

Metadata i datalager

En studie i hur organisationer
hanterar metadata i datalager.

Marcus Lindström¹

Institutionen för Data- och Systemvetenskap
Stockholms Universitet / Kungl. Tekniska Högskolan

September 2002

¹ Uppsatsen motsvarar 20 poäng

Sammanfattning

Metadata är en central del av ett datalager (eng. data warehouse). Rutinmässigt brukar metadata översättas till data om data. En mer preciserad förklaring är att metadata är dokumentation, beskrivningar och förklaringar över vad som finns i datalagret. Metadata skall vara till hjälp för alla i en organisation som direkt eller indirekt arbetar med datalagret.

Vikten av god hantering av metadata i datalager ökar ständigt och orsakerna till detta är flera. Systemen blir allt mer komplexa med den snabba tekniska utvecklingen. Det samlas in mer data och storleken på datalagren ökar. Men den kanske viktigaste faktorn är att det idag är så många fler människor involverade i ett projekt och då krävs metadata för att ett effektivt jobb skall kunna utföras.

Syftet med denna uppsats är att undersöka hur organisationer hanterar metadata. En utvärderingsmodell med fem faktorer har skapats och för varje faktor har ett antal kriterier tagits fram. Genom att tillämpa faktorerna och deras respektive kriterier kan organisationers hantering av metadata analyseras. Med utvärderingsmodellen som grund har en kvalitativ undersökning genomförts där fyra organisationer intervjuats om deras hantering av metadata.

Undersökningen visar att de fyra organisationerna är medvetna om vikten av metadata och att de har vissa strategier för att hantera metadata. En utmaning som många organisationer står inför är att få slutanvändarna mer involverade och engagerade. Från företagsledning finns hårda krav på teknisk dokumentation, så att företaget inte blir sårbart vid avhopp eller sjukdom. Men eftersom viss kunskap är svår att dokumentera innehas en del kunskap endast av ett fåtal personer i några av de undersökta organisationerna.

Abstract

In order for a well functioning data warehouse to succeed many components must work together. Metadata, usually called data about data, is an important part. A more detailed description is that metadata is a documentation, description and explanation about the contents of the data warehouse. Metadata is supposed to be a helping hand to all co-workers in an organization that work directly or indirectly with the data warehouse.

Information systems are becoming more and more complex and with rapid technical development it means that metadata and good management of it is more important than ever. Increasingly more data is collected and stored in today's data warehouses, but maybe the most important factor is that nowadays many more people are involved in a project and, in order for them to work efficiently, metadata is required.

The purpose of this thesis is to investigate how organisations manage metadata. To answer that question a qualitative survey has been carried out. Four organisations from different branches have been interviewed.

The author's conclusion is that organisations are aware of the importance of managing metadata. A challenge that many organisations are up against is how to get the end user more interested and therefore more involved. There are increasingly harder demands from company management for technical documentation so that they do not become vulnerable in case of staff defection or illness. However, as certain knowledge is difficult to document part of it may be known only by one person in some of the investigated organisations.

Tack

Ett stort tack till min inspirerande handledare på DSV Erik Perjons, som alltid har haft tid för att besvara mina frågor och som hjälpt mig framåt i arbetet med sina goda råd och synpunkter.

Jag vill även rikta ett stort tack till Niclas Hansson på RSV som med stort engagemang och intresse skapat förutsättningar för denna uppsats.

Vill även tacka de personer som jobbar på de fem företag som deltagit i undersökningen för att ni tog er tid och besvarade mina frågor. Tack till Nils Behrman - Data warehouseansvarig på Nordea, Göran Lord – Analyschef på Sydvästra Sjukvårdsområdet i Stockholms Läns Landsting, Anders Nilsson –Systemadministratör på Hennes & Mauritz AB, Torsten Sigel - Samordnare i Stockholms Läns Landsting och Rickard Wärnelid - Data warehouseansvarig på Skandia Kapitalförvaltning.

Tackar även Per Yhde som är konsult på SAS Institute för idéer och artiklar.

Slutligen ett tack till studiekamrater som kommit med konstruktiv kritik under resans gång. Tack till Carl Brolin, Britta Eriksson, Lars Gunnarsson, Åsa Johansson, Samy Lidner, Urban Lithner, Katarina Karlsson, Suvi Korpela och Magnus Svan.

Innehållsförteckning

1	Inledning	1
1.1	Bakgrund och problem	1
1.2	Syfte	1
1.3	Mål	2
1.4	Avgränsningar	2
1.5	Uppdragsgivare	2
1.6	Målgrupp	2
1.7	Arbetsmetod	2
1.8	Disposition	4
2	Datalager	4
2.1	Introduktion	4
2.2	Teknik	5
2.3	OLTP och OLAP	6
2.4	Datamart	7
2.5	Fördelar, nackdelar och problem med ett datalager	8
3	Metadata	8
3.1	Data, information och kunskap	8
3.2	Termer och begrepp	9
3.3	Begreppet metadata	10
3.4	Metadata i datalager	10
3.5	Metametadata	11
3.6	Termkatalog	11
3.7	Repository	11
4	Utvärderingsmodell	12
4.1	Metadata Quality	13
4.2	Business Terms	13
4.3	Business Metadata	13
4.4	Administrative Metadata	14
4.5	Operational Metadata	14
5	Empirisk analys	14
5.1	Undersökningen	15
5.2	Intervjuobjekt	16
5.3	Business Terms	17
5.4	Business Metadata	19
5.5	Administrative metadata	21
5.6	Operational metadata	23
5.7	Metadata quality	25
6	Slutsats och Epilog	26
6.1	Slutsats	26
6.2	Utvärderingens och utvärderingsmodellens giltighet	27
6.3	Den empiriska undersökningens giltighet	27
6.4	Framtida forskning	27
6.5	Epilog	28
	Källor	29
	Bilaga 1: Ordlista	31
	Bilaga 2: Riksskatteverket	33
	Bilaga 3: Intervjumall	35

1 Inledning

1.1 Bakgrund och problem

Allt fler företag och organisationer implementerar datalager i sin verksamhet. I ett datalager sammanförs och sammanställs data från flera olika operativa IT-system i organisationen. Med hjälp av olika applikationer kan datalagrets data sedan analyseras för att till exempel upptäcka trender, som nya köpbeteenden eller köpstarka kunder [Söderström 97]. Sådan information kan sedan kan ge stöd i strategiska beslutssituationer. Datalager är en översättning av engelskans data warehouse, och ibland används även termerna informationslager och datavaruhus på svenska.

I ett datalager måste flera komponenter samverka. En viktig sådan komponent är metadata, som rutinemässigt brukar definieras som data om data. Enligt Inmon är detta en definition som i datalagersammanhang är alldeles för tunn, ofullständig och omodern. [Inmon 02]

Metadata är dock inte något enkelt att beskriva. Kimball skriver att metadata är något mystiskt: trots att vi inte vet vad metadata exakt är, eller var det finns, pratar vi mycket om det, oroar oss för det och känner skuld eftersom ingen gör något åt det. När Kimball försöker ge ett svar på frågan vad metadata är, lyder detta svar, överraskande intetsägande, att det är allt som finns i ett datalager utom den data som finns lagrad. [Kimball 02]

Mer innehållsrika definitioner på metadata, som Devlin och Jeffery, brukar definiera metadata som dokumentation, beskrivningar och förklaringar över vad som finns i datalagret, framförallt information om dess data, samt beskrivningar över hur datalagret är uppbyggt och hur det fungerar [Devlin 99], [Jeffery 00]. Metadata skall vara till hjälp för alla i en organisation som direkt eller indirekt arbetar med datalagret, som till exempel analytiker, programmerare, ekonomer och databasadministratörer [Bischoff et al 97].

Vikten av god hantering av metadata ökar ständigt och orsakerna till detta är flera. Systemen blir allt mer komplexa med den snabba tekniska utvecklingen. Det samlas in mer och mer data och storleken på datalagerna ökar [Wiener 02]. Enligt Jeffery är kanske den viktigaste faktorn att det idag är så många fler människor involverade i ett projekt, och då krävs bra metadata för att ett effektivt jobb skall kunna utföras. [Jeffery 00]

Brackett är kritisk till hur metadata hanteras i organisationer. Metadata är ofullständigt dokumenterade och innehåller gammal och icke uppdaterad data. Dokumentationen är slarvigt skriven och svår att förstå för personer som inte är väl insatta i verksamheten. Brackett påstår vidare att metadata ofta är utspridda på olika ställen och svåra att hitta. I stället för väldokumenterad metadata finns kunskapen ofta samlad hos ett fåtal personer, vilket leder till problem då dessa personer blir sjuka eller lämnar företaget. Det är som synes ingen gott betyg som Brackett ger över organisationers hantering av metadata. [Brackett 96]

Även Tannenbaum anser att organisationer måste börja ta metadata på allvar då metadata är en viktig faktor för att en datalager ska bli en framgång, det vill säga ge ett mervärde till en organisation. Hon menar att organisationer har koncentrerat sig allt för mycket på andra komponenter i ett datalager och ignorerat metadata. [Tannenbaum 02]

Brackett och Tannenbaums hårda kritik mot organisationers hantering av metadata leder till ett antal frågeställningar: Har Brackett och Tannenbaum rätt i att organisationer ignorerar metadata? Hur hanterar organisationerna egentligen metadata och finns det olika former av metadata som hanteras på olika sätt? Kan en organisation få ut mer nytta av sitt datalager genom att hantera metadata på ett annat sätt?

1.2 Syfte

Syftet med uppsatsen är att undersöka hur ett antal organisationer hanterar metadata i datalager.

1.3 Mål

Målen med uppsatsen är följande:

- att ta fram en utvärderingsmodell i form av fem faktorer. Faktorerna skall användas för att analysera hur organisationer använder metadata. Utvärderingsmodellen kan också användas för att ge förslag på hur metadata skulle kunna användas och vara en bas för att diskutera hantering och förändringar i ett existerande datalager.
- Empiriskt undersöka hur organisationer hanterar metadata.
- att reda ut och förklara alla begrepp och termer som florerar inom området.

1.4 Avgränsningar

Datalager är stora och komplicerade system. För att ett datalager skall ge ett mervärde till en organisation krävs att alla komponenter fungerar. Uppsatsen koncentrerar sig på hantering av metadata ur en organisatorisk synvinkel, medan tekniska faktorer, som till exempel vilka tekniska lösningar som används för att lagra metadata, inte kommer att behandlas i någon större utsträckning.

1.5 Uppdragsgivare

Uppsatsen skrivs på uppdrag av Riksskatteverket² (RSV) som tillsammans med Skattemyndigheterna och Kronofogdemyndigheterna har ansvaret för folkbokföring, service och information, beskattning, uppbörd och indrivning för medborgarnas räkning.

På RSV:s IT-avdelning arbetar drygt 600 personer och deras uppgift är att tillhandahålla IT-stöd för hela organisationens verksamhet. De utvecklar och förvaltar IT-stöd för skattehantering, folkbokföring, allmänna val och indrivning med hjälp av både ny och beprövad teknik. De huvudsakliga användarna av systemen är skatte- och kronofogdemyndigheternas medarbetare.

RSV har många delenheter som hanterar data från flera olika IT-system. Sedan en tid tillbaka pågår ett projekt där RSV IT skall skapa ett datalager, eller informationslager som de internt benämner det. Tanken är att data från olika IT-system som kan vara intressant vid uppföljning, revision och planering integreras och samlas på ett ställe. I dagsläget har de olika verksamhetsenheterna egna rutiner eller delsystem för lagring av data, vilket innebär svårigheter vid arbete över enhetsgränser och jämförelse av data. Under projektets gång har frågan väckts om hur metadata, och framförallt den del som kallas termkatalog³, skall hanteras. Den frågan har varit en av utgångspunkterna för denna uppsats.

1.6 Målgrupp

Uppsatsen riktar sig till personer som jobbar med IT-frågor i organisationer som har, eller planerar att införa, ett datalager. Uppsatsen är också skriven för studenter vid högskolor och universitet och andra intresserade som vill veta mer om datalager i allmänhet och metadata i synnerhet. En annan viktig målgrupp är uppdragsgivaren.

1.7 Arbetsmetod

Följande avsnitt beskriver kortfattat hur arbetet har framskridit. Närmare presentation om hur undersökningen gått till och en beskrivning av utvärderingsmodellen finns i kapitel 4 och 5. Figur 1.1 visualiserar arbetsstegen.

² Mer om RSV och deras datalagerprojekt i bilaga 2.

³ I bilaga 1 finns en ordlista som förklarar ord och begrepp som förekommer inom domänen.

1.7.1 Förstudier

Kunskap om datalager och metadata inhämtades genom att studera litteratur i ämnet och söka information på Internet. Litteratursökningen har haft en osystematisk karaktär, det vill säga att muntliga källor och tillfälliga litteraturlistor har använts [Halvorsen 92]. Det finns relativt mycket litteratur som beskriver datalager, men ofta är den mycket teknisk och behandlar lagringsstrukturer och söktekniker. Vanligtvis brukar en bok innehålla ett kapitel som ytligt behandlar metadata. På Internet har dock flera intressanta och användbara artiklar inom ämnesområdet hittats. I det inledande skedet av uppsatsarbetet studerades även RSV som organisation och framförallt deras datalagerprojekt.

1.7.2 Metod för framtagning av utvärderingsmodell

För att göra undersökningen togs en utvärderingsmodell fram. Det var inte givet från början hur denna utvärderingsmodell skulle se ut. Flera förslag togs fram och förkastades. Ytterligare litteraturstudier resulterade i en slutgiltig version. Den innehöll fem faktorer som alla hade ett antal kriterier. Tre av faktorerna byggde på Chaudhuri's sätt att kategorisera metadata [Chaudhuri 97], medan de två övriga är komplement för att kunna få en helhetsbedömning och har hämtats från Shanks och Devlin [Shanks et al 02], [Devlin 99]. Mer om den utvärderingsmodellen i kapitel 4.

1.7.3 Metod för datainsamling

Med utvärderingsmodellen som grund gjordes ett antal intervjufrågor⁴ som skulle användas som stöd för insamlingen av primärdata. En kvalitativ metod har använts vid datainsamlingen. Efter en testintervju med Skandia Kapitalförvaltning förändrades utvärderingsmodellen och därmed även frågorna.

Innan intervjufrågorna sammanställdes studerades bland annat intervjuteknik och hur man formulerar bra frågor.

Intervjufrågorna skall ses som ett diskussionsunderlag, det vill säga fungera som en grund för samtalet. Vid själva intervjutillfället har följdfrågor ställs som inte fanns med bland de formulerade frågorna och respondenterna har tillåtits att göra utvecklingar utanför de ställda frågorna när det ansetts befogat.

Undersökningen består av sex intervjuer hos fyra organisationer. Då organisationerna jobbar på olika sätt och datalagren är olika har det varit nödvändigt att ibland ställa djupare, mer detaljerade frågor och ibland att ställa frågorna i en annan ordning. Intervjuformen har med andra ord varit ostrukturerad [Lekvall et al 93]. Mer om undersökningsmetoden i avsnittet 5.1.

1.7.4 Metod för utvärdering

Analys och tolkning av primärdata handlar om att kategorisera och jämföra. Det handlar om att finna mönster och relationer. All analys innebär även förenklingar. [Halvorsen 92]

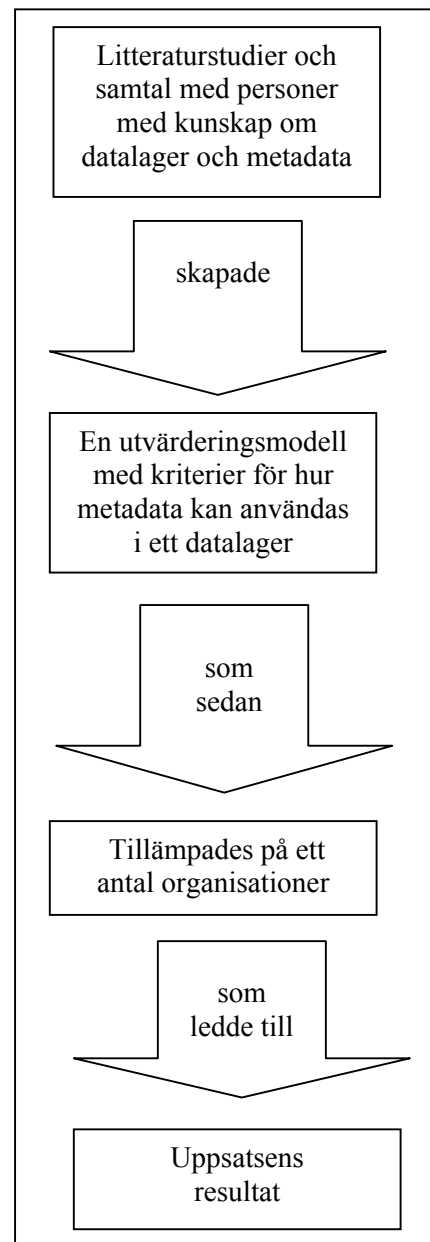


Fig. 1.1 Arbetsmetod för uppsatsarbetet.

⁴ Intervjumallen finns i bilaga 3

Det som sagts på intervjuerna har sammanställts och skrivits ned. När man analyserar insamlat material från en kvalitativ undersökning finns det två huvudformer av analys, nämligen helhetsanalys och delanalys [Halvorsen 92]. Delanalyser förutsätter att exempelvis ett undersökningsresultat går att dela upp i olika påståenden som kan kategoriseras eller räknas. Helhetsanalys innebär att man analyserar ett undersökningsresultat och försöker bilda sig en allmän uppfattning. I denna uppsats har de båda teknikerna använts, delanalyser har gjorts utifrån de kriterier som skapats samt utifrån de fem faktorerna. Halvorsen påpekar att det vid helhetsanalys är lätt att man tappar bort komplexiteten och variationen som uppstår vid en kvalitativ undersökning [Halvorsen 92]. Genom att även använda delanalyser har detta problem hanterats.

1.8 Disposition

Denna uppsats innehåller sex kapitel, samt källförteckning och tre bilagor.

Kapitel 2 - Datalager

I kapitlet diskuteras vad ett datalager egentligen är, här presenteras OLAP och det redogörs för skillnaden mellan ett datalager och ett datamart.

Kapitel 3 - Metadata

I detta kapitel förklaras begreppet metadata, vad det är och hur det används i datalager. Även komponenter i ett datalager som hanterar metadata beskrivs.

Kapitel 4 - Utvärderingsmodell

Här presenteras utvärderingsmodellen som intervjuerna bygger på. Alla faktorer och deras kriterier beskrivs.

Kapitel 5 - Empirisk analys

Undersökningens resultat presenteras utifrån respektive faktor, där respondenternas svar finns sammanfattade.

Kapitel 6 - Slutsats och diskussion

I detta kapitel presenteras undersökningens resultat. I ett diskussionsavsnitt ges tillfälle att reflektera över resultatet.

Källor och bilagor

Till sist finns en förteckning över de källor som används i uppsatsen. De tre bilagorna består av en ordlista, en mer detaljerad redogörelse av uppdragsgivarens datalagerprojekt samt intervjumallen.

2 Datalager

I detta kapitel presenteras grundläggande information om datalager och här kan bland annat läsas om skillnaden mellan OLAP och OLTP. Ett avsnitt beskriver datamart och slutligen beskrivs en del fördelar och nackdelar med datalager. Vissa uttryck som förekommer senare i uppsatsen preciseras och klargörs i detta kapitel. Kapitlet skall ses som en introduktion i ämnet och kan läsas kursivt för den som är väl insatt i domänen.

2.1 Introduktion

Data warehouse är ett engelskt ord som brukar översättas till datalager, ibland används även termerna informationslager och datavaruhus synonymt.

Då termerna datalager och data warehouse används i många olika sammanhang kan deras betydelse vara oklar. En organisation kan ha en definition, medan en annan avser något helt annat när ordet kommer på tal. Det är många som försökt definiera termen och här följer ett urval:

"Datalager är en logisk sammanhållen datamängd, som är avsett för analys och som speglar flera tidsperioder genom att data regelbundet hämtas från andra register." [Söderström 97]

"A subject-oriented, integrated, non-volatile, time variant collection of data in support of management's decisions." [Inmon 02]

"A single, complete and consistent source of data obtained from a variety of sources and made available to end users in a way that they can understand and use in a business context." [Devlin 99]

Det alla författare har gemensamt när de definierar ett datalager är att det på något sätt skall skapa ett mervärde för en organisation. Syftet är att samla ihop all data som finns i en organisation, kombinera denna data med eventuell data som insamlats utifrån och sammanställa. Med hjälp av all data och olika söktekniker kan man sedan få fram rapporter och analyser som kan ligga till grund för att välja rätt handlingsalternativ, ta strategiska beslut samt följa upp tidigare tagna beslut. Det är också viktigt att påpeka att om datalagret inte utnyttjas på detta sätt i en organisation har det inget mervärde [Skalin 98].

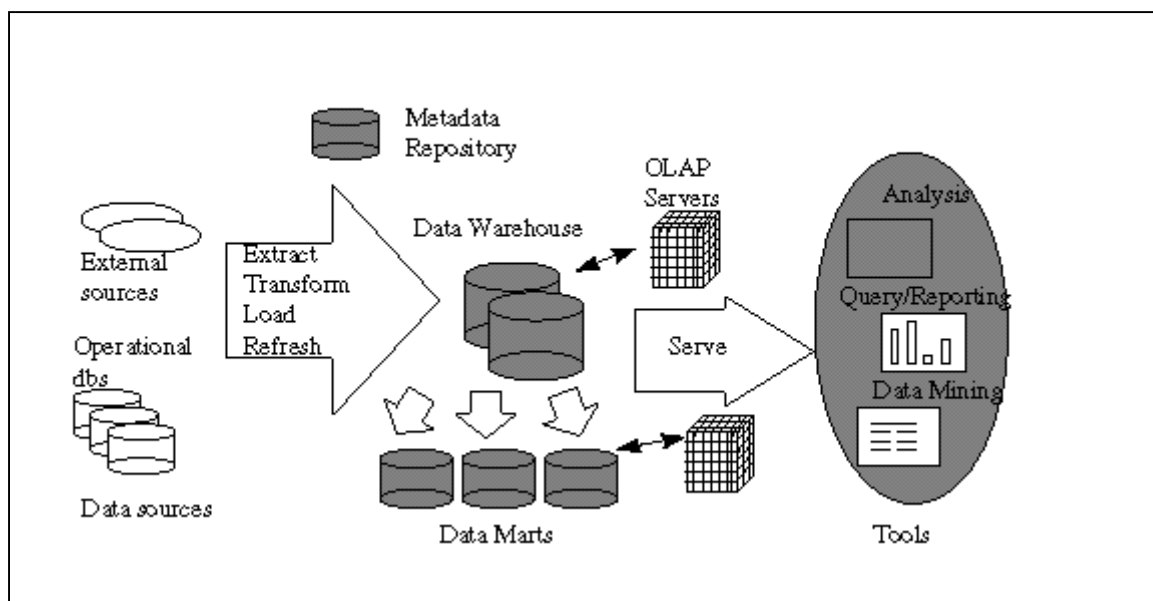
Tanken och idén bakom datalager har funnits en längre tid, men det var först 1990 som W. H Inmon myntade termen data warehouse. Han brukar sägas vara fadern till begreppet och har gett ut ett flertal böcker inom området. [Inmon 02]

Datalager har länge uppfattats som ett mer eller mindre magiskt, hemligt vapen för att vinna fördelar gentemot konkurrenter. Men nu börjar man inse att det är ett nödvändigt redskap och stöd i många verksamhetsorienterade processer och för extern informationservice. [Berlid 97]

Med datalager i denna uppsats avses ett system som kontinuerligt samlar in data från flera olika IT-system över en längre tid. Data som sedan används för rapportframställning och analys, som i sin tur ligger till grund för olika beslut.

2.2 Teknik

Man kan definiera ett datalager som en specialiserad databas. Begreppet VLDB står för Very Large DataBases och är karaktäristiskt för ett datalager. All historik som lagras kan ta väldigt stort utrymme i anspråk. Men hur stort ett datalager måste vara för att kallas datalager råder det olika uppgifter om. Skalin säger att det inte finns någon definition på hur stort ett datalager behöver vara. Ett datalager kan vara väldigt stort, men det kan också – i ett litet företag – få plats på en PC [Skalin 98]. Söderström å sin sida påstod 1997 att det inte finns några riktiga datalager i Sverige, då svenska datalager innehåller för liten mängd data. Han skriver vidare att den amerikanska varuhuskedjan Wal-Mart har världens största datalager på cirka 10 terabyte (10 000 000 000 000 tecken), men att storleken kommer att öka dramatiskt med tiden [Söderström 97]. Enligt branschtidningen Computer Sweden har IBM och Deutsche Telekom



Figur 2.1 Ett datalagers uppbyggnad och struktur. [Redbook 02]

inlett ett samarbete för att ta fram en datalager som är tänkt att lagra 100 terabyte, vilket motsvarar 25 miljarder sidor med text [Computer Sweden 00].

Figur 2.1 visar hur ett normalt datalager är uppbyggt och fungerar. Data kan hämtas från interna databaser men även från externa databaser, som till exempel från Statistiska Centralbyrån (SCB). ETL är en förkortning av de engelska orden extract, transform och load, som kan översättas till hämta, omvandla och ladda. ETL-processen kallas de steg då man överför data till datalagret [Unger 02]. Vid hämtning kan man välja att hämta allt, eller bara det som har ändrats sedan förra hämtningen. Vid transformeringen omvandlas alla datakällornas olika format till ett enda som används av datalagret. I denna process tvättas även data, det vill säga felaktig data korrigeras eller plockas bort. Slutligen laddas data till datalagret. [Wernersson 99]

Hur mycket data som skall hämtas och hur ofta denna process skall utföras är en fråga ett företag bör begrunda. Vissa organisationer överför så mycket data som möjligt och med så hög detaljupplösning som möjligt. Det ger möjlighet att i framtiden svara på de frågor som man i nuläget inte kan förutse, men det kostar i form av lagringsutrymme. Andra organisationer hämtar endast den data som de vet att de skall använda. Organisationer brukar ladda hem data en gång per dygn eller mer sällan, en gång i veckan eller en gång i månaden. Ibland laddas viss data ned dagligen medan det mesta laddas hem en gång i månaden. ETL-processen sker oftast nattetid.

I ett system som stödjer den dagliga operativa verksamheten struktureras data så att den endast lagras en gång. Det kallas för normaliserad struktur och med denna teknik minimeras redundans, det vill säga dubbellagring. Att strukturera data i ett datalager på detta sätt är inte lämpligt, det är bättre att använda sig av stjärnstrukturen (Eng. Star join schema) eller snöflingestrukturen (Eng. Snowflake schema). I båda dessa strukturer normaliseras inte tabellerna fullt ut. Det gör att färre tabeller måste "joins" vid frågor vilket ger snabbare svar. Nackdelen är att lagringsutrymmet ökar på grund av ökad redundans. [Söderström 97]

Stjärnstrukturen bygger på en central faktatabell och flera dimensionstabeller. Faktatabellen innehåller grundläggande mätvärden som används för de analyser som skall göras av systemet. Dessa mätvärden bör väljas med omsorg, eftersom de avgör vilka analyser man kan göra med systemet. Dimensionstabellerna är mycket mindre och innehåller naturliga verksamhetsparametrar, som till exempel kund, produkt, butik och tid. Genom att kombinera dessa verksamhetsparametrar kan man få fram till exempel hur många produkter som en eller en grupp butiker sålt av en viss produkt under viss tid, eller vilka kunder eller kundgrupper som är lönsamma, genom att summera aktuella mätvärden i faktatabellen. Faktatabellen och dimensionstabellerna är kopplade till varandra genom primärnycklar. [Kimball et al 98]

Snöflingestruktur är en variant av stjärnstrukturen som kan användas då man har stora dimensioner. Dessa delas då in i underdimensioner. Dimensionerna blir då mer normaliserade, vilket generellt ger sämre egenskaper vid analys, det vill säga snabbare svar på frågor. Fördelen är att lagringsutrymmet minskar. [Kimball et al 98]

2.3 OLTP och OLAP

För att utnyttja datalagret så flexibelt som möjligt används så kallade OLAP-verktyg. OLAP brukar ofta ställas mot OLTP-system. OLTP (Eng. On-Line Transaction Processing) är de operativa system som finns och används i en organisation för att stödja verksamheten. En organisation har ofta ett flertal olika OLTP-system som bland annat kan hantera ekonomi, lagerhållning, kunder, biljettbokningar eller tidsrapportering. Systemen kan oftast betraktas som navet i ett företag. [Connolly et al 99]

Det som karaktäriserar OLTP-systemen är att de är optimerade för att kunna hantera många transaktioner och användas samtidigt av många användare. Den data som hanteras uppdateras allt eftersom nya uppgifter kommer in och systemen måste hantera en hög genomströmning av transaktioner. [Saber et al 97]

OLAP (Eng. On-Line Analytical Processing) är en benämning på den teknik som används för att analysera data på ett flexibelt sätt. Ett OLAP-verktyg kopplas till ett datalager eller bara en PC. [Donmeyer 02].

Idén med OLAP går ut på att utnyttja en multidimensionell miljö för att på ett snabbt och naturligt sätt kunna ställa avancerade frågor till systemet och kunna vrida och vända på den data som man får till svar. Man kan på detta sätt få bättre förståelse för hur en organisation fungerar. Man kan se OLAP som ett mycket dynamiskt spreadsheet där man själv kan välja vad som skall visas i rader och kolumner. [OLAP-report 02]

I ett litet företag kan man hämta ut vissa rapporter som används vid strategiska beslut ur ett OLTP-system, men det är inte lämpligt eller möjligt i ett storföretag. Data är inte effektivt organiserade och frågorna tar tid, kräver resurser och stör den dagliga, operativa verksamheten och driften. I ett OLAP-system finns även historisk data, som ger möjligheter att göra sökningar och jämförelser över tiden. Om historisk data skulle sparas i ett OLTP-system skulle det bli väldigt stort och stora problem med redundans och inkonsistens skulle uppstå. [Skalin 98]

I tabellen sammanfattas skillnaderna mellan OLTP och OLAP i en tabell:

OLTP	OLAP
Innehåller uppdaterad data	Innehåller historisk data
Innehåller detaljerad data	Innehåller detaljerad och summerad data
Data är dynamisk	Data är statisk, endast läsbar
Repetitiva processer	Ostrukturerade och heuristiska processer
Hög genomströmning av transaktioner	Låg genomströmning av transaktioner
Förutsägbara användningsprocesser	Oförutsägbara användningsprocesser
Transaktionsorienterad	Analysorienterad
Programorienterad	Uppgiftsorienterad
Stödjer de dagliga besluten	Stödjer de strategiska besluten
Används av många användare i den operativa verksamheten	Används av få – tex. beslutsfattare och analytiker.

Tabell 2.1 Skillnader mellan OLTP och OLAP [Connolly et al 99]

2.4 Datamart

Datamart kan sägas vara ett datalager i miniatyr, med data som endast är av intresse för en specifik avdelning eller process i ett företag. Det finns ingen vedertagen översättning av termen, men dataskafferi och dataförråd är två uttryck som förekommer i litteraturen. I denna uppsats används den engelska termen, det vill säga datamart. Det finns två typer av datamart, eller rättare sagt två sätt att använda sig av datamart, dependent och independent datamart.

Dependent datamart finns i stora datalagerprojekt, där man har flera datamart kopplade till det centrala datalagret. Data insamlad från företagets operativa system lagras i datalagret. Från datalagret sänds sedan viss data iväg till ett antal datamart. Varje avdelning eller verksamhetsområde har ett eget datamart, som endast innehåller data som är intressant för just deras verksamhet. När sedan analytiker på avdelningen tar fram rapporter och gör sökningar sker det från det lokala datamartet. [Fuller 02]

Ett independent datamart utvecklas av en avdelning och data samlas endast in från avdelningens källsystem, man arbetar alltså oberoende av andra avdelningar [Fuller 02]. Detta sätt att jobba kräver färre inblandade personer, och blir därmed betydligt billigare. Engagemanget ökar och därför finns förutsättningar för att det verkligen kommer att användas. Givetvis finns det även nackdelar med ett

sådant datamart. Genom att hårt fokusera på en specifik avdelning kan man nå snabba resultat, men det finns en stor risk för suboptimering⁵. Vissa organisationer som har utvecklat flera datamart på avdelningsnivå har sedan kopplat samman dem och skapat något som kan liknas vid ett datalager. Ett vanligt problem är dock att avdelningar har använt sig av olika leverantörer och tekniker och systemen har därför inte varit kompatibla med varandra. [Skalin 98]

Fuller påpekar att storleken inte är avgörande då man skall göra skillnad på ett datalager och ett datamart. Ett datamart är alltid mindre än ett datalager inom samma organisation, men om man jämför mellan organisationer är inte påståendet korrekt. Man kan inte heller säga att om ett datamart är större än x gigabyte så är det ett datalager. Det finns ingen sådan gräns. Skillnaden mellan ett datamart och ett datalager beror på om insamling och användning är avdelningsspecifik eller sker på företagsnivå. [Fuller 02]

2.5 Fördelar, nackdelar och problem med ett datalager

Litteratur i ämnet pekar på att det finns många fördelar med att implementera ett datalager i en organisation. Ett datalager stödjer snabbare beslutsfattning och ger aktuell information [Skalin 98]. Analyser och uppföljning kan göras utifrån ett stort antal infallsvinklar och perspektiv. En stor mängd data kan hanteras och det öppnar för nya affärsmöjligheter och ger affärsfördelar gentemot konkurrenter [Crohns 01]. Analyser kan leda till att resurser och tid sparas genom rationaliseringar, i form av minskad resursåtgång, vilket ger hög ROI (Eng. Return of investment).

Men för att uppnå alla dessa fördelar måste man lägga ned ett stort arbete och på vägen finns många fallgropar. Här nedan följer ett antal problem som kan uppstå då en organisation skall utveckla och underhålla ett datalager. [Conolly et al 99] [Skalin 98]

- Projektet blir alltför omfattande och engagemanget sjunker.
- Projektet blir alltför teknikorienterat.
- Man blir alltför beroende av leverantören av systemet.
- Användarna använder inte systemet eller använder det på fel sätt.
- Dålig kvalitet på data bidrar till felaktiga analyser.
- Svårigheter med att hämta data från gamla system.

3 Metadata

Det råder begreppsförvirring kring vad som är metadata. Två författare beskriver samma sak, men benämner det på olika sätt. IT-samordnare och databasadministratörer ute i företag pratar om olika saker men använder samma namn. Vad är egentligen skillnaden mellan termkatalog och metadata-catalogue, eller mellan front-room-metadata och administrative metadata? Detta kapitel ska försöka bringa klarhet i ämnet. Även i detta kapitel klargörs uttryck som förkommer senare i uppsatsen.

3.1 Data, information och kunskap

Data, information och kunskap är tre begrepp som är starkt relaterade till varandra, men det är inte samma sak. Data bär på information, men data är inte information. Data kan sägas vara symboler utan mening, symboler i en syntax, som blir information och slutligen kunskap i en kontext. [Junfors et al 96]

Data kan sägas vara de minsta byggstenarna vid en sammanställning av information. Data som separerade delar har ingen mening, förutom att de är just data. För att få någon vidare mening måste de sättas in i ett sammanhang.

⁵ Satsning på att uppnå bästa möjliga resultat på avdelningsnivå, vilket ofta kan vara dåligt för företaget i stort.

Information är ett komplexare begrepp än data. När data blir information får det en betydelse eller en mening. Information kräver mer intelligens och analys än data, och människor är bättre på att tolka information än datorer.

Kunskap är det mest komplexa begreppet. Kunskap är den mest förädlade formen av de tre. En person som fått information behöver inte ha förstått meningen med informationen. Det är först då informationen omvandlats till mening som det kan kallas för kunskap. Mottagarens tidigare erfarenheter och referensramar styr hur mottagaren uppfattar informationen och omvandlar den till kunskap. [Tannenbaum 02]

3.2 Termer och begrepp

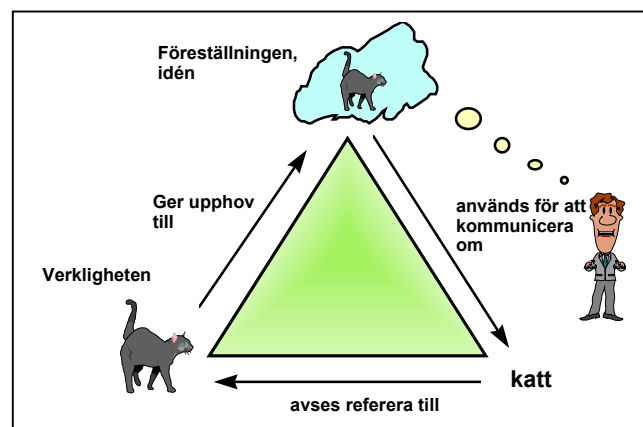
Människor talar olika språk. Inom en organisation talas också olika språk, vilket kan orsaka problem. Ett ord, det vill säga term eller uttryck, kan ha en viss betydelse i en del av organisationen, medan en annan grupp menar något annat då ordet används. Vissa ord kan vara synonyma med andra och det finns ofta lokala varianter av ord. Ibland används dessutom samma ord för olika begrepp, så kallade homonymer. Det är därför viktigt att en organisation kan beskriva och definiera dessa begrepp på ett entydigt sätt, så att alla vet vad alla talar om. [Junfors et al 96]

Ogdens triangel är ett sätt att beskriva sambandet mellan verkligheten, föreställningar och ord. Triangeln beskriver först de företeelser i verkligheten som begreppet betecknar och det ord eller de ord som vi använder för att beteckna begreppet. Begrepp är en abstraktion. Begrepp bildas bland annat genom att välja ut klasser av referenter och knyta dem till begreppet. Slutligen behöver begreppet en beteckning – ett namn som blir ett ord i vårt språk. [Söderström 97]

I ett IT-system har vi ingen möjlighet att via gester och mimik verifiera att vi menar samma sak. Vi måste ha en totalt entydig definition på våra begrepp för att vi skall få en fungerande kommunikation både mellan människor och datorer och mellan datorer. Grunden för en verksamhets effektivitet är kommunikation och det kräver att både sändare och mottagare menar samma sak. [Primdok 02]

Problemet med oklara begrepp och ord inom organisationer har funnits en lång tid. Men det har blivit mer påtagligt den senaste tiden då företag slås ihop och e-business⁶ ökar interaktionen mellan företag. Vi har blivit mer internationella, språkbruket har förändrats och påverkan från engelska språket har ökat. [DW-Institute 02]

Semantik är ett verksamhetsproblem och inte ett datalagerproblem. Men ofta lyfts detta problem upp till ytan då ett datalager införs i en verksamhet [Berlid 97]. Söderström tar upp ett fall från en amerikansk bank som skulle implementera ett datalager. Det allra första som gjordes av IT-avdelningen var att kontakta företagets ledning och be dem ta fram enhetliga begrepp inom banken. Ledningen tillsatte en grupp för att jobba med frågan, och de kontaktade kunniga inom banken för att tillsammans definiera begreppen. När sedan alla var överens togs beslut om dessa begrepp av styrelsen, som var sammansatt av personer från olika delar av banken. På så sätt förankrades beslutet av hela banken. En begreppslista skapades som hela banken stod bakom och det verkliga arbetet med datalagret kunde påbörjas, det vill säga ett datalager med hög datakvalitet. [Söderström 97]



Figur 3.1 Ogdens triangel

⁶ Affärer som sker på elektroniska handelsplatser, tex. Internet.

3.3 Begreppet metadata

Ordet metadata betyder data om data och används i många olika sammanhang, till exempel för att förenkla sökning av webbsidor på Internet, beskriva händelser och processer i en verksamhet eller bara en bildtext till en bild. Med metadata kan man transformera data till information. Bara en siffra eller ett namn säger oss ingenting, men om vi känner till vad siffran eller namnet beskriver så förstår vi mer. Siffran kan till exempel vara antalet barn i en klass och namnet kan vara en person som är lärare för klassen. Utan metadata är data meningslös. [Korzybski 02]

För att förstå hur viktigt det är med metadata kan dessa tre exempel vara till hjälp [Korzybski 02]. Hur tolkar en analytiker dessa tre exempel?

1. 73952 1323288 45554 26488 12544
2. "An Inverness Group report dated 1/9/99 states that the European market for repository tools expanded by 33% in 1998"
3. "Leading gadget vendors: Ports Group 48%, Harris Goods 29%, Zymuryg Inc. 13%"

I det första exemplet är svaret att analytikern inte förstår något. Siffrorna kan vara försäljningssiffror för ett antal butiker i en kedja, invånare i ett antal kommuner eller antalet personer som gjort inköp under en månad. Det finns två sätt att förstå siffrorna:

- Genom kontext: Om data kommer från en SQL-fråga från en viss tabell förstår vi vad vi söker.
- Genom metadata: Om det finns metadata kopplad till dessa siffror berättar den namnet på tabellen, vilket ökar förståelsen och vi kan då förstå vad som menas med siffrorna.

Det andra exemplet är lättare att förstå, med ett undantag. Vilket datum kom rapporten? Läser man enligt det amerikanska sättet så kom den i början av 1999, den 9 januari. Följer man brittisk standard så kom rapporten på den senare delen av året, den 1 september. Skillnaden är som synes väldigt stor, och problemet kan lösas genom att använda metadata, som förklarar vilken datumstandard som används.

Det tredje exemplet innehåller viss metadata, men inte tillräckligt. Ordet gadget⁷ kan betyda olika saker för människor. Procenttalen berättar inte om vilken marknad eller tidsperiod det rör sig om. Ingen information om hur procenttalen är uträknade eller källhänvisning. Slutsatsen blir att exemplet innehåller information, men för att informationen skall betraktas som värdefull krävs metadata som berättar mer.

3.4 Metadata i datalager

Några definitioner för metadata i datalagersammanhang är:

"Data that describes the meaning and structure of business data, as well as how it is created, accessed, and used" [Devlin 99]

"Metadata is everything, except for the data itself" [Kimball 02]

Begreppet metadata i datalager har på senare tid förändrats, förklaringen 'data om data' anses alldeles för enkel och omodern. I äldre litteratur beskrivs ofta metadata som information om tabellernas innehåll, så att slutanvändare kunde få information om vad som finns i datalagret. Men nu har man börjat prata om olika sorters metadata som stödjer olika delar. [Inmon 02]

Vissa skiljer på två sorters metadata, andra talar om tre sorters metadata. Kimball tillhör den första kategorin och kallar de två sorterna för back room metadata och front room metadata. Den första är mer processororienterad och beskriver varifrån data härrör. Denna typ är till stor hjälp för databasadministratören vid arbetet med att ladda ny data och vid tvättning av data. Front room metadata är beskrivande och används framför allt av slutanvändare för att identifiera betydelsen av ord och begrepp. [Kimball et al 98]

⁷ Gadget kan översättas till tillbehör, finess, grej eller pryl

Chaudhuri skriver om tre sorters metadata, business metadata, administrative metadata respektive operational metadata. Ett likhetstecken kan sättas mellan business metadata och Kimballs front room metadata, då de innebär samma sak. Administrative metadata ska underlätta för de som utvecklar och underhåller datalagret. Det kan sägas vara dokumentation över datalagret som är till hjälp i deras arbete. Det kan vara versionshantering, datamodeller och förklaringar av programmeringskod. Den tredje typen är operational metadata som kan beskrivas som statistik på driften av datalagret. Utifrån denna statistik kan datalagrets kapacitet och drift förbättras. [Chaudhuri 97]

Marco presenterar något som kallas technical metadata och vid närmare studier visar det sig vara detsamma som Chaudhuris administrative metadata [Marco 00].

Metadata behövs för att ett effektivt jobb skall kunna utföras [Jeffery 00]. Systemen blir allt mer komplexa med den snabba tekniska utvecklingen. Det samlas in mer data och storleken på datalagren ökar. Vi måste även dokumentera för framtiden, så att inte all lagrad data blir ett hinder för ett företag i framtiden. Hur skall man i framtiden kunna veta var data kommer ifrån eller hur den räknades ut? Svaret på frågan är metadata [Wiener 02].

3.5 Metametadata

Metametadata är ett begrepp som ibland förekommer i litteraturen och betyder data om metadata. Ibland förkortar man det till meta2data [Korzybski 02]. Uttrycket används då man skall beskriva metadata och de processer runt omkring. Det kan vara information om när metadata var skapad, av vem och när senaste uppdateringen skedde eller hur en organisation hanterar metadata. [Tannenbaum 02]

3.6 Termkatalog

Termkatalog är en svensk benämning vars betydelse är olik beroende på kontext. Den enklaste sortens termkatalog kan till exempel vara en lista på ett papper med termer som används i en specifik verksamhet eller organisation [Junfors et al 96]. I datalagersammanhang är dock betydelsen mer komplex.

Vissa författare översätter ordet termkatalog i datalagersammanhang till engelskans metadata. En mer korrekt översättning borde vara metadata catalogue eller kanske ännu hellre business metadata catalouge, eftersom en termkatalog innehåller just metadata om verksamhetsbegrepp.

En termkatalog stödjer sökning av data i ett datalager för slutanvändare, som till exempel analytiker eller beslutsfattare. Den är en grundläggande, och en av de viktigaste, komponenterna i ett avancerat datalager. Den innehåller information om den data som finns lagrad. Genom att använda termkatalogen kan man se vilken data som finns i datalagret, från vilken enhet data härrör samt definitioner. [Jeffery 00]

Tidiga datalager som utvecklades hade inga termkataloger. De var små och innehöll lite data. Med tiden blev informationsflödet större, systemen komplicerade och svårare att överblicka. För att kunna hantera all data, deras betydelse och ursprung används nu termkataloger i alla större datalager. För att ett datalager skulle kunna utnyttjas på bästa sätt är en implementation av en termkatalog en förutsättning. [Bischof et al 97]

Numera har termkataloger ofta ett webbaserat gränssnitt, som skall göra det enkelt för användare att navigera i systemet. Men ibland kan det vara komplicerade verktyg som är svåra för användare att förstå och ta till sig. Ett problem som alla organisationer måste lösa är hur man skall få personer att använda termkatalogen. Det är lätt att det glöms bort i stora projekt. Om termkatalogen inte används av personalen vet de inte vad som finns i datalagret. [Bischof et al 97]

Att utnyttja en termkatalog på bästa sätt är extremt komplext och en svår uppgift. Termkatalogen är en kritisk punkt i ett datalager och bör tas på största allvar. [Connolly et al 99]

3.7 Repository

Repository är oftast ett vidare begrepp än termkatalog i litteraturen. Med repository menas i de flesta fall en komponent som har all information om datalagrets funktionalitet presenterad. Medan en termkatalog endast innehåller information om termer och begrepp, så finns i repository till exempel även datalagrets

uppbyggnad beskriven och hur hämtning från källdatabaser fungerar. [Berlind 97] En termkatalog kan sägas vara en del av ett repository.

Precis som för många andra ord i detta ämne finns det även synonymer för repository. Ett ord som är ganska vanligt i litteraturen är metadata warehouse. I en vanlig ordbok översätts repository till förråd, men det används inte så i datalagersammanhang. I viss litteratur används istället ordet termkatalog, men det är en missvisande översättning då termkatalog i de flesta fall avser något annat. Slutsatsen är att det inte finns något bra svenskt ord att översätta till och därför kommer uttrycket repository att användas i denna uppsats.

Ett repository kan vara aktivt eller passivt. I ett aktivt repository sker uppdateringar automatiskt när ny data läggs in i datalagret. I ett passivt repository krävs att någon person gör uppdateringarna då ny data läggs in i datalagret. [Brackett 96]

Det är lätt att tro att ett repository löser alla problem i ett datalager. För att kunna skapa väl fungerande datalager krävs så mycket mer än ett repository, men det är en av många ingredienser i strävan mot ett datalager med god kvalitet [Berlind 97].

4 Utvärderingsmodell

För att avgöra hur organisationer använder sig av och hanterar metadata har en utvärderingsmodell med fem faktorer skapats. Det finns många sätt att kategorisera metadata, denna utvärderingsmodell följer Chaudhuri sätt att kategorisera. Chaudhuri delar in metadata i tre typer och benämner dem business metadata, administrative metadata och operational metadata. Namnen är tydliga och enkla att förknippa med vad de står för och dessutom blir det mer specifikt med tre sorters metadata, vissa andra författare delar endast in metadata i två typer. I denna utvärderingsmodell kallas dessa tre faktorer för grundfaktorer. [Chaudhuri 97]

Övriga två faktorer i utvärderingsmodellen, metadata quality och business terms, kan sägas vara stödjande då de är en grundförutsättning för de tre övriga. Metadata quality är ett begrepp som Shanks presenterar och behandlar metadata och kvalitet [Shanks et al 02]. De sista faktorn i utvärderingsmodellen, business terms, behandlar gemensamma termer och begrepp inom en organisation [Devlin 99].

För att kunna göra delanalyser och studera hur väl var och en av dessa fem faktorer hanteras har två eller tre kriterier tagits fram för respektive faktor.

Sammanlagt innehåller utvärderingsmodellen fem faktorer. De beskrivs i detalj i avsnitten 4.1 till 4.5, det vill säga vad de innebär och vilka kriterierna är. I tabell 4.1 beskrivs kortfattat faktorerna.

FAKTOR	BESKRIVNING	FRÅGOR
Metadata Quality	är kvalitén på metadata, det vill säga huruvida metadata är uppdaterade med senaste informationen och är korrekta.	49-55
Business Terms	är gemensamma termer och begrepp om verksamheten, det vill säga huruvida det finns gemensamma begrepp och termer i en organisation	7-17
Business Metadata	är metadata för slutanvändare om tabellers innehåll, betydelse och härkomst.	18-27
Administrative Metadata	är metadata om datalagrets uppbyggnad och struktur, främst för utvecklare och personer som jobbar med underhåll.	28-37
Operational Metadata	är statistik om datalagrets drift som kan användas för att skapa ett bättre datalager.	38-48

Tabell 4.1 Beskrivning av faktorerna i utvärderingsmodellen. Siffrorna i kolumnen med namnet FRÅGOR avser vilka frågor i intervjumallen som är kopplade till respektive faktor. Intervjumallen finns bifogad i bilaga 3

4.1 Metadata Quality

Metadata Quality är den ena av de två stödjande faktorerna. Faktorn handlar om kvalitet på metadata, det vill säga huruvida metadata är uppdaterade med senaste informationen och är korrekta [Shanks et al 02]. Systemet måste även vara tillgängligt och fungera utan missöden.

Om metadata innehåller flera fel blir användarna missnöjda och slutar använda systemet. För att metadata skall vinna tilltro bland användarna är det viktigt att den alltid är uppdaterad [Wiener 02]. Organisationen bör därför ha lagt ned resurser på att öka kvalitén och mäta den. Ett mått på kvalitet är antalet fel som mätningar ger.

Kriterier:

- Kvalitetsarbete:
Graden av nedlagda resurser för att öka kvalitén på metadata, samt mäta kvaliteten.
- Fel:
Graden av antalet fel på metadata i datalagret.

4.2 Business Terms

Business terms är den andra av två stödjande faktorer och handlar om gemensamma termer och begrepp om verksamheten, det vill säga huruvida det finns gemensamma begrepp och termer i en organisation. Inom en organisation är det viktigt med ett gemensamt språk som alla känner till [Junfors et al 96]. Alla avdelningar och människor i organisationen måste använda samma termer, begrepp och förkortningar för att kunna jämföra data [Söderström 97]. Alla måste vara klara med vad begreppet eller termen betyder. Ofta blir detta problem tydligt då man implementerar ett datalager i verksamheten [Berlind 97]. Det är kanske först då man samlar in och jämför data från flera olika enheter, som till exempel dotterbolag, olika länder eller avdelningar.

Gemensamma termer och begrepp måste finnas dokumenterade och lätt åtkomliga för användare. Centralt är också att användare använder sig av denna dokumentation om det är nödvändigt. [Devlin 99]

Kriterier:

- Dokumentation:
Graden av dokumentation över termer och begrepp som existerar i organisationen?
- Åtkomst:
Graden av funktionens åtkomst. Hur lätt kan användarna nå funktionen?
- Användning:
Graden av användning. Hur mycket används funktionen? Skulle den kunna användas mer?

4.3 Business Metadata

Business metadata är den första av de tre grundfunktionerna i utvärderingsmodellen. Business metadata skall främst vara till hjälp för slutanvändare, till exempel analytiker eller beslutsfattare, när de söker efter information i datalagret. De skall här till exempel kunna ta reda på vad som finns i datalagret, vad de olika tabellerna betyder och varifrån data ursprungligen kommer. [Chaudhuri 97]

Det ska inte bara finnas en sådan funktion implementerad utan det skall även kunna gå att lätt nå funktionen. För att datalagret skall bli en framgång och lönsamt är det också viktigt att denna funktion används aktivt av slutanvändare [Bischoff et al 97]. Det är de som skall skapa ett mervärde genom att använda datalagret och utifrån insamlad data ta avgörande beslut [Skalin 98].

Kriterier:

- Implementering:
Graden av implementerad funktion för att stödja slutanvändarna vid sökning av business metadata. Finns det någon implementerad funktion?
- Åtkomst:
Graden av funktionens åtkomst. Hur lätt kan slutanvändarna nå funktionen?

- Användning:
Graden av användning. Hur mycket används funktionen av slutanvändarna?

4.4 Administrative Metadata

Administrative metadata är den andra av de tre grundfunktionerna i utvärderingsmodellen. Administrative metadata skall främst vara till hjälp för de som utvecklar och jobbar med underhåll av datalagret, till exempel systemarkitekter och databasadministratörer. Med hjälp av administrative metadata skall de kunna hitta information som rör datalagrets uppbyggnad och tekniska detaljer. Det kan till exempel vara hur hämtning från källdatabaserna fungerar, hur tvättning av data sker, information om datalagrets hierarkiska struktur eller beskrivningar i programkoden. [Chaudhuri 97]

Administrative metadata kan ses som en teknisk dokumentation över datalagret. Att den finns och att den är korrekt är mycket viktigt, till exempel då någon ansvarig blir sjuk eller byter jobb. Om funktionen administrative metadata är väl utbyggd och fungerar bra kan man lätt finna svar på frågor och lösa de problem som uppstår [Marco 00]. Precis som för business metadata är det centralt att funktionen går att komma åt och att den används.

Kriterier:

- Dokumentation:
Graden av dokumentation.
- Aktualitet:
Graden av dokumentationens aktualitet.
- Användning:
Graden av användning.

4.5 Operational Metadata

Operational metadata är den sista av de tre grundfunktionerna i utvärderingsmodellen, och kanske den minsta och mest bortglömda. Operational metadata skall främst vara till hjälp för de som utvecklar och underhåller datalagret, till exempel systemarkitekter och databasadministratörer. Operational metadata kan beskrivas som statistik och information om datalagrets drift. Ständigt samlas data om till exempel vilka tabeller som det söks i mest, felrapporter eller vilken typ av användare som utnyttjar datalagret. [Chaudhuri 97]

Med utgångspunkt från insamlad operational metadata kan man konfigurera datalagret på bästa sätt, programmera bort flaskhalsar och lägga resurser där de bäst behövs. Genom att samla in operational metadata och utvärdera kan man skapa ett perfekt arbetande och högpresterande datalager.

Kriterier:

- Insamling:
Graden av insamling av operational metadata.
- Användning:
Graden av användning.

5 Empirisk analys

I detta kapitel beskrivs hur undersökningen har genomförts och de metoder som använts. Här kan läsas om de organisationer som deltar i undersökningen samt undersökningens resultatet.

5.1 Undersökningen

5.1.1 Undersökningsmetod

Utvärderingsmodellen har legat till grund för att ta fram de frågor som sedan används för att samla in empiriskt material. Detta material har baserats på primärdata, det vill säga jag har själv samlat in all data, och inte på sekundärdata, det vill säga färdiga offentliga undersökningsmaterial. Skälet till det är att jag inte hittat några undersökningar om metadata's hantering för datalager. Valet mellan att göra en kvalitativ eller kvantitativ undersökning var lätt, jag valde den förstnämnda. I den kvantitativa metoden kodas det insamlade materialet och undersöks genom att utnyttja statistiska analystekniker [Lekvall et al 93]. Ämnet och frågeställningen i denna uppsats har en sådan karaktär där den kvantitativa metoden lätt skulle leda till missförstånd, eftersom begreppsförvirringen är stor inom området. Huvudsyftet med den kvalitativa metoden är att förstå hur människor upplever olika situationer och försöka återge detta. Vid en kvalitativ undersökning kan följdfrågor ställas och en annan förståelse uppnås. Den kvalitativa metoden är dyrare att genomföra, men bäst för denna typ av undersökningar. Med den kvalitativa metoden är det oftast inte meningsfullt att kvantifiera insamlad data, det vill säga omvandla dem till sifferform för att sedan analysera och tolka då antalet undersökningsobjekt är alltför få [Lekvall et al 93]. En nackdel med den kvalitativa metoden är att bearbetningen kan präglas av den person som utför arbetet, detta skäl anses dock inte vara så starkt för att välja bort denna undersökningsmetod [Patel et al 94]. De vanliga arbetsmomenten i en kvalitativ undersökning är planering, datainsamling, bearbetning, analys och tolkning och dessa ingår även i denna undersökning [Körner et al 98].

5.1.2 Undersökningsobjekt

Det finns flera företag och organisationer som använder sig av datalager i Sverige. Valet av organisationer som deltagit i undersökningen har i stort sett varit slumpmässigt, med undantag för att organisationerna skulle vara stora i sin bransch och att organisationerna skulle tillhöra olika branscher. Endast stora organisationer har datalager som är så stora att behovet av hantering av metadata blir tydligt. Ett av skälen till att organisationerna skulle tillhöra olika branscher var ett önskemål från en av respondenterna, om att deras konkurrenter inte skulle delta i undersökningen. Fyra organisationer deltog i undersökningen. Fler deltagande organisationer skulle vara önskvärt men bristen på resurser gjorde att det endast blev fyra som undersöktes. Orsaken till att just dessa fyra organisationer har deltagit beror bland annat på personliga kontakter, tips av personer insatta i branschen samt sökande på Internet. Ett antal organisationer har fått en förfrågan om att delta, men tackat nej. Orsakerna har bland annat varit tidsbrist, bristande intresse och några organisationer berättade att de inte hade ett datalager.

5.1.3 Intervjuer

Intervjufrågorna skapades med utvärderingsmodellen som bas. Innan intervjufrågorna sammanställdes studerades intervjuteknik och hur man formulerar frågor. Intervjufrågorna kan ses som ett diskussionsunderlag och som en grund för samtalet, där eventuella följdfrågor gjorts av frågeställaren och där respondenten tillåtits att diskutera angränsande frågeställningar än de som nedtecknats före intervjutillfället. En testintervju⁸ har genomförts och den resulterade i att utvärderingsmodellen förändrades och därmed även frågorna. Resultatet från testintervjun redovisas inte i uppsatsen och ligger inte till grund för någon av uppsatsens slutsatser.

Då organisationerna jobbar på olika sätt och deras datalager är olika har det varit nödvändigt att ibland ställa djupare, mer detaljerade frågor och ställa frågorna i en annan ordning. Intervjuformen har alltså varit ostrukturerad [Halvorsen 92]. Alla respondenter erbjöds att få se frågorna innan intervjun, men inte alla tyckte det var nödvändigt. Sex intervjuer har genomförts med fyra olika organisationer, totalt har sex personer deltagit i undersökningen. Alla intervjuer gjordes på respondenternas respektive arbetsplatser och spelades in på kassetband. Intervjuerna har sammanfattats och det som redovisas i texten i stycke 5.3 till 5.7 grundas på intervjuerna. En organisation har valt att vara anonym.

⁸ Testintervjun genomfördes med Rickard Wärnelid, Datawarehouseansvarig på Skandia Kapitalförvaltning

5.2 Intervjuobjekt

Här nedan presenteras kort de organisationer som deltagit i undersökningen. Den mesta av informationen i detta stycke har inhämtas från respondenterna, men vissa kompletteringar har gjorts. Det är främst uppgifter om antalet anställda och organisationernas omsättning, som i samtliga fall inhämtats från respektive organisations webbplats på Internet.

De som intervjuats i organisationerna har haft titlar som data warehouseansvarig, analyschef, systemadministratör och samordnare. Det har alla varit väl insatta i organisationerna och organisationernas datalagerprojekt. En till två personer har intervjuats i respektive organisation.

5.2.1 Börsbolaget

Ett av företagen i undersökningen har valt att vara anonym. Företaget är noterat på Stockholms Fondbörs och kommer i uppsatsen att kallas för börsbolaget.

Det kan dock avslöjas att cirka tio personer jobbar med datalagret och räknat till datamängd är de ett av de största i Sverige. Datalagret hämtar data från cirka 25 källsystem och hämtningar görs dels en gång per dygn, dels månadsvis. Data samlas inte bara in från Sverige, utan även från ett par andra länder där företaget bedriver verksamhet.

Till datalagret kopplas ett antal datamart och mot dessa jobbar cirka 200 användare med att ta fram och analysera statiska och dynamiska rapporter. Flera analyser hämtas även direkt från datalagret av ett par specialister. Det redovisade materialet i undersökningen bygger på en intervju med två personer som jobbar med datalagret.

5.2.2 Hennes & Mauritz

Hennes & Mauritz (HM) säljer kläder och kosmetik i 14 länder världen över. Koncernen har cirka 34 000 medarbetare och omsättningen år 2001 uppgick till cirka 46 miljarder kronor.

I slutet av år 2000 påbörjades arbetet med att skapa ett datalager hos HM och sedan i mars 2002 har datalagret varit igång. Till datalagret hämtas data från cirka 15 finans- och ekonomisystem som finns i tio olika länder runt om i världen. Uppdateringar görs oftast månadsvis, men i vissa fall även dagligen. Det finns inga datamart kopplade till datalagret, utan man jobbar direkt mot datalagret. Cirka 150 personer, främst ekonomer, över hela världen jobbar aktivt mot datalagret genom ett beslutsstödssystem och har alltså möjlighet att borra sig ned, jämföra och plocka fram just den data som är intressant. Oftast används systemet kring månadsskiftet, sedan är det relativt låg användning under övrig tid. Ett stort antal rapporter produceras även varje månad och levereras till butiker och avdelningar inom koncernen.

På IT-avdelningen i Stockholm jobbar sex personer med utveckling och underhåll av datalagret. Det finns ingen exakt uppgift på datalagrets storlek, men det är mindre än de andra i denna undersökning och även det minst komplexa. All data sparas på en Oracle-databas och Oracle Sales Analyser är namnet på beslutsstödssystemet. På HM har Anders Nilsson, Systemadministratör, intervjuats.

5.2.3 Stockholms Läns Landsting och Stockholms sydvästra sjukvårdsområde

Stockholms läns landsting har ansvar för sjukvård och kollektivtrafik i Stockholms län. Landstingets omsättning var cirka 45 miljarder kronor år 2001. Antalet anställda var vid årskiftet 48 500 och av dessa arbetade cirka 84 procent inom sjukvården.

Landstinget i Stockholm har ett ganska komplext system för insamling av data, eftersom det byggts ut efter hand och det är många system som är sammankopplade. VAL kallas internt den databas där man sedan 1993 samlat data och den är placerad vid Sabbatsbergs sjukhus. Data som samlas kommer från ett flertal källdatabaser, bland annat ifrån alla sjukhus och vårdcentraler i länet, men även från apotek och privata kliniker. Nya källdatabaser kopplas på allteftersom och på så sätt byggs systemet ut. Från VAL-databasen hämtas sedan data till ett antal datamart som finns ute i de olika sjukvårdsområdena. En samordningsgrupp på central nivå inom Landstinget ger stöd till de olika sjukvårdsområdena, men

hanteringen av data, metadata och hur man jobbar, skiljer sig åt i de olika sjukvårdsområdena. På Landstinget har Torsten Sigel, Samordnare, intervjuats.

I denna undersökning har Sydvästra sjukvårdsområdet studerats närmare och det är ett av sex sjukvårdsområden inom Stockholms läns landsting. Sydvästra sjukvårdsområdet omfattar kommunerna Södertälje, Nykvarn, Salem, Botkyrka och Huddinge samt stadsdelarna Hägersten, Liljeholmen, Skärholmen och Älvsjö i Stockholms stad. Sjukvårdsområdet ansvarar för att i cirka 360 000 människor som bor i området får en god hälso- och sjukvård. Vid Huddinge Universitetssjukhus jobbar tre personer i en analysgrupp med att ta fram rapporter och analyser till cirka 55 personer som arbetar med uppföljning av driften. Göran Lord, Analyschef, har intervjuats.

5.2.4 Nordea

Nordea är en koncern inom bank- och finansbranschen och har verksamhet i Norden och Östersjöregionen. I koncernen finns 9,7 miljoner privatkunder och cirka 1 miljon företagskunder. Antalet anställda i Sverige är cirka 7500 personer.

1986 påbörjades på Nordbanken byggandet av ett datalager. Detta projekt har sedan utökats och blivit större. 1999 implementerades en stordator och webbtjänst började användas. Datalagrets storlek kan beräknas på många olika sätt, men en gissning är att det ligger på drygt en terabyte. Nordea har därmed ett av Sveriges största datalager. För tillfället samlas endast data in från Sverige, med något enstaka undantag. I de övriga länderna finns egna datalager, men tanken är att en dag försöka föra dem samman. Cirka 30 personer jobbar med Nordeas datalager och datamarts.

Datalagret hämtar data från ett 80-tal produktionssystem eller affärssystem. Hämtningar kan ske dagligen eller en gång per månad beroende på vilken analys som är intressant. Innan hämtningar sker så översänds data till ett ekonomisystem som stämmer av data och kontrollerar eventuella fel. Sedan överförs data till datalagret. Till datalagret finns sedan cirka 80 datamart kopplade, från datalagret överförs viss data till respektive datamart och det är mot dessa merparten av analyserna och rapporterna skapas. Ytterst få sökningar görs direkt mot datalagret, de som görs är oftast så kallade ad hoc-sökningar, det vill säga sökningar som inte görs regelbundet. En avdelning vill kanske vid ett tillfälle få fram speciell information. Aktuell avdelning kontaktar då någon som jobbar med datalagret och en specialist gör sedan dessa sökningar.

Större delen av de som jobbar i banken använder datalagrets data, fast det sker via de olika datamarten och de uppföljningssystem som är kopplade till datamarten. På Nordea har Nils Behrman, Datawarehouseansvarig, intervjuats.

5.3 Business Terms

5.3.1 Börsbolaget

Börsbolaget har ett antal affärsregler för att skapa likformighet och för att alla inom företaget ska jobba mot samma mål. Här finns även förklaringar på ord och begrepp som förekommer inom företaget, till exempel definieras termen kund. En databas med alla dessa affärsregler och definitioner finns i företagets interna system och kan nås av alla på ett lätt sätt. Listan utvecklas ständigt och på företaget är man medveten om att den är viktig.

Trots att det finns en lista och att den kan nås enkelt, känner inte alla i företaget till att den finns och följaktligen används den inte heller som den borde. Det beror på att det är något relativt nytt, att det är svårt att få ut information om listan till alla och att kanske inte alla förstår hur viktigt det är med gemensamma regler och definitioner. Detta leder till att det ibland uppstår problem och att vissa enstaka siffror som visas på analytikernas rapporter kan uppfattas felaktigt.

När de som jobbar med datalagret inte hittar vad de söker bland affärsreglerna så kontaktar de någon ansvarig person på aktuell avdelning och försöker reda ut begreppet. Ofta sker det då man jobbar med

framtagning av en kravspecifikation för ett visst projekt. När man sedan slagit fast en definitionen dokumenteras det i kravspecifikationen och läggs sedan in bland affärsreglerna.

Respondenterna tror att de som jobbar med datalagret skulle vilja ha fler affärsregler och tror att nyttan med datalagret skulle bli större. Ett projekt pågår inom företaget att skapa ännu tydligare affärsregler och att se till att dessa följs av alla avdelningar.

5.3.2 Hennes & Mauritz

Att alla talar samma språk är viktigt för en fungerande kommunikation, men på HM finns ingen nedtecknad dokumentation över termer och begrepp. Alla ekonomer är "uppfostade" i Sverige och med alla HM-begrepp. De är således bekanta med alla uttryck. Det är även samma begrepp som finns i finans- och ekonomisystemen som finns i datalagret och eftersom alla jobbar med dessa system så förstår de också begreppen som förekommer i datalagret. Ingen i organisationen verkar sakna dokumentation över business terms, men skulle det uppstå problem får man kanske skapa en sådan. Någon enstaka gång händer det kanske att folk blandar ihop två begrepp och får ut ett annat resultat, men det ligger säkert inom felmarginalen. Detta uppfattas inte som ett stort problem.

5.3.3 Stockholms Läns Landsting och Stockholms sydvästra sjukvårdsområde

Inom Landstinget finns en mängd olika yrkesgrupper med många egna definitioner, fackuttryck och medicinska termer. Det finns många viljor och det är en stor organisation så det kan vara svårt att enas om ett begrepp. Om man frågar olika människor inom organisationen vad som menas med ordet patient får man många olika svar.

Det finns en rad olika nationella samlingar med termer inom sjukvården. Bland annat har Socialstyrelsen en begreppssamling som man inom landstinget försöker att följa. En informationsansvarig på Landstinget har lagt upp en termdatabank på intranätet, som är möjlig att nå för de flesta som jobbar inom landstinget. Det finns ingen mätning hur mycket den används, men det troliga är att det sker i begränsad omfattning. Det är svårt att få ut budskapet om dess existens.

För att en rapport skall vara tillförlitlig är det oerhört viktigt att alla jobbar på ett likartat sätt och med samma termer, men det görs tyvärr inte. Man har haft problem med termerna nybesök och återbesök. Det visade sig vara stora skillnader på tolkning och användning av termerna mellan olika vårdcentraler. Vissa hade mest återbesök medan andra hade en majoritet av nybesök. Vid närmare kontroll visade det sig att termerna tolkades på olika sätt på respektive vårdcentral. På vissa vårdcentraler hade man även konfigurerat systemen så att alla var nybesökare och om någon var återbesökare fick man ändra det manuellt, ett moment som var lätt att glömma eller hoppa över.

5.3.4 Nordea

På Nordea finns en framtagen lista med termer, begrepp, produktkoder och förkortningar som används i olika IT-system. Listan kallas för termförteckning och finns inte implementerad som någon funktion eller söktjänst, utan det är helt enkelt en tjock katalog av papper. De som utarbetat listan är controllers på ekonomiavdelningen och de är skiljedomare om tvister skulle uppstå. Hela banken är dock med och utreder och fungerar som remissinstanser då nya termer skapas. Termförteckningen anses mycket viktig, och kallas internt "bibeln", och utan den skulle inte datalagret kunna fungera så bra som det gör idag. Det är ytterst viktigt att uppdateringar sker när förändringar görs.

Termförteckningen finns på de avdelningar i banken som har behov av den. En rådgivare på en bank känner kanske inte till den och behöver inte heller använda den. Men alla som bör känna till den gör det. Den finns inte implementerad på något elektronisk sätt, men hade man idag skapat termförteckning hade man gjort detta.

Användningsområdet för katalogen är relativt stort och beroende på arbetsuppgifter. Till exempel kan man vara intresserad av att få reda på om en sorts information finns i något system. Beroende på tjänst

och arbetsuppgift tittar man i den olika mycket och det går inte att ge något generellt svar på hur ofta den används. Uppfattningen är att de som skall och behöver använda termförteckningen verkligen gör det.

5.3.5 Sammanfattning – Business Terms

Alla respondenter är överens om att ett gemensamt språk är viktigt för en väl fungerande organisation. Organisationerna jobbar alla för att utveckla ett sådant språk, men man jobbar på lite olika sätt. Det största problemet verkar dock vara att få människor ute i verksamheten att använda de funktioner där informationen finns. Att lägga ut informationen på ett intranät är inte hela lösningen, utan det krävs mer för att få medarbetare i en organisation att börja använda det. Analysfel på grund av felaktig användning av termer förekommer, men de flesta faller inom respektive organisations gällande felmarginal.

Kriterier:

- Dokumentation:
Tre av fyra undersökta organisationer hade en förteckning med termer och begrepp som förekommer i organisationen. Den fjärde respondenten påstod att användarna hade denna kunskap i huvudet.
- Åtkomst:
Av de tre respondenter som hade en förteckning med termer och begrepp var det två som hade den på sitt lokala intranät och därmed fanns den åtkomlig för en majoritet av de anställda. Den tredje respondenten hade endast en tryckt förteckning med mer teknisk information som inte var av intresse för gemene man i organisationen.
- Användning:
De organisationer som har en lista med business terms använder dem inte fullt ut och medarbetare vet inte om att den överhuvudtaget finns. Med bättre användning skulle mervärdet av datalagret bli större, tror ett par av respondenterna.

5.4 Business Metadata

5.4.1 Börsbolaget

På börsbolaget finns en applikation där slutanvändare till viss del kan gå in och titta på vilken data som finns i datalagret, det vill säga ge användarna viss information om business metadata (tabellers innehåll, betydelse och härkomst). Hur mycket de kan se beror lite på vilken roll de har i organisationen, eftersom det finns olika behörighetsnivåer för olika roller. Om inte slutanvändaren finner det han eller hon söker så kontaktas de som jobbar med datalagret. De försöker uppfylla önskemålet och ta fram önskad information och skapa en rapport. Efterfrågan av denna information varierar. Vissa användare är mer drivna och söker mer information, medan andra är nöjda med den information de har i dagsläget.

Börsbolaget har just startat ett projekt med att ta fram en webbportal där man har tänkt att lista alla rapporter som skapas varje månad samt information om datalagrets innehåll. Tanken är att slutanvändarna på ett enklare sätt skall kunna finna vad de söker. Det kan sägas vara en innehållsförteckning över vilka rapporter och information som kan hämtas. Det är tänkt att det även ska beskrivas hur varje rapport är framtagen och varifrån data kommer. I webbportalen kommer man troligtvis endast att kunna se det som hämtas ur datalagret just nu, inte all data som finns att hämta. Det skulle troligtvis växa till ett allt för omfattande och komplicerat projekt. När webbportalsprojektet blir klart är ännu lite oklart, men man tror definitivt att affärsnyttan kommer att bli större om det används på rätt sätt.

I analysverktygen finns hjälpfunktioner där man kan se vad tabeller och siffror betyder, men det skulle utvecklas mer och det görs även ständigt. Ibland frågar användare om bakgrunden till vissa siffror och då får de som jobbar med datalagret förklara varifrån data kommer, men det anses inte som något stort problem och det händer inte så ofta. Ibland händer det att flera personer klagar på en viss funktion och då försöker man åtgärda och hitta en bättre lösning.

5.4.2 Hennes & Mauritz

Före datalagrets introduktion i HM, i mars år 2002, plockade ekonomerna ut information från sina ekonomi- och finanssystem och överförde det till ett Excelark. Man använde helt enkelt ”copy-and-paste-tekniken”. I Excel kunde man sedan ta fram eftersökt information. När sedan datalagret introducerades var alla som skulle jobba med det mycket bekanta med siffrorna och termer, vilket gjorde övergången enkel.

Det nya beslutsstödssystemet som är kopplat till datalagret är mycket bättre och effektivare, och det har underlättat arbetet för många. Beslutsstödssystemet introducerades med en endagskurs för alla slutanvändare.

Det bör även tilläggas att all information från finans- och ekonomisystem överförs till datalagret och att alla system är identiska, skillnaden är bara att de innehåller information om olika länders ekonomi. När en ny person kommer in i organisationen får han eller hon först lära sig jobba med företagets finans- och ekonomisystem och när de sedan är bekant med dem får de utbildning om hur datalagret fungerar och dess beslutsstödssystem.

Slutsatsen är den att användarna är bekanta med termer och begrepp från ekonomisystem och att all data finns tillgänglig. Därför läggs det inte ner något större jobb på att utveckla business metadata. Det är inte heller någon större efterfrågan på mer business metadata från användarna.

5.4.3 Nordea

Inom Nordea måste man skilja på de få som jobbar direkt mot datalager och det stora flertalet som jobbar mot ett datamart i ett uppföljningssystem. De förstnämnda är specialister som jobbar direkt mot datalagret och de är väl insatta i hur det fungerar och vad som finns där.

De flesta inom banken jobbar med ett uppföljningssystem som är kopplat mot ett av datamarten. I uppföljningssystemen finns rapporter och information om just deras egen verksamhet, men bankens medarbetare är medvetna om att det finns så mycket mer lagrat. I alla system finns hjälpfunktioner där slutanvändare kan se vad en specifik rad betyder i en rapport eller vad en förkortning står för. Observera dock att denna metadata inte är kopplad till datalagret, utan endast en hjälpfunktion implementerad i applikationen. Den typ av business metadata som finns i uppföljningssystemen når man på lite olika sätt, beroende på hur systemet ser ut. I vissa system klickar man helt enkelt på ordet eller tabellen som efterfrågas, och därmed öppnas ett fönster med information. I andra system klickar slutanvändaren på en hjälpruta, och en beskrivning över hela sidan visas. Det är viktigt den skall vara tillgänglig på ett enkelt sätt, att behöva slå i en manual för hand är inget bra alternativ.

Det är svårt att uppskatta hur stor användningen av denna typ av business metadata är bland det stora antalet slutanvändare. Men man påpekar i internutbildning och via helpdesk-funktioner att det finns en hjälpfunktion där användaren kan få veta mer. Säkert är att om den inte funnits hade det varit mer förfrågningar. Funktionen måste absolut finnas.

När någon avdelning vill ha mer data om vad som finns i det stora datalagret, så får den anställde kontakta någon på som jobbar med datalagret. Detta sker ibland och ofta har de önskemål om specifik data och frågar om den finns tillgänglig. Om det är möjligt så kan man senare gå vidare på olika sätt. Att skapa en applikation där användare kan söka direkt mot datalagret finns det inget behov av och det skulle vara oerhört komplicerat och svårt. Det nuvarande systemet anses fungera bra.

5.4.4 Stockholms Läns Landsting och Stockholms sydvästra sjukvårdsområde

Det finns olika användarnivåer inom landstinget och informationen på respektive nivå skiljer sig. På intranätet finns en förteckning över vad som finns lagrat i den stora VAL-databasen. Det är en dokumentation som kan nås av alla, men som inte används och inte heller skall användas av alla. Dokumentationen ger till en början en enkel beskrivning på hur databasen är uppbyggd och fungerar. Sedan följer ett antal tabeller med beskrivningar över lagrade data, variabelnamn och enklare förklaringar. En samordnare på Landstinget ser till att denna dokumentation är uppdaterad. Detta är ett ganska enkelt

sätt att arbeta på. För de personer som inte är så insatta i systemet kanske det kan vara svårt att förstå vissa delar, men då kan man få hjälp av samordnare på landstinget eller de som jobbar i analysgrupper som finns i sjukvårdsområdena.

I sydvästra sjukvårdsområdet så finns det ingen förteckning över vad som finns i deras datamart och vad som är nåbart för slutanvändare. Det finns inget större behov av detta, men om det uppstår kan de använda sig av landstingets dokumentation över vad som finns i VAL-databasen och dra slutsatsen att motsvarande data finns i det lokala datamartet. Om en fråga skulle uppstå hos en slutanvändare är det vanligare att han eller hon kontaktar analysgruppen som finns i varje sjukvårdsområde för att få besked.

Cirka 55 personer jobbar med rapporter som tas fram av analysgruppen som finns i Stockholms sydvästra sjukvårdsområde. Allt finns att nå via intranätet där man samlat all information på ett ställe. Alla kan se alla rapporter och arbeta med analysverktygen. Det kan vara svårt att förstå alla diagram för den oinvidige, men det är inte ett problem för de som jobbar med det nu, eftersom de väl känner till verksamheten. För att få alla att börja använda rapporterna och verktygen finns flera hjälpfunktioner. Bland annat kan man få se rörliga bilder med ljud som beskriver webbsidan och dess funktioner. Då alla jobbar på samma plats är det även enkelt för slutanvändaren att kontakta analysgruppen då frågor uppstår.

5.4.5 Sammanfattning – Business Metadata

Undersökningen visade att det finns två typer av business metadata, dels traditionell dokumentation som beskriver innehållet i datalagret, dels de hjälpfunktioner och förklarande texter som finns i uppföljningssystemen och analysverktygen.

Endast två av organisationerna hade något stöd i elektronisk form för slutanvändarna för den första typen av business metadata. Börsbolaget var ensamt om att försöka utveckla stödet och de trodde även att funktionen skulle öka mervärdet av datalagret. Ett annat sätt för en slutanvändare att få reda på fakta om vad som finns i datalagret är helt enkelt att de kontaktar någon som jobbar med datalagret, och de som jobbar på detta sätt tycker att det fungerar. En diskussion som uppstod i de flesta intervjuerna var om alla slutanvändare egentligen har tid och om de skall lägga ned tid på att söka efter business metadata. Kanske är det bara specialister eller ett fåtal på varje avdelning som skall göra dessa sökningar.

Samtliga organisationer hade hjälpfunktioner och förklarande texter i sina uppföljningssystem och analysverktyg, men ingen visste hur stor användningen var och vad slutanvändarna egentligen tyckte.

Kriterier:

- **Implementering:**
Två organisationer i undersökningen hade någon form av implementerad funktion där slutanvändare kunde nå business metadata. Parallellt löser de även problem med muntlig kommunikation, ett sätt som även den tredje respondenten använder sig av. Den fjärde respondenten påstod att det inte fanns någon efterfrågan av business metadata.
- **Åtkomst:**
Organisationerna har olika lösningar, men de som har implementerade funktioner nås lätt av användarna. Ett problem som organisationerna har, är att inte alla slutanvändare känner till att funktionerna finns.
- **Användning:**
Användningen och efterfrågan av business metadata bland slutanvändarna är inte så stor enligt respondenterna och därför läggs det inte ned så mycket arbete på att utveckla detta.

5.5 Administrative metadata

5.5.1 Börsbolaget

Börsbolaget använder sig av ett verktyg som heter SAS Data warehouse-administrator (WA), för att hantera administrativa metadata, det vill säga information om ETL-processen och datalagrets struktur och uppbyggnad. Verktyget skulle kunna användas mer. Just nu används det mest för dokumentation, men

verktyget innehåller många fler funktioner som inte används. Ett annat problem är att inte alla system (som tillhör datalagret) finns dokumenterade i WA, eftersom de är äldre. Men företaget jobbar för att lägga in allt i WA, vilket gör det enklare för till exempel en nyanställd. Programmerare och administratörer missar ibland att dokumentera uppdateringar i WA, vilket innebär att systemen kommer i ofas. Ett exempel är att variabelnamn byts i en av källdatabaserna, men inte direkt gör uppdateringen i WA.

Det finns även skriftlig dokumentation som levereras då ett delsystem är klart, som framförallt används i början. Efterhand lär sig medarbetarna och de behöver inte titta så mycket i dokumentationen.

Även i koden finns dokumentation som är till hjälp. När man gör förändringar i någon kod skriver man alltid vad som gjorts, vem som gjort det och tidpunkt.

Det finns en eller två personer i företaget som besitter kunskap om datalagret som ingen annan har. Dessa jobbar mot speciella avdelningar som har en komplicerad struktur. Dessa medarbetare är inte ovärderliga, men viktiga. Den kunskap som de besitter går säkerligen att inhämta genom att studera dokumentation, prata med andra i organisationen och kontakta leverantörer. Detta jobb skulle dock ta tid.

5.5.2 Hennes & Mauritz

Vid HM sker dagligen små förändringar i systemet. Det finns en mängd dokumentation över systemet, hur det till exempel är uppbyggt och denna dokumentation är enkel att förstå. Det finns även ett versionshanteringssystem som används då man gör förändringar. Om en av de sex som jobbar skulle bli sjuk eller sluta skulle det finnas tillräckligt med information dokumenterad för att en ny person skulle kunna sätta sig in i jobbet.

5.5.3 Nordea

När någon ny person skall lära sig mer om datalagret vid Nordea, till exempel en nyanställd så brukar man först presentera datalagrets struktur och uppbyggnad med en hjälp av en PowerPoint-presentation. Sedan finns en termförteckning där de kan läsa mer om alla termer och begrepp. I företaget finns väldigt stränga krav på standardiserad dokumentation av system, vilket är till stor hjälp. Det finns regler för hur allt skall se ut och göras. I programkoden står det noggrant och utförligt vad just en specifik koddell gör. Nordea har diskuterat att implementera SAS Data warehouse-administrator (WA) – ett system som delvis hanterar metadata. Implementeringen skulle dock innebära andra stora förändringar av det befintliga systemet, så detta har valts bort. Hade man påbörjat ett datalagerprojekt idag hade troligen WA varit en komponent som använts.

Så fort en förändring görs i systemet måste detta dokumenteras i systemdokumentationen och noteras i koden. Det finns dock inget system som kontrollerar om någon eventuellt missar det, utan det bygger helt på personligt ansvar.

Givetvis används administrativa metadata, och det används mycket av de som jobbar aktivt med systemen. Ett mål är att minst två personer skall kunna samma sak. Om endast en person har kunskapen blir företaget sårbart och personen i fråga aldrig riktigt ledig. Någon större efterfrågan på ytterligare administrativa metadata finns inte. De som jobbar med datalagret kan sitt jobb och en ny person i organisationen får stöd och hjälp på olika sätt.

5.5.4 Stockholms Läns Landsting och Stockholms sydvästra sjukvårdsområde

Hos Landstinget skiljer sig hanteringen av administrativa metadata åt beroende på vilket delsystem som undersöks och det kan vara svårt att ge en generell bild. Systemet är komplext och data transporteras mellan olika databaser som administreras av olika personer. SAS Data warehouse-administrator (WA) används till exempel i en tidig fas, där data överförs från vissa källdatabaser till VAL-databasen. Detta arbete utförs och administreras av TietoEnator.

På sydvästra sjukvårdsområdet finns ett krav från ledningen att alla måste dokumentera hur de jobbar, så att någon annan kan överta arbetsuppgifterna. Till datamartet finns en manual med den allra viktigaste

informationen för att kunna förvalta systemet. Här kan man läsa om hur hämtningar sker och var man kan finna speciella program.

5.5.5 Sammanfattning – Administrative metadata

Administrative metadata är viktigt, det anser alla respondenter. Precis som andra datoriserade system måste det även för ett datalager finnas systemdokumentation och beskrivningar i koden. SAS Data warehouse-administrator är ett verktyg som många av respondenterna jobbar med eller funderar att börja använda. Mycket av kunskapen finns även hos dem som jobbar med datalagrena, men för att säkerställa att kunskapen finns kvar i organisationerna kräver företagsledningarna att det även finns manualer och annan dokumentation. Administrative metadata är något som organisationerna har kontroll över.

Kriterier:

- Dokumentation:
Alla organisationer har strikta regler på god dokumentation och följer dessa. Det förekommer dock att viss kunskap är svår att dokumentera och endast finns i huvudet på en eller ett fåtal medarbetare.
- Aktualitet:
Dokumentationen har hög aktualitet, när en förändring görs i systemet ändras det också i dokumentation. Någon respondent erkände att det ibland kunde slarvas, men att det inte var allvarligt och inte orsakade några större konsekvenser. Ingen av respondenterna slarvade så mycket att inte någon litade på dokumentationen.
- Användning:
Administrative metadata används på olika sätt av olika personer i organisationen. Många som jobbar med datalager har ofta god kunskap efter många års arbete och behöver därför inte leta i dokumentationen. Men för en nyanställd är det viktigt, men det är lika viktigt att ha någon vid sin sida som man kan fråga.

5.6 Operational metadata

5.6.1 Börsbolaget

Hos börsbolaget består datalagret av flera system och alla delsystem har sina egna rutiner. I varje applikation som används av slutanvändare kan man gå in och se hur mycket de används, det vill säga få tillgång till statistik och information om driften, det som i uppsatsen kallas operational metadata. Statistik sammanställs endast på avdelningsnivå, men om man önskade skulle även mätningar kunna göras på individnivå.

Ansvar för den stordator där all data samlas har produktionsavdelningen och de skall se till att den fungerar på bästa sätt. De som jobbar med datalagret har dock bra kontakt med dem och meddelar om det är något speciellt som skall ske eller om de tycker att något skall förändras.

Alla specialbeställningar på rapporter som görs samlas i en databas, och när man ser att det är stor efterfrågan på en sorts rapport, blir den permanent och möjlig för slutanvändarna att nå direkt.

Just nu pågår en enkät bland slutanvändarna där de tycker till om datalagerprojektet. När enkäten är sammanställd kommer resultatet att analyseras och ett försök att genomföra de förbättringar som efterfrågas. Mycket går på känsla på Börsbolaget, man märker när det behövs förändringar.

5.6.2 Hennes & Mauritz

Datalagerssystemet vid HM togs i drift i mars år 2002 och innan dess hade de under nästan två år utvecklat och testat för att göra ett så bra system som möjligt. För tillfället görs inte mätningar kontinuerligt, men det har fungerat bra hittills. Systemet håller god prestanda och det är relativt få användare, vilket gör att administratörerna har god överblick över hur det fungerar. Om systemets prestanda skulle försämrats är det möjligt att de får börja undersöka och mäta. Det som ligger närmast att mäta är nätverkets trafik och belastning och processorernas arbetskapacitet.

Då systemet sattes i drift mätte man hur användarna utnyttjade systemet, men det görs inte längre. Nu används det av alla som skall använda det.

5.6.3 Nordea

Vid Nordea finns det ett system som kontrollerar hur webbapplikationerna (uppföljningssystemen) används av personalen. Denna information används för att se vem som använder systemen och på vilket sätt de används. Ett exempel är att kontrollera vad som händer efter en kurs, vilken effekt en kurs har haft. Om en viss region eller ett kontor går dåligt kan man kanske se att användning av uppföljningssystemet är dålig och påpeka detta. Dessa uppföljningssystem ger mycket användbar feedback för många och är till stor hjälp.

Ett problem är svarstider på webbapplikationerna. Ibland inkommer det klagomål på att svarstiderna är långa och att sökningar tar tid, ett problem som alltid beror på att det finns gamla datorer ute på kontoren och på dålig överföringsteknik över nätverket. Datakraften i själva datalagret och datamarten är bra, så problemet med långa svarstider är svårt att åtgärda för de som jobbar med datalagret. Lyckligtvis försvinner allt fler gamla datorer och överföringshastigheten blir snabbare. Problemet håller på att försvinna, men det har varit ett stort problem.

Det finns en avdelning inom Nordea som arbetar med datakraft och hårdvara, som kallas för produktionsavdelning. De kontrollerar hur körningar och batchar fungerar och får informationen från stordatorn. De jobbar aktivt med att skapa stabilare och smidigare system i hela koncernen och det gäller givetvis även i datalagret. De gör hela tiden förändringar och förbättrar prestanda och ibland sker det på önskemål från de som jobbar med datalagret.

Fel på data från källdatabaserna upptäcks i 98 % av fallen av ekonomiavdelningen som först kontrollerar data innan det överförs till datalagret. Kontrollen sker genom att siffror jämförs från olika system och om något inte stämmer sker kontroll manuellt.

5.6.4 Stockholms Läns Landsting och Stockholms sydvästra sjukvårdsområde

Det finns ingen gemensam insamling av statistik vid Landstinget. Uppdateringar av statistik sker ad hoc, då något verkar fungera dåligt försöker de lösa det.

I Sydvästra sjukvårdsområdet finns inte heller några kontroller eller mätningar. De ansvariga jobbar nära slutanvändarna och får snabb respons på om något är fel eller svårt att använda. Vid hämtningar från VAL-databasen kontrolleras alltid loggfilen så att allt har fungerat som det skall, och om något skulle vara fel rapporteras det till landstingets samordnare.

5.6.5 Sammanfattning – Operational metadata

Ingen av respondenterna hade någon nedskrivna handlingsplan för hur man skall hantera operational metadata. Organisationerna samlar in viss data över hur systemen fungerar och denna information ligger till viss del till grund för förändringar och förbättringar av systemen, men främst jobbar organisationerna i undersökningen på känsla eller ad hoc. Om man upptäcker att något fungerar dåligt gör man något åt problemet. Alla respondenter tyckte att det nuvarande sättet att hantera operational metadata fungerade på ett tillfredställande sätt.

Kriterier:

- Implementering:
Organisationernas system är stora och komplicerade och består av flera delsystem där insamling av data sker på olika sätt. Man för loggar över överföringar av data mellan system och kontrollerar hur servrar jobbar. Ingen av respondenterna hade någon samlad statistik över hela datalagrets drift.
- Användning:
Organisationerna använder sig av insamlad operational metadata till viss del, men de flesta förändringar görs mer på känsla. Mindre förändringar görs nästan dagligen hos alla respondenter.

5.7 Metadata quality

5.7.1 Börsbolaget

Vid börsbolaget är metadata inget som diskuteras dagligen, men man är medveten om att det är viktigt. Hög datakvalitet kräver hög kvalitet på metadata. Man skulle vilja ägna sig mer åt metadata och tror att mervärdet skulle bli större om man lade ned mer arbete. Men man måste prioritera och det finns viktigare saker, eller rättare sagt mer kritiska saker som måste lösas först.

Det finns ingen handlingsplan för att höja kvalitén på metadata, men det finns en del projekt på gång där kvalitet om metadata finns med på ett hörn. Till exempel vill företaget få in alla system i WA och de jobbar även med en webbportal där slutanvändare kan söka efter innehåll i datalagret.

5.7.2 Hennes & Mauritz

HM:s datalager är relativt nytt och innan det togs det i drift gjordes ansträngningar för att hitta en optimal lösning. Idag är därför alla, båda slutanvändare och administratörer, nöjda med resultatet. Kanske beror det på att det gamla sättet att arbeta var omodernt och krångligt. HM tror inte att de skulle kunna öka mervärdet genom att använda metadata mer. Inte heller skulle användarna kunna använda systemet mer och på så vis få ut mer information. Med tiden kommer kanske mer information in i datalagret, fler människor blir involverade och då ökar efterfrågan på metadata, men inte just nu.

5.7.3 Nordea

Vid Nordea anser man att för att all data i datalagret skall hålla hög kvalitet är det viktigt med kvalitet på all metadata. Det går hand i hand och Nordea anser att det är oerhört viktigt med bra kvalitet på all data. För ett par år sedan skapades ett datamart med namnet Evenemangsspecifikation. Här samlas all information om en kund från olika affärssystem och produktionssystem, information som sedan sänds tillbaka till kunder och som används vid bokföring, årsredovisningar och deklarationer. Tidigare hade varje enhet sänt denna information, men numera görs detta arbete centralt via det aktuella datamartet och datalagret. Att dessa siffror skall vara korrekta är mycket viktigt, många var först tveksamma till projektet och ifrågasatte om all data var tillförlitlig. Det hela blev ett stort kvalitetsprojekt och det föll mycket väl ut. Alla parametrar ansågs viktiga och givetvis även metadata.

Nordea påpekar även att ju fler som använder systemet ju snabbare upptäcks eventuella fel och hela tiden kommer nya idéer hur man kan utveckla ett system. Om endast två tre personer använder sig av datalagret går utvecklingen långsammare.

Varje månad har de som jobbar med datalagret ett avstämningsmöte med ekonomiavdelningen där de diskuterar eventuella fel vid överföringen eller allmänna förbättringar.

Någon direkt handlingsplan för att öka kvalitén på metadata finns inte. Men man försöker ständigt öka hjälpfunktionerna i uppföljningssystem så att slutanvändarna skall få ökad förståelse. Det finns som tidigare nämnts strikta regler på hur dokumentation i systemet skall se ut.

Det finns ingen egentlig statistik på hur ofta fel på metadata uppstår eller vad det är för fel. Om någon upptäcker ett fel så meddelas det till ansvarig och sedan så åtgärdas det.

5.7.4 Stockholms Läns Landsting och Stockholms sydvästra sjukvårdsområde

Bristen på kommunikation är ibland uppenbar inom Landstinget, och det beror på att det är en stor organisation och som är utspridd över många lokaler. Men de försöker bli bättre. Bland annat har analysgrupperna i de olika sjukvårdsområdena börjat jobba i ett nätverk för att kunna förmedla kunskap till varandra på ett bättre sätt. Detta samarbete kommer säkert även att behandla metadatakvalitet och hur de skall kunna arbeta med det på ett bättre sätt.

Kvalitén på data är god, men inte 100-procentig. Ibland överensstämmer inte verkligheten med den data som finns lagrad, men det beror på den komplexa verksamheten och inte på de som jobbar med

datalagret. Ett exempel är vårdtillfälle, där en person kan vara inskriven på två olika vårdinrättningar samtidigt. Detta kan leda till att man kan dra fel slutsatser då man analyserar antalet inskrivna på en vårdinrättning.

5.7.5 Sammanfattning – Metadata quality

Respondenterna är medvetna om att hög kvalitet på metadata är viktigt för att skapa ett bra datalager, men arbetar inte på något sätt aktivt med det. De flesta respondenter anser att dagens datalager fungerar på ett tillfredställande sätt. Att lägga ned mer arbete på att skapa bättre metadata skulle kosta mer än det smakar. Det är en avvägning, anser respondenterna.

Kriterier:

- Kvalitetsarbete:
Organisationerna har ingen samlad plan för att hantera metadata eller för att skapa bättre metadata. Vissa organisationer i undersökningen driver dock mindre projekt för att skapa bättre metadata.
- Fel:
Ingen av respondenterna mäter antalet fel på metadata eller har någon organisation som hanterar fel på metadata. Den allmänna uppfattningen är att det inte är så många fel. Fel upptäcks på olika sätt och åtgärdas på olika sätt. Ett typfall kan vara att en slutanvändare påpekar ett fel eller ett problem för en datalageradministratör. Den sistnämnde kontrollerar och åtgärdar därefter eventuellt felet.

6 Slutsats och Epilog

6.1 Slutsats

Syftet med uppsatsen är att undersöka hur organisationer hanterar metadata i datalager. Slutsatsen av undersökningen är att organisationerna är medvetna om att det är viktigt med metadata och de har vissa strategier för att hantera metadata. Organisationerna är nöjda med sitt arbete, men en allmän uppfattning är att de skulle kunna jobba mer med metadata och därmed få ut större affärsnytta med hela datalagerprojektet.

Ett gemensamt språk är viktigt för att en organisation skall fungera och alla undersökta organisationer jobbar med att förbättra sig inom detta område. Ett hinder är att medarbetare ute i verksamheten inte alltid är medvetna om gemensamma begrepp och var man kan finna denna information.

Alla organisationer har stöd för slutanvändarna, så kallad business metadata, så att de skall kunna förstå vad siffror och tabeller betyder. Vissa organisationer har funktioner som gör att slutanvändarna kan söka efter data som finns i datalagret. Ett problem är att användarna inte vet var den sistnämnda funktionen finns eller hur de skall använda den. Det vanligaste sättet för en slutanvändare som vill veta vad som finns i datalagret är att personligen kontakta någon administratör eller specialist.

Kunskapen om datalagret, det vill säga administrative metadata, finns dokumenterad hos organisationerna på ett, enligt de själva, tillfredställande sätt. Men många av respondenterna är dock sårbara för avhopp, eftersom viss kunskap är svår att dokumentera och endast innehas av ett fåtal personer. Några av respondenterna påpekade att det fanns hårda krav på dokumentation från ledning, men nämnde att det finns luckor.

När det gäller att förbättra och utveckla datalagret samlar organisationerna in viss statistik och information, så kallad operational metadata, över hur datalagret fungerar och används. Ett ord som återkom hos flera respondenter var ad hoc. Administratörer av datalagret jobbar på känsla och då något fungerar dåligt eller orsakar problem så åtgärdas detta.

Ingen av de undersökta organisationerna arbetar aktivt med att förbättra kvalitén på metadata, då detta är kostsamt och kvalitén just nu anses vara god.

I det första kapitlet i denna uppsats, under rubriken bakgrund är författarna Brackett och Tannenbaum kritiska till hur metadata hanteras i organisationers datalager. Resultatet i denna undersökning och

författarnas åsikter stämmer inte överens. Organisationerna i denna undersökning är medvetna om vikten av metadata och de jobbar för bättre hantering av metadata.

6.2 Utvärderingens och utvärderingsmodellens giltighet

Utvärderingen av hur organisationer hanterar metadata har gjorts med hjälp av utvärderingsmodellen presenterad i kapitel 4. Resultatets tillförlitlighet är direkt beroende av utvärderingsmodellens giltighet. Någon undersökning av utvärderingsmodellens giltighet är inte gjord, det vill säga någon utvärdering av själva utvärderingsmodellen är inte gjord. En sådan utvärdering är därför ett nödvändigt nästa steg för att utveckla utvärderingsmodellen ytterligare. Till exempel skulle utvärderingsmodellen kunna testas på ett antal experter inom området, som skulle kunna föreslå förändringar av föreslagna faktorer eller tillägg av ytterligare centrala faktorer.

Den utvärderingsmodell som presenteras i denna uppsats ska alltså ses som ett steg på vägen mot en mer genomarbetad utvärderingsmodell.

6.3 Den empiriska undersökningens giltighet

Undersökningen har gjorts enligt den kvalitativa metoden och det finns många fördelar med denna typ av undersökningar, men även nackdelar. Uppsatsens resultat är i stor utsträckning beroende av hur frågorna ställts och hur de uppfattats av respondenterna. Resultatet kan även ha påverkats av att respondenterna fått frågorna i förhand och därmed kunnat förbereda sig och ge tillrättalagda svar. Fördelen med att sända ut frågorna i förväg är att respondenterna kunnat ge mer utförliga och genomtänkta svar.

Jag har under intervjuerna försökt att ställa detaljerade frågor och följdfrågor, för att få en uppfattning av verksamheten. Men ibland vill kanske de tillfrågade presentera en bättre sida av sitt företag och lämnar därför ett svar som inte helt överensstämmer med verkligheten.

Resultatet kanske även hade blivit annorlunda om man talat med någon annan person i organisationen, beroende på personlighet och ställning inom organisationen. Jag har valt fyra olika typer av organisationer från olika branscher, troligtvis hade resultatet blivit ett annat om jag intervjuat fyra andra organisationer.

En annan källa till fel eller avvikelser är att jag tolkat all insamlad data. Jag kan ha uppfattat ett svar på ett sätt, medan respondenten avsett något annat. Avslutningsvis är det värt att påpeka att uppsatsen visar på tendenser, för att uppnå en högre grad av reliabilitet och validitet krävs en mer omfattande studie.

6.4 Framtida forskning

Under uppsatsens gång har ett antal frågeställningar och problemområden dykt upp som kan vara intressanta att studera i framtida undersökningar:

- Denna uppsats skulle kunna fungera som utgångspunkt för en djupare och mer detaljerad undersökning. En fallstudie där endast en organisation genomlyses och där man intervjuar ett flertal personer på olika nivåer inom organisationen. Denna undersökning skulle ge en god bild av en organisations användning av metadata, kanske en slags förstudie till ett större förändringsarbete.
- The Meta Data Coalition är en grupp bestående av 53 företag. De har skapat en standard för metadata som kallas Metadata Interchange Specification (MDIS). Tanken är att företag i branschen skall jobba mot samma mål så att deras produkter blir kompatibla med varandra. En studie eller en utvärdering av standarden kan vara ett uppslag till en framtida uppsats.
- Metadata är en vital ingrediens för ett välfungerande Internet, det kan man läsa i många böcker. Jag har stött på mycket information i detta ämne och det finns många infallsvinklar för att behandla detta ämne.
- Hur kan man få människor i organisationen att aktivt använda sig av datalagrets information? I all litteratur som behandlar datalager påpekas det att om inte datalagret används av människor i en

organisation så ger det inget mervärde. En uppsats som mer detaljerat tar reda på hur man skall uppmuntra och stödja användandet skulle vara intressant.

6.5 Epilog

Det är oerhört viktigt att administratörer och tekniker lyssnar på användarna. En undersökning som gjorts i USA visade att 84 % av de högre tjänstemännen ansåg att de nuvarande IT-systemen som fanns i företagen inte erbjöd den information som de önskade sig. [Marco 00]

Några av respondenterna påpekade att om data i datalagret användes av slutanvändarna på ett bättre sätt så skulle mervärdet av investeringen öka. Nedan har jag sammanställt en lista med råd för att öka användningen av metadata. Listan baseras på litteratur och på de intervjuer jag gjort i samband med uppsatsen.

- **Fråga användarna**
Bortse aldrig från användarna. De skall använda systemen och därmed är det de som vet bäst. Många slutanvändare har inte tid att engagera sig, men det är viktigt att få just deras kommentarer. Fråga hur de vill att funktionen skall fungera och vilken sorts data de eftersöker. Alltför många projekt har misslyckats på grund av att användarna inte varit med i utvecklingsarbetet.
- **Utbilda och ge information**
Alla respondenter har tryckt på att man måste utbilda för att få folk att använda nya artefakter. Utbildningen skall vara kort och kärnfull. Varva teori med praktik. Följ upp utbildningen ett par veckor senare och kontrollera hur det gått och om de har några frågor. Starta med att utbilda de personer som kan ha störst användning av en funktion och som har visat mest intresse.
- **Samla allt på ett ställe**
Flera organisationer i undersökningen har skapat webbportaler, där slutanvändare kan finna all information som rör informationslagret på en enda plats. På denna plats ska förklaringar finnas till alla funktioner och genvägar till dem. Detta skall vara den naturliga startplatsen för de som skall läsa rapporter och dylikt. Här kan organisationerna även informera om förbättringar eller ge tips.
- **Motivera användning**
Berätta för användarna hur bra systemet är, hur effektivt det kan vara och vilket resultat det kan ge vid rätt användning. Ge konkreta exempel på hur någon person inom organisationen funnit ny data i termkatalogen, kunnat få fram nya analyser och till exempel sparat pengar. Berätta om hur lite tid som lagts ned för att spara mycket pengar.
- **Hjälpfunktioner**
Ett par respondenter påstod att manualer är ute, i alla fall tjocka pappersmanualer. I stället skall hjälpen finnas där den verkligen behövs. Exempel kan vara att det finns hjälpknappar som man kan trycka på för att förstå en viss sida eller att man kan föra med musen över ett ord så visas det en förklarande ruta.

Hur bra organisationerna är på metadata beror givetvis till stor del på administratörerna, det vill säga de som jobbar med datalagret. En ambitiös administratör kan lyckas trots små medel.

När det gäller den typ av metadata som kallas administrativa metadata så verkar det finnas många verktyg ute på marknaden som skall underlätta hanteringen av denna. Många av de företag som marknadsför dessa verktyg beskriver dem som fantastiska, med en mängd bra funktioner. Men mellan verkligheten och reklamens värld kan steget vara långt. Det gäller att vara på sin vakt då valet av verktyg görs.

Källor

- [Berlind 97] Berlind, S. *Data warehouse – en introduktion*, 1997, Svenska Institutet för Systemutveckling (SISU)
- [Bischof et al 97] Bischoff, J., Alexander, T. & Zachman, J.A. *Data Warehouse – Practical advice from the experts*, 1997, Prentice-Hall Inc.
- [Brackett 96] Brackett, M.H. *The data warehouse challenge*, 1996, John Wiley & Sons Inc.
- [Chaudhuri 97] Chaudhuri, S. *An overview of data warehouseing and OLAP technology*, Sigmod Record, Vol. 26, No. 1, March 1997
- [Computer Sweden 00] Computer Sweden, 2000-01-27
- [Connolly et al 99] Connolly, T., Begg, C. & Strachan, A. *Database Systems*, 1999, Addison Wesley
- [Crohns 01] Crohns, R. *Data Warehouse*, D-uppsats, No. 01-11-DSV-SU, 2001
- [Devlin 99] Devlin, B. *Data Warehouse - from architecture to implementation*, 1999, Addison-Wesley
- [Donmeyer 02] Data Mining & Analysis LLC, *MOLAP or ROLAP*, en artikel publicerad på <http://www.donmeyer.com/art3.html> 2002-04-29
- [DW-Institute 02] The Data Warehousing Institute <http://www.dw-institute.com/> 2002-03-26
- [Fuller 02] Fuller, D.R. *Introducing data marts*, en artikel publicerad på <http://www.datawarehouse.com/iknowledge/articles/print.cfm?ContentID=2168> 2002-02-07
- [Halvorsen 92] Halvorsen, K. *Samhällsvetenskaplig metod*, 1992, Studentlittertur
- [Inmon 02] Inmon, B. <http://www.billinmon.com> 2002-02-10
- [Jeffery 00] Jeffery, K.G. *Metadata: The future of information systems*, artikel i en bok av Brinkenkemper, S. *Information Systems Engineering*, 2000, Springer
- [Junfors et al 96] Junfors, A. & Wallin, S. *Vad är ett termsystem för hälso- och sjukvården*, 1996, SPRI
- [Kimball 02] Kimball, R. *Metadata*, en artikel publicerad på <http://www.dbmsmag.com/9803d05.html> 2002-03-02
- [Kimball et al 98] Kimball, R., Reeves, L., Ross, M. & Thornthwaite, W. *The Data warehouse - Lifecycle toolkit*, 1998, Wiley
- [Korzybski 02] Korzybski, A. *What is meta data?*, en artikel publicerad på http://www.computerwire.com/bulletinsuk/212e_1a6.htm 2002-04-13
- [Körner at al 98] Körner, S., Ek, L. & Berg, S. *Deskriptiv statistik*, 1998, Studentlittertur
- [Lekvall et al 93] Lekvall, P. & Wahlbin, C. *Information för marknadsföringsbeslut*, 1993, IHM förlag AB
- [Marco 00] Marco, D. *Building and managing the metadata repository*, 2000, Wiley
- [OLAP-report 02] The OLAP Report, <http://www.olapreport.com/> 2002-04-29
- [Patel et al 94] Patel, R. & Davidsson, B. *Forskningsmetodikens grunder*, 1994, Studentlitteratur
- [Primdok 02] Primdok-projektet, <http://www.primdok.nu> 2002-02-10

- [Redbook 02] Data warehousing, Decision Support & OLAP,
<http://redbook.cs.berkeley.edu/lec28.html> 2002-04-29
- [Saber et al 97] Saber, K. & Teofilovski, T. *Data Warehousing – Metodik och tillämpning på RSV*, D-uppsats, No. 97-37-DSV-SU, 1997
- [Shanks et al 02] Shanks, G. & Corbitt, B. *Understanding Data Quality: Social and Cultural Aspects*, en artikel publicerad på
<http://www.vuw.ac.nz/acis99/Papers/PaperShanks-095.pdf> 2002-04-29
- [Skalin 98] Skalin, H *Datalager*, en artikel i Windowstidningen, 8 november 1998
- [Söderström 97] Söderström, P. *Data Warehouse – Datalager*, 1997, Studentlitteratur
- [Tannenbaum 02] Tannenbaum, A. *Metadata solutions*, 2002, Addison-Wesely
- [Unger 02] Unger, J. *ETL-verktyg för datavaruhus*, en uppsats publicerad på,
<http://www.cs.umu.se/education/examina/Rapporter/JohanUnger.pdf>
2002-05-20
- [Wernersson 99] Wernersson, C. *Datalager i praktiken*, D-uppsats, No. 99-x-102-DSV-SU,
1999
- [Wiener 02] Wiener, J. *Meta data in Context*, en artikel publicerad på
<http://www.datawarehouse.com/iknowledge/articles/print.cfm?ContentID=160> 2002-02-07

Bilaga 1: Ordlista

Det förekommer en rad ord och termer inom området som behandlar datalager. Många begrepp är diffusa och har olika betydelse i olika sammanhang och för olika människor. Vissa begrepp har översatts till svenska och används flitigt, men väldigt ofta är det den engelska termen som används även här i Sverige. Här följer en lista med begrepp och termer som förekommer inom det aktuella området. Där det känts angeläget har begreppet översatts.

Aggregerad data:	Summerad data som sparas i ett datalager. Data kan till exempel vara summerad på veckovis eller över en produktgrupp. På detta sätt sparas utrymme i datalagret och sökningar kan göras snabbare.
Batch:	Batchkörning eller satsvis bearbetning är bearbetning av i förväg insamlat material, ordnat på ett sådant sätt att körningen kan ske utan hjälp av operatör. Innebär ofta att databaser uppdateras i efterhand med den senaste tidens transaktioner.
B2B:	Eng. Business to business. Elektronisk handel mellan företag.
Business Intelligence:	Sv. Omvärldsbevakning. Systematisk insamling av information som hjälper företag och organisationer att förutse och anpassa sig till förändringar av olika slag.
CRM	Eng. Customer Relationship Management. Applikationer som företag använder för att behålla befintliga kunder och göra dem nöjda. Viktigt i sådana system är möjligheter att kartlägga kunderna och dela in dem i grupper, så att man kan sälja rätt sak till rätt person.
Datamart:	Ett datalager i miniatyr som innehåller data avsedd för en specifik avdelning i en större organisation.
Datamining:	Sv. Datautvinning. En teknik som används för att hitta trender och mönster i stora datamängder.
Datatvätt:	Eng. Data Cleaning eller Data Scrubbing, Korrigering av data som kan ske vid överföring av data från de operationella databaserna till datalagret.
Drill down and drill up:	Sv. Borra upp och borra ned. Med hjälp av ett OLAP-verktyg kan man borra sig upp och ner genom hierarkier och söka efter önskad information. Genom att borra ned får man en mer detaljerad vy och en mer generell vy ges om man borrar uppåt.
E-business:	Sv. Elektronisk handel. Affärer som sker på elektroniska handelsplatser, tex. Internet.
ETL-process:	Eng. Extract, Transform och Load. Process som överför data från källdatabaser till datalagret. De engelska termerna kan översättas till hämta, omvandla och ladda.
HOLAP:	Eng. Hybrid On-Line Analytical Processing. En kombination av ROLAP och MOLAP.
Middleware:	Sv. Mellanvara. Ett program som ligger mellan två system och möjliggör kommunikation. Behövs ofta för att heterogena system ska fungera.
MOLAP:	Eng. Multidimensional On-Line Analytical Processing. OLAP som lagrar aggregerad data i multidimensionell struktur.
OLAP:	Eng. Online Analytical Processing. En teknik för att göra så kallade multidimensionella sökningar i databaser medan de är i drift. Används för

	försäljningsanalys och beslutsstöd. Principerna för OLAP formulerades 1992 av Edgar F Codd.
OLTP:	Eng. On-Line Transaction Processing. Databehandling i ett system som är under drift, och där instruktionerna utförs utan större väntetid. Exempel är kassasystem, bankautomater och bokningssystem där databaserna uppdateras direkt efter varje transaktion
Relationsdatabas:	Numera den vanligaste formen av databas. En relationsdatabas består av tabeller med två eller flera kolumner. Värdet i en kolumn, till exempel personnummer, används som nyckel och identifierar en post i tabellen. Teknikens skapare Edgar F Codd har tagit fram tolv strikta regler för vad som får räknas som en relationsdatabas.
Repository:	En samling definitioner och beskrivningar av objekt och datatyper i ett system, avsedd att underlätta systemutveckling.
ROLAP:	Eng. Relational On-Line Analytical Processing. OLAP som lagrar aggregerad data i en relationsdatabas.
SAS Institute:	Amerikanskt företag som levererar affärslösningar och programvara inom bland annat e-intelligence, CRM och data warehousing.
Star schema:	Sv. Stjärnstruktur. En databasstruktur som underlättar analyser. Innehåller en faktatabell och en eller många dimensionstabeller som medger sammanställning av data efter olika indelningsgrunder.
Termkatalog:	Eng. Metadata Catalogue. En funktion som stödjer sökning av data i ett datalager. Till hjälp för slutanvändare.

Bilaga 2: Riksskatteverket

Riksskatteverket (RSV) har tillsammans med Skattemyndigheterna och Kronofogdemyndigheterna ansvaret för folkbokföring, service och information, beskattning, uppbörd och indrivning för medborgarnas räkning. Totalt jobbar cirka 13000 personer i organisationen.

På RSV:s IT-avdelning arbetar drygt 600 personer och deras uppgift är att tillhandahålla IT-stöd för hela organisationens verksamhet. De utvecklar och förvaltar IT-stöd för skattehantering, folkbokföring, allmänna val och indrivning med hjälp av både ny och beprövad teknik. De huvudsakliga användarna av systemen är skatte- och kronofogdemyndigheternas medarbetare.

Datalager

RSV har många delenheter som hanterar mycket information och data. Sedan en tid tillbaka pågår ett projekt där RSV IT skall skapa ett gemensamt datalager, eller informationslager som man internt kallar det, för alla enheter. Data skall samlas in från olika enheter och användas vid uppföljning, revision och planering.

I dagsläget har alla verksamhetsenheter egna rutiner eller delsystem för lagring av data, som kan ses som ett exempel på stuprörstänkande. Detta medför många svårigheter, till exempel vid arbete över enhetsgränser och jämförelse av data. Det är inte heller ekonomiskt och mycket resurskrävande jobba med många små system.

Enligt RSV skall det nya informationslagret utvecklats för att:

- Hantera uppföljningsdata.
- Möjliggöra enkel och snabb förändring och konstruktion av uppföljningssystem.
- Möjliggöra enhetlig leveranshantering för verksamhetssystem
- Möjliggöra enhetlig och väldefinierad process för konstruktion och förvaltning av uppföljningssystem
- Uppföljningsinformationen ska ha hög kvalitet och rätt aktualitet
- Skapa effektivare utveckling och förvaltning av främst system för uppföljning, styrning och statistik
- Leverera all data som är intressant för uppföljningssystem

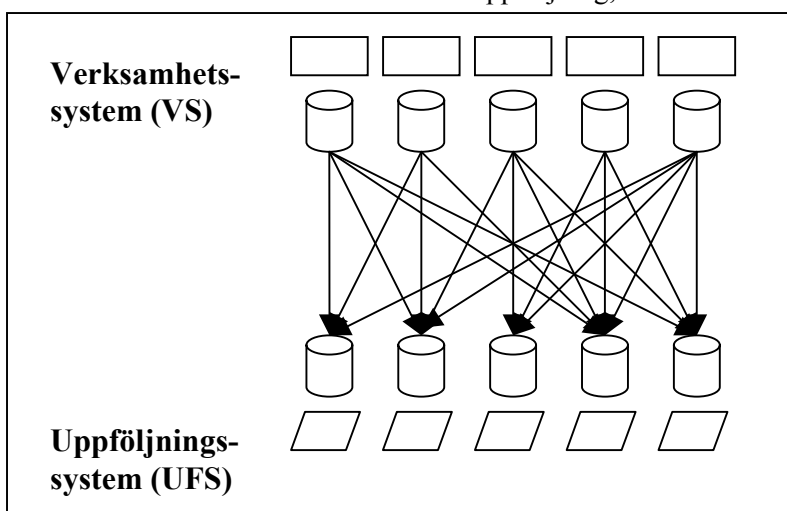
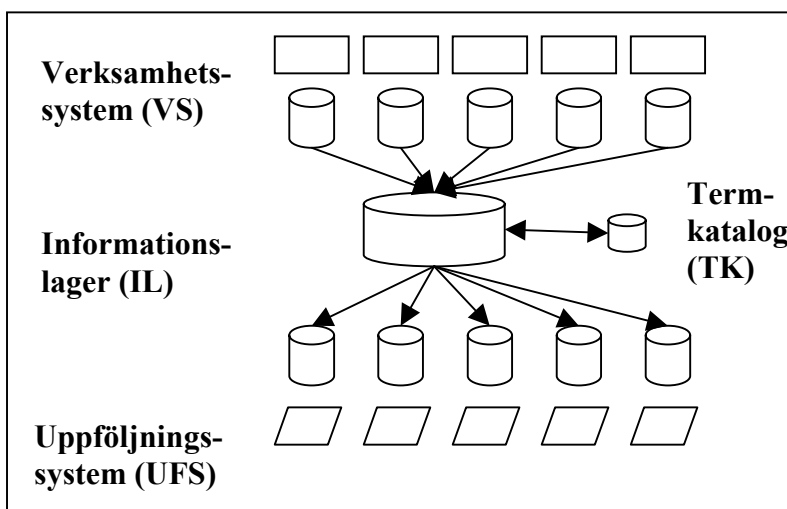


Fig. B.1 RSV:s struktur före informationslagrets (datalagrets) införande.



Figur B.2 RSV:s struktur efter informationslagrets (datalagrets) införande.

- Hantera all information för resultatdialog, resultatredovisning, prognoser och underlag för verksamhetsplanering

På RSV talar man om Verksamhetssystem (VS) och Uppföljningssystem (UFS). Via batchar kommer transaktionsorienterad data från VS till informationslagret (IL). Endast den data som är intressant för uppföljning överförs. Från informationslagret överförs sedan data till UFS.

Teknisk plattform

Det nya informationslagrets tekniska plattformen består av:

- HP-UX B.11.00.A
- Databas Oracle 8.1.7
- PowerMart – verktyg för bearbetning av data
- WebLogic-server för WEB-applikationen

Termkatalog

Termkatalogen är av central betydelse i informationslagret. RSV ger i sin interna dokumentation följande beskrivning:

"En termkatalogs syfte är att framförallt att vara behjälplig i arbetet med att bringa klarhet i användandet av och betydelsen hos termer inom berörda verksamhetsområden."

Vidare kan man läsa att termkatalogen kan beskrivas som en specialiserad databas med information om data. All relevant information rörande en verksamhets termer lagras, uppdateras och görs tillgänglig. En webbaserad applikation skall arbeta mot termkatalogen för att göra den enkel att använda och lätt att nå för användare. Man skall via applikationen kunna se vad en term betyder och från vilket källsystem den kommer. Mått är en speciell typ av uppgift och ofta ett resultat av en beräkning, det skall även vara möjligt att ta fram information i termkatalogen på vilket sätt ett mått är uträknat eller framtaget.

Det finns alltså två syften med termkatalogen. Dels att kunna ta reda på vad termer och begrepp betyder, men även för att kunna ta reda på vilken information som finns i informationslagret. Det sistnämnda skälet är viktigt för att kunna utveckla analyser och ta fram nya rapporter och på så sätt skapa effektivare uppföljning och planering.

Vid en delrevision av RSV:s datalagersystem, utfört av en extern konsultbyrå, framkom att termkatalogen är en framgångsfaktor. Med detta menas att RSV bör satsa på termkatalogen. Om inte termkatalogen fungerar och används på rätt sätt av rätt personer blir inte informationslagret en sådan framgång som det har potential att bli.

Bilaga 3: Intervjumall

Här följer det frågeunderlag som använts vid undersökningen. Detta frågeunderlag har sänts i förväg till de respondenter som så önskat.

Introduktion

Uppsatsen är ett arbete om 20 poäng och en avslutning på min data/systemvetenskapliga utbildning om 160 poäng vid Stockholms Universitet/KTH. Syftet med uppsatsen är att undersöka hur organisationer hanterar metadata i datalager. Jag har för undersökningen skapat en utvärderingsmodell med fem faktorer med ett antal kriterier och det är runt dessa faktorer intervjufrågorna är uppbyggda.

FAKTOR	BESKRIVNING	FRÅGOR
Metadata Quality	är kvalitet på metadata, det vill säga huruvida metadata är uppdaterade med senaste informationen och är korrekta.	49-55
Business Terms	är gemensamma termer och begrepp om verksamheten, det vill säga huruvida det finns gemensamma begrepp och termer i en organisation	7-17
Business Metadata	är metadata för slutanvändare om tabellers innehåll, betydelse och härkomst.	18-27
Administrative Metadata	är metadata om datalagrets uppbyggnad och struktur, främst för utvecklare och personer som jobbar med underhåll.	28-37
Operational Metadata	är statistik om datalagrets drift som kan användas för att skapa ett bättre datalager.	38-48

Observera: Beroende på vilken typ av funktion som finns implementerad i organisationen kan frågorna passa bättre eller sämre. Dessa frågor kan ses som ett diskussionsunderlag och som en grund för samtalet, där eventuella tillägg kan göras av intervjuaren och respondenten under intervjun.

Allmänna frågor

1. Vill ni vara anonyma? På vilket sätt vill ni presenteras i undersökningen?
2. Berätta kort om företaget/organisationen?
3. Berätta kort om ert datalager?
4. När implementerade ni ett datalager i er verksamhet?
5. Hur stort är datalagret?
6. Vem använder datalagret? Hur många använder det?

Business Terms

Allmänt:

Begreppet business terms beskriver jag som begrepp och termer som finns inom en organisation. En förutsättning för ett väl fungerande datalager är gemensamma business terms. Skulle ni vilja lägga till något i definitionen?

Dokumentation:

7. Hur kan en nyanställd veta vad som menas med ett specifikt begrepp?
8. Finns det någon gemensam lista/ordbok över termer och begrepp?
9. Vem har utarbetat listan/ordboken?

Åtkomst:

10. Var finns listan/ordboken?
11. Är den implementerad?
12. Hur lätt kan användarna nå funktionen?
13. Vet alla om att det finns?
14. Kan alla nå den?

Användning:

15. Känner alla till att listan/ordboken finns?
16. Hur mycket används funktionen?
17. Skulle den kunnas användas mer?

Business Metadata

Allmänt:

Begreppet business metadata beskriver jag som dokumentation som främst är till för slutanvändare. Det är information om termer och begrepp, varifrån data ursprungligen kommer och hur de är sammansatta. Data om den data som finns i datalagret. Skulle ni vilja lägga till något i definitionen?

Implementering:

18. En analytiker vill få mer information. Hur kan han få information vad som finns i datalagret?
19. Finns det någon funktion implementerad?
20. Behövs en implementerad funktion? Skulle man inte kunna prata med varandra inom organisationen och på så sätt ta fram fakta?
21. Vem bestämmer vilken ny information skall komma in i datalagret?

Åtkomst:

22. Hur kan en slutanvändare nå business metadata (funktionen)?
23. Är inte en webbaserad gränssnitt det optimala?

Användning:

24. Vem använder sig av funktionen? Hur många?
25. Efterfrågar slutanvändare business metadata?
26. Skall de använda det mer? Vad görs för att slutanvändare skall använda business metadata mer?
27. Får ni många frågor om funktionen skall utnyttjas?

Administrative Metadata

Allmänt:

Begreppet administrative metadata beskriver jag som dokumentation om datalagrets uppbyggnad och struktur som främst är till för de som utvecklar och underhåller systemet. Data om datalagret. Skulle ni vilja lägga till något i definitionen?

Dokumentation:

28. Hur kan en ny person som skall jobba med datalagret lära sig mer om hur det är uppbyggt?
29. Finns det någon funktion implementerad eller nedskrivna dokumentation? Hur fungerar den?
30. Hur detaljerad är dokumentationen? Vilken nivå på kunskap krävs för att kunna förstå den?

Aktualitet:

31. Hur uppdateras förändringar?
32. Hur ofta sker uppdateringar?
33. Ett vanligt problem är att man slarvar med att uppdaterar förändringar, vilket leder till att ingen litar på dokumentationen. Känns det bekant?

Användning

34. Vem använder sig av funktionen? Hur många?

35. Hur ofta används funktionen?
36. Efterfrågas administrative metadata? Av vem då?
37. Vad händer om två personer som jobbar med datalagret byter jobb eller blir sjuka?

Operational Metadata

Allmänt:

Begreppet operational metadata beskriver jag som statistik om datalagrets drift och felrapporter. Exempel kan vara att man tar fram information om vilka tabeller som det söks mest i, driftsstopp eller felrapporter. Utifrån Operational metadata skall man sedan kunna förbättra datalagrets prestanda. Skulle ni vilja lägga till något i definitionen?

Insamling:

38. Hur samlas statistik över datalagrets drift in?
39. Hur ofta samlas statistik över datalagrets drift in?
40. Hur sparas denna information?
41. Hur redovisas felrapporter eller driftstopp?
42. Vem är ansvarig för insamlingen?

Användning:

43. Hur ofta görs förändringar av datalagret med hjälp av insamlad statistik?
44. Vem är ansvarig? Vem bestämmer om eventuella förändringar?
45. Hur går ni till väga då det sker driftstopp eller ni mottar en felrapport?
46. Finns det någon manual eller nedskrivna dokumentation hur man skall gå till väga?
47. Ge konkreta exempel på förändringar som gjorts efter att operational metadata har insamlats?
48. Skulle ni kunna använda er av operational metadata mera? Varför – varför inte?

Metadata Quality

Allmänt:

Begreppet hög metadata quality innebär för mig att all metadata är uppdaterad med senaste informationen och att den är korrekt. Systemet måste även vara tillgängligt och fungera utan missöden. Skulle ni vilja lägga till något i definitionen?

Kvalitetsarbete:

49. Hur kontrollerar ni att den data som överförs från källdatabaserna till datalagret är korrekt (tvättning)?
50. Hur jobbar ni med att öka kvalitén på metadata?
51. Finns det någon utarbetad handlingsplan för att åtgärda felaktig metadata?
52. Hur mycket tid läggs ned för att öka kvalitén på metadata?

Fel:

53. Hur ofta upptäcks fel på metadata i datalagret?
54. Vem upptäcker det?
55. Hur upptäcks det?