



ARKIVVERKET
RIKSARKIVET

Databasebevaring

Arne-Kristian Groven,
KDRS-samling, Trondheim
05.06 2014



For de som ikke kjenner meg,
her er litt bakgrunnsinformasjon

- Har 20 års bakgrunn innen IT-forskning
- Snart 3 år i Riksarkivet
- Arbeider med:
 - Innhenting av bevaringsverdig arkivmateriale (uttrekk/SIP generering)
 - Innovasjoner/FoU/nye metoder og verktøy



Hvorfor starte her?

- Dette er presentasjonen jeg holdt på NorDig workshop i København for to dager siden
 - Men jeg vil bruke halvparten av tiden jeg da brukte: Det vil si en 30 minutters presentasjon på 15 minutter...
 - For deretter å vise et verktøy som bidrar med ett steg på veien til «evig databaseliv»
 - De siste 30 minuttene vil jeg bruke til å diskutere forbedringspotensialet hva gjelder bevaring og tilgjengeliggjøring



Hvorfor databasebevaring er et velegnet sted for meg å starte

- Det meste av det digitale materialet vi ønsker å bevare har en databasekomponent i seg
 - Enten databasene representerer dataene man ønsker å bevare eller de er metadataene
- I det danske Rigsarkivets nye depotsystem er det pr. i dag lagret 500 arkivpakker (AIPer)
 - 90% av dette er bevarte databaser



Hva jeg har drevet med i det siste

- Kartlegging av arbeidsprosesser og programvare for databasebevaring
 - Dagens praksis i arkiv-Norge
 - "State-of-art", internasjonalt
- Bruk av SIARD i prosjekter/piloter siste år:
 - Ny metodikk for bevaring og tilgjengeliggjøring
 - Departementsprosjektet
 - ...



Et eksempel på systemfloraen i en norsk kommune

Acos	Bostøtte / St@rtsak Husbanken	Epanet	Habildata	Kompas	Overformyn deri	Riksgab	Visma Link
Agresso	Capitech	Ephorte	HS-pro	Matrikkel	Pa Kirke	Rita	Visma oppvekst bhg.
Alreg	Cardinal	Familia (avsluttet)	Humanus (avsluttet)	MD Flyktning	PA Kis	Sharepoint	Visma oppvekst skole
Aplhareg	EA Driftssentral	Fond 2000	InteliCAD	MD Voksenopplæri ng	Proaktiv	Sofie (avsluttet)	Visma Unike kulturskole
ArcGis	Ecclesia	Fronter	ISAK	Micromarc (Avsluttet)	Proaktiv kirkegård	Sysvak	Visma Velferd
Arcibus	Ekko innfordringss ystem	G-Prog	KF Infoserie	Mobilprofil	Procasso	VA Gemini	Web-basert opptakssyste m Barnehage
Arkivplan.no	Entro	Gemini	KF Tjenestebesk rivelser	Norkom	Profil	Velferd Familia	WinMap (GAB)
Arx Access	Entry	Gemini Melding	King QM	Notus (avsluttet)	Ps GIS	Visma Cultus	Wintank
AutoCAD Map/Lt	EPA	GNSS	KIS	Online Scale	Reedsoft	Visma familia	Wis skole



Dagens status i Norge

- Mye bevaringsverdig digitalt skapt arkivmateriale forsvinner
 - Systemer og deres informasjonsinnhold dør
 - Spesielt ille for fagsystem: Register, støttesystem...
- **VI TRENGER NOE SOM KAN BEVARE DATABASER:**
 - **Effektivt!**
 - **Med høy kvalitet!**



ARKIVVERKET
RIKSARKIVET

Hvilke alternativer har vi?



Arkadukt/ADDML

Arkadukt

Arkadukt er et dataprogram utviklet for å produsere datasettbeskrivelser som følger Arkivverkets ADDML-standard.

- Verktøy ikke distribuert til arkivskapere

Hva er Arkadukt?

Arkadukt er et verktøy som kan lage eller redigere ADDML-datasettbeskrivelser, og være et hjelpemiddel for arkivskapere som skal overføre elektronisk arkivmateriale til Arkivverket eller for Arkivverket i sin rolle som arkivdepot og tester av mottatt arkivmateriale. I tillegg kan Arkadukt være aktuell for andre arkivdepoter i Norge.

Hva er en datasettbeskrivelse?

En datasettbeskrivelse beskriver strukturen og innholdet i en samling datafiler. Datasettbeskrivelsen er en XML-fil med tekniske metadata, og i noen tilfeller med referanse til andre beskrivende filer som XML-skjemaer. XML-filen hører sammen med datafilene den beskriver og lagres vanligvis sammen med dem.

Versjoner av Arkadukt

- Arkadukt 2.0 er utviklet i Java. Denne versjonen lager beskrivelser i henhold til ADDML 8.2, men kan også lese ADDML 7.3. Denne versjonen er for tiden bare tilgjengelig for Arkivverket.
- På et senere tidspunkt vil nyere versjoner av Arkadukt bli distribuert til aktuelle aktører.

Egenskaper ved Arkadukt

Programmet fungerer i stor grad som en editor som leser og skriver XML-filer, men avviker ved at det er spesialtilpasset gjeldende ADDML-versjon. Hvilken som helst tekst- eller XML-editor kan brukes for å lage datasettbeskrivelser, men Arkadukt forenkler jobben og sørger for at Arkivverkets spesielle bruk av ADDML blir ivarettatt. Følgende egenskaper gjelder fra og med versjon 2.0 og ADDML 8.2:

Arkadukt

- er en Java-applikasjon.
- er plattformuavhengig.
- validerer datasettbeskrivelsene i henhold til ADDML-skjemaet og Arkivverkets tilleggsregler.

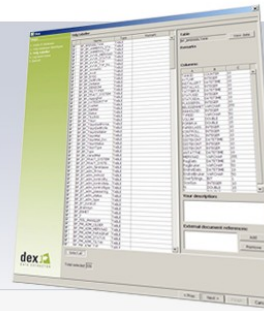


Dex Extractor



En **enkel og effektiv** metode for å ta vare på **digital informasjon**.

[Finn ut mer >>>](#)



Fang, dokumenter og avlever data fra fagsystemer og sak-/arkivsystemer.

Vi befinner oss i en digital mørketid. Digitalt lagrede data er informasjon som i sin natur er i en flyktig form. De blir raskt uleselige for oss dersom vi oppbevarer dem i samme form som i den daglige bruken.

Vi må foreta stadige konverteringer for å gjøre den digitale informasjonen flyttbar og håndterlig på nye teknologiplatformer. Det er behov for standardiserte løsninger som hindrer at hver enkelt konvertering bare binder oss til en ny systemavhengig teknologi.

Gjør hvert steg i avleveringsprosessen mer effektivt med DEX Data Extractor.

- Koble til aktuelt IT-System
- Velg tabell typer
- Velg hvilke tabeller som skal være med
- Ta med tilhørende dokumenter
- Lag en beskrivelse av systemet
- Gjør uttrekket!

Ønsker du mer info eller en **demonstrasjon?**

Aktuelt

- [Stadig flere ønsker uttrekk](#) 05.01.2014
- [Ny DEX - bedre feilhåndtering](#) 24.03.2010
- [DEX henger. Hva gjør jeg?](#) 24.03.2010
- [Hvor mye disk krever uttrekk?](#) 09.03.2010

[Arkiv | Søk](#)

 [English site](#)

```
<EndretDato>2008-01-01</EndretDato>  
<Avdeling>@NULL@</Avdeling>  
<Sakstype>9</Sakstype>  
<SaknrE>@NULL@</SaknrE>
```



- Har av forskjellige årsaker ikke fått stor utbredelse



Chronos

Database archiving with CHRONOS

Release your database!

Achieve a storage reduction and an increase of your performance of your database by exporting old data. Also the users of your database will be happy because of the shorter response times.

Reduce your costs for storage!

Export data, that you seldom or never need, onto a cheaper storage media. Use your existing infra structure and take the ILM (Information Lifecycle Management) concept into account. Of course, you still have access to all your archived data and they can also be re-imported.



Access simply and high-performance to your archived data!

Use the varied possibilities of access of CHRONOS and access to your archived data without problems. CHRONOS can be integrated in your existing applications and so you avoid an unnecessary amount of further training for the access to your archived data. CHRONOS finds your data very quickly by indexing the archive data. You will be surprised!

Automate your generating of test data!

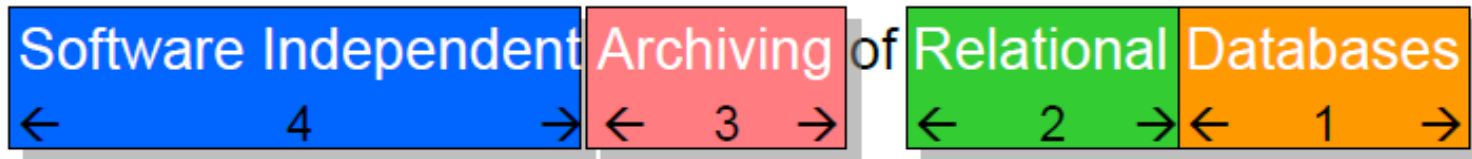
Offer your software-development-department consistent test data by just extracting a part of your database (=Subset) as test data. In



- For omfattende/dyrt



SIARD



1. Type of data: **Databases**
2. Type of database: **Relational** database
3. Purpose: **Archiving** (conversion into archival format)
4. Data is archived **independent** of specific applications (software)



SIARD oppsummert

- Veldefinert bevaringsformat, SIARD-formatet
- Veldefinerte transformasjoner
 - Normalisering fra ulike databaseplattformer (Oracle, MS SQL Server, MySQL, MS Access, DB/2) til et bevaringsformat (SIARD formatet)
 - Fra SIARD-formatet til en databaseplattform (på SQL:1999 formatet)
- Velfungerende verktøystøtte, Siard Suite
 - SiardFromDB, SiardToDB, SiardEdit (GUI)



Om SIARD-formatet

- En SIARD fil er en ZIP fil, på samme måte som DOCX og ODT filer er ZIP filer
 - 64-bit på grunn av databasenes størrelse
 - Bestående av XML filer
 - Muligens også tekst og binærfiler, store objekter
 - UTF-8 tegnsett på tekstfiler og XML filer
- Innholdet i SIARD-filen i henhold til SQL:1999 standarden
 - Ikke bare syntaktisk, men også i henhold til SQL:1999 sine konsistensregler



SIARD Formatet: Filenes struktur

header

```
metadata.xsd  
metadata.xml
```

content

schema1

table1

```
table1.xsd
```

```
table1.xml
```

lob1

```
record1.txt / record1.bin
```

lob2

```
record2.txt / record2.bin
```

table2

```
table2.xsd
```

```
table2.xml
```

schema2 ...

...

File and folder names short, plain ASCII strings.



Dette tas vare på av SIARD

- Database
 - Schemas, Users, Roles
- Schemas
 - Tables, Views, Routines
- Tables
 - Columns, Rows, Keys (Primary, Foreign, Candidate)
 - Constraints
- Rows
 - data records containing primary data
- Views
- Users, Roles

Metadata som tas vare på av SIARD

Database level

version (of SIARD format): 2.0
dbname
description
...
schemas List of schemata in the DB
users
roles
privileges

Users, Roles, Privileges are defined in SQL:1999 and serve informational purposes. *They are not required.*

Only on this level a minimum amount of non-technical, „archivists“ metadata may be stored.

Metadata som tas vare på av SIARD

Schema level

name
folder e.g., `schema0`
description
tables List of tables in the schema
views
routines

Views and Routines are defined in SQL:1999 and serve informational purposes (representing code rather than data). *They are not required.*

Metadata som tas vare på av SIARD

Table level

name
folder e.g., `table0` (in the `schema` folder)
description
columns List of the columns in the table
primaryKey
foreignKeys
checkConstraints
...
rows number of rows in the table

Constraints and Triggers are defined in SQL:1999 and serve informational purposes (ensuring consistency when the database is changed, which never happens to archived data). *They are not required.*

Metadata som tas vare på av SIARD

Column level

name

folder name of the LOB folder, e.g., `lob0`

description

type SQL:1999 column type

typeOriginal

defaultValue

nullable

Nullability and *DefaultValue* are defined in SQL:1999 and serve informational purposes (ensuring consistency when the database is changed, which never happens to archived data). *They are not required.*

Metadata som tas vare på av SIARD

Column level

name

folder name of the LOB folder, e.g., `lob0`

description

type SQL:1999 column type

typeOriginal

defaultValue

nullable

Nullability and *DefaultValue* are defined in SQL:1999 and serve informational purposes (ensuring consistency when the database is changed, which never happens to archived data). *They are not required.*



Metadata som tas vare på av SIARD

Other table level Metadata

- Primary key Metadata
- Foreign key Metadata
- Reference Metadata

...

Other Metadata

- View level
- Routine level
- User level
- Role level
- Privilege level



Eksempel: Transformasjoner fra MySQL datatyper til SIARD datatyper

<i>MySQL</i>	<i>JDBC (java.sql.Types)</i>	<i>SQL:1999 (SIARD)</i>	<i>XML</i>
CHAR	CHAR(1)	CHARACTER(1)	xs:string
CHAR(n)	CHAR(n)	CHARACTER(n)	xs:string
VARCHAR(n)	VARCHAR(n)	CHARACTER VARYING(n)	xs:string
TINYTEXT (up to 255)	VARCHAR(255)	CHARACTER VARYING(255)	xs:string
TEXT (up to 65'535)	LONGVARCHAR (65'535)	CHARACTER LARGE OBJECT	clobType
MEDIUMTEXT (up to 16'777'215)	LONGVARCHAR (16'777'215)	CHARACTER LARGE OBJECT	clobType
LONGTEXT (up to 2'147'483'647)	LONGVARCHAR (2'147'483'647)	CHARACTER LARGE OBJECT	clobType

⋮

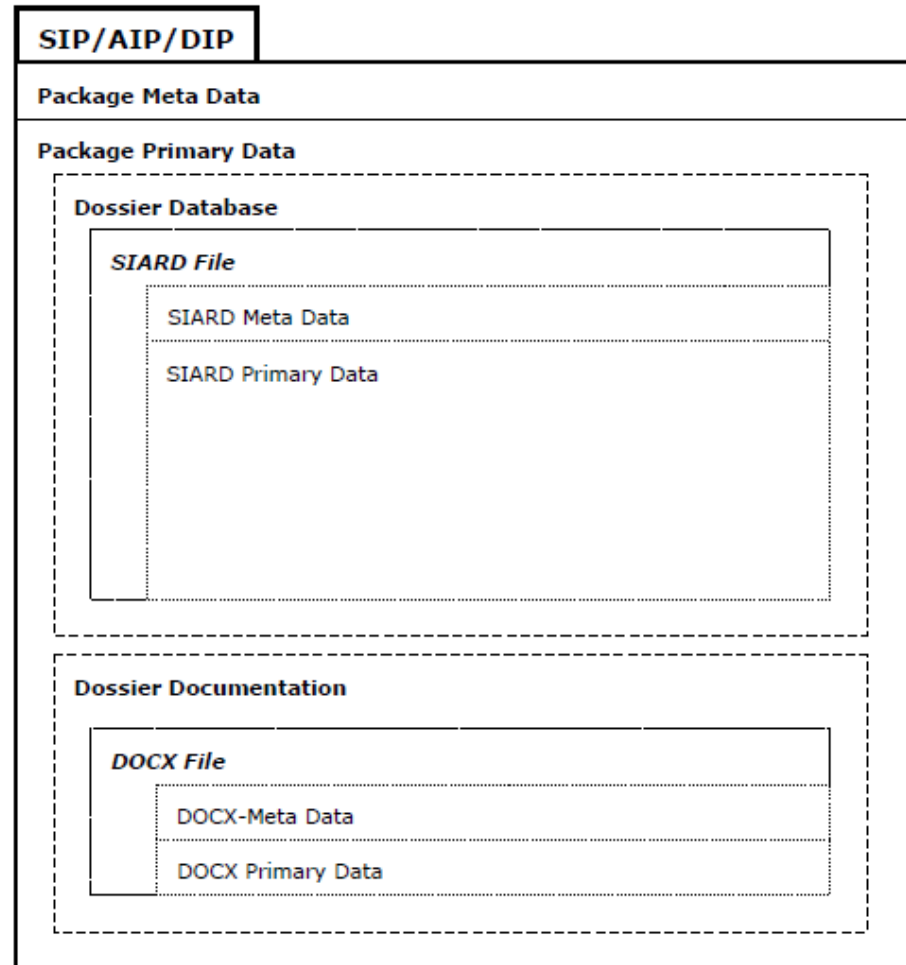


Erfaringer med SIARD

- Enkelt/lett å generere SIARD-filene
 - Oppgir databasens adresse og en databasebruker/passord (leseaksess)
- Databasepersonell hos arkivskapere kan gjøre jobben
- SIARD Suites GUI muliggjør rask inspeksjon/analyse av data
- Automatisert, veldokumentert transformasjon gir økt tillit.



MEN: SIARD-filen er bare en del i arkivpakken





SIARD og arkivaren

- SIARD (Suite) gjør ikke arkivarens jobb
- Arkivarens jobb starter når SIARD har gjort sin, eller aller helst lenge før
- Beskrivelser kan skrives både på databasenivå og tabellnivå
- Data(base)ordbok og annen dokumentasjon bør vedlegges SIARD-filen.



Våre neste aktiviteter

- For databaser til Noark-4 godkjente system
 - Eksperimentelt definere uttrekk, med utgangspunkt i SIARD-filen
 - Et prioritert målformatet vil være en RDF representasjon av Noark 5
- Hente inn flere fagsystem ved hjelp av SIARD Suite
 - Dette gjøres allerede av våre kommunale og interkommunale arkivsamarbeidspartnere