

Specifikation
Tillägg

FGS Databas (Baserad på SIARD) Tillägg

*Förvaltningsgemensam
specifikation (FGS) för
relationsdatabaser baserad
på SIARD*

RAFGS6V1.0A20210628

Kontakta oss

Information om arbetet med
FGS:er
hittar du på vår webbplats:
www.riksarkivet.se/fgs-earkiv

Du kan även kontakta oss via
e-post: ra-fgs@riksarkivet.se

Juni 2021



Riksarkivet

INNEHÅLLSFÖRTECKNING

| | | |
|-----|--|---|
| 1 | Inledning..... | 4 |
| 1.1 | Läsanvisning..... | 4 |
| 1.2 | SIARD-standarden..... | 4 |
| 2 | Programvaror för att använda FGS Databas (Baserad på SIARD, Relationsdatabaser)..... | 5 |
| 2.1 | Generell beskrivning av verktyg | 5 |
| 3 | Att använda FGS Databas (Baserad på SIARD, Relationsdatabaser)..... | 6 |
| 3.1 | Tekniska och organisatoriska förutsättningar för att använda programvarorna | 6 |
| 3.2 | Metadata som tillförs SIARD-filen..... | 6 |
| 3.3 | Vyer och sökningar | 6 |
| 3.4 | Validering av SIARD-fil | 6 |
| 3.5 | Att titta på informationen i exporten..... | 7 |
| 4 | Användningsfall..... | 7 |
| 5 | Att placera informationen i ett informationspaket..... | 7 |
| 5.1 | Fält att ange i FGS Paketstruktur | 8 |
| 5.2 | Placering av filer i ett paket enligt FGS Paketstruktur | 9 |

ÄNDRINGSLOGG

| Version | Datum | Namn | Beskrivning |
|---------------------|------------|----------------------------------|-----------------------------|
| RAFGS6V1.0A20210614 | 2021-06-14 | Kommunalförbundet Sydarkivera | Första version av tillägget |

1 Inledning

Det här dokumentet innehåller teknisk information som är föränderlig men som inte påverkar eller förändrar användningen av FGS Databas (Baserad på SIARD). Dokumentet kommer att versionshanteras.

Dokumentet innehåller en instruktion för hur en SIARD-fil placeras i ett informationspaket enligt FGS-paketstruktur. Det innehåller även en beskrivning av metodiken och de programvaror som finns tillgängliga för arkivering av relationsdatabaser med hjälp av SIARD.

1.1 Läsanvisning

- Läs först dokumentet ”Introduktion till förvaltningsgemensamma specifikationer” där det finns en allmän översikt över FGS:er och hur de kan användas.

FGS Databas (Baserad på SIARD) består av ett antal dokument som bör läsas i följande ordning:

1. ”FGS Databas (Baserad på SIARD)” som specificerar användning av standarden SIARD.
2. ”FGS Databas (Baserad på SIARD) Tillägg som innehåller ytterligare vägledning för att använda FGS:en samt utförligare beskrivningar av grundstandarderna.

Information som exporteras i SIARD-format i enlighet med FGS Databas (Baserad på SIARD) kan paketeras i ett informationspaket i enlighet med FGS Paketstruktur.

FGS-dokumenterna går att nå via länkar från: <https://riksarkivet.se/fgs-earkiv>.

1.2 SIARD-standarderna

SIARD, version 2.1 (Software Independent Archiving of Relational Database)¹, en officiell nationell schweizisk standard som ursprungligen tagits fram av det Schweiziska Bundesarchiv (SFA) för långtidsbevarande av information ur databaser. Standarden är i dag en del av det europeiska byggnadsblocket för e-arkivering.

För en mer omfattande beskrivning av SIARD hänvisas till det Schweiziska Riksarkivet (Swiss Federal Archives, SFA) och Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) som i samarbete ger ut och ansvarar för standarden i sin lättillgängliga form.

SIARD har ursprungligen tagits fram av Schweiziska Riksarkivet (Swiss Federal Archives, SFA) för långtidsbevarande av information ur databaser och är fastställd som en eCH standard². Standarden används i flera europeiska länder och är i dag internationellt vedertagen. Den är också en viktig del av det europeiska byggnadsblocket för e-arkivering.

SIARD bygger på ISO standard SQL (ISO/IEC 9075) vilket är en viktig förutsättning för att ett stort antal databashanterare ska kunna hanteras. SIARD möjliggör export av databaser till ett generellt format som innehåller nödvändiga metadata för att kunna återskapa en relationsdatabas. SIARD Version 2.1 baseras på SQL 2008. Det tidigare formatet SIARD 1.0 baseras på en äldre version av SQL. Formatet definierar tabeller, kolumner, primärnycklar och främmande nycklar. SIARD kan även definiera databasscheman. Förutom SQL används de internationellt erkända standarderna XML, UNICODE och ZIP 64.

¹ <https://dilcis.eu/content-types/siard>

² eCH är den Schweiziska standardiseringsorganisationen för e-förvaltning.

2 Programvaror för att använda FGS Databas (Baserad på SIARD, Relationsdatabaser)

Till standarden finns det ett exportformat i XML framtaget för paketering och arkivering av relationsdatabaser och även ett antal verktyg som automatiskt kan exportera databaser i vanligt förekommande databashanterare som använder programspråket SQL. SIARD-filerna kan importeras som databas med hjälp av vanligt förekommande databashanterare.

Verktyg som hanterar SIARD kan använda standardkopplingar mot databaser vilket ger förutsättningar för transparens i lösningarna. Att använda SIARD resulterar i en total export av en databas. SIARD-formatet gör det möjligt att förutom informationen (innehållet) bevara relationer och viktiga funktioner i relationsdatabaser, det vill säga logiken med relationer och datatyper. SIARD-formatet möjliggör även import till måldatabas vilket ger förutsättningar för avancerade sökningar och mer fullständig funktionalitet.

Att manuellt genomföra en mappning och export av en relationsdatabas är inte något som rekommenderas på grund av komplexiteten. Det behövs därför ett verktyg för att skapa SIARD-filer. En programvara behövs för att hantera följande funktioner:

- Skapa SIARD-fil från databas,
- validera SIARD-fil mot XML-schema, och
- återsöka och visa data och metadata i SIARD-fil

I det här dokumentet beskrivs det generellt hur sådana verktyg fungerar. Verktyg kan exempelvis tillhandahållas av de organisationer som förvaltar standarden och specifikationerna för SIARD vilka är det Schweiziska Riksarkivet Swiss Federal Archives SFA och det europeiska byggnadsblocket för e-arkivering (Building Block eArchiving). Det kan även finnas andra alternativ som exempelvis olika kommersiella produkter på marknaden med stöd för SIARD eller egenutvecklade verktyg.

Varje organisation som väljer att använda SIARD för arkivering av relationsdatabaser behöver testa och utvärdera verktygen och ta ställning till vad som fungerar i den specifika tekniska och organisatoriska miljön. I samband med test och utvärdering behöver tillgänglig information om installation, användning och villkor granskas både utifrån tekniska och juridiska aspekter. Det är inte möjligt att ge några generella anvisningar som fungerar i alla förekommande situationer.

2.1 Generell beskrivning av verktyg

Programvarorna kan ha grafiskt användargränssnitt för olika tekniska plattformar och kan även finnas som kommandobaserade gränssnitt som stödjer schemaläggning av körningar. De organisationer som förvaltar standarden och specifikationerna tillhandahåller programvaror som är licensierade som öppen källkod för att säkerställa tillgänglighet till verktygen över tid. Öppen källkod och fria programvaror innebär att både kod och dokumentation finns publicerad öppet på internet.

Programvaran kan ha stöd för följande funktioner:

- Skapa SIARD-fil genom att ansluta till vald databashanterare.
- Validera SIARD-fil mot specifikation samt mot flera olika kontrollpunkter.
- Lägg till kompletterande beskrivande metadata om tabeller och kolumner.
- Navigera och söka i innehållet.
- Skapa och spara vyer för att presentera innehåll och underlätta sökning, exempelvis kan en handlingstyp uttryckas i en vy.

Verktygen kan ha stöd för att ansluta till vanligt förekommande databashanterarna som baseras på SQL. I dokumentationen anges vanligtvis vilka databashanterare som stöds.

Verktygen hanterar export från relationsdatabas och lagrar informationen i en SIARD-fil för långsiktigt digitalt bevarande i leverantörsberoende format. I övrigt kan programvarorna stödja olika funktioner när det gäller kontroller av SIARD-fil samt att söka och bearbeta data för vidareanvändning och tillgängliggörande. Vid behov skapas en ny databas för att söka och använda informationen som långtidslagras.

3 Att använda FGS Databas (Baserad på SIARD, Relationsdatabaser)

3.1 Tekniska och organisatoriska förutsättningar för att använda programvarorna

Programvaran behöver vara installerad på en PC (alternativt arbetsstation eller server) och det behövs åtkomst till den eller de databaser som programvaran ska ansluta till. För att köra programvaran direkt i producentens tekniska miljö så behöver programvaran stödja den aktuella tekniska plattformen. Det går även att fjärransluta till en databas och köra uttaget på distans. Det är möjligt att arbeta mot en databaskopia (databasdump), men då behöver man läsa in kopian i en databashanterare först som programvaran kan ansluta till.

Det beror på användarens tekniska kunskaper och förutsättningar om det är mest lämpligt att använda programvara med grafiska användargränssnitt eller arbeta kommandobaserat. Grafiska användargränssnitt sänker tröskeln något när det gäller att hantera tekniken för den som ska genomföra uttaget. Det är i vilket fall en förutsättning att användaren har grundläggande kunskaper om SQL, både för att kunna analysera databasen och för att skapa vyer med hjälp av SQL-frågor.

Tekniskt så behöver programvaran känna till port för att nå den aktuella servern, databasinstans och konto med användarnamn och lösenord. Kontot behöver ha rättighet till att läsa alla tabeller i databasen. Det behövs inte skrivrättigheter.

3.2 Metadata som tillförs SIARD-filen

Datafilerna lagras som XML-filer. En databasexport enligt SIARD-standarden utgörs av en paketfil i okomprimerat ZIP-format innehållande datafiler i XML-format samt ett antal metadatafiler i XML-format. SIARD-standarden definierar endast metadata relevanta för uttaget av databasen.

För att skapa SIARD-filen är det ett fåtal uppgifter som användaren behöver ange manuellt: Det handlar till exempel om datum för uttaget, vem som skapat SIARD-filen, klient för att skapa SIARD-filen, databashanterarens produktnamn, och var data kommer ifrån när det gäller tidsperiod och organisation. Dessa fält preciseras i dokumentet FGS Databas, (Baserad på SIARD, Relationsdatabaser).

3.3 Vyer och sökningar

SIARD-formatet och verktygen stödjer att skapa vyer för att visa valda handlingstyper. Vilka handlingstyper som ska sparas som vyer beror på överenskommelsen och vilka handlingstyper som vanligtvis behöver lämnas ut. Ett arbete som arkiv och systemförvaltningsorganisationerna behöver göra tillsammans. På det sättet går det att återskapa vyer från systemet för enklare sökning i den arkiverade databasen.

3.4 Validering av SIARD-fil

Validering (kontroll) av XML-filerna sker med automatik genom att använda det XML-schema som följer med SIARD-filen. Verktgssviten kan användas både för kontroll mot XML-schemat och andra kontrollpunkter som verktygen stödjer. XML-filerna är fullt möjliga att validera direkt mot XML-schema med oberoende verktyg. Valideringen kontrollerar inte den överförda informationen, utan endast att informationen har strukturerats korrekt.

3.5 Att titta på informationen i exporten

De verktyg som skapar exporten innehåller ofta möjligheter att visualisera data och att söka information i SIARD-filen. Till exempel ge ett användargränssnitt för att navigera och söka i innehållet i databasen.

SIARD-filerna kan återanvändas med hjälp av valfri databashanterare. Det gör det möjligt att återanvända data och också visualisera och göra data sökbart med alla de möjligheter som finns tillgängliga.

4 Användningsfall

Eftersom programvaran sköter packning och registrering av metadata med automatik så kräver själva uttaget minimal arbetsinsats. Det som kräver arbetsinsats är att identifiera handlingstyper som behöver sparas som vyer och analys av tabeller som kan gallras. Eftersom processen att skapa SIARD-filer kan automatiseras i hög grad så är det möjligt att planera och genomföra arbetet stegvis.

Exempel på enkel arbetsprocedur:

- Steg 1: Spara en fullständig SIARD-fil.
- Steg 2: Genomföra analyser av databasen. Skapa en ny SIARD-fil, och denna gång gallra de tabeller som inte behöver bevaras i processen.
- Steg 3: Skapa och spara vyer så att det går att spara en SIARD-fil med färdiga vyer för visning och sökning.

Exempel på arbetsprocedur som även innehåller att bearbeta data:

- *Kopiera:* Tag en kopia av den databas (i produktionsmiljö) som ska bevaras och importera den till en överföringsmiljö för vidare behandling och paketering.
- *Denormalisera:* Slå ihop (denormalisera) lämpliga tabeller för att få färre tabeller. Detta görs för att en databas ska blir mer överblickbar för ett mänskligt öga.
- *Skapa en testfil:* Skapa eventuellt en första SIARD-testfil med valfritt IT-verktyg.
- *Granska data:* Kontrollera data och metadata med valfritt IT-verktyg.
- *Förbättra data:* Utför bedömda förbättringar av data lagrade i överföringsmiljön.
- *Skapa slutversion:* Skapa en slutversion SIARD-fil med valfritt IT-verktyg i överföringsmiljön.
- *Granska och komplettera:* Granska data och komplettera SIARD-fil med metadata.
- *Mottagning:* Läs in SIARD-fil till valfritt e-arkiv.
- *Åtkomst:* Montera SIARD-fil vid behov i återsökningsmiljö med valfritt verktyg för att underlätta återsökning av användare.

För spårbarhet behöver de olika stegen dokumenteras i bevarandesystemet, förslagsvis i form av bevarandemetadata i PREMIS-format.

5 Att placera informationen i ett informationspaket

När uttaget har genomförts placeras det i ett informationspaket i enlighet med FGS Paketstruktur. De element som beskrivs i FGS Databas (Baserad på SIARD, Relationsdatabaser) och de element i FGS Paketstruktur som definieras nedan ska användas för överföring, bevarande och återsökning av informationspaket som innehåller relationsdatabaser. Elementen innehåller däremot inte information som gör det möjligt att göra generella informationsåtersökningar i den överförda informationen.

Detta då en relationsdatabas i sig ofta är av sådan storlek att man för att effektivt kunna söka och presentera informationen krävs att man återskapar eller via ett gränssnitt som kan hantera uttagsformatet för relationsdatabasen kan visa informationsinnehållet.

5.1 Fält att ange i FGS Paketstruktur

I paketet används följande fält för att ange att det är en överföring av en relationsdatabas.

Tabell 1: Specifika fält att använda i FGS Paketstruktur version 1.2

| Element namn | Elementets utseende beskriven med hjälp av XPath i METS | Värden |
|------------------------------|---|--|
| Beskrivning | <mets:LABEL="[Beskrivning]"> | Skriv i klartext att detta är ett uttag av en relationsdatabas med hjälp av formatet SIARD |
| Informationstyp | <mets:TYPE="[Informationstyp]"> | Ange värdet "Databases" |
| Informationstypspecifikation | <mets:ext:CONTENTTYPESPECIFICATION="[Informationstypsspecifikation]"> | Ange namnet på denna specifikation, "FGS Databas (Alternativ 2, relationsdatabaser) version 1" |
| Systemtyp | <mets:ext:SYSTEMTYPE="[Systemtyp]"> | Ange värdet "SIARD" |
| Startdatum | <mets:ext:STARTDATE="[Datum]"> | Om möjligt ange datum för första posten som återfinns i uttaget. |
| Slutdatum | <mets:ext:ENDDATE="[Datum]"> | Om möjligt ange datum för sista posten som återfinns i uttaget. |
| System namn | <agent:ROLE="ARCHIVIST" TYPE="OTHER" OTHERTYPE="SOFTWARE"><name>[System Namn] | Ange namnet på systemet som uttaget skett ifrån |
| System version | <agent:ROLE="ARCHIVIST" TYPE="OTHER" OTHERTYPE="SOFTWARE"><note>[System Version] | Om möjligt ange version på systemet som uttaget skett ifrån |
| Levererande System Namn | <agent:ROLE="CREATOR" TYPE="OTHER" OTHERTYPE="SOFTWARE"><name>[Levererande System Namn] | Ange namn på programvaran |

| Element namn | Elementets utseende beskriven med hjälp av XPath i METS | Värden |
|----------------------------|---|--|
| | | som använts för att skapa uttaget |
| Levererande System Version | <agent: ROLE="CREATOR" TYPE="OTHER" OTHERTYPE="SOFTWARE"><note>[Levererande System Version] | Ange version på programvaran som använts för att skapa uttaget |
| Funktion | <file: USE="[Funktion]"> | Ange att funktionen för filen är SIARD |

5.2 Placering av filer i ett paket enligt FGS Paketstruktur

De skapade SIARD-filerna placeras i paketet enligt instruktionerna givna i leveransöverenskommelsen.