

TRIAD Arbetsrapport U2

Krav på IA
Modellering
Utbildning för handledare
Katalogprinciper
■ Uttagssystem
Informationspridning

Rapport U nr 1: HYBRIS i Unix-miljö, förstudie

■ Rapport U nr 2: DEBRIS

DEBRIS

– användargränssnitt i decentraliserad databasmiljö

Peeter Kool, SISU

Spridningsförbehåll:

Denna rapport får endast spridas och användas inom de organisationer som deltar som parter i TRIAD-projektet.
© TRIAD-parterna nov 1992

Rapporten är skriven i och för TRIAD delprojekt Uttagssystem.

DEBRIS – Användargränssnitt i decentraliserad databasmiljö

Innehåll

1. Inledning.....	1
2. Problembeskrivning.....	2
3. DEBRIS demonstratorn.....	3
3.1 Skrivbordet.....	4
3.2 Kopplingsverktyget.....	5
3.3 Selektorn.....	7
3.4 Uppslagsboken (Metadatabasen).....	8
3.5 Kommunikation.....	10
4. Arbetssätt (scenario).....	11
5. Resultat.....	14
6. Vidareutveckling.....	15

1. Inledning

DEBRIS-aktiviteten har syftat till att utveckla ett användargränssnitt, som integrerar verktyg för navigering och sökning, för ett uttagssystem mot en decentraliserad databasmiljö.

Idag finns det verktyg som Hybris som löser problemen med utsökning av information ur en databas effektivt, men dessa verktyg räcker inte till för sökning i flera databaser samtidigt. Men det finns användare idag som har tillgång till flera databaser men saknar möjligheten att rationellt utnyttja dessa tillsammans. DEBRIS är ett försök att skapa en möjlighet för detta.

Många företag har idag investerat stora summor i olika datalager. Att få möjlighet att utnyttja data från flera databaser kan rationalisera och effektivisera utnyttjandet av investeringen i data.

Resultat av aktiviteten är dels en demonstrator som visar principerna för och användandet av DEBRIS-gränssnittet samt en rapport.

2. Problembeskrivning

För rena uttagssystem som DEBRIS finns det tre huvudsakliga problemområden som introduceras av decentraliserade databasmiljöer:

- Frågesplittring och koppling (eng Join)
- Schemaintegration
- Datakvalitet

Frågesplittring och koppling används då en databasfråga innehåller kopplingar som sträcker sig mellan olika databaser. Databasfrågan måste delas upp i en fråga per fysisk databas. Svaren måste sedan kopplas ihop. Detta problem kan tyckas enkelt att lösa, men i själva verket är det svårt att utforma frågesplittring och koppling effektivt.

Schemaintegrationsproblemet uppstår då det finns homonymer och synonymer i de olika databasschemorna. Homonymer innebär att två objekt har samma namn men representerar olika saker. Synonymer är att två objekt har olika namn men representerar samma sak. Förutom detta kan det skilja sig även på datatyper i kolumner. Det är tyvärr vanligt att kolumner som beskriver samma företeelse har olika längd.

Datakvalitetsproblemen som decentraliserade databaser medför kan ha olika orsaker. En av de vanligare är att databaserna har olika uppdateringstidpunkter och uppdateringsfrekvens. Resultat kan bli missvisande då en fråga ställs mot två databaser där den ena uppdateras kontinuerligt och den andra en gång i månaden.

En annan orsak till datakvalitetsproblemet kan vara att två objekt i olika databaser representerar samma sak där den ena är en delmängd av den andra.

3. DEBRIS –en demonstrator

Den grundläggande filosofin bakom DEBRIS är att ge användaren de verktyg och det stöd som behövs för att användaren själv skall kunna göra kopplingar som sträcker sig utanför en databas. Dessutom skall de utnyttjade databaserna inte påverkas. Detta ställer naturligtvis högre krav på användaren än en helt integrerad lösning där den distribuerade databasen representeras som en enhet. Däremot ger DEBRIS lösningen stor flexibilitet för användaren.

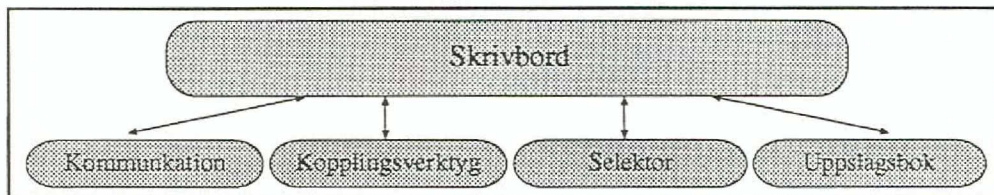


BILD 1: SYSTEMSKISS AV DEBRIS

Syftet med DEBRIS-demonstratorn är att visa på tänkbara implementationer av verktygen och stödet för användarna. DEBRIS-demonstratorn baseras på Hybris och består huvudsakligen av fem huvuddelar skrivbordet, selektor, uppslagsbok, kommunikation och kopplingsverktyget.

3.1 Skrivbordet

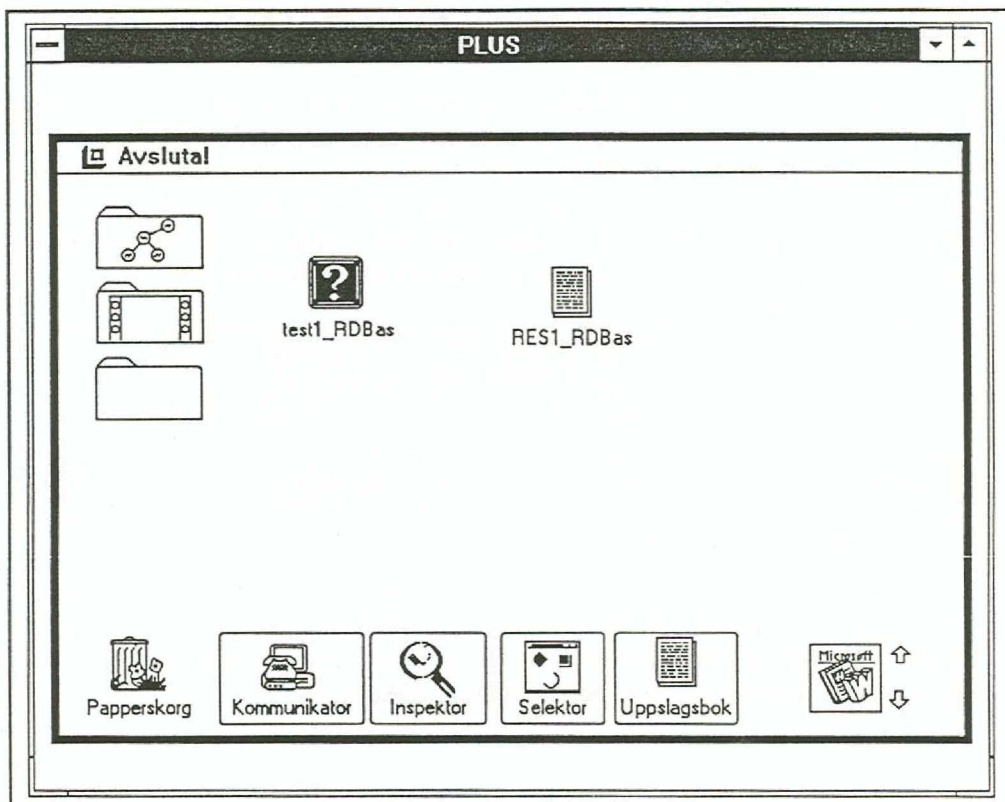


BILD 2: SKRIVBORDET I DEBRIS

Utgångspunkten för all användning av DEBRIS är skrivbordet (se Bild 2). Skrivbordet är uppbyggt av ett antal symboler som symboliserar olika typer av objekt.

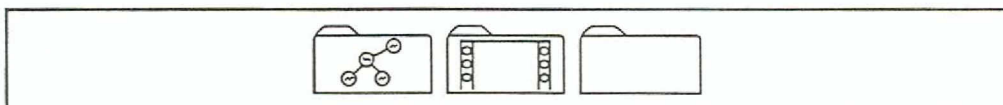


BILD 3: FÖRVARINGSOBJEKT (DATABASMODELLER, RESULTAT OCH FRÅGOR)

Förvaringsobjektens (se Bild 3) huvudsakliga uppgift att fungera som en förvaringsplats för dataobjekt som inte skall vara synliga på skrivbordet. De fungerar på samma sätt som mappar eller kataloger i filsystem.

Databasmodell är tänkt att innehålla databasmodeller som användaren har tillgång till. *Resultat* är där alla databasutsökningsresultat finns. *Frågor* innehåller sparade databasfrågor.



BILD 4: DATAOBJEKT (FRÅGA OCH RESULTAT)

Dataobjekten (se Bild 4) representerar data som kan vidarebearbetas. *Frågeobjektet* representerar en fråga och *resultatobjektet* ett resultat.

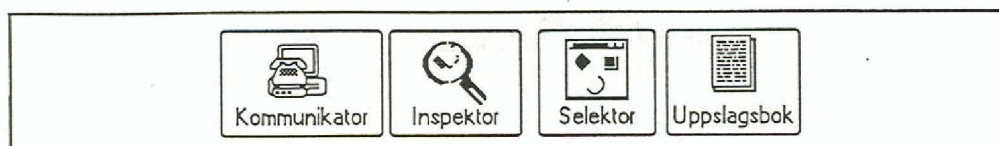


BILD 5: VERKTYGSOBJEKT

Verktysobjektens uppgift är att redigera dataobjekt (dvs frågor och resultat). Kommunikatorn används till att skicka frågor till databaserna. Inspektorn är till för att presentera innehållet i ett dataobjekt. Selektorn är frågeformuleringsverktyget som i DEBRIS är en modifierad variant av Hybris. Uppslagsboken är en uppslagsbok över de ingående databaserna.

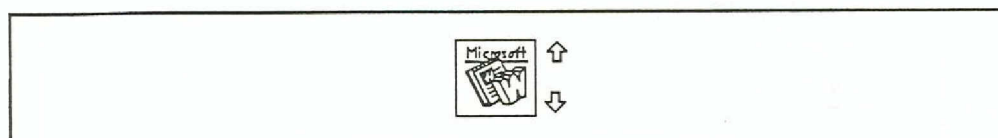


BILD 6: TILLÄMPNINGSOBJEKTET

Tillämpningsobjektet innehåller de tillämpningar som används då man vill vidarebearbeta resultat. Pilsymbolerna används till att rulla mellan de tillämpningar som finns tillgängliga.

Skrivbordet bygger till stora delar på idéer från CUA 91 (Common User Access). Huvudsakligen används dra-och-släpp som interaktionsprincip i skrivbordet. Dataobjekt kan släppas på förvarings- och verktygsobjekt. För att omformulera en tidigare skapad fråga drar man frågan till selektorn och släpper den där. Det är även möjligt att klicka direkt på verktygs- och förvaringsobjekt. För att placera ett resultat från en databasutsökning på skrivbordet klickar man på resultatförvaringsobjektet och får då välja bland de resultat som finns.

3.2 Kopplingsverktyget

En väsentlig del i DEBRIS utgörs av kopplingsverktyget. Kopplingsverktygets uppgift är att utföra olika typer av kopplingar (eng Join) på databasutsökningens resultat.

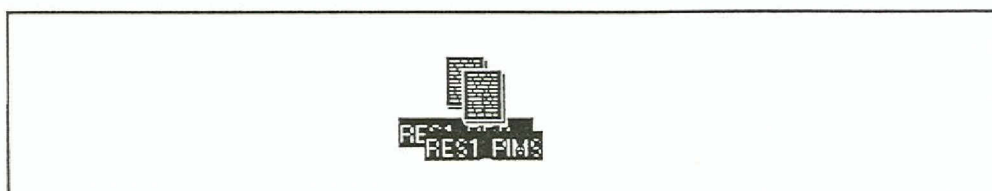


BILD 7: AKTIVERING AV KOPPLINGSVERKTYGET.

Aktivering av kopplingsverktyget sker genom att ett resultat släpps på ett annat resultat i skrivordet (se Bild 7). Kopplingsverktyget presenterar de ingående kolumnerna som finns i de två resultaten (se Bild 8). Kopplingsverktyget presenterar också de kopplingar som finns fördefinierade i metadatabasen

(uppslagsboken). I Bild 7 finns en fördefinierad koppling mellan kolumnen Samlingskundnummer och kolumnen Huvudkundnummer.

Kopplingsverktyget utnyttjar även metadatabasen för att ta reda på om det finns sammansatta nycklar med i kopplingen och för hur sammansättningen av nycklarna ska gå till.

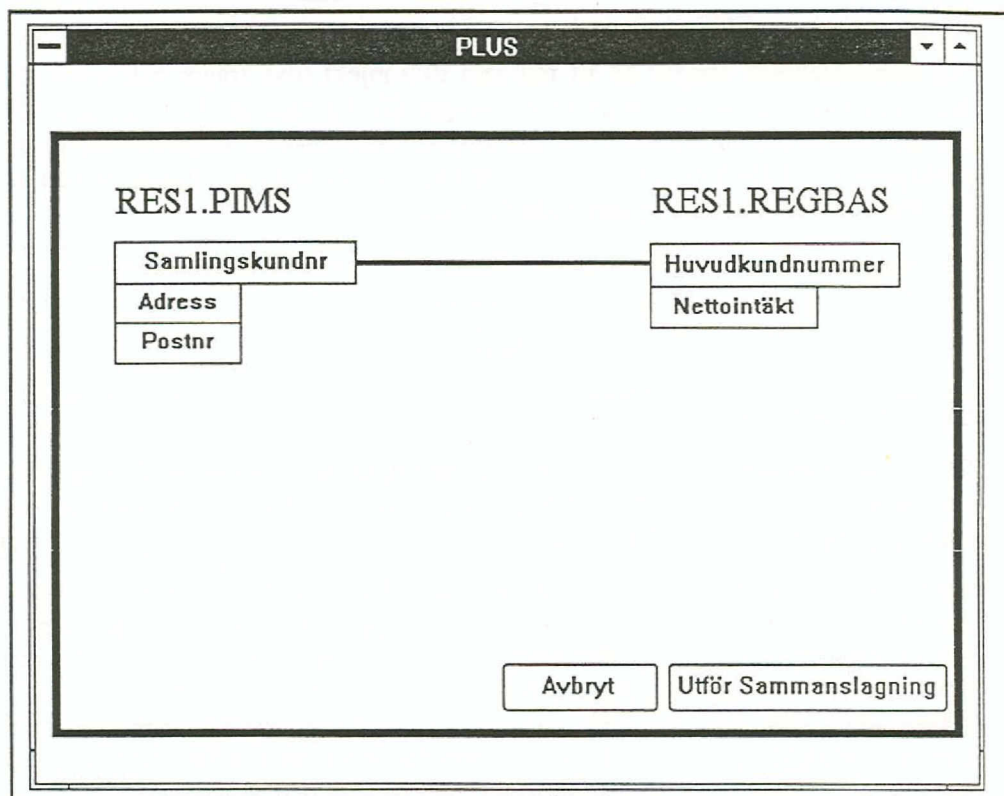


BILD 8: KOPPLINGSVERKTYGET

Naturligtvis är inte kopplingsverktyget låst till att användas enbart för de fördefinierade kopplingarna. Det är möjligt att skapa egna kopplingar genom att klicka på kolumner i vardera resultatet. På varje koppling är det sedan möjligt att definiera kopplingsvillkor av typen "=", "<", ">" osv. Utdata från kopplingsverktyget är ett resultatobjekt.

3.3 Selektorn

Selektorn är en modifierad variant av Hybris. De punkter som tas upp här är de punkter som skiljer DEBRISselektorn ifrån Hybris.

När selektorn aktiveras utan att en fråga släppts på den får användaren bestämma mot vilken av de befintliga databaserna frågan skall formuleras. Släpper användaren en tidigare formulerad fråga på selektorn startar den direkt med den databasmodell som använts för frågan.

Den största skillnaden mellan Hybris och DEBRISselektorn finns vid villkorsställande på ett attribut i ett objekt (se Bild 9)

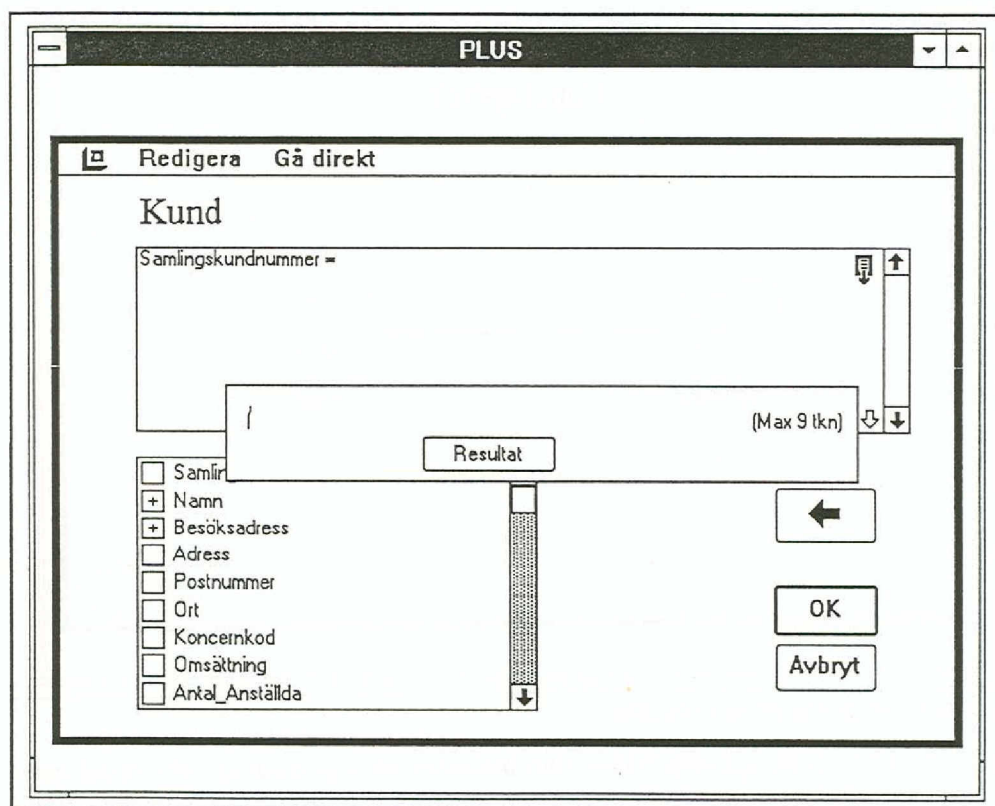


BILD 9: VILLKORSDIALOG I DEBRISSELEKTORN

Det har införts ett nytt alternativ "Resultat" i villkorsdialogen. "Resultat" gör det möjligt att använda en kolumn i ett befintligt resultat som villkor i en fråga. För att det skall vara omöjligt att formulera icke korrekta frågor kontrolleras vilka resultat som verkligen går att utnyttja i villkoret för det aktuella attributet. För kontrollen utnyttjas samma metadata som kopplingsverktyget använder.

"Resultat" har införts för att minimera mängden data som skall hämtas från databasen. Dessutom kan det vara möjligt att minska tiden för kopplingsverktyget att utföra kopplingen.

3.4 Uppslagsboken (Metadatabasen)

Uppslagsboken i DEBRIS är i grund och botten densamma som Hybris, men tillägg har gjorts.

När uppslagsboken öppnas är det möjligt att titta på uppslagsböckerna för de databaser som ingår i DEBRIS eller att titta på de uppslagsböcker som beskriver kopplingarna mellan databaserna (se Bild 10).

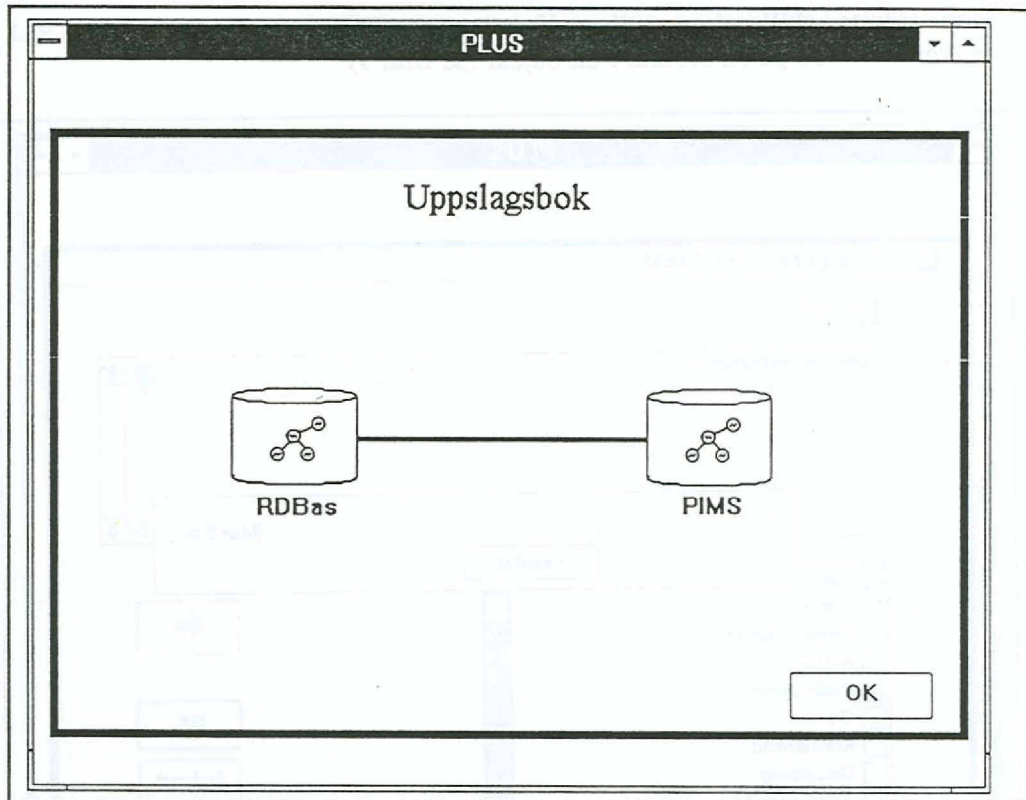


BILD 10: UPPSLAGSBOKEN I DEBRIS

För att komma till uppslagsboken som beskriver en viss databas behövs det bara ett klick på dess symbol. För att komma till den uppslagsbok som beskriver kopplingarna mellan två databaser klickar man på strecket mellan dem.

Kopplingsbeskrivningen innehåller information om vilka objekt som ingår i kopplingen och mellan vilka attribut kopplingen finns (se Bild 11).

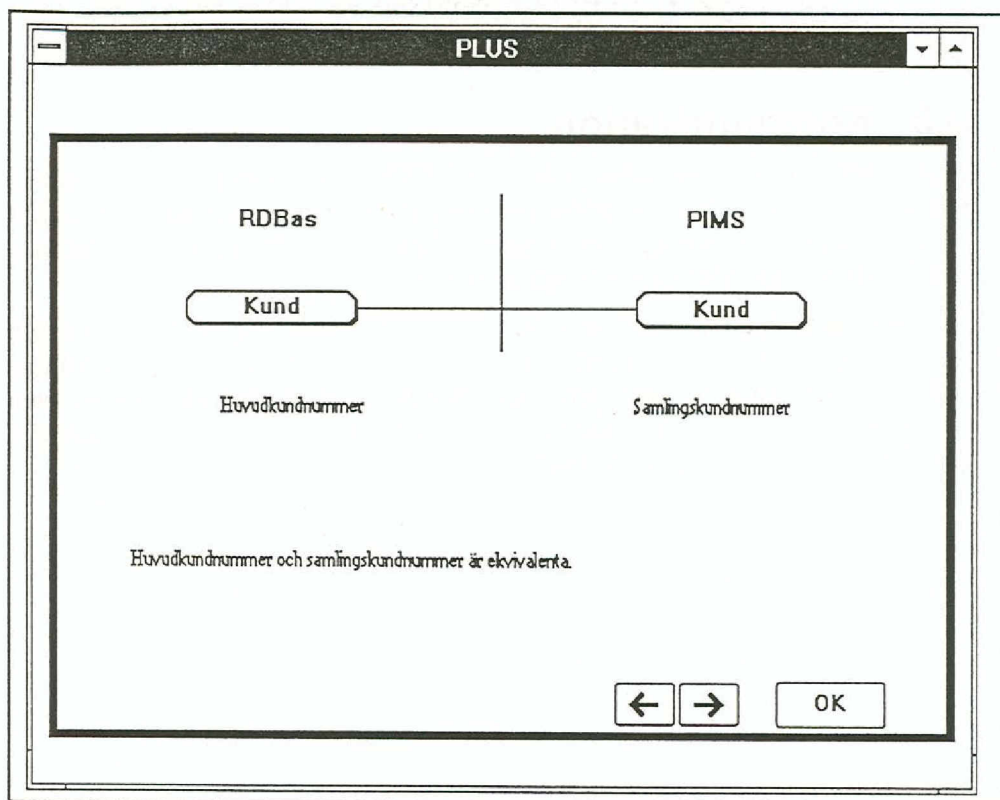


BILD 11: KOPPLINGSBESKRIVNING I UPPSLAGSBOKEN

Dessutom finns en textuell beskrivning på kopplingen. Beskrivningen kan till exempel innehålla begränsningar i kopplingen som att det ena objektet är en delmängd av det andra objektet.

Det metadata som uppslagsboken måste innehålla för varje koppling mellan två databaser i ett fungerande DEBRIS är som följer:

- Namnen på de ingående objekten och till vilken databas de hör.
- Vilka attribut i respektive objekt som används i kopplingen.
- Information om hur eventuella sammansatta attribut byggs upp.
- Datatypskillnader i attributen.
- Textuell beskrivning av kopplingen för användare. Beskrivningen bör innehålla eventuella hänsyn som måste tas på grund av datakvalite, t ex skillnader i uppdateringsfrekvens.

Användbarheten av ett DEBRIS system bestäms i stor utsträckning av den information som finns i uppslagsboken. Det är den enda källan användaren och verktygen kan utnyttja för att se vilka kopplingar som finns och av vilken typ de är. Uppslagsboken innehåller med andra ord hela schemaintegrationen.

Informationen i uppslagsboken är svår för att inte säga omöjlig att generera automatiskt. Orsaken till detta är att det behövs information om skillnader i datatyper samt uppdateringstidpunkter för att säkra datakvaliten. För ett realistiskt användande av DEBRIS kommer det att krävas någon eller några personer som hanterar uppdateringar av uppslagsboken.

3.5 Kommunikation

Kommunikationen är ett olöst problem i DEBRIS-demonstratorn. Däremot är inte kommunikationen ett svårlöst problem, dock med ett undantag. Används en kommunikationslösning som bygger på en modem-förbindelse är det inte möjligt att upprätthålla två samtidiga sessioner.

Utnyttjas istället ett nätverk är kommunikationen inget problem oavsett om terminalemulatorer eller programmeringsgränssnitt (t ex ODBC) används. Nätverkslösningen är troligen den enda realistiska för ett DEBRIS-system även på grund utav att koppling utförs lokalt och det därför kan bli frågan om förhållandevis stora dataöverföringar.

4. Arbetssätt (scenario)

För att beskriva arbetssättet i DEBRIS använder vi oss av ett scenario. Scenariot finns beskrivet i bild 12. Utgångspunkten för detta scenario är en försäljare på Posten som vill ta reda på adresserna till de kunder som har köpt flest varor och tjänster av Posten. Mer specifikt vill han adress och summa inköp i en försäljningsrapport som skall skrivas.

Denna försäljare har tillgång till två databaser:

- REGBAS som innehåller försäljningsstatistik.
- PIMS som innehåller kundinformation.

Försäljaren vet sedan tidigare att informationen om vem som köpt vad av Posten finns i REGBAS. Han öppnar selektorn från skrivbordet, väljer REGBAS som databas och upptäcker att det inte finns någon information om adressen till kunder i databasen. Det finns enbart namn och kundnummer.

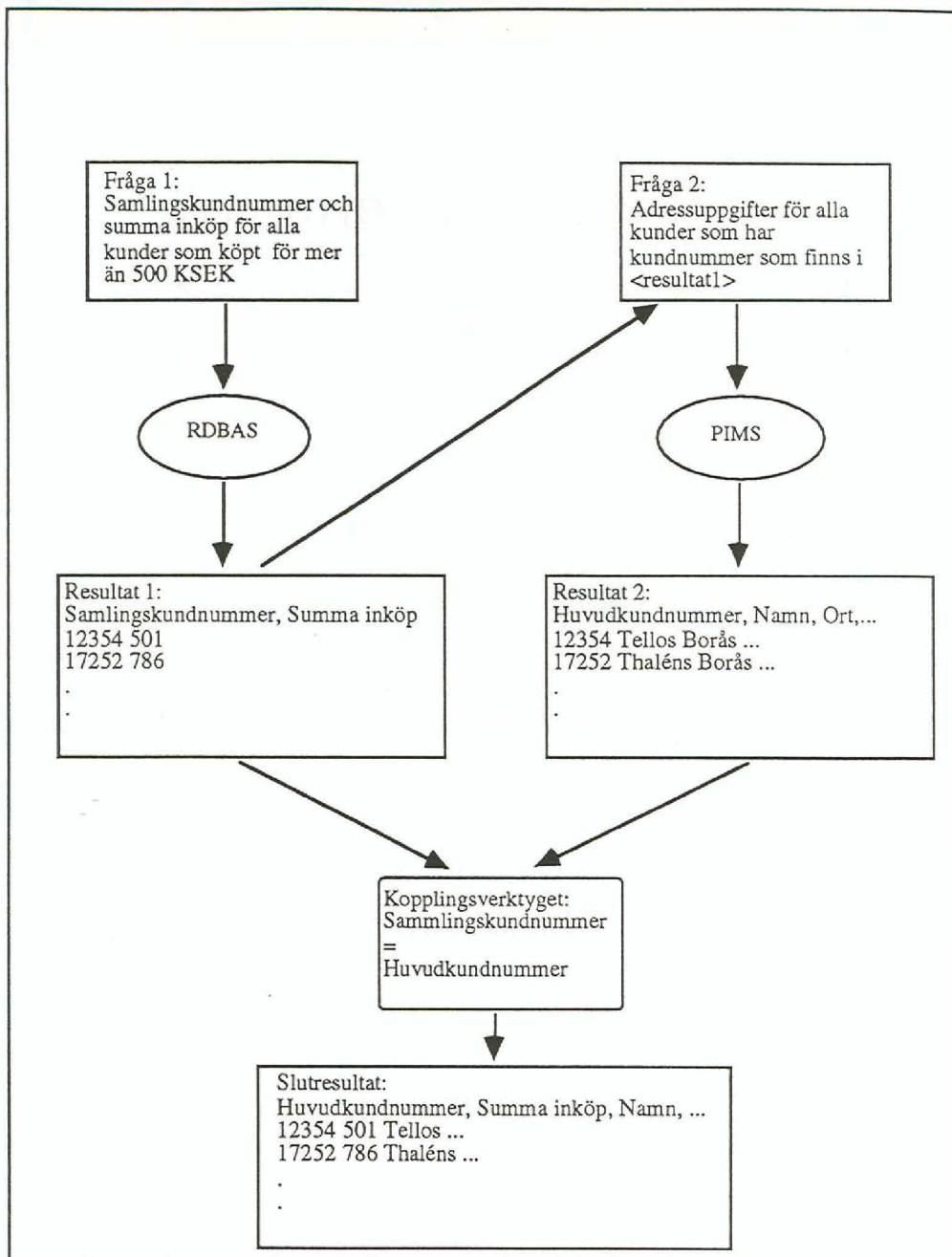


BILD 12: SCHEMATISKT ÅSKÅDLIGJORT SCENARIO

Försäljaren kommer då ihåg att adressuppgifterna finns i PIMS databasen. Han börjar fundera på om det finns någon koppling som han kan utnyttja. Han går vidare till uppslagsboken och tittar på kopplingar mellan REGBAS och PIMS och finner att det finns en koppling, samlingskundnummer är ekvivalent med kundnummer i respektive databas.

Försäljaren går tillbaka till selektorn för REGBAS och formulerar en fråga som kommer att ge som resultat alla de kunder som köpt för mer än 500 KSEK detta år. Han sparar frågan och skickar den med hjälp av kommunikatorn till databasen och resultat kommer tillbaka.

Försäljaren tar fram selektorn för PIMS och börjar formulera en fråga för att få fram adressuppgifter. Han vet att Posten har tusentals kunder och att det finns adressuppgifter för alla dessa i PIMS databasen. För att inte få ett jättelikt resultat tar försäljaren och sätter villkor på attributet huvudkundnummer och väljer "Resultat". Han får upp de resultat som innehåller en kolumn som kan användas som villkor. Han väljer då det resultat som han tidigare skapat ifrån REGBAS.

När frågan exekverats mot PIMS och resultat hämtats aktiverar försäljaren kopplingsverktyget genom att släppa resultatet från PIMS-databasen på resultatet från REGBAS. Kopplingsverktyget föreslår en koppling mellan huvudkundnummer och samlingskundnummer kolumnerna. Han väljer denna koppling och utför den. Han har nu fått ett nytt resultat som innehåller det data han var intresserad av.

Försäljaren tar sitt nya resultat och släpper det på Word-symbolen och använder resultatets data för att skriva sin rapport.

5. Resultat

DEBRIS demonstratorn har visat att det är möjligt att integrera dagens relationsdatabaser genom en ren användargränssnittslösning. Men det ställer större krav på användarens kunskap om databaserna än en helt transparent lösning. Användaren får själv bestämma sökstrategin och minimera resultatmängderna som förs över till den lokala maskinen. Dessutom måste användaren även hantera problemen med datakvalite på egen hand med hjälp av uppslagsboken.

För att få ett väl fungerande DEBRIS system ställs det krav på att uppslagsboken innehåller relevant information och hålls uppdaterad. Det måste finnas någon sk DEBRIS-administratör som har huvudansvaret för detta.

Ett problem som kvarstår olöst i DEBRIS är kommunikationen mot databaserna. Om databaserna finns i ett nätverk är detta problem inte svårt att lösa. Däremot är problemet svårlöst vid en uppringd (modem-) koppling

DEBRIS har demonstrerats för Triads projektledningsgrupp, Televerket Nät och Televerket Data. Reaktionerna har i allmänhet varit positiva. Några frågetecken har emellertid rests.

- Insamling och upprätthållande av informationen i uppslagsboken.
- Kraven som ställs på användaren då det gäller frågeställningsstrategier.

Vad det gäller insamling och upprätthållande av informationen i uppslagsboken är det konstaterat att det behövs en ansvarig person som utför detta.

6. Vidareutveckling

DEBRIS-demonstratorn kan utvecklas vidare till en fungerande prototyp. DEBRIS-demonstratorn är utvecklad med hjälp av programvaran PLUS. PLUS visade sig ha begränsningar som gör det svårt att utveckla en prototyp av DEBRIS i detta verktyg. Istället bör en implementering göras med hjälp av t ex Visual Basic. Huvudsyftet för denna prototyp är att utnyttja den för användartester samt för att utveckla och studera metoder för sökning i flera databaser.

Det är även möjligt att bygga vidare på DEBRIS-demonstratorn. Den skulle även kunna innehålla mer intelligens. Till exempel skulle det kunna finnas någon form av aktiv hjälp som leder användaren fram till det resultat som han vill uppnå och skulle kunna ge råd om frågeställningsstrategin

En annan tänkbar utveckling för DEBRIS-demonstratorn vore att ta delar av den och införa i Hybris. Redan idag är skrivbordet tillgängligt i Hybris, dock enbart i den anpassning av Hybris som kallas PimWin (Ett marknadsinformationssystem inom Posten). Det är möjligt att implementera ett kopplingsverktyg och föra in i dagens Hybris. Samma sak gäller för funktionen att använda resultat i villkor. Men för att få stöd med förslag till kopplingar med mera krävs det även att metadatabasen i dagens Hybris byggs ut med denna information. Dessa två funktioner finns det även nytta av när man endast har en databas. I förlängningen är det möjligt att få ett Hybris som understödjer sökning i mer än en databas.

TRIAD-rapporter per 921130

Verksamhetskrav på informationsadministration

- V 1: IA och verksamhetens krav – erfarenheter från offentlig förvaltning
- V 2: Fallstudie av IA-projektet vid Televerket
- V 3: IA-erfarenheter från företag och myndigheter
- V 4: Frihandel och handlingsfrihet - informationsvägar till 2000-talet (KOMMER jan 1993)

Modellering

- N 1: Modelleringsansatser för begrepps- och datamodellering: – Beskrivning och försök till jämförelse
- N 2: Generering av konceptuella modeller från policydokument
- N 3: Espritprojektet Tempora
- N 4: Prövning av regelbaserad metodik inom Posten
- N 5: En kokbok i remodellering - utkast
- N 6: Datorstöd för modellintegration
- N 7: Modellbaserad kunskapsinsamling
- N 8: Modellkvalitet (NY - nov 1992)
- N 9: Samband mellan dokument och modeller (NY - nov 1992)

Utbildning

- H1 – Handledarutbildning för modelleringsledare, avancerad
- H2 - Slutrapport HUMLA prototyp (NY - nov 1992. 4 sidor)

Katalogprinciper

- K 1: IRDS
- K 2: IRDS Modeller och modellnivåer
- K 3: Kopping begreppsmodell - relationsmodell
- K 4: IBM:s Repository Manager- en Introduktion
- K 5: IBM:s Repository Manager: Datamodelleringsbegreppen
- K 6: IBM:s Repository Manager: Begreppsmodellering i Information Model
- K 7: IBM Repository Manager: Attribut- och värdemodellering i Enterprise Submodel
- K 8: Navigering i Repository
- K 9: TRIAD Newsletter – IRDS inom ISO. Dagsläget
- K 10: TRIAD Newsletter – ISO/IRDS. Händelseutvecklingen 91/92
- K 11: Samverkan mellan resurskataloger – visioner eller behov
- K 12: AD/Cycle I Information Model – Processer och informationsflöden mellan processer
- K 13: AD/Cycle I Information Model – Info Flows inom Processmodellen
- K 14: AD/Cycle I Information Model – Relationsdatabasmodellering
- K 15: AD/Cycle I Information Model – Härlednings-specifikationer i begreppsmodellen
- K 16: IA-prototyp (NY - nov 1992)
- K 17: Repository AD/Cycle (R//AD) - International Users Group (NY - nov 1992)

Uttagssystem

- U1: Hybris i Unix-miljö, förstudie
- U2: DEBRIS (NY - nov 1992)