



# **Databaser och kalkylblad: En guide till god hantering**

Översatt, omarbetat och kompletterat av Ulf Jakobsson

PDF/A-1 (ISO 19005-1), skapat i Microsoft Office 2010 från formatet "docx".

# Databaser och kalkylblad: En guide till god hantering

Översatt från Archaeology Data Service's "Databases and Spreadsheets: A Guide to Good Practice" (<http://guides.archaeologydataservice.ac.uk/g2gp/Main>), varefter det har omarbetats och kompletterats för att bättre passa andra datamaterial med annat ursprung än arkeologi.

## Innehåll

1. Introduktion till Databaser och kalkylblad .....	5
1.1 Vad är databaser och kalkylblad?.....	5
Kalkylblad .....	6
Databaser .....	6
2. Att tänka på när man skapar databaser och kalkylblad .....	8
2.1 Allmänna överväganden.....	8
Undvik infogat material.....	8
Kontrollera överensstämmelse mellan data och dokumentation.....	8
Formatering av data och celler.....	8
Datum .....	9
Boolesk datatyp.....	9
Decimaltal.....	9
Valuta .....	9
3. Arkivering av databaser och kalkylblad .....	10
3.1 Vilka filer ska arkiveras? .....	10
3.2 Hur ska det arkiveras? .....	10
Viktiga egenskaper .....	10
Generella kontroller .....	10
Namngivning av filer.....	11
Filformat för långtidslagring.....	12
Kontroller efter migrering .....	13
Andra format .....	13
3.3 Metadata och Dokumentation .....	14
3.4 Strukturera data .....	15
4 Filformat .....	16
Vanliga filformat för databaser .....	16
Vanliga filformat för kalkylblad .....	22
Databaser och kalkylblad: Bibliografi .....	25
Verktyg .....	25



# 1. Introduktion till Databaser och kalkylblad

Även om databaser och kalkylprogram har olika funktioner så används de olika typerna av applikationer i många fall på liknande sätt för att samla in och lagra data (dvs. i rader och kolumner). Med långtidsbevaring i åtanke blir likheterna mer tydliga när man tittar på de viktigaste egenskaperna i formaten. För både databaser och kalkylblad är dessa egenskaper dels de datavärden som lagts in, dels själva strukturen (tabeller och blad) i vilka data lagts in. Ur det här perspektivet kan både databaser och kalkylblad hanteras (och långtidsbevaras) på liknande sätt.

Den här guiden syftar till att ge en överblick över vad man bör tänka på inför arbetet med att bevara de vanligaste funktionerna i databaser och kalkylblad. Guiden belyser deras likheter, och hur de kan hanteras på liknande sätt, men också deras skillnader och vilka ytterligare element, funktioner och processer som måste dokumenteras. Guiden ger inte detaljerad information om designen av databaser och kalkylblad bortom det som påverkar deras bevarande.

## 1.1 Vad är databaser och kalkylblad?

Definitionen på vad en databas är varierar beroende på vem man pratar med. Klart är att en databas består av en samling information som hör ihop och är organiserad på så sätt att det är lätt att söka efter och hämta enskilda bitar av information men också att kunna ändra på informationen. För dem som arbetar med databasteknik ska databasen också ha ett schema (en explicit beskrivning av vad det är för data) och vara konsistent eller logiskt koherent (inte innehålla motsägelser). Ordet databas kan beteckna dels informationen som finns lagrad, dels den programvara (databashanterare) som ska användas för att tolka datastrukturen som finns lagrad. Många databashanterare är väldigt komplexa och kan bestå av hela system med olika program. Exempel på databashanterare är Microsoft Access, MySQL, Microsoft SQL Server, Oracle och PostgreSQL. Fördelen med den här typen av databaser är bland annat att användaren själv bestämmer över struktur och funktion. Utan att gå in i det i detalj så kan man som användare dels ändra på den logiska strukturen (t.ex. utöka mängden kolumner) utan att för den sakens skull behöva skriva om programmet. Man kan också ändra på den fysiska lagringsstrukturen utan att behöva skriva om de program man använder. Några andra fördelar med att använda den här typen av databashanterare är t.ex. att flera användare kan arbeta samtidigt med samma databas; information som man matar in hamnar direkt i databasen och inte i datorns primärminne (som kan tappas om strömmen går); olika gränssnitt kan skapas beroende på användarnas behov; det går att ge olika användare olika behörigheter för att därigenom skydda mot obehörig åtkomst.

I detta dokument tar vi huvudsakligen upp den typ av databaser som har fördefinierade programvaror där man som användare matar in de uppgifter man vill ska finnas där varefter man använder de funktioner som finns fördefinierade. Användandet av dessa funktioner kan antingen ske via kommandorad eller via menyer. Data lagras sedan i filer vilka kan exporteras och importeras allt efter behov. Bland dessa program finns allt från den enklaste typen av så kallade kalkylblad (Excel) till dataprogram för statistik analys (SPSS, STATA m.fl.).

Även om det på den mest grundläggande nivån finns likheter mellan databaser och kalkylblad, i det att de innehåller tabulerad data med värden organiserade i kolumner och rader, så finns det en stor skillnad när det kommer till funktion. Kalkylbladsapplikationer, med sitt ursprung i redovisningskalkyler på papper, är tänkta att användas för matematiska data (t.ex. ekonomiska redovisningar) med snabba uträkningar och bearbetningar. Databasapplikationer är designade för att

lagra stora mängder data av olika typ med avancerade sök-/analys<sup>1</sup> och rapportfunktioner för dessa data.

### **Kalkylblad**

Kalkylblad är den enklaste formen av databaser och består vanligtvis av ett eller flera blad med tabulerad data. Kalkylblad kan, förutom data, innehålla formler, bilder, diagram och tabeller. Till exempel kan ytterligare värden (t.ex. totalsumman av en kolumn) skapas genom olika formler. Diagram och tabeller kan genereras genom att använda data där diagram/tabell sedan kan placeras i ett existerande blad eller existera enskilt som ett nytt blad. Formatering av celler eller av deras värden kan också vara ett viktigt element i kalkylbladet och kan användas för att förmedla information eller för att markera delar av information. Inmatning av data och användandet av kalkylbladet kan begränsas till viss del genom att låsa celler och cellspecifika beräkningar (t.ex. avrundning av värden eller genom att värdena anges i bestämda format som valuta eller tid).

Kalkylprogram har, liksom många ordbehandlingsprogram, under de senaste åren fått mer och mer stöd för XML-baserade filformat. Program som Microsoft Office, OpenOffice och WordPerfect Office har numera stöd för både Office Open XML (OOXML) och OpenDocument Format (ODF) filformat.

### **Databaser**

Till skillnad från kalkylblad tillåter, och ofta kräver, de flesta databasprogram specifikeringar av fältens längd (antal tecken) samt typen av data som ska registreras (numerisk osv.). I motsats till kalkylblad, vilka till stor del har samma grundläggande konstruktion och hantering av data, kan databaser delas in i ett antal olika typer baserade på deras arkitektur.

*Hierarkisk databas* är en äldre typ av databasmodell där data lagras i en trädstruktur, dvs. förälder-barn-förhållande, där varje post består av en "förälder" som sedan kan ha ett eller flera "barn" med liknande strukturer. Systemet är mycket snabbt på att ta fram enskilda poster och det är enkelt att lägga till/ta bort information, men modellen är minneskrävande, långsam på fri datasökning samt relationsmässigt stelbent.

*Rektangulära databaser* liknar kalkylblad i det att tabulerad data är organiserad i horisontella rader innehållande data om det undersökta objektet, och vertikala kolumner som representerar en bestämd typ, ett värde eller ett attribut som ska registreras för objektet. I rektangulära databaser kan det finnas en friare definition av vad data är men även hur data registreras i systemet. Detta samtidigt som det kan finnas duplicering av information i de olika värdena.

*Relationsdatabaser* löser dessa och liknande problem genom att kräva en datastruktur som är fördefinierad genom att gruppera ihop data med liknande attribut till separata tabeller vilka sedan länkas ihop genom vissa s.k. nyckelfält. Kombinationen av ett eller flera nyckelfält kan generera en

---

<sup>1</sup> Om det kallas sökfunktioner eller analysfunktioner varierar beroende på programvara. Hur avancerade dessa funktioner är varierar också mellan olika program.

nyckel. Det finns olika typer av nycklar t.ex. primära<sup>2</sup> och främmande<sup>3</sup> nycklar. Till skillnad från kalkylblad och många rektangulära databaser kräver de flesta databasapplikationer en strikt definiering av fältens längd samt typen av data (numerisk osv) som ska registreras.

*Objektorienterade databaser* skapas för lagring av komplexa objekt, som t.ex. multimediafiler och CAD-objekt. Data lagras normalt inte i databasen utan som attribut i objekt. När ett program gör ett anrop till en objekt-databas görs sökningen inte alltid direkt i databasen utan programmet ber objektet att utföra en bestämd rutin och leverera ett resultat. Även vanlig data lagras i objekt-databasen för att underlätta sökningar.

Liksom tabeller som genereras från kalkylblad kan databaser bestå av mer än enbart data-värden och metadata. Formulär, som används för att lägga in data eller för att göra beräkningar, är ofta det enda sättet användare är i kontakt med databasen och kan därför ses som en del av densamma, men separerad från själva data. På samma sätt kan frågeställningar och resultat eller rapporter, som resultat av användningen av databasen, ses som "icke-data"-delar av databasen. Vissa delar av dessa material kan därför vara viktiga att lagra tillsammans med databasen.

---

<sup>2</sup> En primär nyckel är den minsta uppsättning kolumner i en databas som man behöver veta värdet på för att hitta en unik post (tupel) i en databas. Vilka kolumner som ingår i nyckeln ska definieras. T.ex. kan nyckeln "identitet" bestå av information från kolumnerna Namn, Födelseår, Månad och Dag. Den kan även, som i Sverige, bestå av ett värde: personnumret. Detta då det är unikt för varje enskild person.

<sup>3</sup> Om det finns en främmande nyckel (d.v.s. en hänvisning till en primärnyckel i en annan tabell/databas) ska det anges. T.ex. kan nyckeln för identitet från exemplet i noten ovan användas som främmande nyckel i en annan databas, varvid uppgifterna för Namn, Födelseår, Månad och Dag inte behöver anges igen.

## 2. Att tänka på när man skapar databaser och kalkylblad

### 2.1 Allmänna överväganden

Nedan följer ett antal punkter som man bör titta på för att säkerställa att data är konsistenta och enkelt återanvändningsbara, samt att de förblir så genom hela bevarandeprocessen.

- Använd, där det är möjligt, kontrollerade vokabulärer och etablerade nyckelordlistor vid inmatning av data för både databaser och kalkylblad.
- Tänk på att vara konsekvent och ha meningsfulla namn för tabeller/kalkylblad och rader/kolumner. Var medveten om att det inte är säkert att tabeller eller kalkylblad kommer att lagras i en enskild fil. Tänk också på att vissa applikationer har begränsningar i vad man kan ange för namn på fält (t.ex. i ORACLE kan man inte använda namn på tabeller som börjar med en siffra eller fält vars namn börjar med "desc" eller "date") eller att använda mellanslag vid namngivning av fält eller kolumn. Även om vissa applikationer accepterar detta så är det bättre att undvika det eftersom det kan innebära problem för framtida datamigrering<sup>4</sup>.
- Vid användning av kalkylblad, undvik att använda formatering och sidlayout för att markera betydelsen av vissa värden då det kan gå förlorat när data migreras eller då man använder olika program.

#### Undvik infogat material

Många databas- och kalkylbladsapplikationer tillåter användaren att lägga in andra typer av filer i filen (t.ex. bilder). Medan man i databaser oftast lägger in länkar till externa filer snarare än filerna själva, kan man i kalkylbladsapplikationer som Microsoft Excel och olika versioner av OpenOffice Calc även lägga in tabeller och diagram skapade av data tillsammans med andra bilder. Det är rekommenderat att sådant innehåll lagras och arkiveras som separata filer i samma mapp som databasen/kalkylbladet. Därigenom säkerställer man att man inte tappar kvalitén på filerna. Det är viktigt att ha en lagringsstrategi för formatet av databasen/kalkylbladet så att den kan återskapas i framtiden.

#### Kontrollera överensstämmelse mellan data och dokumentation

Kodad eller inkonsekvent inlagd data innebär problem vid återanvändning när det gäller både databaser och kalkylblad. Kodade fält och kodad data måste vara tillräckligt dokumenterade och dokumentationen måste sparas tillsammans med databasen eller kalkylbladet för att betydelsen av kodningen inte ska förloras. Inkonsekvent inlagd data (vilket på ett enklare sätt kan kontrolleras i en databas än i ett kalkylblad) kan resultera i att betydelsen av data förloras (t.ex. är "A" detsamma som "a"?) och innebära problem när man ställer frågor till databasen.

#### Formatering av data och celler

Databaser och kalkylblad kan innehålla mer än tabulerad data, diagram och bilder. Särskilt kalkylbladsapplikationer tillåter formatering (t.ex. olika färg på text och celler) av data och de celler data ligger i. Den typen av formatering används vanligtvis för att markera eller förenkla läsningen av data (t.ex. kolumnsumma, negativa värden m.m.) men även för att markera information. Denna typ av formatering kan gå förlorad vid överföring till andra format, särskilt till ren text.

---

<sup>4</sup> Migrering kan innebära överföring mellan olika medier men också överföring mellan olika filformat.



## Datum

Det finns olika varianter att skriva datum, i vilken ordning dag/månad/år skrivs samt om man skriver årtalet med två eller fyra siffror. Rekommendationen är att skriva enligt ISO-standard dvs. åååå-mm-dd samt tid enligt hh:mm:ss. Se även ISO 8601.

## Boolesk datatyp

En typ av data där värdet representerar något som antingen är sant eller falskt, ja/nej, markerat/icke markerat osv. Enklast vid export är att göra om dessa värden till enklare text-värden som Y eller N, eller siffervärden som 1 eller 0.

## Decimaltal

En del program rundar per automatik av tal till två decimaler vilket kan innebära förlust av viktig data. Exempelvis kan en koordinat som är skriven som  $x=123,456$  avrundas till 123,46. Det går att ställa om antalet decimaler så att avrundning inte sker. Problemet med detta är att vissa program ändå inte exporterar värdena som de faktiskt är. Till exempel såg tidigare versioner av Microsoft Access export wizard alla siffervärden som decimaltal med två decimaler, och avrundade därmed alla tal per automatik till två decimaler. Detta problem går att komma förbi genom att manuellt tydliggöra att decimaltal inte ska avrundas utan exporteras i ursprunglig form.<sup>5</sup> Ytterligare instruktioner för hur man ska hantera decimaltal hos vissa filformat finns under respektive format i sektion 4.

## Valuta

Vid import av data där man använt sig av valutasymboler kan symbolen, när man öppnar upp en datafil, bli modifierad till den regionala valutan där man befinner sig istället för den ursprungliga valutasymbolen. Detta gör till exempel att dollartecken omvandlas till euro etc. Vid export är det därför bättre att göra om valuta till värde utan valutasymbol för att vid import förändra tillbaka till valuta. Dock ska man dokumentera detta i sin metadatadokumentation.<sup>6</sup>

---

<sup>5</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s.36

<sup>6</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s.37

## 3. Arkivering av databaser och kalkylblad

### 3.1 Vilka filer ska arkiveras?

Liksom textdokument förblir databaser och kalkylblad normalt i samma format genom hela skapelseprocessen. De är också huvudsakligen fristående då de oftast inte inkluderar importerade bilder eller andra typer av media inom filen. När så ändå är fallet är rekommendationen att det infogade materialet istället lagras separat så att långtidslagringen av detta sker på bästa sätt.

### 3.2 Hur ska det arkiveras?

Det centrala i databaser och kalkylblad är själva datatabellerna/bladen tillsammans med dokumentation och metadata som beskriver innehållet i och relationerna mellan tabeller och blad. Ordningen och/eller layouten för rader och kolumner kan också vara viktig (se nedan). Formulär, rapporter, frågeställningar och makron kan också vara värda att bevara.

#### Viktiga egenskaper

De grundläggande element i en fil som ska bevaras och underhållas är:

- **Värden** – De faktiska data inom kalkylblad eller databaser inklusive cellrubriker och själva värdet i cellerna. Det kan finnas flera blad med data eller tabeller.
- **Bilder** – Figurer, diagram och tabeller som finns i kalkylblad. Databaser tillåter sällan infogat material. Det är däremot möjligt att länka till externa filer. Det är dock viktigt att vara medveten om att den typen av funktioner kommer mer och mer med nya typer av databaser (t.ex. i Microsoft Access accdb-filer).
- **Layout** – För framförallt kalkylblad där formatering och användandet av färger och vissa formatmallar används för att lyfta fram betydelsen hos delar av datamaterialet, blir det viktigt att bevara denna ytterligare information på något sätt. Att använda layoutfunktioner är vanligt när man med hjälp av kalkylblad vill ge form åt tabulär data. Om man i kalkylblad använder sig av funktioner som formatering, färger och formatmallar bör alternativa format, som t.ex. PDF/A, även användas parallellt för att bevara utseendet på data då formateringen oftast går förlorad när man exporterar data till textbaserade format.
- **Relationer** – Hos databaser, men även för kalkylblad, är det viktigt att relationer mellan tabeller/blad är väl dokumenterade och begripliga.

#### Generella kontroller

Bortsett från att säkerställa att viktiga egenskaper hos en fil bevaras vid konvertering från ett format till ett annat, så finns det ett antal kontroller som måste göras innan konverteringen sker. Dessa kontroller säkerställer att viktiga karakteristika hos ett kalkylblad eller en databas bibehålls och inte går förlorade under konverteringsprocessen.

- **Layout och formatering** – Som nämnts tidigare kan, framförallt i kalkylblad, användandet av formatering, färger och formatmallar innebära problem när man sedan ska migrera data till t.ex. komma- eller tab-separerade textfiler. Detta om funktionerna har använts för att ge ytterligare mening till data. Vid kontroll bör man titta efter rubriker som går över flera rader eller kolumner, men också efter information som lyfts fram genom användandet av färg, ramar eller annan font. Beroende på typen av formatering så måste antingen data bli redigerad för hand innan migrering för att säkerställa att man inte förlorar information (för t.ex. kalkylblad så måste sammanslagna celler delas och texten dupliceras till varje ny cell) eller så måste man hitta ett alternativt migreringsformat.

- **Tabeller och blad** – Även om man ska anta att databaser och kalkylblad ska bevaras i sin helhet bör man ändå gå igenom varje enskild fil för att bedöma vilka tabeller eller blad som ska bevaras och migreras. Kalkylblad innehåller som grundinställning ett antal tomma blad och användare skapar ofta ytterligare blad – eller tabeller i en databas – för att tillfälligt lagra data som inte är tänkta att bevara.
- **Formler, queries (urval/förfrågningar), makron** – Om en fil innehåller komplexa formler eller urvalsförfrågningar som måste bevaras som fristående resultat måste dessa definieras som uttagsversioner av databasen eller kalkylbladet. Det som då bör bevaras i uttagsfilerna är själva resultaten medan formler och urvalsförfrågningar bör bevaras separat i en textfil så att kalkylbladets funktionalitet kan återskapas vid ett senare tillfälle.
- **Kommentarer och anteckningar** – Vid konvertering från ett format till ett annat kan det hända, att kommentarer och anteckningar som finns i en fil inte kommer med. Detta gör att dessa måste identifieras och sparas i en separat fil med klar information om vilken fil och vilken cell kommentaren/anteckningen tillhör.
- **Dold eller skyddad data** – Ibland innehåller kalkylblad dolda eller skyddade celler. Dessa celler måste identifieras så att informationen om dessa och vad de innehåller inte går förlorad.
- **Särskilda tecken eller avgränsare** – Ibland kan en databas eller ett kalkylblad innehålla specialtecken eller vanliga avgränsare inom datasetet. Avgränsare som vertikalstreck, kommatecken eller tabbseparatorer kan innebära problem i en fil när den ska migreras till textdokument med avgränsare varför den typen av tecken måste identifieras tidigt så att en strategi kan tas fram så att de inte går förlorade vid en konvertering. Bortsett från olika avgränsare kan specialtecken och icke-latinska tecken som t.ex. "&"-tecken, tankestreck ("—") och omvända citattecken (" ") påverka konvertering av data och därigenom hur data sedan ser ut. Databaser och kalkylblad kan också innehålla främmande tecken vilka inte kan exporteras till en textfil om inte en specifik teckenuppsättning anges (exempelvis UTF-8).
- **Länkar** – För databaser är det viktigt att relationen mellan olika tabeller är väl dokumenterad (se även 3.3) och stämmer. Kontroll för att säkerställa att dubbla eller "föräldralösa" poster inte förekommer måste ske. Arbetsblad i kalkylblad kan länkas samman genom att värden i celler hämtas från andra arbetsblad. Både databaser och kalkylblad kan innehålla länkar till, eller namn på, filer som lagras separat utanför databasen/kalkylbladet. Om dessa externa filer är en del av projektet så måste det säkerställas att filnamn och filer lagras korrekt.

### Namngivning av filer

Beroende på hur data ska lagras kan det vara nödvändigt att döpa om tabeller i databaser och arbetsblad i ett kalkylblad. Om filer ska lagras som textdokument med avgränsare så kommer varje tabell och varje arbetsblad resultera i en separat textfil. Detta gäller för både databaser och kalkylblad. I möjligaste mån ska man försöka att bibehålla ursprungsnamnet på filerna (filändelsen kan variera). Om flera arbetsblad eller tabeller konverteras så bör de nya filernas namn referera till ursprungsfilen men också till det/den enskilda arbetsbladet/tabellen, t.ex.:

- *[databasnamn]\_[tabellnamn].txt* t.ex. fynddatabas\_stenålder.txt
- *[kalkylbladsnamn]\_[arbetsbladsnamn].txt* t.ex. platsregister\_foton.txt

Då extra filer skapas för att innehålla bilder, urvalsfrågningar, anteckningar eller formler bör även dessa namnges på ett logiskt sätt så att det blir enkelt att koppla samman dem med ursprungsfilen de hör samman med t.ex.:

*[kalkylbladsnamn]-[arbetsbladsnamn]-[diagramnamn].tif.*

I vissa fall kan det vara nödvändigt att byta namn på tabeller genom att antingen korta ned dem eller genom att ta bort tecken som inte kan vara i ett filnamn. I de fallen ska man eftersträva att hålla sig så nära ursprungsnamnet som möjligt.

### Filformat för långtidslagring

För majoriteten av databaser och kalkylblad är det rekommenderade formatet för långtidsbevaring avgränsad text (tab, csv, etc). Det finns emellertid ofta stilistiska eller funktionella element i databaser och kalkylblad som bara kan bevaras i vissa format. I sådana fall är rekommendationen att ett öppet XML-baserat format, som .ods eller .xlsx, används eller att de delarna av filen dokumenteras och sparas tillsammans med en textbaserad export av data.

Format	Beskrivning
<b>Avgränsad Textfil (t.ex. .tab, .csv)</b>	<p>Avgränsade textfiler är generellt sett det rekommenderade formatet för långtidslagring av databaser och kalkylblad och man kan i många applikationer göra exporter till detta. Cellernas innehåll särskiljs med en avgränsare (t.ex. kommatecken) och ev. med citattecken runt värdet och varje rad indikeras med radbrytning. Vanliga avgränsare är kommatecken (csv-filer), tabb och vertikalstreck ( ). Dessa format är de vanligaste och har fördelen att de kan öppnas direkt av de vanligaste applikationerna som t.ex. MS Excel. Detta gör att formatet fungerar både för tillgängliggörande och arkivering. Som påpekats ovan lagras formatet enbart data medan annan information (bilder, formler mm) bör lagras separat.</p> <p>Det finns ingen specifikation för csv och det är lätt att hitta skillnader mellan olika csv-format. Mjukvara, system och regionala inställningar påverkar hur csv-exporten av en databas eller ett kalkylblad kommer att se ut. Exportguider och väl dokumenterade definitioner av csv-kodningen är nödvändiga för att all data som lagras konverteras på samma sätt. För att undvika problem med diakritiska<sup>7</sup> tecken, interpunktion och liknande så är en export med Unicode rekommenderad.<sup>8</sup></p>
<b>.ods</b>	Fungerar som arkivformat för kalkylblad men inbäddade diagram och tabeller bör lagras separat.
<b>.xlsx</b>	Fungerar som arkivformat för kalkylblad men inbäddade diagram och tabeller bör lagras separat.

<sup>7</sup> Diakritiska tecken är små skrivtecken som läggs till en bokstav (över, under eller ovanpå) som har en särskiljande funktion (det vill säga bokstavens uttal ändras).

<sup>8</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s. 36

## Kontroller efter migrering

Efter migrering av filer till andra format är det viktigt att genomföra ett antal kontroller för att säkerställa att data inte har förlorats eller blivit korrupta under processen. Dessa kontroller inkluderar:

- Kontrollera antalet rader efter konvertering.
- Kontrollera längden på textfälten så att informationen inte har kapats på grund av begränsning i teckenmängd.
- Kontrollera att alla arbetsblad och tabeller har exporterats.
- Kontrollera att specialtecken inte har gått förlorade.

## Andra format

PDF (helst PDF/A) kan vid vissa tillfällen användas för att tillgängliggöra kalkylbladsdata. Detta ska endast göras då mycket information finns via formatering och layout vilka inte på ett godtagbart sätt kan återskapas via en csv-fil och som går förlorade när man konverterar till ett XML-baserat format (ods eller xlsx). I vissa fall kan det vara nödvändigt att bifoga en pdf tillsammans med en csv-fil för att på det sättet visa för användaren hur kalkylbladet ursprungligen såg ut och samtidigt kunna arbeta med materialet.

XML är ett vanligt format för kalkylblad. Det kan öppnas i vanliga kalkylprogram och program som läser text. XML erbjuder också ett potentiellt pålitligt format för att lagra databasdata.

För att beskriva hur ett XML-dokument är strukturerat kan man använda sig av ett eller flera XML-scheman (eller tidigare DTD, Document Type Definition) vilka används för att klargöra vilka element och attribut som är tillåtna eller obligatoriska i en XML-applikation. Båda formerna möjliggör automatisk validering av innehållet i ett XML-dokument men då det finns begränsningar i DTD:s syntax kan man inte göra valideringen lika noggrann som om man använder den modernare tekniker. Båda kan innehålla kommentarer med instruktioner ämnade för mänsklig läsning där förväntat användningsområde för ett element eller ett attribut kan anges. Dessa instruktioner går dock inte att kontrollera maskinmässigt. En annan skillnad mellan DTD och XML-scheman är att i det senare finns möjligheten att ange *namnrymd*<sup>9</sup> samt datatyp.

Ytterligare ett alternativ för databaser är databasverktyget SIARD<sup>10</sup> (Software Independent Archiving of Relational Databases). SIARD är skrivet med standarder som XML, SQL:1999 och Unicode och stödjer ett antal databaser inklusive Oracle, Microsoft SQL Server och Microsoft Access. Swiss Federal Archives tillgängliggör SIARD utan kostnad men licensavtal krävs.

---

<sup>9</sup> Namnrymd (från engelskans *Namespace*): möjliggör användandet av fördefinierade element och attribut i XML-dokument. Då det finns olika XML-scheman kan tvetydigheter mellan de olika schemans definitioner av element och attribut uppstå och behöva klargöras. Detta genom att man för ett enskilt element/attribut anger vilket XML-schema man använt.

<sup>10</sup> <https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html>

### 3.3 Metadata och Dokumentation

För både databaser och kalkylblad behövs det metadata och dokumentation på olika nivåer för att säkerställa att data kan bevaras och återanvändas. Följande element bör dokumenteras och lagras inom varje dataset.

Element	Beskrivning
Projektets titel	
Namn på databas/kalkylblad	

Följande information ska finnas för varje arbetsblad/tabell i kalkylbladet/databasen:

Element	Beskrivning
Namn på arbetsblad/tabell	
Syfte med kalkylblad/tabell	
Antal rader med data	
Primär nyckel (databaser)	Minsta uppsättning kolumner i en databas som man behöver veta värdet på för att hitta en unik post. Vilka kolumner som ingår i nyckeln ska definieras.
Främmande nyckel (databaser)	Om det finns en främmande nyckel (primärnyckel i en annan tabell/databas) ska det anges.

Följande information ska finnas för varje kolumn/fält i kalkylbladet/databasen:

Element	Beskrivning
Namn	Namn på fält (databas) eller kolumn (kalkylblad).
Beskrivning	Fullständig beskrivning av fält samt de koder eller terminologier som använts. Koder som använts i ett dataset kan bifogas i ett separat dokument.
Typ av data samt längd på fältet (gäller databaser)	

För framförallt databaser är det även nödvändigt att beskriva relationerna mellan olika tabeller antingen med ord eller genom att skapa och bifoga ett relationsdiagram.

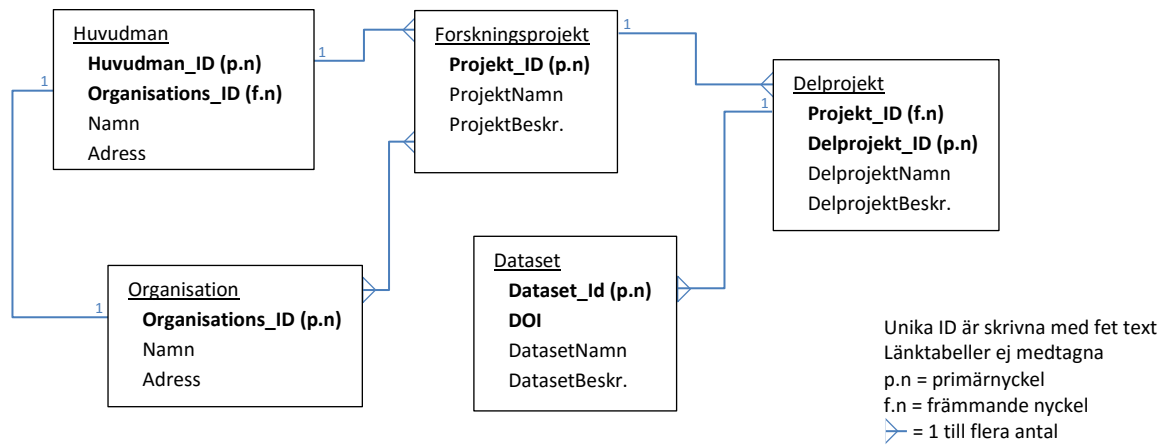


Fig. 1: Exempel på relationsdiagram.

Dokumentationen bör också inkludera information om annat som bör bevaras som t.ex. formler, urvalsförfrågningar, makron och kommentarer. Denna information kan lagras separat i t.ex. en textfil.

### 3.4 Strukturera data

Även om relationen mellan filer ska vara tydlig med hjälp av filnamnen så är det en fördel att lagra relaterade filer i en mapp. Exporterade bilder/diagram m.m. kan sedan lagras i en undermapp inom strukturen.

## 4 Filformat

I tabellerna nedan beskrivs några vanliga filformat som används vid skapandet av databaser och kalkylblad. Även filformatens tillhörande programvaror och hur, eller om, de kan användas för långtidslagring tas upp.

### Vanliga filformat för databaser

dBASE	
<b>Filformat/-ändelse</b>	DBF/.dbf (även .dbt och .ndx)
<b>Format</b>	Ursprungligen använt av dBASE men också använt av andra (t.ex. ESRI och deras ArcInfo).
<b>Beskrivning</b>	Formatet utvecklades för dBase men har sedan utvecklats till ett mer generellt format med ett antal (tyvärr inkompatibla) varianter vilka ofta refereras till som xBase. Även om strukturen är ganska enkel och väl dokumenterad <sup>11</sup> så är det viktigt att dokumentera vilken mjukvara man använt för att skapa filen.
<b>Rekommendationer</b>	Formatet kan användas för långtidslagring men på grund av inkompatibla varianter (se ovan) är det viktigt att veta exakt hur filen har skapats. dBase stöder, liksom de flesta applikationer, export av bättre lämpade format.

Filemaker Pro	
<b>Filformat/-ändelse</b>	FP7/.fp7 och tidigare versioner (MF5/.fp5, FM3/.fp3, FM/.fm)
<b>Format</b>	Filemaker Pro 11 och tidigare versioner. Proprietärt <sup>12</sup> filformat.
<b>Beskrivning</b>	Filer skapade i FileMaker Pro (.fp7) 11, 10, 9, 8, 7 kan öppnas direkt i och konverteras med Filemaker Pro 12 och senare versioner (.fmp12). Tidigare versioner än så kräver konverteringar i flera steg. <sup>13</sup>  Filer skapade med Filemaker Pro 6, 5, 4 eller 3 måste först konverteras till .fp7-format innan de kan konverteras vidare till .fmp12-format. <sup>14</sup>

<sup>11</sup> <http://www.clicketyclick.dk/databases/xbase/format/index.html>

<sup>12</sup> Proprietärt filformat är filformat som har restriktioner (vanligtvis satta av ägaren) vad gäller att använda, modifiera eller kopiera.

<sup>13</sup> [http://help.filemaker.com/app/answers/detail/a\\_id/10571/](http://help.filemaker.com/app/answers/detail/a_id/10571/)

<sup>14</sup> [http://help.filemaker.com/app/answers/detail/a\\_id/10571/](http://help.filemaker.com/app/answers/detail/a_id/10571/)



	Är filerna skapade med Filemaker Pro 2.x eller Filemaker Pro 1 så måste filerna först konverteras till Filemaker Pro 6 innan de konverteras till .fp7-format för att därefter konverteras till .fmp12-format. <sup>15</sup>
<b>Rekommendationer</b>	Filformatet .fp7 och tidigare är inte rekommenderade för långtidslagring men Filemaker Pro stöder genereringen av ett antal bättre lämpade format. <sup>16</sup>
<b>Format/-ändelse</b>	FMP12/.fmp12
<b>Format</b>	Filemaker Pro 12 och senare versioner. Proprietärt filformat.
<b>Beskrivning</b>	FileMaker Pro 14- och FileMaker Pro 14 Advanced-lösningar fungerar bara med de versioner av FileMaker som stöder .fmp12-formatet, dvs. version 12, 13, och 14. Lösningar från FileMaker Pro 11 och tidigare måste konverteras till det senaste filformatet (se ovan) för att kunna öppnas och delas med FileMaker Pro 14 och FileMaker Pro 14 Advanced. <sup>17</sup>
<b>Rekommendationer</b>	Filformatet .fp12 är inte rekommenderat för långtidslagring men Filemaker Pro stöder genereringen av ett antal bättre lämpade format. <sup>18</sup>

<b>Microsoft Access</b>	
<b>Filformat/-ändelse</b>	MDB/.mdb
<b>Format</b>	Proprietärt Microsoftformat som används för Access-databaser (2003 och tidigare varianter).
<b>Beskrivning</b>	Proprietärt Microsoftformat som nu ersatts av formatet .accdb. Filer sparade i filformatet .mdb öppnas med den version av Access som det sparades i, dvs. med den gällande databasmotorn (Jet Database Engine). Det har skett flera uppdateringar av Jet (Access 2.0 använder version 2.5, Access 95 använder version 3.0, Access 97 använder version 3.5, Access 2000 – 2003 använder version 4.0) vilket innebär att det har skett flera förändringar av .mdb-formatet. Problem har märkts framförallt mellan Jet-version 2 och 3. Det är möjligt att filer skapade i Access, baserade på versioner av Jet tidigare än version 3, inte går att öppna.
<b>Rekommendationer</b>	Filformatet .mdb är inte rekommenderat för långtidslagring men Access stöder genereringen av ett antal bättre lämpade format.

<sup>15</sup> [http://help.filemaker.com/app/answers/detail/a\\_id/10571/](http://help.filemaker.com/app/answers/detail/a_id/10571/)

<sup>16</sup> [http://help.filemaker.com/app/answers/detail/a\\_id/6167/~/exporting-data-from-filemaker-pro](http://help.filemaker.com/app/answers/detail/a_id/6167/~/exporting-data-from-filemaker-pro)

<sup>17</sup> <https://www.filemaker.com/se/products/filemaker-pro/pro-14-specifications.html>

<sup>18</sup> [http://help.filemaker.com/app/answers/detail/a\\_id/6167/~/exporting-data-from-filemaker-pro](http://help.filemaker.com/app/answers/detail/a_id/6167/~/exporting-data-from-filemaker-pro)

<b>Format/-ändelse</b>	ACCDB/.accdb
<b>Format</b>	Proprietärt Microsoftformat som används för Access-databaser (2007 och framåt).
<b>Beskrivning</b>	Filformatet .accdb introducerades med ACCESS 2007 och har även använts i ACCESS 2010. Även om formatet är grundformatet för Access 2007 och 2010, så är filer skapade i Access 2010 inte helt kompatibla med Access 2007 <sup>19</sup> .
<b>Rekommendationer</b>	Filformatet .accdb är inte rekommenderat för långtidslagring men ACCESS stöder genereringen av ett antal bättre lämpade format <sup>20</sup> .

<b>OpenDocument Base</b>	
<b>Filformat/-ändelse</b>	ODB/.odb
<b>Format</b>	Ett XML-baserat databasformat med ursprung i OpenOffice.org. Standarden definierades först i och med OpenDocument 1.2 även om den användes innan dess i OpenOffice.org
<b>Beskrivning</b>	En .odb-fil består av ett antal komprimerade mappar med xml-filer. En av filerna innehåller själva data, andra filer innehåller metadatadokumentation, information om applikationen som använts och en dokumentmall. En .odb-fil består även av ett maskinellt gränssnitt, som kan hämta information från annan data, liksom en databasmotor (i det här fallet en HSQL databasmotor <sup>21</sup> ).
<b>Rekommendationer</b>	Filformatet .odb är inte rekommenderat för långtidslagring men flera applikationer som använder formatet (t.ex. OpenOffice/LiberOffice) stöder genereringen av bättre lämpade format.

<sup>19</sup> <http://office.microsoft.com/en-us/access-help/convert-a-database-to-the-accdb-file-format-HA010341552.aspx>

<sup>20</sup> Missing Values konverteras till tomta fält vid export till dBASE  
(<https://support.sas.com/documentation/cdl/en/acpcref/63184/HTML/default/viewer.htm#a003102702.htm>)

<sup>21</sup> <http://hsqldb.org/>

<b>SAS</b>	
<b>Filformat/-ändelse</b>	SAS/.sas
<b>Format</b>	SAS (Statistical Analysis System). Proprietärt och plattformsoberoende filformat. <sup>22</sup> SAS är inte ett program utan en samling datorprogram för databearbetning, statistisk analys och rapportering.
<b>Beskrivning</b>	Kommando- och menystyrt program.  SAS filnamn kan vara upp till 32 tecken långa. Mellanslag och icke-alfanumeriska tecken (alfanumeriska: A-Z samt 0-9) är inte tillåtna bortsett från understreck ("_"). Filnamn måste starta med en bokstav eller med understreck, varefter siffror och bokstäver kan följa. <sup>23</sup>
<b>Rekommendationer</b>	Filformatet .sas är inte rekommenderat för långtidslagring men SAS stöder genereringen av ett antal bättre lämpade format. <sup>24</sup> Däribland export till ASCII varvid man samtidigt skapar en setupfil där bl.a. definiering av vad som är variabler görs. Säkerställ efteråt att exporten fungerat som det skall.
<b>Filformat/-ändelse</b>	XPORT/.xpt
<b>Format</b>	SAS Transport
<b>Beskrivning</b>	När man skapar en SAS transportfil (.xpt) följer inte missing data med utan raderas om man inte använder sig av SAS alpha missing codes. Detta i sin tur innebär ett problem eftersom skillnaden mellan olika typer av missing data (som tex "naturligt bortfall" kontra "vägrar att svara") inte följer med då alla missing data-värden slås ihop. Att särskilja de olika missing data-värdena kan vara viktigt vid sekundäranalys. Vid förberedelserna inför att skapa en SAS transportfil är därför rekommendationen att inte använda sig av SAS alpha missing codes utan att man skapar separata filer. <sup>25</sup>  Ett annat problem är det som kallas för SAS Proc Formats (value labels), som inte lagras i en SAS transportfil. SAS Proc Formats kan däremot sparas i separata programfiler eller lagras i SAS katalogfiler, vilka är operativsystemspecifika. Bästa sättet att lösa det på är att bifoga användardefinierade "SAS Proc format" och "format" i separata filer. <sup>26</sup>

<sup>22</sup>

<https://support.sas.com/documentation/cdl/en/movefile/59598/HTML/default/viewer.htm#a000986099.htm>

<sup>23</sup> <https://support.sas.com/documentation/cdl/en/lrcon/62955/HTML/default/viewer.htm#a00098953.htm>

<sup>24</sup> <https://support.sas.com/documentation/cdl/en/acpcref/63184/HTML/default/viewer.htm#a003102702.htm>

<sup>25</sup> <http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter6.html> Portable software-specific files.

<sup>26</sup> <http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter6.html> Portable software-specific files.

<b>Rekommendationer</b>	<p>Ej avsett för långtidsbevaring utan som medium för transport mellan olika datamiljöer.<sup>27</sup></p> <p>SAS support desk föreslår konvertering till ASCII (CSV).<sup>28</sup></p> <p>CESSDA föreslår konvertering till ren text med Unicode encoding.<sup>29</sup></p>
-------------------------	--

<b>SPSS</b>	
<b>Filformat/-ändelse</b>	POR/.por
<b>Format</b>	SPSS-portabelt format. Proprietärt filformat.
<b>Beskrivning</b>	<p>Portabelt format som kan läsas av andra versioner av SPSS oavsett operativsystem.</p> <p>Variabelnamn är begränsade till 8 bytes och konverteras, om nödvändigt, automatiskt till 8 bytes. Man kan inte spara data i .por-format med inställning för Unicode.</p> <p>Data sparat med IBM SPSS kan inte läsas av versioner äldre än version 7.5. Data sparat med Unicode-kodning kan inte läsas av versioner av IBM SPSS äldre än version 16.0.</p> <p>Om man öppnar datafiler med variabelnamn som är längre än 8 bytes i version 10.x eller 11.x, skapas unika versioner av variabelnamnen som är 8 bytes långa. De ursprungliga namnen sparas automatiskt för att kunna användas i version 12.0 eller senare. I versioner äldre än 10.0 så går det ursprungliga variabelnamnet förlorat. När man använder data med text-variabler som är längre än 255 bytes i versioner som föregår version 13.0 bryts text-variabeln ner till flera text-variabler som är maximalt 255 bytes långa.<sup>30</sup></p> <p>SPSS bibehåller missing data-värden vid genereringen av portabla filer.<sup>31</sup></p>
<b>Rekommendationer</b>	<p>Ej avsett för långtidsbevaring utan som medium för transport mellan olika datamiljöer.<sup>32</sup></p> <p>CESSDA föreslår konvertering till ren text med Unicode encoding.<sup>33</sup></p>

<sup>27</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s. 34

<sup>28</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s. 34

<sup>29</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s. 34

<sup>30</sup> <http://www->

[01.ibm.com/support/knowledgecenter/SSLVMB\\_20.0.0/com.ibm.spss.statistics.help/savedatatypes.htm](http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/savedatatypes.htm)

<sup>31</sup> <http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter6.html> Portable software-specific files.

<sup>32</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s. 34

<sup>33</sup> [http://ppp.cessda.net/doc/D10.4\\_Data\\_Formats.pdf](http://ppp.cessda.net/doc/D10.4_Data_Formats.pdf) s. 34

<b>Filformat/-ändelse</b>	SAV/.sav
<b>Format</b>	Proprietärt SPSS-format. SPSS betydde ursprungligen Statistical Package for the Social Sciences.
<b>Beskrivning</b>	Inledningsvis kommandobaserat men har sedan utvecklats till att använda grafiskt användargränssnitt (GUI). Syntax-kommandon, SAX Basic och Python är tre andra sätt att använda sig av SPSS om man inte vill använda sig av menyer.
<b>Rekommendationer</b>	Filformatet .sav är inte rekommenderat för långtidslagring men SPSS stöder genereringen av ett antal bättre lämpade format. Däribland konvertering till ASCII varvid man samtidigt skapar en setupfil där bl.a. definiering av vad som är variabler görs. Säkerställ efteråt att exporten fungerat som det skall.

<b>STATA</b>	
<b>Filformat/-ändelse</b>	DTA/ .dta
<b>Format</b>	Utvecklat av STATA som ett proprietärt och plattformsoberoende filformat avsett för rektangulär data. Programmet kan endast arbeta med ett dataset i taget.
<b>Beskrivning</b>	<p>Kommandobaserat användargränssnitt. Dataformatet har förändrats över tid men senare versioner kan läsa in tidigare versioners data. Data kan sparas i gällande version, samt i senast föregående versions format. Detta gäller dock fram till Stata 13. Senare version av Stata (Stata 14, filformat .dta_118) kan endast spara i sitt eget gällande format<sup>34</sup>.</p> <p>Text kodas med UTF-8 i Stata. Variabelnamn kan vara upp till 32 tecken långa där varje variabelsträng automatiskt avslutas med en numerisk nolla (/0). Varje UTF-8 tecken tar 1-4 bytes utrymme, varför <math>32 \cdot 4 + 1 = 129</math> bytes måste allokeras för varje variabelnamn.<sup>35</sup></p>
<b>Rekommendationer</b>	Filformatet .dta är inte rekommenderat för långtidslagring men STATA stöder genereringen av ett antal bättre lämpade format. Däribland konvertering till ASCII varvid man samtidigt skapar en setupfil där bl.a. definiering av vad som är variabler görs. Säkerställ efteråt att exporten fungerat som det skall.

<sup>34</sup> <http://www.stata.com/help.cgi?dta> "description" och "2. Versions and flavors of Stata"

<sup>35</sup> <http://www.stata.com/help.cgi?dta> "3. Representations of strings"

## Vanliga filformat för kalkylblad

Lotus 1-2-3	
<b>Filformat/-ändelse</b>	123/.123 och .wk* (t.ex. .wk4, .wks, etc.)
<b>Format</b>	Filer skapade med kalkylprogrammet Lotus 1-2-3. Binärt och proprietärt format.
<b>Beskrivning</b>	Lotus 1-2-3 var ett populärt programpaket på 1980- och 1990-talen men utvecklades inte vidare efter 2002. Filformatet kan läsas av MS Excel 2000 och OpenOffice Calc där det kan konverteras till bättre lämpade format.
<b>Rekommendationer</b>	Formatet är inte rekommenderat att användas för långtidsbevaring. Det kan dyka upp problem när man konverterar filer med hjälp av andra program (t.ex. kan formler som används räknas på ett annat sätt eller t.o.m. inte fungera alls). Det är därför viktigt att ha dokumentation som förklarar centrala beräkningar men också andra ytterligare egenskaper och funktioner.

Microsoft Excel	
<b>Filformat/-ändelse</b>	XLS/.xls
<b>Format</b>	Ett proprietärt binärt format som används för Microsoft Excel (t.o.m. Excel 2003).
<b>Beskrivning</b>	Även om formatet är proprietärt och ägs av Microsoft så är .xls-formatet vitt använt och kan importeras av ett antal tredjepartsapplikationer (t.ex. OpenOffice och Google Docs). Formatet har ersatts av ett XML-baserat format (.xlsx) i versioner av Microsoft Office 2003 och senare.
<b>Rekommendationer</b>	Även om filformatet är kompatibelt med andra öppna applikationer så är .xls inte standardformat för Excel-filer (se .xlsx nedan) och är inte rekommenderat för långtidslagring.
<b>Filformat/-ändelse</b>	XLSX/.xlsx
<b>Format</b>	Del av Office Open XML (OOXML)-formatet skapat av Microsoft. En ECMA <sup>36</sup> och ISO <sup>37</sup> -standard.
<b>Beskrivning</b>	Ett relativt nytt format från Microsoft som släpptes med Office 2007. Microsoft valde att utveckla en egen specifikation (OOXML) snarare än att

<sup>36</sup> [ECMA-376](#)

<sup>37</sup> [ISO/IEC 29500-1:2008](#)

	använda det existerande ODF-formatet. Formatet består av xml-filer avsedda för mänsklig läsning (human readable) med olika typer av information som ligger i en zippad fil.
<b>Rekommendationer</b>	.xlsx är en öppen och väl dokumenterad standard som fungerar för långtidsbevaring men infogat material bör lagras separat. .xlsx-filen är förenklat ett zippat arkiv men man bör spara innehållet i ett okomprimerat format. Migrering till rena textformat bör övervägas för maximal tillgänglighet, man bör dock säkerställa att decimaltal <sup>38</sup> förs över på ett korrekt sätt så att man inte tappar värdefull information.

<b>OpenDocument Spreadsheet</b>	
<b>Filformat/-ändelse</b>	ODT/.ods
<b>Format</b>	Del av OpenDocument-paketet av olika filformat, vilka är ISO-standarder (ISO/IEC 26300:2006 <sup>39</sup> ) för XML-baserade dokumentformat. Formatet stöds och används av flera olika kontorsapplikationer.
<b>Beskrivning</b>	Liksom OpenDocument Text-formatet (.odt) är .ods-filen i princip en packad fil innehållandes flera separata filer med dokumentmall, text (som XML) och inbäddade filer (t.ex. bilder).
<b>Rekommendationer</b>	Eftersom formatet är ett öppet XML-baserat format är det lämpligt för långtidslagring men bör lagras i okomprimerad form. Om ett dokument innehåller bilder och annat tillagt innehåll bör detta lagras separat i lämpliga format. Formatet .ods kan användas som långtidslagringsformat i de fall då .csv inte på ett tillräckligt bra sätt kan bevara all viktig information som finns i en fil. Det kan vara värt att prova konvertering till .xlsx för att se vilket som fungerar bäst. Värt att notera är att Calc erbjuder möjligheten att bevara information om teckenuppsättning vid export till .csv. Detta är ett bra alternativ vid konvertering från .xls till .csv då UTF-8 kodning är nödvändig.

<sup>38</sup> När man ska konvertera en Excel-fil till CSV-format bör man först notera vilka cell-värden som är i text- respektive numeriskt format osv. När detta är gjort gör man om alla värden till textformat varefter man sparar datamaterialet som CSV. För att säkerställa att import från en CSV-fil blir rätt: Öppna en ny Excel-fil, under Data-fliken välj Från Text, leta upp rätt fil, välj Import, följ instruktionerna som kommer upp i importfönstret. Observera att en CSV-fil blir större än en XLSX-fil på grund av att Excel komprimerar sina filer.

<sup>39</sup> [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=43485](http://www.iso.org/iso/catalogue_detail.htm?csnumber=43485)

OpenOffice/StarOffice...	
<b>Filformat/-ändelse</b>	SXC/.sxc
<b>Format</b>	Formatet är en del av OpenOffice XML-paketet och används i OpenOffice 1.0 Calc. Senare ersatt av OpenDocument Format (.ods).
<b>Beskrivning</b>	.sxc-formatet (numera ersatt av .ods) är som sin efterföljare ett XML-baserat format. Formatet stöds fortfarande av ett antal olika applikationer, däribland OpenOffice. Formatet består av ett antal komprimerade XML-filer där de faktiska data finns i en fil, medan bilder och annan information finns i ett separat register.
<b>Rekommendationer</b>	Även om det är ett öppet format och textbaserat rekommenderas användandet av det nyare formatet .ods.

Quattro Pro	
<b>Filformat/-ändelse</b>	WQ/.wq*, QPW/.qpw
<b>Format</b>	Filer skapade med kalkylprogrammet Quattro Pro (nu del av kontorspaketet WordPerfect Office).
<b>Beskrivning</b>	Ett proprietärt format använt av kalkylprogrammet Quattro Pro. Äldre filer kan öppnas med Microsoft Excel 2000 (för Quattro Pro för Windows versionerna 1-5 behövs en installation av Quattro Pro-konverterare) men det kan uppstå problem med import av filerna till Excel. Diagram, inbäddade objekt och makron som finns i Quattro Pro-filerna kan gå förlorade vid import till Excel. Nuvarande versioner av Quattro Pro har stöd för OOXML-format.
<b>Rekommendationer</b>	Inte att rekommendera för långtidsbevaring. Filerna bör exporteras till XML- eller textbaserade format.



## Databaser och kalkylblad: Bibliografi

OASIS (2005) *Open Document Format for Office Applications (OpenDocument) v1.0.*

<http://www.oasis-open.org/committees/download.php/12572/OpenDocument-v1.0-os.pdf>

## Verktyg

Pronom: <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

GDFR: [http://library.harvard.edu/preservation/digital-preservation\\_gdfr.html](http://library.harvard.edu/preservation/digital-preservation_gdfr.html)

CESSDA Tools: <http://cessda.net/CESSDA-Services/Resources/Tools>