

**Datalager – endast för storföretag?**

**(HS-IDA-EA-01-317)**

**Tomas Johansson (a98tomjo@student.his.se)**

*Institutionen för datavetenskap  
Högskolan i Skövde, Box 408  
S-54128 Skövde, SWEDEN*

Examensarbete på det systemvetenskapliga programmet under  
vårterminen 2001.

Handledare: Linda Furåker

### **Datalager – endast för storföretag?**

Examensrapport inlämnad av Tomas Johansson till Högskolan i Skövde, för Kandidatexamen (B.Sc.) vid Institutionen för Datavetenskap.

**2001-06-07**

Härmed intygas att allt material i denna rapport, vilket inte är mitt eget, har blivit tydligt identifierat och att inget material är inkluderat som tidigare använts för erhållande av annan examen.

Signerat: \_\_\_\_\_

## **Datalager – endast för storföretag?**

**Tomas Johansson (a98tomjo@student.his.se)**

### **Sammanfattning**

Syftet med denna rapport är att undersöka om datalager, en sorts databas som fungerar som beslutsstödssystem, kan användas av mindre företag och organisationer inom en snar framtid. Datalager har tidigare varit en fråga främst för stora företag. Mindre företag nämns sällan i datalagersammanhang. Datalagerteknologin är ny vilket gör att den fortfarande utvecklas i hög takt.

Svaret söks genom en litteraturstudie samt en empirisk studie. I den senare tillfrågas ett antal datalagerutvecklare. Litteraturstudien fokuserar på fyra synvinklar: en storleksmässig synvinkel, en nyttomässig synvinkel, en ekonomisk synvinkel och framtida möjligheter. Ur dessa synvinklar studeras mindre företags möjligheter att använda datalager. Den empiriska studien fokuserar på spridningen av datalager i Sverige och mindre företags möjlighet att använda datalager.

Resultatet visar att det redan idag finns möjligheter för mindre företag att använda datalager. Även om möjligheterna finns är det en komplex process att implementera ett datalager och arbetet med datalagret upphör inte efter implementationen.

**Nyckelord:** Datalager, Data Warehouse, beslutsstöd, användning, mindre företag.

# Innehållsförteckning

<b>1 Inledning</b>	<b>1</b>
<b>2 Bakgrund</b>	<b>2</b>
2.1 Historik	2
2.2 Vad är ett datalager?	4
2.2.1 Fördelar och nackdelar med datalager	6
2.2.2 Datalagrets arkitektur	8
2.3 Utvecklingsprocessen	9
2.4 Dataförråd	12
2.5 Datautvinning	13
2.6 Tidigare arbeten och undersökningar	14
<b>3 Problembeskrivning</b>	<b>16</b>
3.1 Problemprecisering	16
3.2 Avgränsningar	17
3.3 Förväntat resultat	18
<b>4 Metod</b>	<b>20</b>
4.1 Metodvalsprocessen	21
4.2 Relevanta synvinklar	21
4.2.1 Vald synvinkel	23
4.2.2 Relevant litteratur	24
4.2.3 Validering	25
4.3 Relevanta undersökningsmetoder	26
4.3.1 Vald undersökningsmetod	27
4.3.2 Validering	27
<b>5 Genomförande</b>	<b>28</b>
5.1 Litteraturstudien	28
5.1.1 Förberedelser	28
5.1.2 Genomförande	28
5.1.3 Värdering av material	29
5.2 Den empiriska studien	29
5.2.1 Förberedelser	29
5.2.2 Svansfrekvens	30
5.2.3 Värdering av material	30

<b>6 Analys och resultat.....</b>	<b>31</b>
6.1 Material från litteraturstudien.....	31
6.1.1 Storleksmässig synvinkel.....	31
6.1.2 Nyttomässig synvinkel.....	32
6.1.3 Ekonomisk synvinkel.....	34
6.1.4 Framtida möjligheter .....	35
6.2 Material från den empiriska studien .....	38
6.2.1 Fråga ett .....	38
6.2.2 Fråga två.....	39
6.2.3 Sammanställning av den empiriska studien.....	40
6.3 Jämförelser med förväntat resultat.....	42
<b>7 Slutsatser .....</b>	<b>44</b>
<b>8 Diskussion.....</b>	<b>46</b>
8.1 Allmän diskussion.....	46
8.2 Vad arbetet har tillfört .....	47
8.3 Förslag på framtida arbete .....	47
<b>Referenser .....</b>	<b>49</b>
<b>Bilaga 1 – Företagsfrågor .....</b>	<b>1</b>

# 1 Inledning

Databaser existerar överallt omkring oss. Databaser lagrar och hanterar information med syftet att underlätta behandlingen av stora mängder data. Forskningsområden inom databasområdet har på senare tid kretsat kring utvecklandet av stöd för nya applikationer såsom CAD/CAM, e-handel, digital publicering, geografiska informationssystem (GIS), sjukvårdsapplikationer och multimediaapplikationer (Elmasri & Navathe, 2000), (Silbershatz m.fl., 1996). Alla dessa applikationer kan hanteras av olika databaser.

Ett annat användningsområde inom databasområdet är att använda databaser för att stödja beslutsfattare i deras beslut, så kallade *Decision support systems* (DSS), vilka även kallas *Executive information systems* (EIS) (Elmasri & Navathe, 2000). Dessa system har till uppgift att enbart bistå beslutsfattare med viktig och intressant information varpå dessa effektivare kan fatta beslut. Dessa databaser som skapats för att stödja beslutsstödsystem kallas datalager och är uppbyggda med tanken att klara få, ovanliga och komplicerade frågor. Denna arkitektur skiljer sig kraftigt från den traditionella databasstrukturen, vilken skall klara av många och enkla frågor. Detta innebär att datalagren har en annan uppbyggnad, vilket resulterar i att de blir dyra att tillverka, är komplicerade och kräver större lagringsutrymmen i jämförelse med traditionella databaser (Connolly m.fl., 2000). Att de är dyra att utveckla gör att de i dagens läge främst används av stora företag med stark ekonomi. Detta krävs då de är resurskrävande att äga och underhålla då det innebär en extra, separat, databas vilken används enbart som datalager. Nackdelen med datalager är att det är svårt att få dem att bli lönsamma. Att få dem lönsamma kan även ta lång tid. Fördelen med datalager är att de utgör en stor och effektiv konkurrensfaktor för de företag som verkligen lyckas.

Kapitel två ger en övergripande förklaring till vad ett datalager är samt en kort historik samt hur datalager utvecklats. Kapitel tre beskriver detta arbetes problembeskrivning och problemprecisering. Denna tar upp mindre organisationers möjligheter till att använda datalager i dagsläget och inom en snar framtid. Kapitel fyra beskriver vilka metoder och synvinklar problemställningen angrips med. Detta arbete utförs med en kombinerad litteraturstudie och empirisk studie. I litteraturstudien identifieras fyra synvinklar ur vilka problemställningen studeras. I den empiriska studien tillfrågas svenska datalagerutvecklare om förhållandena på den svenska marknaden. Kapitel fem beskriver arbetets genomförande och resultatet från arbetet presenteras och analyseras i kapitel sex. Analysen tyder på att det finns olika typer av datalager och att mindre företag kan använda ett av dessa datalager. I kapitel sju presenteras slutsatserna från arbetet. Slutligen diskuteras erfarenheter och framtida arbete och dylikt i kapitel åtta.

## 2 Bakgrund

Detta kapitel ger en introduktion till vad ett datalager är. Först ges en kort historik till databaser i allmänhet och varför datalager utvecklats över huvud taget och hur datalager kan definieras. Olika för och nackdelar beskrivs och datalagers övergripande arkitektur visas även. En vanlig utvecklingsprocess av ett datalager beskrivs varpå dataförråd, vilket är en mindre version av ett datalager, beskrivs. Sedan tas datautvinning upp, vilket är den kanske mest intressanta tekniken vad det gäller finndet av ny information i datalager. Slutligen beskrivs tidigare arbeten och undersökningar som hittats och har bidragit till detta arbete.

### 2.1 Historik

*"Data Warehousing is near the top of most companies' strategic IT initiatives."*

(Watson m.fl., 2001, sid. 47)

Connolly m.fl. (1999) skriver att organisationer sedan 1970-talet har koncentrerat sina investeringar i nya datorsystem som automatiserar deras affärsprocesser. Organisationerna skapade på detta sätt konkurrensfördelar gentemot sina konkurrenter. Arbetet kunde bedrivas snabbare och effektivare och nya och bättre tjänster kunde erbjudas kunderna. Över årens lopp har företagen samlat på sig bland annat stora mängder data i sina databaser, data om kunderna, data om sina egna varor och tjänster, data om organisationen själv, ekonomisk data etc. (Connolly m.fl., 1999).

Företags data användes tidigt enligt Strouse (1999) och Jarenko & Vahlgren-Wall (1986) främst i det operationella arbetet och administrativt stödjande funktioner som exempelvis budgetering. Användandet av denna data betonade kontroll och uppföljning av kostnader. Detta innebar att datafunktionerna i företag främst tjänade ekonomiavdelningarna. Företags data används i detta läge inte till att analysera marknader eller till att skapa beslutsstöd för företagsledningen. Nästa trend var att använda data i den operativa verksamheten och datasystemen utvecklades för att hantera mycket operativ information. Datoriseringen av den operationella verksamheten gav nytt stöd till bland annat produktionsplanerare som nu kunde effektivisera sitt arbete. Utvecklingen fortsatte i riktning mot att stödja mer strategisk verksamhet. Detta innebar att data användes för att stödja företagsledningar och beslutsfattare och bistå dessa med viktig information om den egna organisationen, information om konkurrenter och dylikt. Datoriseringen bredde ut sig mer och mer och tidigare stora funktioner som bokföring fick stiga tillbaka till förmån för nyare funktioner såsom marknadsföring och marknadsanalysering (Strouse, 1999). Den viktigaste faktorn för att skapa dessa konkurrensfördelar är att förbättra sin kundkännedom. Det som krävs är mer ingående kunskaper om kunderna, inte endast hur många de är och var de bor. Denna analysering av kunder och deras rutiner kan datalager stödja. Strouse (1999) skriver vidare att användandet av datalager kan bli ett

nödvändigt konkurrensmedel, precis som automatiseringen blev under det tidiga 1900-talet.

Att överhuvudtaget samla data på ett ställe var nödvändig för att kunna överblicka den och använda den till någonting. Stora datamängder samlades tidigare i separata filer, så kallade filbaserade system (Connolly m.fl., 1999). Dessa system handlade sin egen data och var oberoende av omvärlden. Nackdelen med dessa filbaserade system var bland annat att informationen var separerad och inte kunde hållas konsekvent. Olika parter hade sin egen information om samma objekt. Exempelvis kunde information om en kund finnas på flera olika platser inom ett företag. Detta kunde göra att olika information om en och samma kund existerade.

Ur de filbaserade systemens brister växte de första databassystemen upp (Connolly m.fl., 1999). Dessa byggde på principen av en centraliserad förvaring av all information. Datan kunde nås av alla i organisationen och då datan endast fanns på en plats fick alla samma information. Informationshanteringen blev genom denna utveckling bland annat effektivare och snabbare. En av de största nackdelarna med databassystemen var att deras komplexitet ökade kraftigt jämfört med de filbaserade systemen (Connolly m.fl., 1999). Databassystemen har utvecklats sedan introduktionen men huvudprincipen med en centraliserad förvaring av datan är fortfarande kvar.

Huvudtanken med databassystemen är precis som med de filbaserade systemen att effektivt kunna nå och arbeta med data av olika slag. Databassystemens uppgift är alltså att snabbt och effektivt hjälpa användare att nå den information de behöver. Den information användare oftast behövt är enkel information, exempelvis data ur register av olika slag som exempelvis namn och nummer. Informationen skulle även kunna nås relativt snabbt. Detta gjorde att databassystemen specialiserats på att bli just snabba på att främst arbeta med enkel information.

Då önskemål kom på att databassystemen även skulle utföra avancerade uppgifter, som analysering av data, uppstod problem i och med att de utvecklats för att vara snabba och effektiva på enklare uppgifter. De nya kraven som datalagerfunktionerna ställde innebar analysering på avancerade datastrukturer som exempelvis flerdimensionella datastrukturer, även kallade datakuber (Connolly m.fl., 1999).

Då företag på senare tid har insett möjligheterna i att använda sin samlade data som grund för att skapa beslutsunderlag och på så sätt skapa en ny konkurrensfördel (Connolly m.fl., 1999). Detta innebar att datalager utvecklades, det vill säga datasystem som skapats med tanke att utföra den mer avancerade analyseringen av data som de ursprungliga databassystemen ej var skapade för att utföra.

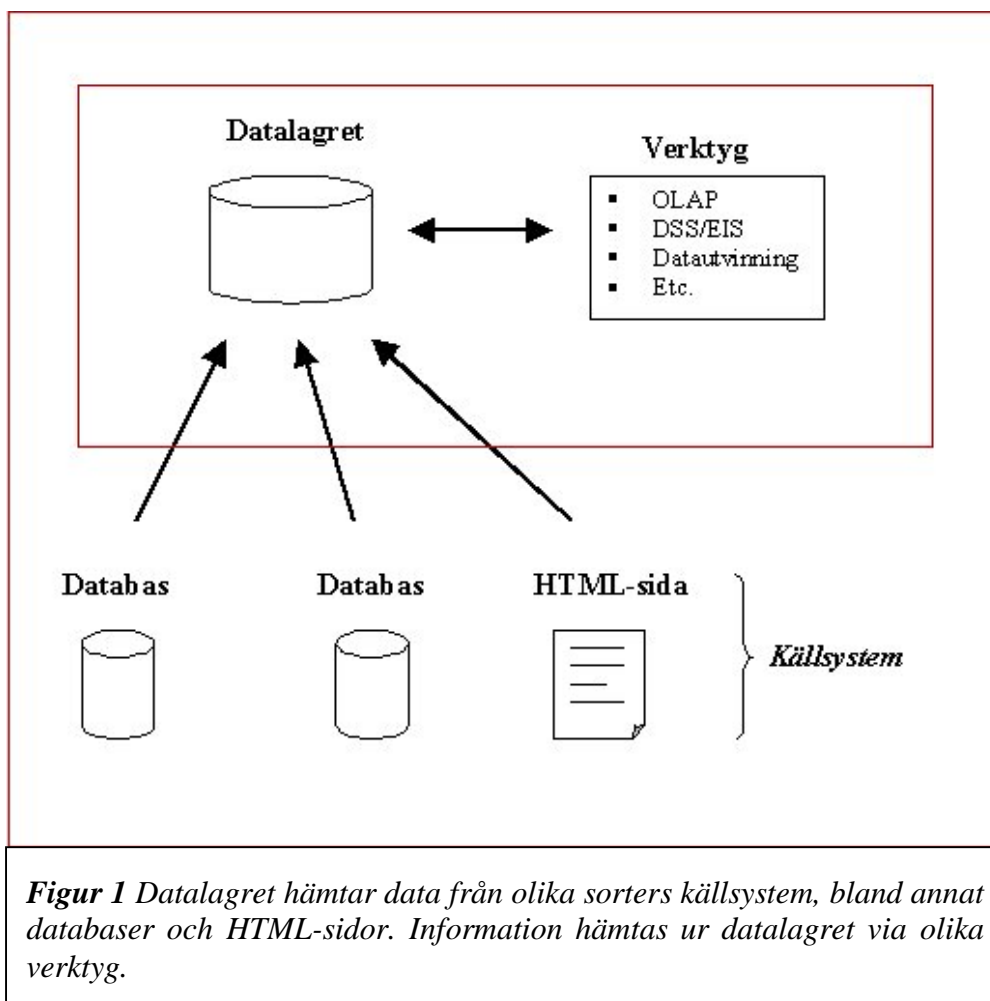
Datalager är enkelt beskrivet ett datasystem med funktioner och verktyg som främst skapats för att stödja organisationers strategiska mål. Detta innebär att datalager inte är skapade för att ge konkurrensfördelar för exempelvis operationella datasystem som arbetar på lägre nivåer, det vill säga operationella och taktiska nivåer. Datalager fyller alltså ingen funktion på verkstadsgolvet. Däremot finns det ingen strikt skiljelinje för hur organisationer verkligen använder sina datalager.

Ökad lagringskapacitet och minskade priser på lagringsmedier har på senare tid även gjort det möjligt att ha data on-line och göra on-line-analys av denna data (Srivastava & Chen, 1999).



## 2.2 Vad är ett datalager?

Organisationer som hade använt stora datasystem och databaser länge hade samlat på sig mycket information som lagrades i deras databaser. Under det tidiga 1990-talet började de största organisationerna visa ett intresse för att använda den lagrade datan för att skapa beslutsunderlag, beslutsunderlag som kanske skulle kunna hjälpa dem i den allt mer hårdnande konkurrensen på marknaden. Organisationerna försökte då analysera data med hjälp av sina befintliga system, men att anpassa de befintliga systemen till de nya uppgifterna var inte problemfritt (Connolly m.fl., 1999). De system som fanns var ej byggda för att utföra dessa nya funktioner. Dessutom fanns det ofta flera olika operativa system inom organisationerna, system som arbetade enskilt och ej alltid på samma sätt. Detta faktum försvårade ytterligare detta nya informationsökande mellan de olika databaserna. För att lösa dessa problem växte datalager, även kallat informationslager, fram. Dessa skulle kunna hämta data från flera olika instanser, främst databaser, och använda den sammanställda datan som stöd till beslutsfattande. Datalager är ett beslutstödssystem vilket Poe m.fl. (1998) beskriver som ett system som hjälper användare att analysera olika saker och situationer för att hjälpa dem att ta beslut. Systemet kan förse beslutstagare med färdiga rapporter sammanställda av användare, samt hjälpa användarna själva i deras vardagliga ställningstaganden.



Beslutsfattandet som systemet stödjer kan användas till både långsiktig och kortsiktig planering. Långsiktig (strategisk) för att analysera och studera marknadstrender över längre tidsperioder, till exempel för att identifiera behov av att utveckla nya produkter, tjänster etc. Beslutsfattandet kan i vissa fall även vara kortsiktigt (taktiskt och operationellt) för att studera om aktuella beställningar och orderkvantiteter från leverantörer kan göras på ett effektivare sätt, etc. Beslutsstödssystem består av ett antal applikationer för att bland annat skapa rapporter och utföra analyser. Dessa applikationer är skapade, i förväg eller på plats, av interna experter inom organisationen. Rapporterna är ofta dynamiska, det vill säga de kan till viss del anpassas efter de krav användaren har och den situation användaren befinner sig i. En annan sorts system som liknar beslutsstödssystem är så kallade *Executive information systems*. Dessa är system som skapats för att förse beslutsfattare med fördefinierade rapporter, vilka innehåller sammanfattad information om organisationen.

Datalager definieras ofta som:

*"A subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process."*

(Inmon & Hackathorn, 1994, sid. 2).

Denna definition av datalager kommer att användas i detta arbete. Nedan ges en kort förklaring till vad definitionen betyder (Inmon & Hackathorn, 1994).

- Datalagret är ämnesorienterat (*subject-oriented*) då det skapas av organisationer för att stödja organisationernas arbetsuppgifter.
- Det är integrerat eftersom all data behandlas lika och lagras på ett sätt i datalagret oavsett hur det lagras i källsystemen där datan hämtas ifrån.
- Det är föränderligt över tiden (*time-variant*), vilket innebär att all data i datalagret har en anknytning till den tid den lagrades. Detta till skillnad från operationella databaser vilkas värde är aktuellt just nu, men vars historiska värden ej är av intresse.
- Datan i datalagret är ej förändringsbenägen (*non-volatile*) då data lagras i datalagret en gång och därefter aldrig kommer att förändras.

Det finns även andra definitioner av datalager, men det övergripande syftet med dem är alltid att skapa ett system för hela organisationen där användare kan ställa frågor, skapa rapporter och utföra analyser. Data hämtas från olika instanser inom organisationen och sammanställs för att kunna användas som beslutsunderlag inom organisationen (Connolly m.fl., 1999).

Datalager gör att beslutsfattandeprocesser snabbas upp. Söderström (1997) skriver att beslutsordningen i äldre hierarkiska organisationer ofta är långsam. Detta gör att organisationer agerar och reagerar långsamt på förändringar i omvärlden. I och med datorernas intåg har denna process snabbats upp avsevärt, men problemet att få fram rätt information kvarstår. På 80-talet försökte organisationer lösa dessa problem genom att skapa *information centers* där enklare register sammanställdes och beslutsfattare kunde hämta dessa rapporter när de behövde dem. Dessa register

klarade dock inte av att ge sammanställningar över avdelningsgränser, eller att jämföra olika tidsperspektiv. Datalagerapplikationer används bland annat till bedrägeriupptäckter, vinstanalyser, som hjälp för att bibehålla kunder, inventariestyning, kredit-risk-analysering och prissättning (Söderström, 1997). Watson m.fl. (2001) skriver att datalagers främsta fördelar jämfört med att inte ha något datalager alls, är att de ökar antalet besvarade frågor och ökar antalet analyser som hinner göras innan ett beslut fattas. Datalager ökar även värdet på resultatet av frågor och analyser, och ökar i förlängningen hastigheten och kvaliteten på beslutstagandeprocessen. Datalager används idag främst till att producera rapporter och inte till att hitta samband mellan data. Att hitta samband mellan data görs genom datautvinning, vilket är en teknik som arbetar med att, bland annat, hitta dolda samband mellan data. Denna teknik är dock under utveckling och har ännu inte nått sin fulla potential. Datautvinningsapplikationer behöver enligt Watson m.fl. (2001) bli mer användarvänliga innan de kan komma att användas i det vardagliga arbetet. I en undersökning fann Watson m.fl. (2001) att datalagerhantering som först utvecklades i stora företag inom detaljhandeln, telekommunikation och bank- och försäkringsbranschen, på senare tid även har spridits till andra industrier och till mellanstora och till och med små företag (Watson m.fl., 2001).

Connolly m.fl. (1999) skriver att traditionella databassystem är konstruerade för OLTP (*on line transaction processing*), det vill säga ett kontinuerligt inflöde av många små och enklare frågor (transaktioner) av typen ”vad var medelvärdet på vår försäljning i våra butiker 1997?”. Större organisationer har ofta flera separata system av denna traditionella typ. Datalager, däremot, har som främsta uppgift att hantera komplexa, stora och ofta specialskrivna frågor, vilka ibland kan ta flera dagar att göra färdiga. Exempel på sådana frågor skulle kunna vara ”vilken typ av fastigheter säljs för priser över medlet i de största städerna i Storbritannien och hur står detta i förhållande till demografiska data?”. Analys av data som motsvarighet till OLTP brukar kallas OLAP (*on line analytical processing*). Målet är att organisationer endast skall ha ett datalager som alla kan använda. Även om de två typerna är olika är de nära besläktade då de traditionella OLTP-systemen tillhandahåller den data datalagren använder sig av (Connolly m.fl., 1999).

### 2.2.1 Fördelar och nackdelar med datalager

Det finns ett flertal fördelar med datalager. Connolly m.fl. (1999) påpekar att de eventuellt ger hög återbäring, trots att det är en stor investering organisationerna måste göra. Datalager ger även konkurrensfördelar och ökar produktiviteten hos beslutsfattare. Söderström (1997) och Poe m.fl. (1998) betonar fördelen med att införandet av ett datalager skapar en övergripande vy av företaget. På detta sätt undviks *silos of information*, det vill säga informationen finns inom avdelningens väggar men lämnar sällan avdelningen. En enhetlig syn på organisationen skapas, om alla siffror hos alla avdelningar tas från datalagret kan faktumet att siffror från olika avdelningar betyder olika saker undvikas. Söderström (1997) betonar även att datan får en jämn och känd kvalitet. Detta innebär att data från olika källor kan jämföras och att beslutsfattare själva kan, utan hjälp, hämta och sammanställa data vid behov. Poe m.fl. (1998) påpekar även vikten av marknadsanalysering. Det bästa sättet att förutspå framtiden är att se bakåt, något som är enkelt med datalager som lagrar historisk information. Genom detta tankesätt kan olika kundgrupper enkelt ringas in. Att skapa ett beslutsstödssystem utan att behöva störa de operationella systemen är även det en viktig fördel. Detta sker då datalagret läggs som en egen enhet ovanpå de

operationella systemen och lägger sina rötter i de operationella databaserna, vilka datalagret vill hämta data ifrån. Denna separation gör hela datasystemet, både de operationella databaserna och datalagret, stabilare än om man försöker bygga ihop dem till en enhet (Poe m.fl., 1998). Många av fördelarna är av organisatorisk karaktär, de skapar en enhetlig syn på organisationen, utöver datalagrets strategiska huvudfunktion, att vara ett beslutstödssystem.

Nackdelar med datalager är enligt Connolly m.fl. (1999) bland annat underskattning av förberedning och preparering av datan. Denna kan ta upp till 80 % av den totala utvecklingstiden (Inmon, 1990, i Connolly m.fl., 1999). Connolly m.fl. (1999) beskriver även andra nackdelar med datalager. Den data som används i ett datalager hämtas från ett eller flera källsystem. Om det finns gömda problem med källdatan från dessa system kan problemen följa med till datalagret. Är problemen gömda är det naturligtvis mycket svårt att upptäcka dem i tid, vilket gör att datan kan skapa problem längre fram. Om data från flera olika avdelningar i en organisation skall sammanställas i ett datalager så måste all data vara tillgänglig från början. Detta kanske inte alltid är fallet då olika instanser inte samlar in likadan eller jämförbar data. Detta gör att data som inte finns i ett källsystem måste ersättas på något sätt eller så får datalagret innehålla mer eller mindre komplett data på vissa områden. Allteftersom systemets användare blir mer och mer vana och kapabla att använda systemet, ökar dessutom kraven på systemet. Detta gör att datalagret kanske behöver byggas ut eller kompletteras med ytterligare verktyg etc. Att underlätta denna framtida utbyggnad är viktigt att ha i åtanke då datalagret byggs. Verktyg eller tekniker för att underlätta liknande förändringar i datalagret måste skapas för att ej uppdateringar av hela vyer skall krävas när en förändring sker (Rundensteiner m.fl., 1999; Samtai m.fl., 1999). Datalager tar mycket plats, vilket gör att hårdvaruresurserna förutom att kunna lagra information även måste vara snabba och effektiva. Detta gör att kostnaderna för hårdvara ofta blir hög. Ett annat viktigt problem är vem som anser sig äga datan när data från olika delar av verksamheten samlas på ett och samma ställe. Datalager kräver även mycket underhåll, de tar lång tid att konstruera och installera. Även komplexiteten av integreringen av datalagret i existerande system kan orsaka problem. Sammantaget innebär detta att det är många saker som kan gå fel eller underskattas då ett datalager utvecklas. Lyckas utvecklingen av datalagret skapar det en förvaringsplats för data som är optimerad för beslutsunderlag. Det kan enligt Watson m.fl. (2001) användas till rapportering, fördefinierade och *ad hoc* frågor, DSS/EIS (*Decision Support System/Executive Information System*), datautvinning, eller andra applikationer som kräver speciellt förberedd data (Watson m.fl., 2001).

Söderström (1997) påpekar svårigheterna med att utvärdera nyttan med datalager. Detta beror enligt Söderström (1997) främst på två orsaker; för det första att definitionen av datalager är oklar. Även om det finns definitioner på och förklaringar till vad datalager är, kanske en organisation som har ett datalager per definition inte själva anser sig ha ett datalager. Motsatt så kan organisationer som per definition inte har ett datalager, säga sig ha ett, då detta ger ett bra sken av hög status och kompetens på en hård marknad. Det andra problemet med att utvärdera datalager är svårigheten med att hänvisa ett visst resultat eller idé till datalagret. Är det datalagrets förtjänst att en medarbetare som använder datalagret fick en bra idé eller är det medarbetarens förtjänst? Jag tror att det ofta är en kombination av de båda, men det är viktigt att påpeka att datorer, hittills i alla fall, endast är till för att hjälpa människor i deras arbete. Åtminstone så länge det handlar om arbete som drar slutsatser från stora mängder osorterad information.

Srivastava & Chen (1999) anser att en av de viktigare frågorna i utvecklandet av datalager är problemet med dataintegration, vilket kan vara ett problem som uppstår då datalagret skapas och skall laddas med data. Problemet innebär att data som tas från flera oberoende databaser inte är helt kompatibla med datan från andra databaser. Hur skall till exempel posterna "KundID: kund001" och "KundID: 004-12-993" hanteras? Vilket av alternativen skall lagras i datalagret? Dock har forskning om dataintegration bedrivits i ett flertal år i samband med förenade databaser och heterogena distribuerade databaser, vilket till många delar liknar skapandet av datalager (Srivastava & Chen, 1999). Även synkronisering och förenade databasstrukturer är viktiga områden. Synkroniseringsproblem innefattar bland annat utvecklandet av algoritmer för att effektivisera transaktioner, det vill säga uppdateringar av datan etc. Detta är ett område som mycket forskning om *materialized views maintenance* berör i dagsläget (Srivastava & Chen, 1999). Förenade databasstrukturer är ett alternativ då den data som är intressant för hela organisationen är liten och den data som lagras skall vara relativt aktuell. I detta fall placeras datan i olika nivåer med olika grad av abstraktion, vilket gör att mindre data behöver behandlas och att responstider i genomsnitt blir snabbare.

### 2.2.2 Datalagrets arkitektur

En fråga som arkitekturens vikt när ett datalager utvecklas besvaras olika beroende på vem som tillfrågas. Just datalagers arkitektur är tillsammans med frågeoptimering troligtvis den fråga som behandlas mest inom den akademiska världen för tillfället. Hur viktig är datalagrets arkitektur om datalagret utvecklas inom en organisation som enbart är intresserad av att få datalagret att fungera tillfredsställande? Även dessa utvecklare inser säkert vikten av en övergripande struktur för arbetet, men vikten av en optimal arkitektur kanske inte är av lika stor vikt som den är inom den akademiska världen. För att dock beskriva hur datalager kan se ut är det givetvis lättast att vända sig till den akademiska världen.

Enligt Connolly m.fl. (1999) består datalager av följande komponenter: operationell data, en laddningsmodul, en datalagermodul, en frågemodul, detaljerad data, lite respektive mycket sammanfattad data, arkiverad- och backupdata, metadata och användarverktyg för att komma åt datan. Att datalager är ett område under utveckling, och att flera datalager utvecklas på olika håll, gör att det finns många olika uppfattningar om hur ett datalagers arkitektur bör se ut för att fungera bäst för just de aktuella utvecklarnas organisation. Den mest övergripande arkitekturen av datalager är dock de flesta överens om. Denna arkitektur omfattar följande delar:

- Ett eller flera källsystem varifrån data hämtas.
- Något sätt att förbereda datan för datalagret.
- Datalagret självt med dess metadata.
- Program för att administrera och sköta datalagret.
- Verktyg för att kunna utvinna information ur datalagret.

Datalager är till stor del skräddarsydda system, det vill säga precis som de flesta informationssystem är. Detta ter sig naturligt då datalager är en del av organisationers informationssystem. Är det möjligt att köpa färdiga produkter som kan hantera alla olika organisationers behov, eller kommer det att vara nödvändigt att lägga mycket tid

och kraft på att integrera dessa inköpta system i datasystemet? Kanske det blir effektivare i längden att bygga datalagret från grunden utifrån befintliga datamodeller. Kommer det även i fortsättningen att vara för dyrt och komplicerat att skräddarsy datalager? Detta skulle innebära att datalager skulle vara exklusivt till för de största företagen.

### 2.3 Utvecklingsprocessen

Datalager är enligt Srivastava & Chen (1999) grovt sett uppbyggda av följande delar:

- Operationella datasystem, det vill säga vanliga databaser, oftast relationsdatabaser, vilka används till dagliga operationer. Dessa databaser är kontinuerligt uppdaterade och innehåller den mest aktuella datan. Även om databaser oftast är källsystem för datan, kan även andra sorters system användas för att hämta data ifrån. Operationella databaser brukar oftast vara från omkring några tiotals megabyte till några gigabyte stora, alltså mycket mindre än datalager som kan bli så stora som många terabyte och även ända upp till petabyte stora.
- Datalagret är den huvudsakliga lagringsplatsen för historisk data. Data från detta hämtas med regelbundna intervall från de operationella datassystemen. Ibland kan datan delas upp i mindre delar för att användas till analyser i dataförråd.
- Dataförråd är mindre versioner av datalager. Dessa är ofta mer specialiserade än datalagren då de ofta finns i mer avgränsade företagsfunktioner och då riktar in sig på en mer begränsad del av organisationens totala data. Datat hämtas intervallvis från det större datalagret.
- Analysverktyg används för att verkligen komma åt den data som är intressant. Verktyg finns för bland annat trendanalyser, datautvinning, simulering, prognostisering och OLAP.

Oavsett vilken arkitektur ett datalager har genomgår det ett utvecklingsarbete. Söderström (1997) beskriver införandet av datalager ungefär som en vanlig systemutvecklingsprocess, vilken i stora drag består av framtagande av kravspecifikation, konstruktion och införande. Konstruktionen bör enligt Srivastava & Chen (1999) inledas med att en övergripande datamodell över organisationen skapas. Detta görs antingen genom att analysera organisationens processer och flöden, men mer ofta genom att använda redan existerande modeller från befintliga system. Arbetet med att identifiera ett företags processer och informationsflöden liknar till stor del det inledande skedet i informationsystemsutveckling. I båda fallen kartläggs det aktuella läget inom organisationen, för att sedan rätt krav och villkor skall kunna ställas innan själva konstruktionen tar vid. Srivastava & Chen (1999) fortsätter att beskriva konstruktionsprocessen, för ett datalager, genom att en övergripande datamodell skapas utifrån de framarbetade modellerna varpå många avväganden och kompromisser måste utföras. Att entiteter och attribut beskrivs och lagras på olika sätt i olika databaser, trots att de beskriver samma sak, är ett exempel på problem som måste lösas. Restriktioner och begränsningar på datan, som exempel att maximal ålder på en entitet är olika i olika databaser, är ett annat problem som konstruktörerna måste ta ställning till. Widom (1995) och Samtai m.fl. (1999) påpekar att all data inte är intressant att lagra i evighet. Datalager har en tendens att växa sig mycket stora relativt snabbt. Att all data inte är intressant för tillfället har Widom (1995) rätt i, men

vem avgör vilken information som är intressant just nu och vilken information som kan komma att bli intressant i framtiden? Givetvis beror detta på vad datan planeras att användas till. Olika organisationer i olika branscher behöver olika data att analysera. Frågeställningar som dessa är både intressanta och komplicerade, men är ledningsfrågor som ledningarna i aktuella organisationer måste ta ställning till. Effektiva metoder behövs skapas, metoder som avgör när data är föråldrad och som sköter en effektiv borttagning av dessa tupler (Samtai m.fl., 1999).

När datalagret modellerats och konstruerats skall det fyllas med information. I och med detta kan andra problem uppstå, exempelvis hur information som beskriver samma objekt men som kommer från två helt olika databaser skall länkas ihop. Ett annat problem är hur hanteringen av olika värden angående samma objekt skall hanteras. Hur skall dataexpansionen behandlas då datalagren efter en tid kraftigt ökat i storlek så att alla operationer tar längre tid då mer information måste hanteras? Kanske måste nya algoritmer utvecklas (Srivastava & Chen, 1999). Widom (1995) beskriver datalagers arkitektur på ett liknande sätt och skriver att en annan viktig sak är hur ”wrappers” och ”monitors” skall fungera på ett så bra sätt som möjligt. Detta är moduler som ansvarar för omkonvertering och uppdatering av datan mellan källorna, de vanliga databaserna, och datalagret. Dessa ”wrappers” och ”monitors” befinner sig alltså mellan källsystemen och datalagret.

När ett informationssystem skall skapas måste ett antal krav ställas på hur systemet skall fungera och se ut. Jämfört med ett traditionellt informationssystem är det enligt Söderström (1997) svårare att ställa upp kraven på ett datalager. En fråga som dyker upp i samband med detta är svårigheterna med att ställa krav på något man ej förstår. Vilka krav skall en användare ställa på ett system då denne ej vet vilka möjligheter och förutsättningar systemet har? Detta skulle kunna försvåra arbetet då datalagertekniken är ny och inga riktiga referenser till existerande detaljerade användningssätt finns. Å andra sidan kanske denna begränsning även kan vara positiv då oinsatta användare ej är begränsade av tekniska ramar. Användarna kan då komma med nya idéer och infallsvinklar som systemutvecklare inte tänkt på.

Söderström (1997) förespråkar en gradvis övergång från ett eventuellt gammalt beslutsstödssystem till det nya datalagret. Detta skapar enligt Söderström (1997) förtroende för den information som ett datalager förmedlar, ett förtroende som inte uppstår över en natt. En fråga som uppstår är då om en strävan efter en gradvis övergång inte kan vara både på gott och ont? En gradvis övergång i anpassning av termer och definitioner på begrepp är troligtvis ofrånkomligt. Fackspråk inom olika avdelningar i större organisationer har uppstått av en orsak och fyller ett viktigt syfte. Att lära ekonomer, konstruktörer och IT-anställda att tala samma fackspråk är antagligen både omöjligt och onödigt. Att däremot ha en snabb övergång från ett gammalt arbetssätt till ett nytt kanske kunde vara ett lättare sätt för användarna att anpassa sig. Det blir så att säga ett rent snitt mellan det gamla och det nya. Sedan är frågan hur stora förändringar en utvecklingsprocess verkligen för med sig? Införande av ett datalager kanske inte behöver påverka så många anställda att deras arbetsuppgifter behöver ändras nämnvärt. Naturligtvis påverkas olika delar av organisationen olika mycket, om de påverkas över huvud taget. Trots detta tror jag att användare behöver tid att anpassa sig, speciellt till större förändringar som införandet av nya informationssystem, som exempelvis ett datalager, kan föra med sig. Införandet av informationssystem förknippas ofta med effektiviseringar, vilket i många fall under 1990-talet betydde avskedande av många anställda. Detta innebär att många anställda ofta känner osäkerhet och obehag inför införande av nya

informationssystem. Går utvecklingen långsammare och stegvis kanske de anställda får tid att anpassa sig till det nya varpå deras inställning till nya system kan gå lättare.

Poe m.fl. (1998) poängterar vikten av att det är organisationen, inte tekniken, som måste vara i fokus i utvecklingsarbetet. Den kanske viktigaste framgångsfaktorn är dock involverandet av slutanvändare i utvecklingsarbetet eftersom det är användarna som kommer att använda systemet. Har de varit med och utvecklat systemet blir det antagligen bättre anpassat till verksamheten, användarna vet vad de får och blir ej negativt överraskade vid införandet. Detta är speciellt viktigt då användarna måste vara nöjda. Även om produkten som sådan är bra, så misslyckas arbetet ändå om användarna ej är nöjda och ej vill arbeta med systemet (Poe m.fl., 1998). Även Söderström (1997) anser att kraven på datalagret bör ställas av användarna. Söderström (1997) och Lehman & Jaszewski (1999) påpekar vikten av förankring i organisationen för att ett projekt som införandet av ett datalager skall vara genomförbart. Införandet av ett datalager påverkar ofta hela eller stora delar av organisationen. Det krävs mycket pengar, pengar som måste tas någonstans ifrån. Många anställda bör vara inblandade i utvecklingsarbetet för att få ett bra resultat. Vidare måste eventuellt ett enhetligt införande av termer utföras för att alla skall prata samma språk då data matas in i de olika databaserna. Lehman & Jaszewski (1999) påpekar att det krävs en naturlig relation mellan datan och företagsfunktionerna för att datan i datalagren skall få någon mening för slutanvändaren. En bra förståelse för hur datan är placerad i datalagret och hur den är uppbyggd är enligt Lehman & Jaszewski (1999) nödvändig då användarna snabbt och effektivt skall kunna tillgodogöra sig den information de söker. Även utvecklandet av nya rutiner och arbetssätt påverkar de anställda. Om många användare krävs för att utvecklingen av ett datalager skall bli bra kan detta vara till nackdel för mindre organisationer. Att undvara anställda till ett utvecklingsarbete innebär att någon annan måste göra de anställdas arbetsuppgifter. Att släppa anställda till ett sådant arbete kanske mindre organisationer inte har råd med. Blir detta fallet är tillgången till enkla, förskapade och lättanvända datalagerprodukter nödvändiga för att mindre organisationer skall kunna få tillgång till datalager.

Connolly m.fl. (1999) konstaterar dock det faktum att inga mjukvaruleverantörer erbjuder en komplett verktygslåda, med CASE-verktyg (*Computer Aided System Engineering*-verktyg) eller liknande, för att utveckla datalagret från början till slut. Detta innebär att flera utvecklingsverktyg och delar från flera olika leverantörer måste införskaffas. Detta ger upphov till nya problem, som exempelvis kompatibilitetsproblem, mellan dessa produkter. Att få alla de olika delarna att fungera bra tillsammans är ett stort men nödvändigt problem att lösa (Connolly m.fl., 1999). Vad gäller själva datalagret, och även vissa mjukvaruprogram för att sammanställa data och administrera själva datalagret, finns det ett antal produkter på marknaden (Widom, 1995). Denna brist på utvecklingsverktyg som sträcker sig över hela utvecklingsprocessen är naturligtvis ett stort hinder för mindre organisationer, vilka kanske inte innehar den kunskap eller de erfarenheter som krävs för att utveckla dessa komplicerade system från början till slut.

En annan viktig sak enligt Söderström (1997) är uppföljningen av utvecklingsarbetet. För att datalagret skall kunna användas måste nya organisatoriska enheter med fasta ansvarsroller skapas. Exempel på dessa är enheter med ansvar för förvaltning av själva datalagret, enheter för framtida begreppshantering inom hela organisationen, enheter för användarstöd och utbildning etc. (Söderström, 1997). Detta skulle innebära stora kostnader, relativt sett, för mindre organisationer. Frågan är om ett datalager verkligen motiverar sin existens om de mätbara effekterna även fortsatt är



små? Naturligtvis måste datalagerhanteringsverktyg kunna vara mycket kapabla och lättförstådda för att mindre organisationer skall kunna utföra detta själva. Å andra sidan slipper mindre organisationer eventuellt stora problem med saker som samordning, brist på nyckelpersoner som saknar övergripande kunskap om organisationen etc. då de anställda kanske jobbar tätare ihop och har en bättre uppfattning om hur organisationen fungerar och arbetar som helhet. Om det finns färre enheter, avdelningar, eller dylikt, är sannolikheten större att terminologi och definitioner är mer enhetliga, vilket skulle göra att data från olika avdelningar skulle hålla en jämnare kvalitet och inte behöva så stora förberedelser innan de lagras i datalagret.

### 2.4 Dataförråd

Eftersom datalager är en stor investering att göra för vilket företag som helst är naturligtvis mer ekonomiska lösningar välkomna.

Connolly m.fl. (1999) beskriver att utvecklandet av datalager är svårt, dyrt, tar lång tid och att det inte finns några garantier för ett lyckat arbete. Detta är orsaker till att dataförråd har utvecklats. Dataförråd, även kallat dataskafferi, är mindre versioner av datalager. Connolly m.fl. (1999) definierar dataförråd som:

”A subset of a data warehouse that supports the requirements of a particular department or a business function.”

(Connolly m.fl., 1999, sid. 933)

Dataförråd kan alltså fungera självständigt inom en avdelning eller vara sammanlänkat med det större datalagret. Dataförråd utvecklas oftast i avgränsade avdelningar inom företag. Anledningen till att utveckla dataförråd istället för datalager är enligt Connolly m.fl. (1999) bland annat att ge användare tillgång till den data de oftast använder. Svarstider minskas då dataförråd innehåller mindre data att gå igenom. De hanterar mindre data, vilket gör att implementationen och integrationen av dataförrådet med de övriga systemen går smidigare. Dataförråd är oftast mindre än datalager, vilket gör att kostnaden för ett dataförråd också är lägre än för ett komplett datalager. En annan viktig fördel med de mindre dataförråden är att det är lättare att nå rätt användare. Dataförråd riktar sig oftast till mer begränsade områden och på så sätt blir det lättare att konstruera förrådet som sådant, men det blir även lättare att rikta utbildning och support till användarna. Dataförråden fokuserar alltså enligt Connolly m.fl. (1999) endast på de aspekter och intressen som finns i en enskild avdelning inom en organisation. Dataförråd innehåller inte operationell data i lika stor utsträckning som datalager. Dataförråd innehåller även mindre information än datalager, vilket gör dem mer lättförståeliga och mer lättnavigerade.

Som beskrivits ovan finns det en mängd frågor och svårigheter med skapandet av datalager. Viktiga frågor och svårigheter med införandet av dataförråd är enligt Connolly m.fl. (1999) att dataförråden ständigt växer, vilket gör dem långsammare och långsammare. Stora dataförråd växer sig så småningom lika stora som små datalager. Ett problem med dataförråd som inte får förbises är att de visserligen är praktiska och bra att utveckla inom avgränsade enheter i en organisation, men dessa dataförråd måste även vara en del av ett större datalager för hela organisationen. När

flera avdelningar har implementerat sina dataförråd måste de i ett senare skede kunna samköras eftersom de är en del av det större allomfattande datalagret. Detta ger upphov till frågor som måste besvaras innan alla avdelningar sätter igång att skapa sina egna dataförråd, så inga oöverkomliga integreringsproblem uppstår i ett senare skede.

Huruvida problemet med flera dataförråd är intressant ur mindre organisationers synvinkel är tveksamt. Dessa organisationer kanske endast skapar ett dataförråd som för dem fungerar som ett helt datalager.

Bland fördelarna med dataförråd, utöver de som finns för datalager i stort, kan nämnas att virtuella dataförråd kan skapas. Detta innebär att specialsydda vyer från flera dataförråd skapas för olika användartyper, vilket i vissa fall kan vara en bra lösning (Connolly m.fl., 1999)

### 2.5 Datautvinning

Att ha ett effektivt och kraftfullt datalager är helt ointressant om det inte går att hämta ut information ur datalagret. Detta kan ske bland annat genom att skapa rapporter, analysera data etc.

En av de mest intressanta möjligheterna med datalager är att kunna använda all sin samlade data för att hitta dold kunskap i den. Detta kallas datautvinning. Det finns flera definitioner av datautvinning. Connolly m.fl. (1999) beskriver datautvinning som:

*”The process of extracting valid, previously unknown, comprehensible, and actionable information from large databases and using it to make crucial business decisions.”*

(Simodius, 1996, i Connolly m.fl., 1999, sid. 960).

Datautvinning har alltså det övergripande målet att hitta information som kan bidra till bättre beslutsfattande genom att hitta ny intressant information som man inte kände till tidigare. När information väl lagrats behövs det ett sätt att hitta den information som är intressant. Detta kan ske genom att användare, via applikationer, ställer frågor till datalagret eller att genom att användare skapar rapporter och sammanställningar om viktiga fenomen. Mest intressant skulle det vara att hitta den kunskap som tidigare varit okänd. Ett sätt att utföra denna jakt på ny intressant information kan vara med just datautvinning. Trots att datautvinning är relativt nytt används datautvinning enligt Connolly m.fl. (1999) redan i ett flertal industrier inom bl.a. försäljnings-, bank-, försäkrings- och läkemedelsbranschen. Watson m.fl. (2001) anser dock att datautvinningsapplikationer ännu är för svåränvända och måste bli mer och mer användarvänliga innan de till fullo kan komma att användas i organisationers dagliga arbete.

En intressant fråga är om datautvinning går att använda till mer än att stödja beslutsfattares strategiska mål. Naturligtvis beror datalagrets möjligheter på vilken data som finns att gräva i och med vilken avsikt systemet skapats. Ett dataförråd som skapats för att underlätta ekonomiavdelningens beslutsstödssystem kan naturligtvis

inte producera intressant information om kunders köpbeteende i västra Frankrike då den informationen ej finns lagrad etc.

Det finns fyra tekniker inom datautvinning (Connolly m.fl., 1999). Den första kallas *Predictive modeling*, vilket innebär att en förståelsemodell av världen skapas. Genom denna förståelsemodell görs sedan förutsägelser om framtiden grundade på historiska erfarenheter. Den andra tekniken kallas *Database segmentation* och syftar till att identifiera olika segment av en population baserad på liknande värden av olika slag. Den tredje är *Link analysis*, vilken försöker hitta länkar och samband mellan olika poster eller grupper av poster i en databas. Den fjärde kallas *Derivation detection* och denna försöker hitta avvikelser från mängden. Den sista tekniken anses vara den kanske viktigaste eftersom den hittar avvikelser från tidigare normer, vilket är intressant då ny kunskap söks.

### 2.6 Tidigare arbeten och undersökningar

Det förekommer många arbeten inom datalagerområdet, men färre som tar upp hur datalager används av företag. Detta är en av orsakerna till att jag inriktat mig på detta område. Jag har ej hittat något tidigare arbete som tar upp samma problemställning eller synvinkel som den som beskrivs i detta arbete.

I en undersökning av Watson m.fl. (2001) undersöktes 106 företag och organisationer som använder datalager världen över med koncentration på USA. De frågor författarna främst var intresserade av var av typen vem finansierade datalagret, vem drev på arbetet, vilka använde datalagret, vad används det till, har datalagret utvärderats, ansågs datalagret vara ett lyckat projekt eller inte, etc. Författarna kom bland annat fram till att det främst är stora organisationer som använder datalager, men att det även finns mindre företag som använder dem. Datalagren används främst till att förbättra organisationernas informationsflöden och ej till att hitta ny kunskap. De flesta organisationerna ansåg att utvecklandet av datalagret varit en stor framgång trots att det är en ny teknik och de flesta datalagren utvecklats inom organisationen utan större hjälp utifrån.

Datalagren i undersökningen (Watson m.fl., 2001) hade en medelålder av 2,6 år. Detta verkar vara lite och frågan är om ett datalager verkligen kan utvärderas på ett riktigt sätt innan det fått tid på sig att växa in i organisationen. Omorganisationer av organisationer och införandet av nya informationssystem kräver av erfarenhet relativt lång inkörningstid innan resultat av arbetet börjar synas. Frågan är alltså om utvärderingen av datalagren kan ge en rättvis bild då medelåldern på dem endast är 2,6 år. Företagen i undersökningen fick ange den totala framgången med datalagret, från ett till fem. Ett innebar eventuellt problematiskt, två innebar att vissa delar var framgångsrika, tre innebar måttlig framgång, fyra innebar framgång och fem innebar att datalagret var extremt framgångsrikt. Trots att medelåldern på datalagren var låg ansåg största delen av organisationerna, runt 83 %, att deras datalager redan var relativt framgångsrika eller bättre (grupp tre, fyra och fem i undersökningen). Endast 4,8 % ansåg att de hade problem med sina datalager (grupp ett i undersökningen). Detta tycker jag verkar vara en otrolig siffra. Informationssystemprojekt brukar i vanliga fall ha en relativt hög procent av misslyckade projekt. Frågor som uppkommer när denna undersökning studeras är: hur kommer det sig att datalagren är så lyckosamma? Beror det på frågeställningen i undersökningen? Beror det kanske på att det främst är stora företag som hittills konstruerat datalager? Stora företag har troligtvis mer och bättre erfarenheter av att hantera stora IS/IT-projekt.

Här måste ändå beaktas att endast en fjärdedel av företagen i undersökningen hade räknat med att få igen det satsade kapitalet. Detta indikerar att organisationerna inte hade några enorma förväntningar på datalagret även om de hade insett dess fördelar om konstruktionen och införandet lyckades.

Ett problem med undersökningen som Watson m.fl. (2001) själva också poängterar är problemet med att kvantifiera återbäringen på satsat kapital. Detta gör att det oftast är väldigt svårt att bedöma om och hur lönsamt datalagret verkligen är.

## 3 Problembeskrivning

Möjligheterna med datalagerteknologi är enorma. Tanken bakom dem, att kunna utvinna ny information ur existerande information, är mycket intressant. Dagens datalager används till största delen av företag och inom den akademiska världen. Skulle datalager till exempel kunna användas i ideell forskning, av privatpersoner via Internet, av småföretag som med sin begränsade konkurrensförmåga tvingas nischa sig till specialmarknader etc. Eller är kanske datalager inte av intresse för dessa över huvud taget? Att kunna hitta samband mellan liknande saker skulle vara värdefullt, inte bara för företag utan även för allmänheten då forskning förhoppningsvis kommer allmänheten till gagn. Forskning och programvaruutveckling driver ständigt ner priset på mjukvara, vilket gör att applikationer som tidigare har varit alltför dyra plötsligt blir tillgängliga för mindre företag och även privatpersoner. Utvecklingen av Internet har gjort att fler och fler tjänster kan utföras via Internet. Kan kanske datalagerhantering användas via Internet och finns det ett behov av detta?

Frågan är om datalagerteknologin verkligen klarar av att uppfylla alla dessa högt ställda förväntningar, eller om förväntningarna snart kommer ner på jorden igen?

### 3.1 Problemprecisering

Datalagerhantering är en intressant teknologi, men den är utvecklad för stora företag och organisationer. Dess fördelar skulle även vara intressanta för mindre företag och andra beslutsfattare. Problempreciseringen lyder således:

”Kommer datalagerhantering vara enbart större organisationer förbehållna eller kommer datalagerhantering inom den närmsta framtiden även att kunna användas av mindre företag eller enskilda användare?”

I och med denna problemprecisering finns det följdfrågor som är intressanta: Finns det problem som gör att datalager inte kan anpassas till mindre organisationer, eller skulle det vara en fördel att vara liten? Andra frågeställningar som är intressanta i anslutning till detta är om datalagerhantering även kommer att kunna användas av mindre företag och privatpersoner att använda själva, parter som inte har samma ekonomiska förutsättningar och andra krav jämfört med större företag. Om jag finner att mindre organisationer kan ha användning av datalager, i så fall till vad? Finns det verktyg som skulle kunna hjälpa dem att använda datalager? Dessa följdfrågor kanske kan hjälpa till att besvara problempreciseringen. Dessa frågeställningar tas ej upp i detta arbete. Eventuellt berörs några av dem indirekt då synvinklar vaks i metodkapitlet. Angående tidsbestämningen av den närmsta framtiden som anges i problempreciseringen avses cirka tre till fem år från dags datum. Denna formulering är medvetet gjord vag men den tros inte påverka resultatet nämnvärt.

Utöver ovanstående problemprecisering har jag för avsikt att utföra en enklare undersökning avseende vilka företag och organisationer som använder datalager i Sverige.

### 3.2 Avgränsningar

Jag vänder mig inte främst till de största företagen då de uppenbarligen redan anammat datalagerhantering. Detta är naturligt då utvecklingen av datalager i huvudsak drivs av dessa stora företag tillsammans med den forskning som bedrivs inom den akademiska världen. Då utvecklandet av datalager ännu är en stor, omfattande och dyr process kanske införandet av datalager ännu inte är av intresse för de mindre och mellanstora organisationerna. Det finns antagligen många områden som behöver tid och pengar att utvecklas inom dessa organisationer, innan ett datalager blir aktuellt. De företag som hittills utvecklar datalager är främst de väldiga företagen i USA, vilka vi inte kan jämföra oss med i Sverige. Datalager utvecklas även i Sverige, men i mindre utsträckning. Sverige ligger även några år efter USA i att ha insett möjligheterna med datalager (Söderström, 1997).

Var lägger jag min avvägning av var ett stort respektive litet företag är? Denna bedömning är naturligtvis godtycklig och beror på vilket syfte arbetet har. Datalager är ännu relativt ovanliga, vilket gör att jag antar ett relativt stort och brett perspektiv för att verkligen hitta någon information att arbeta med. Det vore ogjort arbete att göra en analys av datalager med inställningen att titta på svenska företag med 25 till 50 anställda. Med stora företag avser jag ur ett större perspektiv de multinationella företagen med 1000 eller fler anställda. Med mellanstora företag avser jag företag med 250 till 1000 anställda, och små företag är de med 250 eller färre anställda. Ur andra perspektiv är företag med upp till 250 anställda naturligtvis relativt stora respektive företag med minst 1000 anställda relativt små.

I sina register delar Statistiska centralbyrån (SCB) in företag i storleksklasser. Klass 1 är den minsta med noll antal anställda, klass 16 är störst med över 10 000 anställda. Skulle dessa ramar användas skulle storleksklasserna 7 och 8 vara av intresse i detta arbete. Klass 7 omfattar företag med 200 - 499 anställda medan klass 8 omfattar företag med 500 - 999 anställda.

Eventuellt är inte ett företags personella storlek ett i alla lägen lämpligt mått. Eftersom det är stora datamängder som datalagret är byggda att arbeta med, kan ett företags mängd data även vara ett bra sätt att identifiera målgruppen. Storleksfokuseringen i detta arbete bygger på att datalagren utvecklades av och för de största organisationerna. Ett av de lättast identifierbara kännetecknen hos dessa företag är deras stora antal anställda.

Även om jag avser att undersöka vilka typer av verktyg mindre organisationer kan ha användning av, har jag inte för avsikt att utvärdera vilka eventuella mjukvaruprodukter som i dagsläget finns och om dessa är lämpliga för företag att använda. Det finns redan verktyg på marknaden, men på vilket sätt och hur bra dessa fungerar anser jag mig inte tillräckligt kunnig för att avgöra.

Med avseende på den begränsade undersökningen jag avser att utföra beror resultatet till stor del på hur tillgängliga och villiga företag är att lämna ut uppgifter om sin verksamhet.

### 3.3 Förväntat resultat

Den litteratur jag studerat till bakgrundskapitlet tyder på att utveckling och forskning inte ges det utrymme det behöver. Det bedrivs mycket forskning inom ett flertal områden, men finansiering av forskning och utveckling till dagens databasforskning, relativt sett, är lägre än för annan forskning inom områden med jämförbar vikt (Silbershatz m.fl., 1996). Kanske på grund av de stora företagens vinstjakt. De stora företagen har hittills bedrivit en viktig del av forskningen på många områden. Företags forsknings- och utvecklingsavdelningar får enligt Silbershatz m.fl. (1996) mindre pengar att röra sig med då företag har högre och strängare lönsamhetskrav. Detta gäller främst för amerikanska företag men bör även gälla för övriga världen. Om detta är fallet skulle det kunna tyda på att den forskning som behövs inom datalagerhanteringsområdet inte kommer att bedrivas i den takt som omvärlden vill. Om så blir fallet, är detta en stor förlust för främst de mindre företagen som inte kommer att få tillgång till nya produkter och forskningsrön i lika stor utsträckning, då de ej har möjlighet att utveckla avancerade system själva. Av den forskning som ändå bedrivs skriver både Silbershatz m.fl. (1996) och Bernstein m.fl. (1998) att denna forskning går mot kortare lönsamma forskningsprojekt, så kallade Delta-X-projekt. Denna nya forskning sker med främsta avsikt att skapa nya produkter direkt till marknaden. Tidigare har forskningen till större del bestått av forskning med främsta avsikt att öka kunskap och erfarenheter genom skapande av prototyper och testprojekt.

Användandet av datalagerprodukter verkar inte ligga så långt fram som jag trodde när jag inledde mitt arbete. Detta har gjort att jag blivit mer pessimistisk vad det gäller tillgången till tekniken och dess möjligheter. Främst ekonomiska och kunskapsmässiga krav är höga för att utveckla ett datalager. Det krävs även stöd från alla inblandade parter, kanske främst ledningen, för att lyckas med ett åtagande av denna magnitud.

Trots dessa stora hinder är det inte enbart storföretag som använder datalager (Watson m.fl., 2001). Frågan är då vad dessa företag har för hemlighet? De kanske har insett de stora fördelar som datalager kan inbringa, samtidigt som de har ett överskott i kapitalet som de vill satsa på informationssystemsutveckling. De kanske får ta del av en större organisations utvecklingsarbete som innefattar just datalagerutveckling. De kanske inte definierar datalager lika strikt som den gängse termen, vilken används i detta arbete, och har därför inte ett komplett datalager även om de använder det som ett, eller de kanske inte använder alla de möjligheter datalagret erbjuder. Dessa frågeställningar är i och för sig intressanta, men för breda och skulle ta alltför mycket tid för att jag skall ta upp dem i detta arbete.

Huruvida privatpersoner kommer att kunna få tillgång till datalager är antagligen upp till om någon leverantör erbjuder en sådan tjänst till allmänheten där datalagerteknologi ingår. Den explosionsartade utvecklingen av Internet har drivit på utvecklingen inom många områden. En trend är att företag och myndigheter lägger ut sina tjänster på Internet där privatpersoner själva får göra jobb som företagen och myndigheterna gjort tidigare. Detta kan tyda på att stora databaser med självbetjäning i framtiden kommer att skapas i större utsträckning än vad som är fallet idag, men om dessa kommer att använda datalagerteknologi är svårt att säga.

Ovanstående gör att jag tror att små företag enligt min definition kommer att få det svårt att hävda sig på datalagermarknaden, även om några få mellanstora och små företag uppenbarligen lyckats med införandet av någon sorts datalager. Om

privatpersoner kommer att få tillgång till datalager är jag mer skeptisk till, trots att allt mer tjänster läggs ut på Internet. Att datalager kommer att nå ut till stora och små företag och även till privatpersoner i någon form är jag övertygad om, men efter inledande studier på området tror jag inte att datalagerteknologin kommer att utvecklas tillräckligt för att komma dessa till gagn inom de närmsta åren.

Avseende den undersökning angående vilka företag och organisationer som använder datalager i Sverige, förväntar jag att erhålla ett resultat som det går att dra några slutsatser ifrån. Jag förväntar mig att finna att det verkligen är övervägande stora företag och organisationer som använder datalager.



## 4 Metod

För att svara på problemställningen behövs en metod att följa för att skapa ett strukturerat arbetssätt. Genom ett strukturerat arbete är det lättare att kunna arbeta fram ett bra resultat. Ett strukturerat arbetssätt och en beskrivning av det är även nödvändigt för att en läsare skall kunna se hur arbetet har utförts samt hur resultatet har identifierats. Denna beskrivning underlättar läsarens bedömning av arbetet. Detta kapitel beskriver hur arbetet kan genomföras och på vilket sätt det fortsatta arbetet planeras att genomföras.

För att kunna besvara problembeskrivningen i detta arbete kan olika metoder väljas. De metoder jag identifierat som kan användas för att svara på problembeskrivningen är intervjuer, enkätundersökningar eller en litteraturstudie.

Metoder som jag inte anser är lämpliga är bland annat fallstudier, observationsstudier, *surveys*, experiment och implementeringar. Fallstudier studerar en mindre avgränsad grupp, en population, varpå generaliseringar om en större grupp görs (Patel & Davidsson, 1994). En *survey* är en undersökning på en större avgränsad grupp med exempelvis enkäter eller intervjuer. De två ovanstående är omfattande undersökningar gjorda på hela populationer eller ett slumpmässigt urval ur populationer. Populationen skulle i detta arbete bestå av de företag som möjligtvis använder datalager, de företag som använder datalager och de företag som utvecklar datalager. Denna population är antagligen för stor eller för svår att identifiera för att de skall utgöra möjliga grunder för arbetet. Observationsstudier är en teknik som används för att studera ett händelseförlopp (Patel & Davidsson, 1994). I detta arbete skall inget händelseförlopp studeras. Experiment och implementationer är inte aktuella då det i detta arbete ej finns något att implementera.

För att skapa en god grund för arbetet kan ibland tidigare arbeten användas, detta arbete grundas inte i tidigare arbeten. Pilotstudier är ett annat sätt att skapa en god grund för sitt arbete. Av naturliga skäl finns ingen tid att utföra pilotstudier inför detta arbete.

Av aktuella alternativ har jag valt att i huvudsak genomföra en litteraturstudie. Denna skall, som beskrivits tidigare, kompletteras med en empirisk studie.

Hart (1998) definierar litteraturgenomgångar eller litteraturstudier som:

*”The selection of available documents (both published and unpublished) on the topic, which contain information, ideas, data and evidence written from a particular standpoint to fulfill certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed.”*

(Hart, 1998, sid. 13)

Litteraturgenomgångens syfte är i förlängningen att förmedla kunskaper vidare till andra (Hart, 1998). Författaren måste förstå vilken roll hans eller hennes arbete har och vart det passar in för att kunna göra detta på bästa sätt. När ett verk som någon annan skrivit studeras eller om en litteraturgenomgång inleds bör läsaren eller

författaren vara medveten om hans eller hennes egna värdegrunder och försöka undvika brist på en akademisk respekt för andras idéer (Hart, 1998).

## 4.1 Metodvalsprocessen

Patel & Davidsson (1994) skriver att sättet problemet angrips på grundas i vilket problem som studeras. Undersökningen planeras, läggs upp och genomförs utifrån problemställningen. Utöver detta måste även andra yttre ramar och resurser som tid, pengar och datorresurser beaktas. Hur problemet angrips har stor betydelse för hur bra och tillförlitligt resultatet blir. Vilka tekniker som är att föredra och vilka som är möjliga att använda för att samla in information skall identifieras. Teknikerna måste vara lämpliga för att några slutsatser skall kunna dras och för att problemställningen skall kunna besvaras.

Problemställningen i detta arbete löd: "Kommer datalagerhantering vara enbart större organisationer förbehållna eller kommer datalagerhantering inom den närmsta framtiden även att kunna användas av mindre företag eller enskilda användare?" För att besvara problempreciseringen kommer en litteraturstudie att genomföras. För att göra arbetet ännu mer intressant kommer även en mindre kartläggning över vilka organisationer och företag som använder datalager i Sverige idag, att göras. Arbetet består således av två delar, en litteraturstudie och en empirisk studie. Målet för litteraturstudien är att skapa en djupare och bredare förståelse inom området för att sedan kunna dra riktiga och relevanta slutsatser för att kunna besvara problempreciseringen. Målet med den enklare empiriska studien är att försöka hitta företag och organisationer inom Sverige som idag använder datalager.

Möjliga metoder att lösa problemet på, utöver litteraturstudie, är bland andra test av produkt och fallstudie. Ett test av en produkt skulle kunna utföras genom att införa ett litet datalager i en liten organisation och ta del av de erfarenheter som gavs. Detta alternativ skulle ge en bra inblick i ett specifikt företags möjligheter och problem och ge en mycket produktspecifik vinkel på arbetet. Detta alternativ var inte aktuellt då ingen lämplig organisation var tillgänglig och då jag ej har erforderliga kunskaper för vara med och utveckla ett datalager. Ett alternativ som angränsar till detta skulle kunna vara att läsa dokumentation från ett liknande, redan utfört, utvecklingsarbete. Sådan dokumentation är troligtvis mycket svår att komma över. Att utföra fallstudier på ett eller flera organisationer som utvecklar datalager är en annan möjlig lösning. Även detta alternativ kräver tillgång på lämpliga organisationer. Att få tillgång till intressanta organisationer är troligtvis väldigt svårt då datalager är en strategiskt viktig del i företaget. En så potentiellt viktig del i en organisation är det troligtvis svårt att som utomstående få ta del av.

Nedan beskrivs metodvalsprocessen för de båda delarna i arbetet. Först beskrivs litteraturstudien och vilka synvinklar problemställningen kan studeras ifrån och vilken eller vilka som valts. Sedan beskrivs hur den empiriska studien kan utföras.

## 4.2 Relevanta synvinklar

Problemställningen skall främst besvaras genom en litteraturstudie av aktuella källor. Eftersom problemställningen skall besvaras genom en litteraturstudie blir inte metodvalet till denna del något större problem. En litteraturstudie genomförs genom en ingående studie av relevant litteratur. Det går dock att genomföra en litteraturstudie

på olika sätt beroende på vilket perspektiv och synsätt man utgår ifrån. Olika synvinklar som studerar samma problem får sannolikt olika resultat. Problemställningen har redan antagit synvinkeln att studera användandet av datalager ur mindre företags perspektiv. Utifrån detta perspektiv finns det ytterligare synvinklar att studera problemställningen ifrån. Några intressanta synvinklar beskrivs nedan.

Teoretisk eller praktisk synvinkel – Eftersom arbetet främst är en litteraturstudie kommer litteratur i området att studeras. Den litteratur som finns skriven inom datalagerområdet kan delas in i två huvudgrupper, teoretisk och praktisk. Den teoretiska litteraturen skapas i huvudsak inom den akademiska världen medan den mer praktiska litteraturen skapas inom näringslivet. De olika synvinklarna kompletterar varandra. Akademisk litteratur behandlar ofta tekniska frågor på ett mer övergripande, strukturellt plan, som systemarkitektur och frågeoptimering. Det är denna litteratur som är lättast att hitta. Mer praktisk litteratur som handlar om hur datalager verkligen används och vem som använder dem, är svårare att hitta. Vilken synvinkel litteraturstudien skall anta, teoretisk eller praktisk, är relevant då den litteratur som är mest intressant för att bäst kunna svara på frågeställningen bör väljas.

Stora eller små organisationer – Olika företag har olika förutsättningar. Detta kan bero på vilken bransch de arbetar i, hur konkurrensen ser ut eller hur stora de är. Ett företags storlek kan mätas i antal anställda, omsättning, marknadsandel etc. En intressant fråga är om företagets storlek påverkar dess inställning till datalager, eller dess möjligheter att införa ett datalager.

Ur organisationernas synvinklar är det troligtvis andra resonemang som är mer intressanta som exempelvis:

- Storleksmässig synvinkel - Hur stor måste en organisation vara, personalmässigt eller ekonomiskt, för att ha nytta av ett datalager?
- Ekonomisk synvinkel - Vilken ekonomisk situation befinner sig organisationen i? Vad är de beredda att satsa på ett datalager och vilken avkastning tror de att datalagret skall ge? Är det ekonomiskt försvarbart att använda ett datalager?.
- Teknisk synvinkel - Hur stor kunskap krävs för att kunna skapa och använda ett datalager? Är det tekniskt möjligt eller krävs ombyggnad eller nyinvesteringar i teknik? Vilka krav på befintlig eller ny teknik finns det?
- Tidsmässig synvinkel - Hur lång tid tar det att utveckla datalagret och hur lång tid tar det att få det att fungera tillfredsställande?
- Nyttomässig synvinkel - Har organisationen någon verklig nytta av ett datalager. I så fall vilken.
- Praktisk genomförbarhet - Är det överhuvudtaget genomförbart att starta ett datalagerprojekt? Vilka kunskapskrav krävs av användarna för att de skall kunna använda ett datalager?
- Framtida möjligheter – Hur utvecklas teknologin? Kommer det nya användningsområden, nya tekniska möjligheter, som till exempel webbteknologi, som ökar organisationers möjligheter att använda tekniken?

Mer jordnära synvinklar som dessa borde vara mer intressanta för organisationer som inför eller planerar att införa datalager. Hur de skall få datalagret att fungera är för

dem troligtvis viktigare än att de har den bästa arkitekturen eller den effektivaste frågeoptimeringen.

#### 4.2.1 Vald synvinkel

En snäv synvinkel med många begränsningar skulle ge en detaljerad inblick inom ett begränsat område. Att exempelvis intervjua ett antal mellanstora eller mindre företag om deras inställning till datalager skulle visserligen ge en djup inblick i dessa företags tankar. Det är dock osäkert om dessa tankar skulle vara representativa för hela landet eller hela branschen. Risken finns att många intressanta, men kanske för detta arbete ej relevanta, fakta skulle fångas. Ett intervjuat företag kan endast ge sin syn på frågan ur sitt perspektiv, detta är hela idén med intervjuer. Frågor och synvinklar som inte är av intresse för företagen men som kanske kunde vara av intresse för detta arbete skulle då gå förlorade. En litteraturstudie ger en bredare överblick över hela datalagerområdet på ett bättre sätt än intervjuer, men missar i sin tur mer handfasta bevis och åsikter från företag.

Införandet av datalager är en komplex process som i vissa fall kan omfatta hela organisationer. Att införandet berör stora delar av organisationer gör att en bred syn på området krävs. Denna bredd tror jag bäst kan fås genom en litteraturstudie. ”Case Studies”, det vill säga exempel på verkliga införande av datalager och vilka erfarenheter de erfarit, är en litteraturtyp som dessutom kan visa vilka problem som uppstått hos verkliga organisationer. Denna typ är erfarenheter skulle missas vid intervjuer av mindre organisationer som inte innehar dessa kunskaper.

För att kunna svara på problemställningen anser jag att det krävs en synvinkel som tar upp flera perspektiv. Det är många faktorer som spelar in då ett datalager utvecklas och jag tror inte att det går att svara på problemställningen genom att endast fokusera arbetet ifrån en enskild synvinkel.

Att studera problemet ur **små organisationers synvinkel** är genom problempreciseringen redan fastställt. Utöver detta anser jag det viktigt att även beakta dessa organisationers **nyttomässiga vinst** av att använda datalager. Finns det inga fördelar med att införa datalager är det ointressant att undersöka varför dessa organisationer inte använder datalager.

Det kanske största hindret för införskaffandet av datalager är den ekonomiska kraftansamling som krävs för att kunna utveckla datalager. Därför avser jag även att ta med den **ekonomiska aspekten**. Större företag har ofta större kapital att investera i projekt liknande datalagerutveckling. Mindre företag har troligtvis mindre pengar att lägga ut på investeringar, som datalager, som kan vara en viktig strategisk satsning men som inte ger en snabb eller säker avkastning. Denna osäkerhet kan göra att mindre företag är försiktigare med att starta ett datalagerprojekt. Även om mindre företag har tekniska och kunskapsmässiga möjligheter att använda datalager men inte har ekonomiska möjligheter att införskaffa det, är ett datalager ändå inte av intresse för mindre företag.

Tekniken utvecklas i snabb takt inom många områden. Två av de områden som utvecklas snabbt är datalagerteknologin och Internetrelaterad teknologi, även kallad webbtologi. Därför kommer även **framtida möjligheter** kopplade till webbtologi att tas upp som en synvinkel. Detta därför att webbtologin kan öka organisationers möjligheter att använda tekniken, och kanske främst för de mindre organisationerna som inte tidigare sett möjligheter med datalagerteknologin.

Övriga synvinklar kommer inte att studeras ingående i arbetet. Detta innebär inte att de är ointressanta eller inte påverkar organisationer i beslutet av att utveckla eller avstå från datalager. Motivet till att dessa synvinklar ej tas upp är att behålla fokus på några viktiga synvinklar istället för att försöka vrida synfältet och ta upp för många aspekter inom ett komplicerat område. Jag anser att de valda synvinklarna är viktigast för detta arbete.

Sammanfattningsvis betyder detta alltså att jag valt att anta flera, men långt ifrån alla, synvinklar att studera problemet ifrån. Den viktigaste synvinkeln är den storleksmässiga, det vill säga, att studera problemställningen ur mindre organisationers synvinkel. Denna var, som beskrivits tidigare, till stor del redan given i problemställningen. Jag avser även studera problemet ur de som jag ser det, viktigaste omkringliggande beröringspunkterna, det vill säga nyttomässiga, ekonomiska och framtida möjligheter, vilka jag tror är de aspekter som främst påverkar en organisations ställningstagande till om de skall försöka utveckla ett datalager.

#### **4.2.2 Relevant litteratur**

Vilken litteratur som är relevant beror på vilken synvinkel problemet studeras ifrån. Antar man en snäv synvinkel blir även urvalet av litteraturen mer begränsad. Detta är i sig inget problem om det finns mycket och intressant litteratur inom området. Mycket litteratur, kan dock tyda på att området är väl studerat av många personer som valt att dela med sig av sina åsikter. Bedrivs mycket forskning inom ett område kan problemet som studeras redan vara studerat av andra ur samma eller liknande synvinklar. Detta skulle göra att det egna arbetet förlorar betydelse och till stora delar förlorar sitt värde.

Litteratur från näringslivet, exempelvis *case-studies*, produktinformation och så vidare, kan vara partisk på så sätt att den skrivs ur syftet att sälja en produkt, en idé eller dylikt. Detta kan göra att litteratur från liknande källor inte ger en objektiv syn på hur saker förhåller sig. Exempel på detta kan vara vilken plattform och operativsystem som fungerar bäst med en viss datalagerprodukt etc. Litteraturen kan även vara mycket teknisk med många termer på en låg nivå, som till exempel vilket operativsystem som kan användas till vilket datalagerhanteringssystem, vilka datakommunikationslösningar som passar till dessa system och så vidare. Akademisk litteratur å sin sida tenderar att vara teknisk på ett mer generellt plan, exempelvis vilken arkitektur som skall användas, varför just denna arkitektur är bättre än en annan arkitektur, vilka algoritmer som skall användas för att optimera frågor och transaktioner i distribuerade databaser etc. Litteratur från näringslivet är ofta specifika för enskilda fall medan akademisk litteratur ofta generaliserar frågor. Om litteratur från näringslivet kan vara skriven för att sälja en produkt kan akademisk litteratur å sin sida vara skriven för att stödja en speciell teori eller modell. Detta är något läsaren måste tänka på då han eller hon kritiskt granskar en källa. Akademisk litteratur har dock en stor nackdel, och det är att den sällan tar upp företags syn på och användning av datalager. Det är denna syn som är intressant för detta arbete.

För att få en bra bild av vad som kan påverka användandet av datalager för mindre organisationer tror jag att en på sina ställen djup men ändå relativt bred litteratur krävs. Detta innebär att främst akademisk litteratur kommer att studeras därför att den är lättåtkomlig och oftast skriven utan ekonomiska vinklingar. Den litteratur som är skriven ur mer teknisk synvinkel kommer jag inte att lägga större vikt vid i

litteraturstudien eftersom den främst tar upp aspekter som jag inte anser är intressanta för att besvara frågeställningen. Exempel på dessa synvinklar är sådana som behandlar arkitekturdesign och frågeoptimering. Denna litteratur tar upp mycket programmeringsteori och variabeloptimering för hur ett datalager skall kunna användas på ett effektivt sätt. Denna kunskap anser jag inte är relevant för att kunna besvara problemställningen.

#### **4.2.3 Validering**

För att veta att all relevant litteratur inom området studerats när litteraturstudien väl genomförts måste författaren vara expert inom området. Då detta ej är fallet här måste andra sätt användas för att kunna säkerställa att arbetet är tillräckligt komplett för att några relevanta slutsatser skall kunna dras. Att rådfråga erfarna personer inom området om litteratur är ett sätt. Att leta så mycket litteratur som möjligt och studera dess referensförteckningar för att se när referenserna går i loopar och refererar till varandra, är ett annat. Möjligheten att hitta all relevant litteratur är troligtvis relativt liten. Anledningen till detta är att litteratur främst måste sökas och hittas under arbetets inledande skede. Litteratursökning tar ofta lång tid (Dawson, 2000), vilket gör att all litteratur som skulle kunna vara intressant med största sannolikhet inte kommer hittas.

Att veta om alla relevanta synvinklar för studien identifierats och att relevant synvinkel sedan verkligen valts kan ej helt säkerställas då min erfarenhet och mina kunskaper inom området fortfarande är relativt begränsade. Detta förstärks ytterligare då datalagerområdet dels är ett komplext område, dels ett brett område som påverkas av många faktorer och många delar inom en organisation vid exempelvis utvecklingsarbeten. Jag tror att de viktigaste synvinklarna berörts och att litteraturen som kommer studeras är tillräckligt komplett för att relevanta slutsatser skall kunna dras.

En fallgrop som måste undvikas i litteraturstudier är att läsaren inte får kategorisera en text utifrån de ramar han eller hon själv har eller bedöma en text enbart utifrån det ämne läsaren själv är mest bekant med (Hart, 1998). En läsare måste kunna placera en text i dess kontext i det ämne det skrevs. Det är därför viktigt att vara medveten om ens egna begränsningar och kunskaper.

Kritisk granskning av den litteratur som studeras är ett krav för ett väl genomfört arbete. Att kritiskt granska bygger på att läsaren inte bara läser och förstår det han eller hon läser utan även tänker på vad han eller hon läser (Dawson, 2000).

### 4.3 Relevanta undersökningsmetoder

Hur omfattande den empiriska studien skall vara kan diskuteras. Frågor som hur ingående den bör vara och hur noggrann den måste vara för att resultatet skall vara intressant och giltigt måste besvaras. Då denna studie endast skall komplettera den litteraturstudie som görs parallellt, kommer studien endast att ha en biroll. En ingående studie är då ointressant och heller ej genomförbar då arbetet har förutbestämda ramar främst i form av förbestämd tid. Målet med den empiriska studien är dessutom översiktlig och kommer troligtvis ej att finna alla de företag och organisationer som använder datalager i Sverige.

En undersökning som visar vilka företag och organisationer som använder datalager skulle ge en intressant inblick i vilka som verkligen använder datalager, och om det uteslutande är de största. Utländsk litteratur (Watson m.fl., 2001) visar att det främst är stora företag och organisationer som innehar datalager, eller närmare bestämt frånvaron av mindre företag som innehar datalager. Kunde undersökningen visa på att samma resultat gäller för Sverige skulle detta vara ett intressant konstaterande, även om denna undersökning ej fullständigt skulle kunna säkerställa detta faktum. Skulle resultatet peka mot att det även i Sverige är större organisationer som använder datalager skulle detta vara en ytterligare motivation för detta arbete.

En litteraturstudie kan ej utföras på så många olika sätt, däremot ett antal olika synvinklar. En undersökning kan däremot utföras på åtskilliga sätt. Det finns ett antal olika sätt att genomföra en empirisk studie på och nedan beskrivs de som skulle kunna vara intressant i detta arbete.

Intervjuer skulle antagligen ge den mest uttömmande och bästa bilden av mindre organisationers inställning till datalager. Nackdelen med intervjuer är att de kräver mycket planering, tid och tillgång på personer att intervjua. Ytterligare en nackdel är att resultatet är helt beroende på ett fåtal intervjuer. Skulle dessa intervjuer inte uppfylla de tänkta målen blir det svårare att dra några riktiga slutsatser.

Formulär är ett alternativ till intervjuer. Formulär skickas ut till de personer vars åsikter kan vara intressanta, varpå de förhoppningsvis fyller i formuläret och skickar tillbaka det. Ett alternativ till utskick är att dela ut dem personligen för att kunna vara med vid ifyllandet och förklara eventuella otydligheter och förtydliga frågeställningar. I denna studie skulle formulär vara ett intressant alternativ som skulle kunna ge bra resultat. Nackdelen med formulär är svårigheterna med att skapa ett tydligt formulär som ej innehåller några ledande frågor (Patel & Davidson, 1994). Dessutom tillkommer det faktum att det är den svarandes tolkning av frågorna som besvaras, inte alltid skaparens tolkning av frågorna.

Vid insamlandet av data förekommer det alltid en viss osäkerhet. Denna resulterar i att det inte helt säkert går att säga att det som skulle sökas har sökts och det som hittats var det som skulle hittas. Detta är en faktor jag anser inte går att undvika. Då studien skall vara mycket begränsad och endast fungera som komplettering till litteraturstudien får den inte ta för lång tid. Resultatet skulle bli mer tillförlitligt om exempelvis intervjuer utfördes, men detta är ej praktisk genomförbart då jag anser att ett större antal organisationer skulle behöva intervjuas för att få ett tillförlitligt resultat.

Behandlingen av insamlad information är en viktig del av arbetet. Även om resultaten av en undersökning, exempelvis svarsfrekvensen för ett formulär är bra, måste informationen behandlas och sammanställas på ett korrekt sätt för att ha ett värde och

gå att använda. Mycket information kräver mycket tid att bearbeta (Patel & Davidson, 1994). Detta innebär att studien måste hållas begränsad då det endast finns lite tid till att bearbeta resultaten.

#### **4.3.1 Vald undersökningsmetod**

Då den empiriska studien endast skall ge en övergripande bild av vilka organisationer som använder datalager tror jag att några enkla frågor i ett e-postmeddelande bör fungera optimalt utefter de förutsättningar som råder. E-post är ett medium jag förutsätter att organisationer som innehar eller utvecklar datalager, använder sig av. Frågorna bör ställas till organisationer som själva har datalager eller till organisationer som utvecklar och underhåller datalager åt andra.

Anledningen till att inte intervjuer valts är att jag anser att de ej är praktiskt genomförbara eller mycket svåra att utföra på grund av tidsbrist om de måste göras efter eller i kombination med litteraturstudien. Anledningen till att inte mer avancerade formulär används är att de kräver mycket förarbete, sammanställningen efter kräver mycket arbete samt att de inte finns någon garanti för att för omfattande utskick verkligen besvaras.

#### **4.3.2 Validering**

Det finns flera sätt att validera resultatet från en undersökning. För att veta att rätt saker studeras och att det görs på rätt sätt kan förstudier utföras. I dessa förstudier bör även frågor och formulär testas så att de uppfyller de krav som ställs på en bra och korrekt undersökning (Patel & Davidsson, 1994). För att säkerställa resultatet kan kontrollgrupper användas, kompletterande undersökningar kan göras och så vidare. Då studien är relativt enkel och övergripande anser jag inte att någon avancerad komplettering av resultaten krävs.

Det största problemet jag kan förutspå är svårigheterna med att få svar från organisationerna och företagen. Datalager beskrivs ofta som ett mycket avancerat konkurrensmedel, vilket gör att vissa organisationer kanske inte gärna vill skylta med att de har och använder sig av ett datalager. Att organisationer inom IT-området är mycket upptagna, gör att de kanske ej anser sig intresserade eller ha tid att svara på utskick och undersökningar. Detta faktum går heller ej att förebygga eller skydda sig från.

Ett annat problem som ej får förbises är det faktum att det inte finns någon universell definition på vad ett datalager egentligen är. Detta gör att olika organisationer definierar datalager på olika sätt. Detta problem är svårt att komma till rätta med i en översiktlig studie som denna. För att lösa detta problem behövs en djupare intervju om vad organisationerna använder sina datalager till och hur de är uppbyggda.



## 5 Genomförande

Denna del av rapporten beskriver själva genomförandet av arbetet. Då arbetet är uppdelat i två huvuddelar har jag valt att även dela upp genomförandekapitlet i dessa delar. För båda dessa delar inleder ett stycke om vilka förberedelser som gjorts inför själva genomförandet. Därefter beskrivs genomförandet. Resultatet av arbetet beskrivs i senare delar av rapporten. Slutligen görs en kort värdering av materialet från genomförandedelen.

### 5.1 Litteraturstudien

Denna del beskriver hur litteraturstudien genomfördes.

#### 5.1.1 Förberedelser

De förberedelser som gjordes inför litteraturstudien var att den litteratur som antogs kunde besvara den problemställning som ställts upp i problempreciseringskapitlet, samt de synvinklar som valts i metodkapitlet söktes. En del av källorna som använts i bakgrundsdelen kunde även användas i genomförandedelen. Material som användes i bakgrundsdelen gav där en översiktlig bild över området. I de fall samma litteratur var lämplig att studera ytterligare i litteraturstudien, gjordes detta ur en ny synvinkel, det vill säga de som valts i metodkapitlet. Processen att söka material pågick kontinuerligt under hela arbetet, vilket gjort att chanserna att hitta värdefull litteratur ökade.

#### 5.1.2 Genomförande

Genomförandet av litteraturstudien utfördes genom att studera ett flertal arbeten. Dessa var huvudsakligen vetenskapliga artiklar från olika artikeldatabaser. Anledningen till att jag valt artikeldatabaser som främsta källa är att dessa är relativt lättöverskådliga i kombination med att de oftast innehåller olika artiklar från många olika områden. Artikeldatabaser är även ett bra medium för att hitta de senaste arbetena som skrivits inom olika områden, detta då tillgängligheten till databaserna via Internet är stor var man än befinner sig i världen.

I litteraturstudien har även traditionella utgivna verk studerats, det vill säga verk i bokform. Dessa är inte lika aktuella, då det ofta tar lång tid att skriva en bok, men ger istället ofta en bredare bild av området som studeras, något som i sig kan vara värdefullt.

### 5.1.3 Värdering av material

En studie av vilken litteratur som helst innebär en tolkning av det lästa materialet. Att ett material tolkas kan innebära att läsaren uppfattar materialet på ett sätt som författaren inte tänkt sig. Detta kan både vara en fördel och en nackdel. Feltolkningar av en text kan undvikas genom att läsaren skapar sig en ingående kunskap om området. På detta sätt kan läsaren sedan bättre förstå hur författaren ser på saker och ting, varför argumenten ser ut som de gör etc. Det vore fel att påstå att jag under arbetets tid har kunnat skaffa mig en erfarenhet som är tillräcklig för att kunna förstå och tolka alla författare och arbeten på ett korrekt sätt. Det tar lång tid och mycket arbete för att utveckla en bra kompetens i att studera forskningslitteratur (Hart 1998). Jag tror att jag har tillräckliga kunskaper för att någorlunda kunna tolka artiklarna på ett riktigt sätt. Jag anser därför att det insamlade materialet är tillräckligt för att kunna svara på problemställningen.

## 5.2 Den empiriska studien

Denna del beskriver hur den empiriska studien, det vill säga hur utskicket av företagsfrågorna, genomfördes.

### 5.2.1 Förberedelser

Studiens huvudmål var att skapa en kompletterande bild av hur det förhåller sig i verkligheten, det vill säga om verkligheten ser ut som den beskrivs i artiklar och dylikt material. I metoddelen beskrevs att utskicket skulle ske till företag och organisationer som utvecklar datalager. Det vore även intressant att få deras syn på frågor utöver om de känner till företag i Sverige som innehar datalager. Då studien är kompletterande och inte får ta för lång tid är en mer uttömmande undersökning utesluten, men jag ansåg det dock intressant att få företags syn på mindre företags chanser att få tillgång till datalagerteknologi. Detta skulle ge utvecklarens åsikter på min frågeställning. Vilka frågor som skall ställas är en balansgång mellan att göra en enkel och lättbehandlad undersökning mot att göra en mer givande men även mer krävande undersökning. Ovanstående innebär att jag utöver den ursprungliga frågan, att söka efter svenska datalagerinnehavare, lade till en ny fråga. Jag anser att studien trots denna breddning är tillräckligt begränsad för att inte störa det övriga arbetet, det vill säga litteraturstudien.

De frågor som slutligen valdes var:

- **Vilka Svenska företag har ni hjälpt att utveckla datalager?**
- **Tror ni att mindre organisationer kommer att kunna använda datalager inom en snar framtid?**

När frågorna var fastställda valdes lämpliga företag ut och kontaktades. Att hitta rätt företag som utvecklar datalager och att sedan hitta lämpliga personer i dessa företag var den svåraste delen av den empiriska studien. Då jag inte varit i kontakt med datalagerutvecklare tidigare samt att datalagerteknologin är relativt ny var dessa inledande kontakter inte helt självklara.

### **5.2.2 Svarsfrekvens**

Elva förfrågningar skickades ut till företag som frågade om de ville delta i studien. Av dessa skickades sedan själva frågorna ut till sju av företagen. Av de utskickade frågorna returnerades fem. Detta gav en total svarsfrekvens på cirka 54 %.

Det kan påpekas att det var svårast att få svar från de största multinationella företagen. Hos dessa var det även svårast att hitta adresser direkt till personer. Dessa organisationer kontaktades således genom mer allmänna e-postadresser.

### **5.2.3 Värdering av material**

Viktigt att konstatera är att frågorna innehåller en viss del tolkning vilket innebär att frågorna tolkas olika av olika personer. Då frågorna ställs av en person som ej arbetar med datalager till personer som är experter på datalager, uppstår naturligtvis ibland tolkningsproblem. Som beskrivits i metodkapitlet kan detta problem lösas genom att utföra intervjuer eller mer avancerade formulär där frågeställningar kan klargöras och förtydligas. Detta är inte möjligt i ett kort e-postmeddelande. Detta tolkningsproblem är jag medveten om då resultaten av svaren senare skall analyseras. De insamlade svaren gav, även om de ej kvantitativt var många, några intressanta och överraskande svar. Jag anser att jag har erhållit information som kan användas för att besvara problemställningen.

## 6 Analys och resultat

I detta kapitel presenteras och analyseras informationen från den genomförda litteraturstudien och den empiriska studien. Kapitlet inleds med en presentation av materialet från litteraturstudien och följs av materialet från den empiriska studien. Kapitlet avslutas med jämförelser mellan förväntat resultat och verkligt resultat.

### 6.1 Material från litteraturstudien

Efter att ha genomfört litteraturstudien kan det konstateras att det finns ytterst lite litteratur som tar upp aspekter som direkt kan svara på frågeställningen. Detta innebär att olika verk måste studeras för att jag skall kunna dra egna slutsatser av hur författarnas åsikter sedan påverkar min problemställning.

Denna brist på litteratur som var direkt ”lämplig” för eller direkt applicerbar på mitt arbete, det vill säga litteratur som beskriver hur mindre organisationer kan påverkas av ett datalager, var en av anledningarna till att jag fann problemställningens vinkel intressant till att börja med.

Det är svårt att särskilja på de olika synvinklarna när materialet beskrivs. Synvinklarna har lika aspekter och rör ofta närliggande synvinklar. Jag har försökt att placera materialet i lämpligaste stycke.

#### 6.1.1 Storleksmässig synvinkel

Till för bara några år sedan kunde endast de största företagen bygga och använda stora databaser. I dag kan även de minsta företagen använda stora databaser (Fayyad, 1998). Men till skillnad från de största företagen som kan utföra dataanalyser inom den egna organisationen har de mindre företagen inte samma förutsättningar när även de måste utföra analyserna inom organisationen (Fayyad, 1998). Fayyad (1998) skriver att dessa mindre företag inte anser att de verktyg som finns på marknaden är lämpliga för ickeanalytiker. Sannolikt är de flesta anställda på mindre företag ickeanalytiker.

Alla organisationer behöver information för att ta viktiga beslut, och denna information kommer oftast från analyser av historisk data och identifierandet av trender (Gallegos, 1999). Gallegos (1999) beskriver två frågor som kan identifiera ett företags behov av ett datalager. Den första är om befintlig data kan nås direkt och effektivt. Den andra frågan är om data från olika befintliga system matchar. Då ett datalager främst är ett strategiskt instrument måste ett företag noga analysera och utvärdera sina egna mål innan de kan ta ställning till om ett datalager är vad de verkligen behöver (Gallegos, 1999). Olika organisationer har naturligtvis olika förutsättningar och olika mål.

Det uttrycks sällan i litteraturen hur organisationers storlek påverkar själva användandet av datalager. Detta faktum leder till att större fokus lagts på de nyttomässiga och ekonomiska aspekterna av datalageranvändning. Frågan är då varför företags storlek inte anses vara relevant då datalager diskuteras. Kanske detta beror på att det är en vanlig uppfattning att mindre företag inte har lika mycket, om någon, nytta av datalager och litteratur därför skrivs med stora företag i tankarna. Kanske

storleken på organisationen inte spelar någon roll över huvud taget. Orsaken till att företags storlek sällan omskrivs i akademisk litteratur har ej gått att fastställa i den litteratur som studerats.

### 6.1.2 Nyttomässig synvinkel

Om det var svårt att hitta litteratur ur storlekssynvinkeln så var det desto lättare att hitta litteratur som berörde vilken nytta ett datalager kan göra för en organisation.

Att datalager främst är till för dataanalys och beslutsstöd är konstaterat. Att datalagret även har potential att fungera som en central informationskanal i företag (Jarke m.fl., 1999) och verkligen gör detta (Watson m.fl., 2001) är ett intressant konstaterande. Detta innebär att datalagret i många fall inte används för att förändra företagets uppträdande på marknaden, som exempelvis att förändra produktmixen utefter köptrender, utan att det används till att förbättra informationsflödet i företaget (Watson m.fl., 2001). Detta nya användningsområde kan eventuellt bero på att tekniken är ny och att företag ser nya möjligheter med datalagret. En annan möjlighet är att datalagret är så komplicerat att företagen inte kan använda det på det sätt det var tänkt att användas.

Att datalager gör mycket nytta där de fungerar är de flesta överens om, men som med många nya tekniker och produkter måste man vara medveten om alla risker som följer med. Informationsexplosionen som skedde på 1990-talet har påverkat alla delar av samhället. Utvecklingen av Internet möjliggjorde många nya användningsområden av befintliga applikationer. Att Internet, det globala tele- och datanätet, innebar nya möjligheter för distribuerade databaser, och diverse kommunikationsmedium, innebar möjligheter till stora ekonomiska och tidsmässiga vinster. Harmon (1998) anser att datalagret och Internet tillhör de viktigaste delarna som utvecklats de senaste åren. Medan de flesta företag redan blivit varse om de fördelar, men även de problem Internet för med sig, är säkerhet i datalagersammanhang ett inte lika debatterat område. Datalagertillverkares oro för att det verkligen finns allvarliga brister kanske bidrar till att dessa frågor inte lyfts fram. Identifieras allvarliga brister måste tillverkarna åtgärda dessa, något som kanske kan bli dyrt för dem. Även om inte mjukvarutillverkare försöker tysta ner problem som upptäcks, så är tekniken ny och då upptäcks nya problem och möjligheter oftare än i mogen och väl utvecklad teknik och mjukvara. Många områden behöver utvecklas och säkerheten är ett av de områden som ännu inte har fått tillräcklig uppmärksamhet i litteraturen. Detta är anledningen till att texten nedan fokuseras på säkerhetsaspekter även om de ej direkt åsyftas i valet av synvinklar.

Att datalagers är till nytta för företag är fördelen med dem, men man får ej glömma att datalagret i olyckliga fall även kan orsaka skada för ett företag. Datalagret bygger på att det är ett slags centrallager för all intressant data inom en organisation. Detta innebär att just datalagret är det ideala målet för eventuella företagsspioner och hackers (Harmon, 1998). Viktig information kan även läcka ut via anställda som har tillgång till information de inte skall ha tillgång till. Datalagrets ena hörnsten består i det nätverk som all data skickas via. Detta innebär att alla säkerhetsproblem som finns för nätverket även påverkar datalagret. Datalager gör viktig information mer lättillgänglig, inte endast för auktoriserade användare utan även för inkräktare.

Innan Internets intåg var de stora företagens nätverk praktiskt taget skilda från omvärldens datanät samtidigt som omvärlden inte bestod av lika många potentiella hackers som idag. Även företags ekonomiska läge har gjort att konkurrensen drivits

till sin spets där alla konkurrerande företag ligger och bevakar varandras rörelser. Harmon (1998) beskriver detta som en teknologisk barriär mellan människor med ohederliga avsikter och för företagen värdefull information. Denna barriär är idag kraftigt reducerad. Företag har idag problem med inkräktare i sina nätverk även om många av olika skäl väljer att aldrig offentliggöra sina elektroniska inbrott. Chen & Frolick (2000) anser även de att säkerhetsfrågorna angående datalager bör ses över. Chen & Frolick (2000) föreslår att de säkerhetsfrågor som berör Internet även bör vidgas till att omfatta företags databaser, webbservrar och datan som transporteras i deras nätverk.

Ofta när denna säkerhetsdebatt tar fart ställs administratörer, utvecklare och liknande personer för problemet med avvägningen mellan säkerhet och tillgänglighet. Ett helt säkert datasystem blir näst intill oanvändbart med alla sina säkerhetskontroller. Ett datasystem som är helt och hållet anpassat efter användbarhet blir istället en mardröm för systemansvariga att försöka skydda från elektroniska inbrott. Organisationer behöver ta ställning till vilken information och vilka analysverktyg som kan delas. Detta kräver i sin tur en omfattande analys av vilken information ens användare, både internt i det egna nätverket och externt i främmande nätverk, bör och behöver ha tillgång till (Chen & Frolick, 2000). Exempel på data och verktyg som stödjer det egna företagets strategiska beslut bör enligt Chen & Frolick (2000) inte göras tillgängliga för yttre användare.

Harmon (1998) beskriver diverse säkerhetsanordningar som redan existerar hos databaser, exempelvis brandväggar, vyhantering som säkerhet och in- och utloggning i systemen. Harmon (1998) menar dock att dessa inte är tillräckliga för system med höga säkerhetskrav. För system med högre säkerhetskrav är kryptering av datan nödvändig och det enda egentliga alternativet. Kryptering innebär att information krypteras då den skickas ut i nätverket så att eventuell uppsnappad data inte skall kunna tolkas av en tredje part. Dessvärre förekommer det problem med kryptering av data i databaser (Harmon, 1998). Detta då data i databaser måste bibehålla sin interna struktur då den förflyttas, vilket innebär att krypteringen inte kan användas fullt ut.

Att säkerhetsfrågor är högprioriterade i företag borde vara en självklarhet. Det är däremot inte helt säkert att alla företag känner till problemens storlek. Informationstekniken omfattar ett brett spektrum av områden och är ofta komplicerat. Ett datalager kan visserligen vara till stor nytta, men det är viktigt att inte glömma eventuella fallgropar, som dålig säkerhet i datasystemet, som inte direkt har anknytning till datalagret.

Företags datasäkerhet bör öka med dess storlek. Detta skulle betyda att stora företag har högre säkerhet då de har mer information att skydda. Hur medvetna mindre företag är om sin datasäkerhet vet jag ej, men den kan misstänkas vara lägre än för de största företagen. Hur viktigt datasäkerhet är för ett datalager är upp till företaget självt. Det bör dock vara intressant för mindre företag vilka risker ett osäkert nätverk kan utgöra för ett eventuellt datalager. Att behöva förbättra säkerheten i nätverket tillhör kostnader som förmodligen är en oförutsedd och en inte välkommen kostnad för ett företag som planerar att införa ett datalager.

Skulle liknande problem med säkerheten uppstå i framtiden kan detta bidra till att höja priserna på mjukvara och hårdvara samt även kanske kräva ett nytt säkerhetstänkande. För att skydda sig behövs mer utvecklad programvara och kanske även nya tekniska hjälpmedel för att säkra nätverken. Detta är något som då inte bara skulle påverka datalagret utan även andra nätverksberoende funktioner.

Säkerhetsproblem är dessutom oberoende av företagets storlek även om problemen antagligen ter sig olika för olika företagstyper.

En ökning av datalagerrelaterade datainfrång förutsätter att det verkligen finns någon som anser det lönsamt att spionera i andra datasystem. Detta beror naturligtvis på företagets storlek, vilken bransch företaget verkar i etc. Industrispionage kanske blir mer och mer inbjudande om offret redan sammanställt all viktig information och lagt upp den på fat.

Alla kostnadshöjningar av datalagerkomponenter slår hårdast mot de minsta företagen som har mindre resurser att lägga på liknande saker. Detta gör säkerhetsfrågor även till en ekonomisk fråga.

### 6.1.3 Ekonomisk synvinkel

I litteraturen beskrivs det faktum att det finns datalagerprodukter på marknaden, men det beskrivs litet om vilken, hur omfattande dessa produkter är eller vad de kostar. Det var de största multinationella mjukvaruföretagen som var först med att utveckla produkter till datalager, men senare följde även specialanpassade mindre företag efter. Baum (1997) skriver att datalagerprojekt är rena guldgruvan för återförsäljare av datalagerprodukter. Datalagerprojekt kräver många olika sorters specialiserade kunskaper och produkter för att utveckla, implementera och underhålla ett datalager. Även om det börjar komma ut mer eller mindre kompletta paketprodukter så är de flesta datalager skräddarsydda produkter sammanfogade av diverse hårdvaru-, mjukvaru- och datakommunikationskomponenter (Baum, 1997). Kan inte en komplett datalagerlösning erbjudas ett företag så blir den ekonomiska bedömningen naturligtvis mycket svårare.

För att bättre kunna säkerställa att ett projekt kommer att lyckas används ofta erkända metoder eller processer. Dessa underlättar arbetet och erbjuder ett strukturerat arbetssätt, vilket gör att arbetet blir lättare att följa. Srivastava & Chen (1999) anser att det finns en stor brist på processer och verktyg för att utveckla datalager och ladda dessa med data. Om det inte kommer till bra processer och verktyg tror Srivastava & Chen (1999) att den breda spridningen av datalager kraftigt kommer att hindras. Srivastava & Chen (1999) anser att det finns bra tekniker för att lagra själva datan samt för att utföra analyser av datan, men att det saknas tillräckligt många och bra produkter för att skapa datalagret som helhet. De säger att utvecklingsarbeten av datalager som skett *ad hoc* oftast tagit lång tid, varit mycket problemfyllda och ofta resulterat i återvändsgränder. Detta skulle kunna underlättas genom skapandet av övergripande metoder med kompletterande tekniker. Finns det få eller inga erkända metoder att följa blir datalagerprojekt mer riskabla att utföra, något som kan bidra till att de ekonomiska intressenterna drar sig ur.

En intressant aspekt ur den ekonomiska synvinkeln är hur datalager skulle kunna hjälpa ekonomiavdelningar i deras arbete. Det som egentligen var av intresse ur den ekonomiska synpunkten i detta arbete, var den hur kostnaderna för att utveckla ett datalager skulle påverka om ett företag väljer att fortsätta utvecklingsarbetet eller ej. Jag väljer ändå att nämna denna vinkling då den är relativt okänd och lite litteratur finns skriven om den, men som jag tror har stor potential.

Att datalager kan användas till att analysera trender, köpbeteenden etc. samt att det kan fungera som en speciell informationskanal mellan avdelningar har beskrivits tidigare. Det har även beskrivits att avdelningar som tidigare stod för huvuddelen av

datasystemet var ekonomiavdelningarna som använde datasystem till budgetering, kalkylering och bokföring. Sedan dess har ekonomiavdelningarna antagit en låg profil, om ej oviktig, ur datasystemssynpunkt. Att datalager skulle kunna användas för att effektivisera ekonomiavdelningarnas arbete är en möjlighet som inte utforskats nämnvärt (Olsen & Cooney, 2000). Olsen & Cooney (2000) skriver att ekonomiavdelningar mer än någonsin integrerar informationstekniken i sitt arbete. Trots detta har inte datalageranvändning riktigt tagit fart inom ekonomivärlden. Detta stöds till viss del av Watson m.fl. (2001) som i sin undersökning beskriver att det av de avdelningar som berör ekonomi främst är marknads- och försäljningsavdelningar som stödjer utvecklande av datalager. Efter marknads- och försäljningsavdelningar är det finansieringsavdelningar och därefter bokföringsavdelningar som oftast är pådrivande krafter i utvecklingsarbeten av datalager (Watson m.fl., 2001).

Skulle ett datalager kunna hjälpa till att effektivisera och förenkla arbetet för ekonomiavdelningar i större utsträckning än det sker idag skulle detta säkert vara välkommet, oavsett om det gäller stora eller små företag.

#### **6.1.4 Framtida möjligheter**

Här beskrivs framtida möjligheter med teknologin. Fokus har legat på webbteknologi och dess möjligheter att påverka utvecklingen av och användandet av datalager.

En av de aspekter av datalager som ofta berörs inom litteraturen är hur datalager kan kombineras med Internet. Detta område sträcker sig från hur webbteknologi kan användas för att bygga upp och distribuera tjänsterna till användarna till mer handfasta områden som hur webbläsare kan användas för att underlätta användares åtkomst till datalagret.

Anledningen till att jag tar upp webbaserade datalagerteknologi är att webbteknologin idag är relativt utbredd och är något som används av många företag runt om i världen. Web-teknologi är något som jag tror kan komma att påverka datalagerteknologin starkt framöver. Att använda delar av befintlig teknik skulle innebära att användare redan är bekanta med tekniken samt att kostnaderna för att utveckla blir lägre om redan befintliga delar i system kan användas. Kan befintliga delar användas är detta ett intressant alternativ för företag som inte anser sig ha råd att utveckla ett datalager från grunden.

Chen & Frolick (2000) skriver att de flesta företag idag är rika på data men fattiga på information. Web-baserade datalager är en av de hetare trenderna idag (Chen & Frolick, 2000). Web-baserade datalager är datalager där bland annat tillträde till datalagret, analysering och distribution av informationen sker genom nätverken med hjälp av webbläsare. Den explosionsartade utvecklingen av datalagerteknologi och Internetbaserad teknologi ökar organisationers intresse för webbaserad datalagerteknologi (Chen & Frolick, 2000). Gray & Watson (1998) säger att webbläsare är det huvudsakliga verktyget för användare att nå datan medan Chen & Frolick (2000) anser att dagens datalager arbetar i en isolerad miljö eller client/servermiljö med grafiska användargränssnitt. Utnyttjande av webbteknologi utnyttjar även all forskning och utveckling som sker inom webbläsarområdet. Genom detta blir gränssnittsteknologi som Java, ActiveX etc. tillgängliga, vilket kan underlätta användandet av datalager.



Anledningen till att webbaserad datalagerteknologi är intressant är enligt Chen & Frolick (2000) att:

Client/server-infrastrukturen som dagens datalager använder sig av är dyr att utveckla och underhålla.

Ett enda sorts användargränssnitt är inte längre tillräckligt, speciellt som andelen mobila användare ökar kraftigt och beräknas öka även framöver.

Systemkompatibilitet är ett ständigt problem i dagens client/serversystem. Flera olika system ökar kostnaderna för administration och underhåll. Web-baserad datalagerhantering skulle kunna lösa dessa problem genom att erbjuda en lösning som innehåller en liten webbläsare på klientsidan medan serversidan består av en webbserver som står för allt in- och utflöde av information. Till webbservern är även en applikationsserver och datalagret kopplat.

En sådan lösning skulle enligt Chen & Frolick (2000) ge flera fördelar jämfört med vanliga datalagerlösningar:

Bättre åtkomst för främst mobila användare som då inte behöver ha direktkontakt med aktuellt nätverk. Även åtkomst för kunder, leverantörer och andra intressenter underlättas.

Plattformsberoende datalager, vilket skulle underlätta distributionen av datan.

Lägre initialkostnader och skötselkostnader då klienterna endast behöver stödja en webbläsare, vilket ger lägre kostnader. Denna lösning skulle ge en centraliserad lösning, vilket skulle minska underhåll av klienterna etc.

Web-baserad datalagerteknologi är dock inte helt problemfri. Chen & Frolick (2000) beskriver några problem som måste lösas med webbaserade datalager.

Att erbjuda datalagertjänster via Internet skulle tynga system, vilket gör dem långsammare. Utöver detta är det många faktorer som påverkar överföringshastigheter över Internet så som hur mycket trafik det är på nätet, elektronisk överhörning mellan ledare, elektroniskt brus, avståndet signalerna skall färdas etc.

Det förekommer en viktig säkerhetsaspekt när webbaserad datalagerhantering skall diskuteras. Datasäkerhet i interna nätverk, så kallade intranät, är en viktig fråga för många företag, men om datalagerteknologitjänster skall erbjudas via Internet blir säkerhetsfrågan både större och mer komplicerad. Dessa problem anser jag inte främst ligga inom datalagerområdet utan mer inom datakommunikationsområdet.

Datalager är relativt nya och ännu under utveckling. Att kunna bygga ett datalager för att kunna vidareutveckla och anpassa det i framtiden utefter nya behov anser jag är självklar del i arbetet. Även Gallegos (1999) anser att ett datalager måste utvecklas med en arkitektur som är tillräckligt flexibel för att möta både nuvarande och framtida krav. Även om Gallegos (1999) främst menar hur datan i datalagret är strukturerad, är en flexibel arkitektur viktig ur andra synvinklar. Att utforma system som enkelt kan utvecklas i framtiden kanske ännu inte är av intresse då dagens företag fokuserar på att få sina datalager att fungera idag. Kanske kommer många av dagens datalager att skrotas inom några år för att ge plats för en ny generation datalager.

Flexibilitet behövs även på lägre nivå. Användares krav på datalagret förändras och användare byts ut i företag. En lösning på detta är ett effektivt intranät, vilket enligt Gallegos (1999) kan stödja saker som användarsupport och förbättra kommunikationen mellan människorna i företaget. Att använda nätverket för att förbättra kommunikationen mellan de anställda är en kommunikationsfråga som måste lösas av företagsledningen.

## 6.2 Material från den empiriska studien

Nedan presenteras resultatet av den empiriska studien. Studien utgjordes av två frågor. Då det är ett relativt begränsat antal svar väljer jag att presentera resultatet på varje fråga för sig för samtliga företag. Företagen benämns Företag 1 till 5.

### 6.2.1 Fråga ett

- **Vilka Svenska företag har ni hjälpt att utveckla datalager?**

*Lämnar ni (datalagerutvecklarna) ut sådana uppgifter eller är de klassificerade? Känner ni till andra företag eller organisationer som använder sig av datalager i Sverige som ni ej varit med och utvecklat?*

Företag 1 har utvecklat cirka sextio stycken datalager men de lämnar ej ut namnen på dessa. Företag 1s kunder utvecklar något som de betecknar som datalager.

Företag 2 lämnar av princip ej ut uppgifter om enstaka företag. Under de senaste tre åren har dock företag 2 genomfört sex datalagerprojekt. Av dessa har fyra bestått av att designa och bygga datalagret, medan de två sista har omfattat analys av befintliga datalager. Fyra av dessa projekt har berört börsnoterade företag.

Företag 3 levererar både produkter och tjänster inom datalagerområdet. De lämnar ej ut uppgifter om kunder utan deras medgivande.

Företag 4s datalagerprodukter används av bland annat SEB, Folksam, Skandia, Handelsbanken, Ericsson, SKF, Alfa Laval, ABB, Ceralia, Pripps/Falcon, Spendrup, Lindex, DHL, Bonnier, Fritidsresor, Stena Line, Vägverket, CSN, SAF, diverse kommuner och landsting, AstraZeneca, Pfizer, LSAB, Tamro, Apoteket, SCA, Korsnäs, StoraEnso, Telia, Glocalnet och Eniro. Dessa är stora företag med upp till tusentals användare. Utöver dessa finns det även fler mindre företag som använder Företag 4s produkter. Företag 4 lämnar inte ut mer specifika uppgifter av typen vilka applikationer som kunderna använder och vad de används till. Att ytterligare uppgifter ej lämnas ut är till stor del konkurrensmässiga skäl. Företag 4 beskriver att det förekommer en hård konkurrens på datalagerområdet, både mellan datalagerutvecklare, men kanske främst mellan deras kunder.

Företag 5 lämnar ej ut enskilda företag men säger att de utvecklat datalager åt större svenska fastighetsbolag. En organisation som enligt Företag 5 använder datalager, men som de ej varit med och utvecklat, är Investor.

Företag 6 är en av de större datalagerleverantörerna i Sverige. Företaget använder ett flertal olika datalagerprodukter till sina datalager. Företagets kunder befinner sig främst bland landets 200 –300 största företag, men de har även mindre kunder. Företaget lämnar inte ut namnen på sina kunder utan deras medgivande. Anledningen till detta är att datalager ses som en affärsstrategisk investering för deras kunder.

### 6.2.2 Fråga två

- **Tror ni att mindre organisationer kommer att kunna använda datalager inom en snar framtid?**

*Krävs det för avancerade kunskaper för att mindre organisationer skall kunna använda datalager? Kostar det för mycket pengar att utveckla eller köpa produkter till dem? Finns det ingen mjukvara som är anpassad för mindre organisationer?*

Företag 1 tror att mindre organisationer kommer att kunna använda datalager inom en snar framtid. De anser att det inte krävs speciellt avancerade kunskaper för att kunna använda ett datalager, men att det behövs en bra konsult som hjälper företaget med datalagret. Det finns enligt Företag 1 många program som är anpassade för mindre organisationer och dessa är inte dyra.

Företag 2 anser att mindre organisationer egentligen har större relativ nytta av datalager än stora företag. De anser att det finns mycket prisvärda lösningar, speciellt för företag som själva kan lägga ner arbete på datalagret. Ett exempel ges som beskriver att ett väl fungerande datalager kan skapas för en organisation med omkring 300 anställda för några tiotusentals kronor i licenser utöver befintliga system. Utöver detta tillkommer konsulttid, vilket varierar stort från fall till fall. För detta system skulle cirka fem analytiker behövas för att handha datalagret.

Företag 2 identifierar två centrala frågor för en liten organisation som planerar att införa datalager:

- Hur ser min organisation och dess kommersiella verklighet ut?
- Hur ser jag till att intern data går att samköra med extern data (ekonomisk utveckling, geografiska data etc.)?

Företag 3 anser att det finns mjukvaror som passar de flesta företag. Detta då mjukvaran inte anpassas efter kundens storlek utan mer efter hur stora datamängder som skall hanteras och vilka frågeställningar kunden vill ha svar på.

Företag 4 anser att alla företag kommer att behöva någon form av datalager i framtiden. Företag 4 påpekar några synpunkter från en rapport om problem som finns med informationsbearbetning i svenska organisationer.

- Det är svårt att motivera personal att dokumentera kundinformation.
- IT-stöd för analys av feedback från kunder saknas.
- Alltför stora resurser läggs på att manuellt sammanställa rapporter med data från olika källor.
- Personalen kan inte själva ta fram information från ERP-systemet (*Enterprise resource planning system*).
- ERP-systemet kan inte leverera lättöverskådliga rapporter som ger en sammanhållen bild av hela verksamheten.

Företag 5 tror inte att mindre företag kommer att utveckla och använda datalager inom en snar framtid. Anledningen till detta är enligt företag 5 att de är för dyra att utveckla och företagsanpassa. Man tror att det är tveksamt att standardlösningar kan fylla de behov ett företag har avseende ett datalager. Företag 5 anser dock att det inte krävs några avancerade kunskaper för att sedan använda datalagret.

Företag 5 påpekar även problemet med definitionen av datalager. Är det enbart databasen och OLAP-verktygen som omfattas av definitionen eller skall även gränssnitt omfattas etc.?

Företag 6 säger att det redan i dagsläget finns en rad mindre organisationer som antingen har eller håller på att bygga upp datalager. Med mindre företag menar företag 6 organisationer med 250-500 anställda. Kostnaden för ett datalager är olika i varje fall men kan sägas ligga mellan 250-500 tusen kronor i de minsta utföranden. I detta ingår då konsultarvoden, hårdvara och mjukvara. Företag 6 säger även att kostnaden för datalager, och främst för dess mjukvara, har minskat dramatiskt de senare åren. Ett exempel som företag 6 tar upp är Microsofts nya datalagerprodukter som kan införskaffas för cirka 25 000 kr.

Företag 6 råder inte organisationer utan erfarenheter inom datalagerområdet att utveckla datalager på egen hand. Förkunskapskraven är höga och fallgroparna många. Företag 6 poängterar att produkterna inom området är utvecklingsverktyg och inte applikationer som är färdiga att installera och köra. Företag 6 säger även att det inte krävs någon större erfarenhet för att lära sig att använda ett datalager, personer med IT-vana kan lära sig grunderna på några dagar. För mer ingående kunskaper och ett mer effektivt utnyttjande av datalagret krävs mycket erfarenhet. Detta beror på att alla datalagerlösningar är företagsspecifika och unika då de är skapade efter organisationens unika förutsättningar och krav.

### 6.2.3 Sammanställning av den empiriska studien

En viktig erfarenhet är den att datalagerutvecklare är försiktiga med att lämna ut uppgifter om sina kunder. Detta är i sig inte helt överraskande då datalager är en av de allra senaste verktyg som kan ge stora konkurrensfördelar. Ett företag som använder sig av ett datalager som bidrar till företagets välgång är antagligen inte lika snar att skylta med detta faktum för sina konkurrenter. Denna erfarenhet är viktig då det var denna faktor som mest skulle kunnat påverka resultatet av den empiriska studien.

Av de uppgifter de lämnar verkar det dock vara relativt många företag i Sverige som använder sig av någon sorts datalagerlösning. Även om det visserligen verkar som om det huvudsakligen är större företag som använder datalager, verkar det även som om det är relativt många mindre företag som använder någon form av datalagerlösning.

Utöver de företag som angivits i studien används datalager av bland andra ICA, KF, Gambro, SAS, Ericsson och Gröna Konsum enligt WM-data<sup>1</sup> och Posten (CSA, 1998).

Det verkar som om det i dagsläget finns mjukvaruprodukter som gör det möjligt för mindre företag att använda sig av datalager. Även priserna på dessa produkter verkar enligt svaren vara relativt låga, vilket tyder på att priset på åtminstone mjukvaran inte

---

<sup>1</sup> Enligt pressmeddelande nr. 184, publicerat 1998-03-20. (Tillgängligt på [www.wmdata.se](http://www.wmdata.se))

verkar vara ett av de större hindren i ett utvecklingsarbete. En datalagerutvecklare anger att ett datalager till ett företag med cirka 300 anställda kan byggas för några tiotusentals kronor utöver befintliga system plus konsulttid. En annan datalagerutvecklare anger att priser inom datalagerbranschen minskat kraftigt de senaste åren och att ett datalagers mjukvara i dagsläget kan köpas för omkring 25 000 kr. Den mjukvara som existerar verkar ej vara storleksberoende, ur organisatorisk synpunkt, utan verkar vara applikationsberoende, det vill säga, det beror på vilka tjänster en kund vill att datalagret skall utföra.

Å andra sidan tror andra utvecklare att datalager, på grund av den i vissa fall omfattande företagsanpassningen, inte kommer att komma mindre företag till nytta. Beroende på hur datalagret skall vara uppbyggt och vad det skall användas till, behövs det olika omfattande utvecklingsarbete. Specialbyggda datalager tar naturligtvis mer tid och pengar att utveckla, vilket antyder att mindre företag har mindre chanser att starta ett sådant projekt.

Det verkar inte krävas några expertkunskaper för att kunna använda ett implementerat datalager. Detta är ett av huvudkraven för att datalager skall slå igenom på bred front. Att det däremot krävs viss expertkunskap utifrån då och då verkar vara nödvändigt. Även om det finns billiga lösningar går det, för ett företag som utvecklar datalager, att tjäna mycket på att själva lägga ner tid och resurser på datalagret.

Ett av de stora frågetecknen inför detta arbete var om mindre företag verkligen hade något att vinna på att utveckla ett datalager. Jag trodde att det fanns en risk att mindre företag lade ner stora resurser på något som kanske inte skulle vara tillräckligt stor nytta ställd mot den insats företaget var tvungen att uppbringa. Den empiriska studien tyder däremot på motsatsen. Utvecklarna som deltog i studien anser att mindre företag dels har möjligheter att inom en inte alltför avlägsen framtid och dels att de kommer att ha stor nytta ett datalager, även om de inte kommer att vara helt beroende av det.

Det är viktigt att påpeka att det inte angavs någon definition av datalager i studien. Detta har lett till att det har varit de tillfrågade företagens definitioner på datalager som använts. Då ingen universell definition finns, även om Inmon & Hackathorns (1994) definition oftast används, kan detta leda till att olika sorters produkter anges som datalager. En liknande definitionsfråga anges i studien. Denna fråga påpekar problemet om datalagers utsträckning och om även gränssnittet skall ingå i datalagret eller om datalagret endast skall omfatta databasen och analysverktygen.

Något som inte kommit fram i detta arbete är vilka statliga instanser som använder datalager i Sverige. Om detta beror på att det är få statliga instanser som använder datalager eller om statliga instanser använder sig av utvecklare och konsulter som inte nåtts i denna studie, är okänt.

Sammantaget tyder resultaten från den empiriska studien, angående om mindre företag kommer att kunna använda datalager i framtiden, på två saker:

- Enklare datalager är något mindre företag kan använda i framtiden och redan gör. Dessa bygger till största delen på hyllprodukter som i dagsläget är relativt billiga och lättanvända.
- Mer avancerade datalager är skraddarsydda. Detta tyder på att de både blir dyrare och tar längre tid att bygga. Dessa specialiserade datalager är omöjligt att köpa i någon affär därför att alla organisationer är unika till sin uppbyggnad och i sina behov.

### 6.3 Jämförelser med förväntat resultat

Då arbetet inleddes trodde jag att datalager skulle vara en relativt utbredd företeelse. Efter att bakgrundskapitlet skrivits insåg jag att detta inte verkade vara fallet. Detta gjorde att jag förväntade mig finna att användandet av datalager verkligen inte var så utbrett som jag trott innan arbetet påbörjades. Detta berodde till stora delar på att den forskningslitteratur som studerades till mycket liten del tog upp hur användandet ser ut i näringslivet. Akademisk litteratur behandlar nästan uteslutande ren forskning och utveckling, och nästan inte alls hur datalager används i näringslivet och vilka som använder dem. I litteraturen ges intrycket av att mycket arbete kvarstår innan datalagerteknologin mognar. Det finns ännu många barnsjukdomar och områden som behöver utvecklas ytterligare. Detta skulle då tyda på att datalager ännu inte är redo för den breda marknaden, och då speciellt inte till mindre företag.

Den empiriska studien gav dock en nästan motsatt bild av verkligheten. Studien gav bilden av att många företag i Sverige, både stora och mindre, använder datalager. Det är inte helt fastställt att alla dessa företags datalager stämmer överens med den definition av datalager som används i detta arbete, men det kan dock ge en någorlunda bild av spridningen av datalager.

De olika resultaten speglar troligtvis två sidor av verkligheten. Bilderna från både litteraturstudien och den empiriska studien stämmer troligtvis, men jag anser att resultatet från den empiriska studien är det resultat som är viktigast för att svara på problempreciseringen. Anledningen till detta är att litteraturen visade sig vara mer generell till sin bild av datalager och att den inte precis tog upp organisationers verkliga användande av datalager. Även om den litteratur som använts i detta arbete får anses vara relativt ny tar det tid att publicera verken. Informationen i litteraturen är oavsett publiceringsdatum inte lika aktuell som de svar som givits av datalagerutvecklarna i den empiriska studien. Litteraturen som använts i litteraturstudien är främst från USA. Detta gör att bilden som beskrivs i litteraturen omedvetet får en amerikansk prägel. Även om förhållandena på den svenska datalagermarknaden kan antas likna de som gäller för den amerikanska datalagermarknaden är ändå uppgifter direkt från svenska datalagerutvecklare mer intressanta för detta arbete. Aktuell svensk litteratur skulle med andra ord lämpat sig bäst för detta arbete men någon sådan litteratur hittades inte.

Här måste även det faktum beaktas att det var utvecklare som tillfrågades i studien. Möjligheten finns att detta skulle kunna ge en skev bild av spridningen av datalager. Detta är i och för sig naturligt då det ligger i datalagerutvecklarens intresse att påvisa möjligheterna med datalager och om så inte dölja men ändå inte skylta med bristerna och att ge sken av att många företag redan använder datalager. Detta skulle i sin tur kunna sporra företag som inte har datalager att starta ett sådant utvecklingsprojekt, vilket skulle ge mer arbete åt datalagerutvecklare.

I det förväntade resultatet trodde jag att ekonomiska och kunskapsmässiga krav skulle vara för höga för att mindre organisationer skulle ha möjlighet att använda datalager. Arbetet tyder på att detta inte är fallet. För ett litet datalager krävs inga större ekonomiska summor. För ett mer avancerat och skräddarsytt datalager krävs troligtvis en större insats. Litteraturen menar att det fortfarande finns brister i den mer organisatoriska delen av utvecklingsarbetet. För att lösa detta krävs lämpliga metoder, något som det råder brist på än så länge (Srivastava & Chen, 1999).

I det förväntade resultatet beskrev jag min tveksamhet till om privatpersoner kommer att komma i kontakt med datalager inom den närmaste framtiden. Som förväntat

berörde studerad litteratur inte alls privatpersoners behov av eller tillgång till datalager. Detta är naturligt då datalager inte skapats som stöd för privatpersoner utan som stöd för organisationer.

Studien visar på att det i huvudsak är större företag som använder datalager. Detta stämmer överens med det förväntade resultatet.

I det förväntade resultatet trodde jag att mindre företag skulle ha små möjligheter att hävda sig på datalagermarknaden. Arbetet tyder delvis på motsatsen. Att införa ett litet och enkelt datalager är inget större problem. Skall däremot ett specialbyggt datalager införas krävs mycket mer resurser. Ett specialbyggt datalager är därför troligtvis endast ett alternativ för större organisationer. Istället för frågan om mindre organisationer har möjlighet att utveckla ett datalager blir frågan då istället vad dessa mindre organisationer själva ställer för krav på ett datalager. Den frågan svarar inte detta arbete på.



## 7 Slutsatser

Att införa ett datalager är en process som inleds av det intresserade företaget självt. Alltså måste företaget själv nå insikt om att ett datalager kan vara en bra hjälp för just dem. Om mindre företag verkligen har nytta av datalager är inte lätt att veta vare sig för dem själva eller andra. Datalager utvecklades som ett storföretagsfenomen för att stödja de stora företagen. Dessa stora företag har mer utvecklade och mogna organisationer med många analytiker och stor erfarenhet på olika områden, inte minst inom informationsteknik och databashantering.

Datalager är relativt nya, vilket gör att många företag troligtvis inte vet vad det är. Datalager gör mest nytta hos företag i branscher med hög konkurrens, exempelvis tillverkningsindustrin, detaljhandeln och försäkringsbranschen (Watson m.fl., 2001). För företag som anser att de kan ha nytta av ett datalager finns idag möjligheter att införskaffa ett sådant.

Datalager kräver mycket ansträngningar även efter implementationer. Ett av de tillfrågade företagen i den empiriska studien menar att ”ett datalager ger inga svar, det ger nya frågor”.

### Resultat från litteraturen

Det viktigaste resultatet från litteraturstudien är att det i huvudsak är de stora företagen som använder datalager men att även mindre företag använder dem. De datalager som existerar används främst till dataanalys, men de används även till informationsdistribution inom företagen.

Det bedrivs mycket forskning och utveckling inom datalagerområdet och det kommer att behövas fortsatt forskning då många områden behöver uppmärksamhet.

Litteraturen om datalager fokuserar främst på mer tekniska och abstrakta aspekter, ej så mycket på organisatoriska frågor. Trots detta är datalagerutveckling främst en organisatorisk fråga, både ur behovssynpunkt och utvecklingssynpunkt. I dagsläget finns det få metoder som innefattar den organisatoriska vinkeln av datalagerutveckling.

### Resultat från den empiriska studien

Även om den empiriska studien endast innehåft en stödjande uppgift i detta arbete har flera intressanta synvinklar framkommit. Det är många svenska företag som idag använder datalager. Trots detta är definitionen av datalager oklar. Det viktigaste resultatet från studien är det att det redan idag finns möjligheter för mindre företag att utveckla och använda datalager.

Baserat på ovanstående resultat kan två olika sorters datalager identifieras, dels ett datalager som utvecklas unikt i en organisation utefter organisationens unika behov och krav, dels ett datalager som köps mer eller mindre färdigutvecklat och som sedan anpassas efter företagets krav. Den första sorten är ett stort, dyrt och omfattande arbete som kräver en stor insats och uppoffring. Den andra sorten är ett mindre budgetalternativ, vilket inte tar lika lång tid att utveckla men inte heller blir lika avancerat och effektivt som ett skräddarsytt datalager.

### **Slutsats**

Litteraturstudien gav en något oklar bild av om mindre företag skulle kunna använda datalager inom en snar framtid. Undersökningen gav dock en stark bild av att mindre företag kan få tillgång till ett datalager.

Beroende på vilka krav ett mindre företag har på ett datalager kan följande slutsats dras utifrån det insamlade materialet:

- Mindre företag kommer i en nära framtid att kunna använda datalager. Det finns redan i dagsläget mindre företag som använder datalager.

## 8 Diskussion

Detta kapitel tar bland annat upp erfarenheter, vad arbetet tillfört i ett större perspektiv, diskussion kring resultatet och framtida arbete.

### 8.1 Allmän diskussion

Då arbetet inleddes gjorde jag det i tron att datalagerhantering var ett brett, välutvecklat och väldigt specialiserat område, vilket skulle ha bidragit till att det ej var så välkänt. Jag trodde även att de existerande operativa datalagren skulle vara väldigt få i antal. Under arbetet har jag insett att området inte är så välutvecklat, men att det är mer utbrett än jag trott. Datalagerhantering är en teknik som är någorlunda utbredd inom näringslivet. Kunskaperna om datalagerhantering finns även inom den akademiska världen, men lärs dock inte ut i den utsträckning som branschen skulle vilja.

Material som var svårt att hitta var arbeten som var gjorda i och om svenska förhållanden. Det material jag har hittat är nästan uteslutande internationella arbeten med fokus på USA.

När jag började arbetet trodde jag även att datautvinning (*data mining*) var en stor och viktig del av datalagren, men den var inte lika utvecklad som jag trott. Precis som datalagren i stort är datautvinning ett pågående utvecklingsprojekt som visserligen existerar men som behöver utvecklas främst vad gäller mognad och användarvänlighet.

En viktig del i detta arbete var valet av synvinklar ur vilka litteraturen skulle studeras. Att jag valde de jag gjorde berodde på att jag trodde att de var de som främst påverkade möjligheterna till att införa ett datalager för mindre företag. Jag trodde även att det skulle vara svårt att helt särskilja dem från varandra då alla synvinklar spelar in mer eller mindre när ett datalager skall utvecklas. Angående valet av de synvinklar jag valde att studera problemet ifrån, anser jag att det valet var korrekt. En kraftigare avgränsning skulle dels ha gett mig mindre material att studera, kanske inte tillräckligt material för att kunna dra några riktiga slutsatser, dels skulle detta ha fokuserat arbetet på enstaka punkter varvid det bredare perspektivet förlorats. Detta innebär inte att jag har haft möjlighet att på något sätt täcka hela området eller berört alla intressanta synvinklar. Storleken på och vidden av ett datalagerprojekt är en intressant erfarenhet. Även om datalagret i sig är en separat enhet som kan placeras ovanpå befintliga system och databaser behövs mycket mer arbete för att optimera ett datalager. Att arbetet omfattar hela organisationens strategiska beslutsfattande, samordning av terminologi, till den enskilde användarens användargränssnitt, gör att det i de största fallen blir ett enormt stort och komplicerat arbete att genomföra.

## 8.2 Vad arbetet har tillfört

Huvudmålet med detta arbete var att försöka hitta var utvecklingen inom datalagerområdet står idag. Tanken bakom att kunna analysera gammal data för att hitta ny kunskap är både intressant och spännande. Att dessa tankar än så länge till stora delar bara är tankar och omogen teknik var en besvikelse. Detta till trots har detta arbete kunnat påvisa att företag, både större och mindre, idag använder datalager, vilket innebär att de måste bidra till företagets framgång.

Arbetet har gett en översiktlig bild av den akademiska världens och näringslivets syn på hur datalager används och hur de kan komma att användas i framtiden. I litteraturen beskrivs det att näringslivet har arbetat med datalager längre än den akademiska världen, något som kanske kan förklara varför man får en något omogen bild av datalager i akademisk litteratur, medan näringslivet ger bild av en etablerad datalagermarknad. Dessutom måste det faktum beaktas att den litteratur som studerats i huvudsak varit utländsk medan undersökningen fokuserat enbart på Sverige. Även om världen i dagsläget är globaliserad vad det gäller idéer, tankar, hårdvara och mjukvara, kan detta faktum ha påverkat resultatet om än endast nämnvärt.

Detta arbete kan läsas av de som är intresserade av hur datalager används idag och av de som vill ha en översiktlig beskrivning av vad ett datalager är, inklusive mindre företag som funderar på vad ett datalager kan göra för dem.

## 8.3 Förslag på framtida arbete

Här beskrivs framtida arbeten som skulle vara intressanta, med koppling och vidareutveckling av detta arbete. Detta arbete gav en övergripande bild av möjligheterna för mindre företag att använda datalager. När detta konstaterats vara möjligt, kan fler intressanta och utvecklande frågor ställas.

- Det vore intressant att närmare undersöka några mindre företag som tagit ställning till att införa ett datalager eller inte. Detta skulle ge en djupare bild av mindre företags möjligheter, krav och förväntningar på en datalagerlösning.
- En mer ingående kartläggning av hela Sveriges datalageranvändning, liknande Watson m.fl. (2001), skulle vara intressant om än för stort och omfattande för ett arbete i denna storlek. Har de ändrat uppfattning om dess kapacitet, jämfört med förväntningar? Var införandet lyckat? Watson m.fl. (2001) påvisar att de flesta företag som infört datalager är nöjda, stämmer detta i Sverige?
- Datalager kan användas både som ett analysverktyg och en informationskanal. Varför görs detta och vilka fördelar och nackdelar finns det med detta?
- Mer ingående intervjuer med datalagerutvecklare skulle ge en bättre bild av hur datalager används i näringslivet jämfört med vad en studie av den akademiska världen kan ge.
- Införande av datalager kan vara en mycket stor och komplicerad process. Det vore då intressant att undersöka hur omfattande ett datalagerinförande verkligen är i ett företag och hur mycket organisationen påverkas eller behöver påverkas för att ett datalager skall fungera optimalt.

- Definitionsfrågan av vad ett datalager är, är svårlöst. En undersökning om hur företag och utvecklare definierar datalager i förhållande till hur de ser ut och vad de används till skulle vara värdefull.
- Ett problem som många poängterar är bristen på verktyg för att mäta och värdera ett datalagers nytta. Ett sätt att göra detta skulle vara mycket värdefullt, men antagligen även mycket svårt att skapa.
- Vilka krav ställer mindre organisationer på datalager? Hur står detta i jämförelse med större företags krav på datalagren. Hur väl kan datalagren anpassas till de mindre företagen och deras krav?

Dessa är några uppslag till framtida arbeten inom datalagerområdet.

## Referenser

Baum, D. (1997), Planning and implementing a data warehouse, *Byte Magazine*, **22**(6), sid. 120-124.

Bernstein, P., Brodie, M., Ceri, S., DeWitt, D., Franklin, M., Garcia-Molina, H., Gray, J., Held, J., Hellerstein, J., Jagadish, H., Lesk, M., Maier, D., Naughton, J., Pirahesh, H., Stonebraker, M. & Ullman, J. (1998), *The Asilomar Report on Database Research*, ACM Sigmod Record.

Chen, L. & Frolick, M.N. (2000), Web-based Data Warehousing – Fundamentals, Challenges and Solutions, *Information Systems Management*, **17**(2), sid. 80-86.

Connolly, T., Begg, C. & Strachan, A. (1999), *Database Systems*, andra upplagan, England, Harlow: Addison Wesley Longman.

CSA. (1998), *Informationslager ger Posten bättre kundanalyser*, Tillgänglig på: <http://domino.idg.se/> (Tillgängligt datum: 2001-02-20).

Dawson, C.W. (1999), *The Essence of Computing Projects: A Student's Guide*, England: Prentice Hall.

Elmasri, R. & Navathe, S. (1999), *Fundamentals of database systems*, tredje upplagan, England: Harlow, Addison Wesley Longman.

Fayyad, U. (1998), Taming the Giants and the Monsters: Mining Large Databases for Nuggets of Knowledge, *Database Programming and Design magazine*, **1998**(Mars), sid. 1-8.

Gallegos, F. (1999), Data Warehousing: A Strategic Approach, *Information Strategy: The Executive's Journal*, **16**(1), sid. 41-47.

Gray, P. & Watson, H.J. (1998), Present and Future Directions in Data Warehousing, *The Data Base for Advances in Information Systems*, **29**(3), sid. 83-90.

Harmon, C. (1998), Safeguarding the Data Warehouse, *Computer Fraud & Security*, **1998**(6), sid. 16-19.

Hart, C. (1998), *Doing a Literature Review*, Sage Publications, England: London.

Inmon, W.H. & Hackathorn, R.D. (1994), *Using the Data Warehouse*, USA: John Wiley & Sons.

Jarenko, L. & Vahlgren-Wall, M. (1986), *Information som konkurrensmedel*, Göteborg: IHM Läromedel.

Jarke, M., Jesufeld, M.A., Quix, C. & Vassiliadis, P. (1999), Architecture and quality in data warehousing: An extended repository approach, *Information Systems*, **24**(3), sid. 229-253.

Lehman, P. & Jaszewski, J. (1999), Business Terms as a Critical Success Factor for Data Warehousing, *Design and Management of Data Warehouses*.

Olsen, D. & Cooney, V. (2000), The Strategic Benefits of Data Warehousing: An Accounting Perspective, *Information Strategy: The Executive's Journal*, **16**(2), sid. 35-40.

Patel, R. & Davidson, B. (1994), *Forskningsmetodikens grunder. Att planera, genomföra och rapportera en undersökning*, Lund: Studentlitteratur.

Poe, V., Klauer, P. & Brobst, S. (1998), *Building a Data Warehouse for Decision Support*, andra upplagan, England: London, Prentice Hall.

Rundensteiner, E.A., Koeller, A. & Zhang, X. (2000), Maintaining Data Warehouses over Changing Information Sources, *Communications of the ACM*, **43**(6), sid. 57-62.

Samtai, S., Mohania, M., Kumar, V. & Kambayashi, Y. (1998), Recent Advances and Research Problems in Data Warehousing, *ER Workshops*, sid. 81-92.

Silbershatz, A., Stonebraker, M. & Ullman, J. (1996), *Database Research: Achievements and Opportunities Into the 21st Century*, ACM Sigmod Record.

Srivastava, J. & Chen, P-Y. (1999), Warehouse Creation - A Potential Roadblock to Data Warehousing, *IEEE Transactions on Knowledge and Data Engineering*, **11**(1), sid. 118-126.

Strouse, K. (1999), Weapons of mass marketing, *Intelligence and software*, **237**(9), sid. 26-28.

Söderström, P. (1997), *Data Warehouse – Datalager: Verksamhet, Metod, Teknik*, Studentlitteratur, Lund.

Watson, H.J., Annino, D.A., Wixom, B.H., Avery, L. & Rutherford, M. (2001), Current practices in data warehousing, *Information Systems Management*, **18**(1), sid. 47-56.

Widom, J. (1995), Research Problems in Data Warehousing, *4th International Conference on Information and Knowledge Management CIKM*, sid. 25-30.



## Bilaga 1 – Företagsfrågor

Denna bilaga innehåller det e-postmeddelande som skickades ut till utvalda företag. Före detta e-postmeddelande skickades även ett introducerande e-postmeddelande för att få tillåtelse att skicka dessa frågor. Följande text skickades i ett e-postmeddelande.

---

Hej,

och tack för att ni tar er tid att svara på mina frågor.

Jag heter Tomas Johansson och studerar vid det Systemvetenskapliga programmet på Högskolan i Skövde. Jag arbetar för närvarande med mitt examensarbete som behandlar användandet av datalager (Data Warehouse). Jag skriver nu till er och har några korta frågor angående just datalager, som jag tror ni som utvecklare bäst av alla kan svara på. Jag skulle vara tacksam om ni har tid att svara på nedanstående två frågor.

Problemställningen för hela examensarbetet lyder: ”Kommer datalagerhantering vara enbart större organisationer förbehållna eller kommer datalagerhantering inom den närmaste framtiden även kunna användas av mindre företag eller enskilda användare.” Arbetet består av en litteraturstudie samt en mindre undersökning. Syftet med den enkla undersökning är att få en övergripande bild av vilka företag som använder sig av datalager i Sverige i dagsläget.

Frågorna är utformade för att vara enkla och för att inte ta lång tid att besvara. Svara genom att skicka tillbaka ett nytt e-mail med svaret eller genom att återsända (reply) detta e-mail med infogade kommentarer.

Frågorna lyder som följer:

- **Vilka Svenska företag har ni hjälpt att utveckla datalager?**

Lämnar ni ut sådana uppgifter eller är de klassificerade? Känner ni till andra företag eller organisationer som använder sig av datalager i Sverige som ni ej varit med och utvecklat.

Mitt arbete har fokus på hur mindre organisationer skulle kunna använda datalager.

- **Tror ni att mindre organisationer kommer att kunna använda datalager inom en snar framtid?**

Krävs det för avancerade kunskaper för att mindre organisationer skall kunna använda datalager? Kostar det för mycket pengar att utveckla eller köpa produkter till dem? Finns det ingen mjukvara som är anpassad för mindre organisationer?

Det skulle vara mycket värdefullt för mitt arbete om ni ville svara på dessa frågor. All information kommer att behandlas anonymt i undersökningen. Tack på förhand.

Med vänlig hälsning Tomas Johansson

Systemvetenskapligt Program årskurs 3

Högskolan i Skövde

E-postadress 1: a98tomjo@student.his.se

E-postadress 2: tk.johansson@hjo.mail.telia.com

---