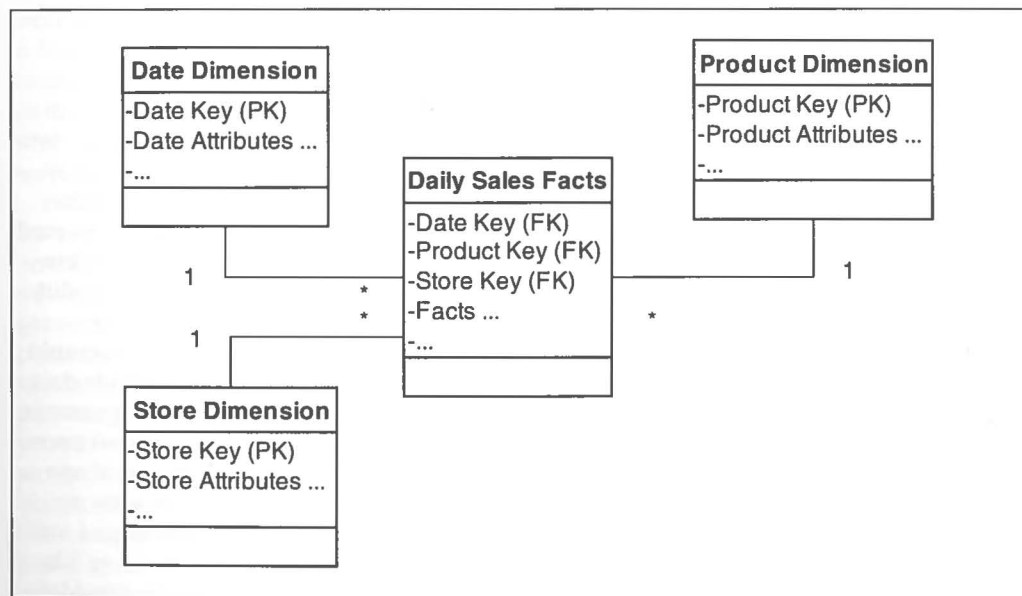
karakteriserende applikationer, der er formaliserede, fastlåste og kun vanskeligt kan ændres. Imidlertid er det centralt for designet af data warehouse, at det netop fanger de centrale beskrivende dimensioner og knytter disse til tællelige facts uden at fastlægge de faktiske applikationer, der benytter disse oplysninger. Eksemplet kan være et salg indeholdende nogle få facts (fx pris og antal) som i øvrigt er forbundet med få dimensioner som produkt (hvad), marked (hvor) og tidspunkt (hvornår) (Kimball, 2002, s. 10). (Se fig. 2).

De tre dimensioner kan visualiseres til en kubus med placering af den atomare enhed: Det enkelte salg med angivelse af hvad, hvor og hvornår. Herudfra vil der uden begrænsninger kunne understøttes mængder af applikationer. Med "slice-and-dice" kan der udskæres større aggregerede områder, hvilket giver et intuitiv brugerinterface.

Usikkerhed og kompleksitet i organisationen

I teorien om den informationsbehandlende

Fig. 2. Facts og dimensioner i stjerne-skema



organisation gives usikkerhed en afgørende plads: "Uncertainty is the difference between the amount of information required to perform the task and the amount of information already possessed by the organization" (Galbraith, 1977, s. 37). Usikkerhed udtrykker mængden af manglende information. Vi er usikre pga. det, vi ved, vi mangler. Videre angives "the greater task uncertainty, the greater the amount of information that must be processed among decision makers during task execution in order to achieve a given level of performance". Jo mindre formaliseret og fastlagt opgaven er, des mere information må skubbes opad i hierarkiet for at opnå en afgørende beslutning. Galbraith ser således usikkerheden som en hel central parameter i organisationen og udnævner usikkerhed til "the core concept upon which the organization design framework is based".

Heroverfor stiller Galbraith opgavens kompleksitet som "the amount of information to be processed in decision making". Kompleksitet anskues som "a function of the division of labor, the goal diversity, and the level of goal performance required". (Galbraith, 1977, 39). Organisationen vil således være indstillet på at indhente yderligere information, men organisationen må balancere dette overfor sine begrænsede muligheder for at behandle informationen. Således bliver balancen mellem usikkerhed og kompleksitet til et dilemma for organisationen, som er i stor risiko for at overskride sin begrænsede kapacitet ved at blive overfyldt af information. Det er umuligt at finde værdifuld information uden at støde på værdiløs information. (Sproull og Kiesler, 1991, s. 170).

Kompleksitet fremstår hos Galbraith med en simplificeret sammenhæng: Når der kræves meget information, er det en kompleks opgave! Et mere nuanceret syn findes hos Axelrod og Cohen, som ved introduktionen af kompleksitet bemærker, at "complexity does not simply denote many moving parts" (1999, s. 15). Kvantiteten er altså ikke en tilstrækkelig grund. Axelrod

og Cohen mener, at kompleksitet kræver opstående egenskaber ("emergent properties"), dvs. nye egenskaber, som ikke fandtes i de adskilte og enestående dele. Eksemplet er "No single neuron has consciousness". Et system med mange bevægelige dele vil Axelrod og Cohen benævne som kompliceret, men ikke som komplekst. Dette kan forekomme som en væsentlig skelnen, men det er en noget usikker skelnen, idet den er subjektivt baseret eller kræver ekspertusikkerhed. Det kan argumenteres, at såfremt det er muligt at udlede systemets egenskaber fra systemets dele, er det et kompliceret system. Men for en mindre kyndig person – som ikke er i stand til at foretage udledningen – vil systemet fremtræde som komplekst. Selv når eksperterne fastslår, at noget faktisk er komplekst, kan der senere komme andre mere vidende eksperter, der kan reducere kompleksiteten til komplicerethed.

Begrebet usikkerhed kan ligeledes raffineres. Galbraith skriver senere, at en mere komplet model ikke blot kan betragte usikkerhed internt i organisationen, men også må omfatte de eksterne omgivelser (Galbraith, 1994, s. 29). En sådan systematisk undersøgelse findes i modellen hos Burton og Obel (1998, s. 174), hvor omgivelsernes usikkerhed opdeles i "uncertainty" (hvor værdierne af variablene er ukendte) og "equivocality" (hvor variable er ukendte eller ignoreres). I denne sammenhæng defineres omgivelsernes kompleksitet ved antallet af relevante variable og deres sammenhæng. Når kompleksitet forbindes med usikkerhed, er der igen tale om en renere kvantitativ betragtning, mens equivocality må betragtes som et kvalitativt tiltag.

Fra et informationssystemisk synspunkt er det interessant om det succesfulde data warehouse kan lægge grunden til systemer, der fungerer som intelligent sentinel overfor usikkerhed, som fx opkomsten af nye væsentlige variable. Gennem analyse og "exception reports" kan forandringer i omgivelserne nok overvåges, men det er ikke tænkeligt, at nye variable direkte kan iden-

tificeres. Så resultatet vil være som for Dylans Mr. Jones: "You know something is happening, but you don't know what it is"⁵. Men denne advarsel kan tjene til at skærpe opmærksomheden og kreativiteten overfor, at nye variable må indhentes.

De fremlagte definitioner af disse dimensioner er ikke helt overensstemmende, men på den anden side synes de forskellige indgange klart at understrege betydning af de centrale dimensioner i organisationsteori- en:

- Usikkerhed
- Komplexitet

Usikkerhed og kompleksitet i systemudvikling

Begreberne usikkerhed og kompleksitet er ikke kun forbeholdt organisationsteorien. I forbindelse med udvikling af informationssystemer er det interessant at genfinde samme begreber i teorier om udvikling af software: "The degree of complexity in a given situation is a measure of the amount and diversity of relevant information needed to solve the problem" skriver Dahlbom og Mathiassen (1993, s. 103) og støtter dermed mængdebetraktningen i Galbraiths definition ovenfor. Samtidig inddrages også diversitet og antyder dermed en opdeling som hos Burton og Obel. Dahlbom og Mathiassen definerer usikkerhed som værende mindre, jo lettere adgang, der findes til informationen, og jo mere pålidelig og relevant, informationen er.

Dahlbom og Mathiassen argumenterer for et paradoks. På den ene side vil en rationel konstruktivistisk approach i systemudvikling kunne reducere kompleksiteten gennem abstraktion og dekomponering. I parentes bemærket må det være klart, at når kompleksitet i sin definition er determineret gennem problemet og den givne situation, så kan en reduktion af kompleksiteten kun fremkomme ved reduktion blandt disse to variable. Ofte vil problemet blive redefineret gennem det konstruktivistiske synspunkt. Et eksempel kunne være udvælgelse og destillation af problemer fra

mange kilder. Et systems funktionalitet fastlægges ved gennemgang af de mange krav, der specificeres fra adskillige sider. Tilbage bliver nogle få operationer, der formaliseres som systemets egentlige funktioner. Ved hjælp af konstruktivistisk formalisering bliver kompleksiteten nedsat, men dermed øges usikkerheden i forhold til den omgivende virkelige verden, som ikke kan indpasses i den simple model. Reduktionen medfører ikke blot ignorering af variables størrelse, men også den voldsommere equivocality, hvor variable ignoreres. Den konstruktivistiske tilgang vil således kunne resultere i en systemudvikling, hvor systemet fra ét synspunkt kan rubriceres som en teknisk succes, hvor de dekomponerede og abstrakte specifikationer opfyldes af et logisk sammenhængende design. Men fra et andet synspunkt fremstår systemet som en organisatorisk fiasko på grund af ignorering af den virkelige verdens komplekse organisatoriske sammenhænge.

Fra det metodemæssigt diametrale udgangspunkt vil den evolutionære tilgang forsøge at håndtere usikkerhed gennem eksperimenter med forsimplede opstillinger. I systemudviklingen vil det ske gennem proto-typer, der hurtigt sættes sammen – og måske i første omgang udformes som papirbaserede modeller ("mock-ups"). Proto-typer er netop udformet med henblik på, gennem demonstrationer og muligheden for hurtigt at kunne foretage forbedringer, at kunne opnå en stor grad af involvering af brugerne af systemet. Modellerne er eksperimenter og hvert nyt eksperiment vil producere ny information – i form af ekstra variable – således at den evolutionære model vil inddrage stadigt flere variable. Dermed nedsættes usikkerheden, men kompleksiteten øges. Den evolutionære tilgang vil således kunne opnå en høj umiddelbar succes blandt brugerne; en succes som primært tilskrives et grundigt testet brugerinterface. Omvendt vil systemet kunne vise sig at indeholde alvorlige uhensigtsmæssigheder, idet det usammenhængende og knop-skudte system ikke vil være fleksibelt

og robust (dvs. hverken velegnet overfor ændringer og udvidelser efter ibrugtagning (Mathiassen, 2001, s. 176) eller velegnet overfor forretningsændringer (Agosta, 2000, s. 34)).

En balance omkring usikkerhed og kompleksitet vil forekomme i projekter som både udviser stor opmærksomhed overfor anvendelse af brugerindflydelse ved udformningen af bruger-interface og som lykkes at trænge gennem virksomhedens kompleksitet og anvende de mest centrale forretningssideer som grundlag for designet af data warehouse. Dahlbom og Mathiassen (1993, s. 99) peger elegant til Herbert Simon's term om "satisficing" snarere end "optimizing", fordi systemet (og problemet) vil undergå yderligere udvikling i en forsat fortolkning og genfremsættelse. Balancen mellem usikkerhed og kompleksitet betyder, at begreberne betegnes som modpoler på en og samme dimension:

- Usikkerhed og kompleksitet

Dimensionen eksploration – eksploitation

I 1991 fremsatte March, at organisationen må påtage sig såvel eksplorative som eksploitative anstrengelser. På den ene side anskues det som nyttelyst af eksplorere, hvis den opnåede viden ikke senere omsættes i eksploitation. Internt i virksomheden vil diskussionspunktet i forbindelse med eksplorative projekter ofte være spørgsmålene om med hvilken sikkerhed, der kan opnås viden, og hvornår denne viden kan udnyttes (udmøntes!). På den anden side vil en ubalanceret udnyttelse af eksisterende ressourcer (eksploitation) efterlade organisationen i en ødemærket, der ikke længere vil kunne danne basis for udnyttelse.

Dimensionen som udspændes mellem eksploration og eksploitation genfindes også i tidlig litteratur om systemteori. Således argumenterer Buckley mod Parsons enøjede synspunkt om tilpasning: "In the case of the social system, the issue is thus whether conformity and order, on the one

hand, and deviance and disorder, on the other, are to be considered on a par as system characteristics or products" (Buckley, 1967, s. 27). Senere (s. 58) henviser Buckley til begreberne morphostasis og morphogenesis, som på den ene side de åbne systemers evne til at bevare og vedligeholde et systems givne form, organisation eller tilstand og på den anden side (morphogenesis) even til at ændre form, struktur og tilstand. Dimensionerne af konformitet/forstyrrelse og morphostasis/morphogenesis har relation til dimensionen omkring eksploration/eksploitation. Organisationen bevæger sig i sin stræben efter overlevelse i en evig balance mellem de to modpoler. For data warehouse kan den velplanlagte støtte til ledelsesbeslutninger og feedback til operationelle systemer ses som eksploitation, mens nogle typer af data mining bevæger sig ind på helt ukendt land og er klart eksplorativ.

- Eksploitation og eksploration

Centralisering og decentralisering

Kanaliserings af information til det rette sted i organisationen, og evnen til at kunne overskue informationen i organisationen betragtes som afgørende for organisationens funktion. Med informationssystemer opnår organisationen værktøjet til at placere informationen på rette sted. Teknisk set kan informationen flyde og omgå organisationens traditionelle hierarkiske opbygning, og i stedet understøtte lateral organisering (Galbraith, 1977 og 1994). Således kan informationsteknik understøtte en beslutningskraft placeret helt tæt på problemområdet. Om mulighederne for decentralisering udnyttes i organisationen er væsentlige designmæssige ledesspørgsmål.

Et data warehouse samler data og opfattes som en data central. Centraliseringstendensen modvirkes dog af, at en grundig validering af data i data warehouse hviler på inddragelse af den ekspertise, der eksisterer blandt ansatte med detaljeret kendskab til transaktionssystemet, og som kender "egne" data i detaljen. Der kan således

være tale om en aktiv decentral data-indhentning, som yderligere vil involvere og aktivere mange afdelinger i virksomheden.

Ud over den tekniske udformning og indhentningen af data fra afdelingerne har et data warehouse karakter af et softwaresystem under stadig udvikling, hvilket også indebærer, at afdelinger har mulighed for at promovere anvendelsen af netop deres data i data warehouse. Afdelingen vil også have interesse i at udnytte de validerede og velkombinerede data i data warehouse (feedback). Afdelingen kan altså i rollen som donor af data til en central instans virke som understøttende den organisationelle centralisering, og øjeblikket efter have den modsatte rolle som den decentrale bruger og beslutningstager.

Balancen er en gensidighed: "Do ut des". Således leverer afdelingen kopi af sine data, for at den centrale forsyning (data warehouse) skal give adgang til de rensede og relevante data for hele organisationen, herunder også den decentrale afdeling.

- Centralisering og decentralisering

Succes dimension

Informationssystemers succes kan fremstilles som bestående af mere uhåndgribelige fordele, mens de økonomiske omkostninger er særdeles presente og ekstremt håndfaste (Laudon & Laudon, 2000, s. 354). Derfor er det vanskeligt at opstille budgetmodeller med cost-benefit og endnu vanskeligere at verificere benefits (fx "Bedre kontakt med kunderne vil generere 600.000 kr.!!"). Indførelsen af ny teknik har desuden effekter på et højere niveau. Således kan sovebyen ses som den samfundsmæssige effekt af jernbanen (Sproull og Kiesler, 1991, s. 5). Informationssystemer vil ligeledes have sekundære effekter med forbindelse til opmærksomhed, social kontakt og samhørighed (Sproull og Kiesler, 1991, s. 147). Informationssystemers effekter kan således både være vanskelige at fastsætte i deres nærhed og have langt mere vidtrækkende effekter.

I en empirisk undersøgelse af data ware-

house (Wixom, 2001) fokuseres på systemets succes ved inddragning af samlende faktorer som data kvalitet og system kvalitet, mens udgangspunktet for dette oplæg er dimensioner, der befinder sig i virksomheden snarere end i informationssystemet. Et generelt virksomhedsproblem er den begrænsede kapacitet til informationsbehandling. En del databehandling kan automatiseres – denne effektivisering frigør menneskelig kapacitet – men virksomheden kan også udnytte sin faktiske kapacitet ved at udbrede beslutningstagen, og informationssystemet kan dermed aktivere flere menneskers kapacitet.

Forskning inden for den økonomiske produktivitet for informationssystemer (Brynjolfsson og Hitt, 1998) har som forklaring fundet, at de succesfulde informationssystemer følges af en høj grad af anvendelse af kostbar konsulentassistance, træning, redesign af forretningsprocesser og andre organisatoriske forandringer. Udviklingen af informationssystemet følges således af en organisationsudvikling og omvendt. Forekommer der meget høje afkast, skyldes det ofte, at succes udelukkende relateres til de direkte omkostninger til informationsteknologien (hardware og den standardiserede software), og at udgifterne hertil kun udgør en mindre del af de totale omkostninger.

Flere forfattere inden for systemudvikling, har udarbejdet lister for, hvorledes succes eller kvalitet for et informationssystem kan kategoriseres og måles (Mathiasen, 2001, s. 174; Pressman, 2000, s. 93; Laudon & Laudon, 2000, s. 403).

Sådanne længere lister bør medtages og videreudvikles i en komplet vurdering af et data warehouse området omtaler ligeledes både kvantificerbare og kvalitative målinger:

"Brug af systemet". Hvor mange brugere og hvor meget brug er der af systemet? Almindeligvis logges anvendelsen af et data warehouse minutløst. Helt parallelt til anvendelsen af data warehouse til at

opnå viden om kunderne og deres adfærd ud fra transaktionssystemerne, vil brugerne af data warehouse blive overvåget med henblik på forbedringer af data warehouse systemet selv.

“Brugertilfredshed” eller information om brugernes accept af systemet (Kimball, 2002, s. 27). For ikke at sammenblende accept med den faktiske brug bør accept måles gennem interviews eller spørgeskemaer. Det er vigtigt at medtage potentielle brugere og ikke kun udgå fra de faktiske brugere, som allerede med adfærd har udvist en vis accept. I modsætning til transaktionssystemet vil data warehouse have frivillige (optionale) brugere - hvis systemet ikke er “godt nok”, bliver det ikke brugt.

“Beslutninger”. Et data warehouse er primært et system til beslutningsstøtte. Kernespørgsmålet er derfor om virksomheden efterfølgende træffer bedre beslutninger på grundlag af indførelsen af data warehouse. (Kimball, 2002, s. 4). En kvan-

tificeret operationalisering kan fx inddrage beslutningstid og antallet af beslutninger.

“Bundlinien”. Spørgsmålet om beslutningers kvalitet vil kunne besvares så mangetydigt, at en naturlig følge er en strikt kvantitativ operationalisering i form af forandring af virksomhedens værdiskabelse.

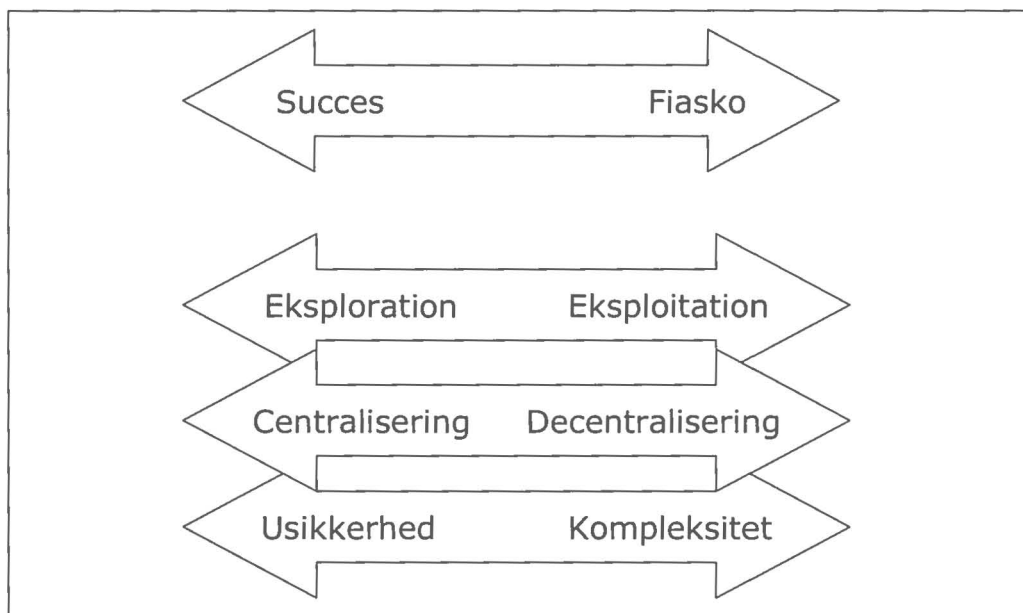
Selv i denne forenklede opstilling omfatter succes i sig selv flere dimensioner (økonomiske, brugsmæssige, organisatoriske forandringer). Disse kan kombineres gennem et formativt indeks til en enkelt afhængig variabel som operationaliserer resultatet af data warehouse projektet:

- succes – fiasko

Samling

En forsimplet samling af de berørte dimensioner opstiller den umiddelbart ukomplicerede opgave at fastlægge, hvilken placering på hver af de tre dimensioner, der vil være optimal for opnåelse af en høj grad af

Fig. 3. Succes og tre væsentlige dimensioner.



succes. Med 3 dimensioner kan der forekomme 8 ekstreme udfaldskombinationer (fx et udfald: Stærkt eksploitativt, stærkt centraliseret, stor usikkerhed).

Der forventes imidlertid ikke så entydige sammenhænge, idet spørgsmålet om succes – som det er fremstået ved behandlingen af dimensionerne – netop er et spørgsmål om opnåelse af balance inden for dimensionen. Eksempelvis vil udelukkende eksploration og udelukkende eksploitation begge være ødelæggende. Frem for at udvælge en enkelt udfaldskombination til at være den korrekte og eneste måde for opnåelse af et succesfuldt data warehouse vil en undersøgelse formentlig kunne fastlægge, at flere sammenhænge blandt de tre dimensioner vil kunne føre til succes. Billedlig talt vil organisationen balancere samtidige dimensioner ved at have succesfulde applikationer placeret flere steder i det tredimensionale rum udspændt af 1) eksploration/eksploitation 2) centralisering/decentralisering og 3) usikkerhed/komplexitet.

Konklusion

Ønsket om en tryllestav er forståeligt. Men som udredningen viser, bør der inddrages adskillige centrale organisatoriske variable, deres sammenhæng kortlægges og afbalanceres, og først på den baggrund vil det være muligt at opnå den organisatoriske succes. En succes som netop består i balancen, og som vi oprindeligt forventede at data warehouse i sig selv ville levere. Det enkle sving

med tryllestaven – gennem anskaffelse af hardware og software – må erstattes med en kompliceret balanceret udvikling. Indførelsen af et succesfuldt data warehouse i virksomheden bør ikke betragtes som anderledes end andre udviklingsopgaver og pga. balanceringen af de organisatoriske variable anbefales systemudvikling efter “unified software” modellen (Jacobsen, Rumbaugh, Booch, 1999).

Udviklingen bør være “use-case driven”, dvs. drevet af anvendelse af data warehouse gennem undersøgelse, udvikling og fastlæggelse af bruger-behov med basis i forretningsideer. Hermed frarådes således udvikling af data warehouse som data-drevet (Inmon, 1996, s. 25). En kopi og samling af operative data resulterer alene i et lager, der ikke danner ny værdi. “Unified” systemudvikling er også centreret omkring arkitekturen. For data warehouse betyder det anvendelsen af Kimballs dimensionale design. Det dimensionale design og dets udvikling er yderligere understøttende for de to andre karakteristika for “unified process”: Iterativ og inkrementel. Vi gør ikke det hele på en gang: Arkitekturen (formen) fastlægges i en balance med funktionalitet i brugerbehov, og i en løbende proces forbedres data warehouse iterativt og udvides trinvis til at være stadigt mere omfattende og støtte virksomhedens værdiskabelse på stadig flere forretningsområder.

Trylleri erstattes af hårdt arbejde. Hårdt, men ikke umuligt.

Summary

Hey presto! The idea of a magic wand is appealing. Can we in one stroke make the firm more profitable and attain greater satisfaction among staff, customers and suppliers? It may seem naïve to believe that things may be done that simply but, conversely, if at all possible, there is no reason to complicate simple solutions - if they can be found! Can we change the organisation by applying new measures for management or implementing new software? "Data Warehouse" is such a magic wand concept marketed by several organisations such as software, hardware and consultancy services, providing solutions to the problems of the organisation.

This article describes the theoretical framework for a study of the introduction of data warehouse in organisations. Through literature studies various theories are discussed which are relevant to the data warehouse. To begin with

there is the theory on the data-managing organisation where central parameters can be discovered and then rediscovered within both organisational and systems development theory. The discussion of these parameters leads to a description of key dimensions for the perception of data warehouse. The intention is to test these dimensions in large enterprises, after proposing hypotheses on the successful introduction of data warehouse.

Data warehouse success is achieved through skilful handling of essential organisational dimensions: 1) exploration/exploitation 2) centralisation/decentralisation and 3) uncertainty/complexity. But data warehouse is no magic wand. Instead, successful development of data warehouse should be attained through successful systems development.

Noter

1. Denne artikel er udarbejdet på grundlag af et working paper fremlagt ved "Workshop on Information and Organizational Design", EIASM, Bruxelles, March 2001.
2. Simon angiver tre faser i beslutningstagen: "Intelligence, design, choice" (Simon, 1960, s. 2).
3. Data warehouse for analyse af virksomhedens internet findes fx som "Data webhouse" (Kimball, 2000).
4. Blandt applikationerne har specielt data mining påkaldt sig interesse. Det skyldes primært, at med data mining understreges kraftigst forskellene mellem et data warehouse og det transaktionelle operative system i organisationen (se Berry og Linoff, 1997 og Rasmussen, 2001).
5. Citat fra "Ballad of a Thin Man" fra Bob Dylans "Highway 61 Revisited".

Litteratur

- Axelrod**, Robert & Cohen, Michael: *Harnessing Complexity*, Free Press, 2000.
- Agosta**, Lou: *Essential Guide to Data Warehousing*, The, 1/e, Prentice Hall, Upper Saddle River, NJ, 2000.
- Berry**, Michael J.A. & Linoff, Gordon: *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, 1997.
- Brynjolfsson**, Erik & Hitt, Lorin M.: *Beyond the Productivity Paradox*, ACM, August 1998/ Vol. 41. No. 8, 1998.
- Buckley**, Walter: *Sociology and modern systems theory*, Prentice-Hall, Englewood Cliffs, NJ, 1967.
- Burton**, Richard & Obel, Børge: *Strategic Organizational Diagnosis and Design* (2. ed), Kluwer Academic Publishers, Norwell, MA, 1998.
- Checkland**, Peter and Holwell, Sue: *Information, Systems and Information Systems - making sense of the field*, John Wiley & Sons, 1997.
- Dahlbom**, Bo & Mathiassen, Lars: *Computers in Context: The Philosophy and Practice of Systems Design*, Blackwell, Cambridge, MA, 1993.
- Dickson**, Gary W. & DeSanctis, Geradine (eds.): *Information Technology and the Future Enterprise. New Models for Managers*, Prentice-Hall, 2001.
- Galbraith**, Jay R.: *Organization Design: An Information Processing View*, Interfaces, Vol4, No. 3, May 1974, 1974.
- Galbraith**, Jay R.: *Organization Design*, Addison-Wesley, Reading, MA, 1977.
- Galbraith**, Jay R.: *Competing with Flexible Lateral Organizations* (2.ed.), Addison-Wesley, 1994.
- Inmon**, W.H.: *Building the Data Warehouse* (2.ed.), John Wiley & Sons, 1996.
- Jacobson**, Ivar & Booch, Grady & Rumbaugh, James: *The Unified Software Development Process*, Addison-Wesley: Reading, MA, 1999.
- Keen**, Peter G.W.: *Relationships*, (s. 163-185) i Dickson ovenfor, 2001.
- Kimball**, Ralph & Merz, Richard : *The Data Warehouse Toolkit. Building the Web-Enabled Data Warehouse*, John Wiley & Sons, 2000.
- Kimball**, Ralph : *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*, John Wiley & Sons, 1996.
- Kimball**, Ralph & Ross, Margy : *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses* (2. ed.), John Wiley & Sons, 2002.
- Laudon**, Kenneth C. & Laudon, Jane Price : *Management Information Systems: Organization and Technology in the Networked Enterprise* (6.ed.), Prentice-Hall, Upper Saddle River, NJ, 2000.
- March**, James G. : *Exploration and Exploitation in Organizational Learning*, Organization Science, Vol. 2., No. 1, February 1991, 1991.
- Mathiassen**, Lars et al. : *Objektorienteret Analyse og Design* (3. ed.), Marko, 2001.
- Pressman**, Roger S. & Ince, Darrel : *Software Engineering: A Practitioners Approach. European Adaptation.*(5th ed.), McGraw-Hill, 2000.
- Rasmussen**, Karsten Boye: *Data Mining – er der guld i virksomhedernes data?*, Ledelse & Erhvervsøkonomi nr. 2, 2001.
- Simon**, Herbert A.: *The New Science of Management Decision*, Harper & Row, New York, 1960.
- Simon**, Herbert A.: *Administrative Behaviour* (4. ed.), Free Press, New York, NY, 1997.
- Spruill**, Lee & Kiesler, Sara: *Connections: New Ways of Working in the Networked Organization*, MIT Press, MIT Press, 1991.
- Welbrock**, Peter R.: *Strategic Data Warehousing Principles Using SAS Software*, SAS Institute Inc., Cary, NC, 1998.
- Wixom**, Barbara & Watson, Hugh J.: *An Empirical Investigation of the Factors Affecting Data Warehousing Success*, MIS Quarterly, Vol. 25, No. 1, March 2001, pp 17-41, 2001.