SIARD 2.0 standarden og litt om EU-prosjektet E-ARK

Arne-Kristian Groven,
Riksarkivet
arngro@arkivverket.no

KDRS-samling, Trondheim, 7.juni 2016

Hva presentasjonen skal handle om

- Gi en oppsummering rundt SIARD (v 2.0), som har blitt spesifisert i EU-prosjektet E-ARK hvor blant annet Riksarkivet/Arkivverket deltar
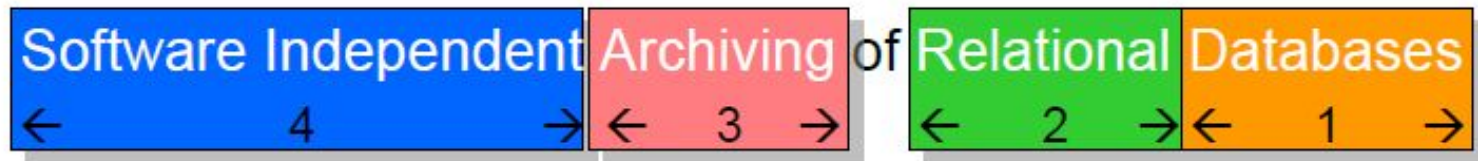- Fortelle litt om EU-prosjektet E-ARK, generelt

Hva er nytt i…

SIARD 2.0

SIARD 1.0

- Versjon 1.0 av SIARD formatet ble utviklet av Swiss Federal Archives i 2007.

- I 2013 be SIARD 1.0 en formell sveitsisk standard, eCH-0165

- Standarden er brukt i flere titalls land ved langtidsbevaring av databaseinnhold

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)

Kilde: Save Your Databases! | Urs Meyer, SFA | ECA 2010, April 2010

SIARD-formatet
(hvordan SIARD filer ser ut)

```
content
    schema1
        table1
            table.xsd
            table.xml
            lob1[11]
                record1.txt / record1.bin
            lob2
                record1.txt / record1.bin
            ...
        table2
            table.xsd
            table.xml
        ...
    schema2
    ...
header
    metadata.xsd[12]
    metadata.xml
```

- Metadata og data fra databasen blir lagret sammen…

- …i en ukomprimert ZIP64-fil.

- Filens suffiks er ".siard"

SIARDs plass i arkivpakkestrukturen
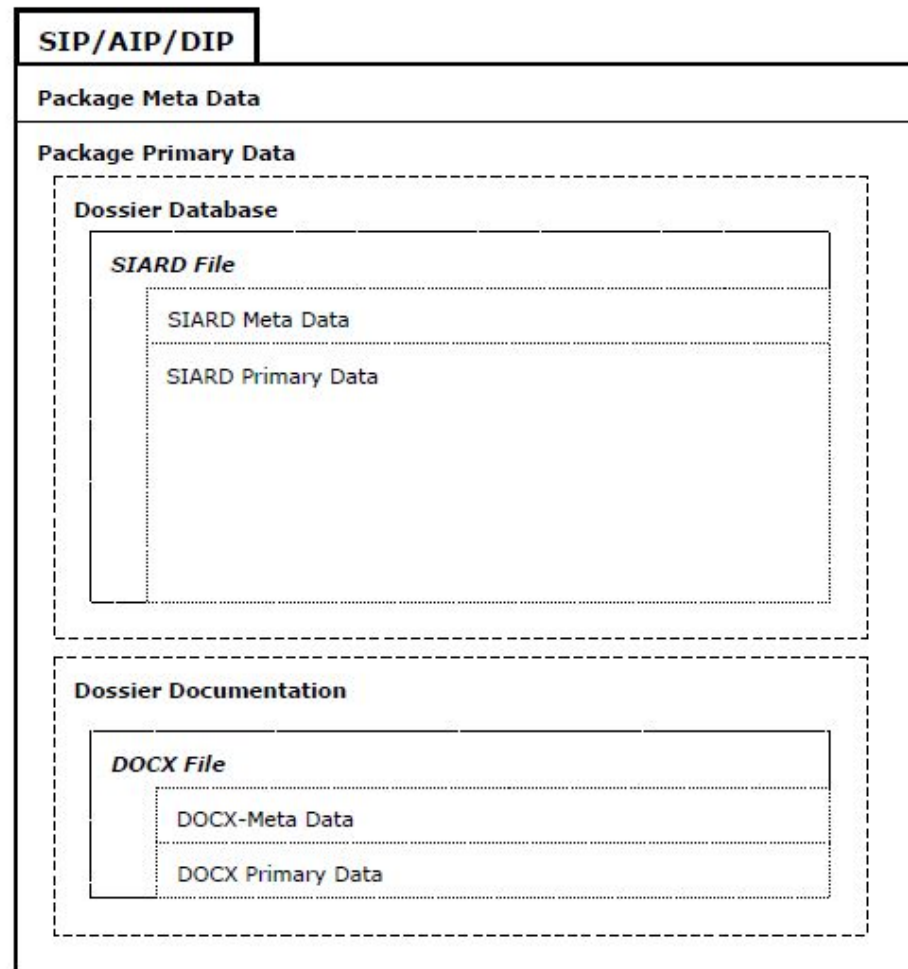
Hva skjer når SIARD-filen genereres?

Oracle	JDBC (java.sql.Types)	SQL:1999 (SIARD)	XML
CHAR	CHAR(1)	CHARACTER(1)	xs:string
CHAR(n)	CHAR(n)	CHARACTER(n)	xs:string
VARCHAR2(n)	VARCHAR(n)	CHARACTER VARYING(n)	xs:string
LONG	LONGVARCHAR	CHARACTER LARGE OB-JECT	clobType
CLOB	CLOB(4000)	CHARACTER LARGE OB-JECT	clobType
NCHAR	OTHER(1)	NATIONAL CHARACTER(1)	xs:string
NCHAR(n)	OTHER(n)	NATIONAL CHARACTER(n)	xs:string
NVARCHAR2(n)	OTHER(n)	NATIONAL CHARACTER VARYING(n)	xs:string
NCLOB	OTHER(4000)	NATIONAL CHARACTER LARGE OBJECT	clobType
XMLTYPE	OTHER(2000)	XML	clobType
INTEGER	DECIMAL(22)	DECIMAL(22)	xs:decimal
NUMBER(n)	DECIMAL(n)	DECIMAL(n)	xs:decimal
NUMBER(p,q)	DECIMAL(p,q)	DECIMAL(p,q)	xs:decimal
FLOAT(p)	FLOAT(p)	FLOAT(p)	xs:float
REAL	FLOAT(63)	FLOAT(63)	xs:float

- Datatypene i Oracle oversettes til datatyper i SQL:1999 standarden når SIARD-filen genereres.
- En tilsvarende, men motsatt operasjon skjer dersom SIARD-feilen eksporteres tilbake til Oracle.
  - Men Oracle spesifikke termer blir da erstattet med mer genersiske

Flere av dere kjenner programvaren SiardSuite og har brukt denne

eCH-0165 SIARD-Formatspezifikation

SIARD 2.0

Name	SIARD-Formatspezifikation
eCH-Nummer	eCH-0165
Kategorie	Standard
Reifegrad	Experimentell
Version	2.0
Status	
Genehmigt am	
Ausgabedatum	2016-02-29
Ersetzt Version	1.0
Voraussetzungen	-
Beilagen	metadata.xsd
Sprachen	Deutsch (Original), Französisch (Uebersetzung), Englisch (Uebersetzung)
Autoren	Luis Faria, KEEP SOLUTIONS, LDA, lfaria@keep.pt
	Anders Bo Nielsen, Danish National Archives (Rigsarkivet), abn@sa.dk
	Krystyna Ohnesorge, Schweizerisches Bundesarchiv, krystyna.ohnesorge@bar.admin.ch
	Claire Röthlisberger-Jourdan, KOST, claire.roethlisberger@kost.admin.ch
	Hartwig Thomas, Enter AG, hartwig.thomas@enterag.ch
	Andreas Voss †, Schweizerisches Bundesarchiv, andreas.voss@bar.admin.ch
Mitarbeitende	Karin Bredenberg, National Archives of Sweden, karin.bredenberg@riksarkivet.se
	Hedi Bruggisser, Staatsarchiv Thurgau, hedi.bruggisser@tg.ch
	Georg Büchler, KOST, georg.buechler@kost.admin.ch
	Janet Delve, University of Portsmouth, janet.delve@port.ac.uk

	Boris Domajnko, Slovenian National Archives, boris.domajnko@gov.si
	Alain Dubois, Staatsarchiv Wallis, alain.dubois@admin.vs.ch
	Bruno Ferreira, KEEP SOLUTIONS, LDA, bferreira@keep.pt
	Arne-Kristian Groven, National Archives Norway (Riksarkivet), arngro@arkivverket.no
	Martin Kaiser, KOST, martin.kaiser@kost.admin.ch
	Lambert Kansy, Staatsarchiv Basel Stadt, lambert.kansy@bs.ch
	Markus Lischer, Staatsarchiv Luzern, markus.lischer@lu.ch
	Zoltán Lux, National Archives of Hungary, lux.zoltan@mnl.gov.hu
	Lauri Rätsep, National Archives of Estonia, lauri.ratsep@ra.ee
	Hélder Silva, KEEP SOLUTIONS, LDA, hsilva@keep.pt
Herausgeber / Vertrieb	Verein eCH, Mainaustrasse 30, Postfach, 8034 Zürich T 044 388 74 64, F 044 388 71 80 www.ech.ch / info@ech.ch

Hva er nytt i SIARD 2.0?...

- This version 2.0 is based on version 1.0 and defines a format that is backward-compatible with version 1.0.
- Version 2.0 has been developed jointly by the Swiss Federal Archives, eCH (who promote, develops and approves eGovernment standards in Switzerland) and the EU project E-ARK.

Hva er nytt i SIARD 2.0?...

- Upgrade of SQL:1999 support to SQL:2008 support
- Support for all SQL:2008 types, in particular user-defined data types (UDTs)
- More explicit validation rules for data type definitions using regular expressions

Hva er nytt i SIARD 2.0?

- Small modification of the definition, showing when to store large objects inline as part of the table XML

- Support for storing large objects outside of the SIARD file using "file:" URIs

- Support for "deflate" as a compression mechanism

I SIARD 2.0 kan BLOBene dyttes ut og gis unike identifikatorer

SIARD 2.0 representerer et nytt nivå av modenhet

SIARD 1.0 (Sveits + drøyt 50 land)

DBML (Portugal ++)

SIARD DK (Danmark)
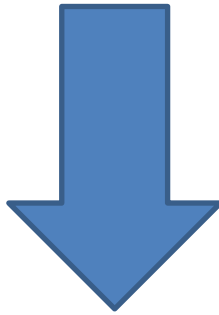
SIARD 2.0

- Dette åpner også opp for flere tilgjengelige verktøy (programvare) for databasebevaring:
  - SiardSuite
  - Database preservation toolkit
  - M.fl.

Flere verktøy blir tilgjengelige, f.eks

Database Preservation Toolkit

The Database Preservation Toolkit allows conversion between Database formats, including connection to live systems, for purposes of digitally preserving databases. The toolkit allows conversion of live or backed-up databases into preservation formats such as **SIARD**, a XML-based format created for the purpose of database preservation. The toolkit also allows conversion of the preservation formats back into live systems to allow the full functionality of databases.

This toolkit was part of the RODA project and now has been released as a project by its own due to the increasing interest on this particular feature. It is now being further developed in the EARK project together with a new version of the SIARD preservation format.

The toolkit is created as a platform that uses input and output modules. Each module supports read and/or write to a particular database format or live system. New modules can easily be added by implementation of a new interface and adding of new drivers.

EARK and SIARD 2.0

A new version of the this tool, together with a new version of the SIARD preservation format, is currently being designed and developed on the EARK project. Meanwhile, if you'd like to know more and even send us use cases and requirements, contact us.

Litt om…

E-ARK PROSJEKTET

E-ARK

- E-ARK prosjektet eksisterer innenfor rammene av EUs forsknings- og innovasjonsaktiviteter
- Det er et såkalt piloteringsprosjekt, hvor uttesting er i fokus
- Prosjektet er delfinansiert av EU-kommisjonen etter gitte regler
- Prosjektet er et samarbeidsprosjekt mellom partnere fra ulike EU/EØS-land
- Prosjektet avsluttes i 2017

Search... GO

E-ARK

Site Map   Register   Login   SHARE

HOME   ABOUT   NEWS   EVENTS   COMMUNITY   USING E-ARK   PARTNERS   RESOURCES   CONTACTS   STORIES

E-ARK

Deleveloping Open Technical Products (tools, services, metadata specifications)

Find out more

HTTP://ARCHIVES.LADROME.FR/

@EARKProject on Twitter

EARKProject @JohnGlenMP, will you speak to @bisgovuk to stop the #privatisation of the #landregistry?
May 26 · reply · retweet

EARKProject RT @Sarah_DPC: Drumroll please! The Digital Preservation Awards are back for 2016. Learn how you could be a winner: https://t.co/3uTIEZJowf...
May 11 · reply · retweet

Welcome to the E-ARK Project

(European Archival Records and Knowledge Preservation)

E-ARK is a multinational big data research project that aims to improve the methods and technologies of digital archiving, in order to achieve consistency on a Europe-wide scale.

Tackling a range of problems associated with independent record-keeping technologies, systems and practices, E-ARK aims to impact the development of internationally accessible archives through: the provision of technical specifications and tools, the development of an integrated archiving infrastructure, the demonstration of improved availability, access and use, and the rigorous analysis of aggregated sets of archival data.

Running from 1st February 2014 to 31st January 2017 it is co-funded by the European Commission under its ICT Policy Support Programme

Latest News

E-ARK Deliverable D5.3 Published
03-05-2016

E-ARK Project Deliverable D5.3 - E-ARK Pilot DIP Specification has now been published. The...

E-ARK Partner DLM Forum announce their member meeting and AGM. The Hague 8 - 9 June 2016
27-04-2016

The DLM Forum Foundation

http://www.eark-project.com/

Prosjektdeltagerne

- University of Brighton (E-ARK Project Co-Ordinator)
- AIT Austrian Institute of Technology GmbH
- Archives of the Republic of Slovenia - Arhiv Republike Slovenije
- The DLM Forum Foundation
- Statens Arkiver: The Danish National Archives
- The Digital Preservation Coalition
- Universität zu Köln: The University of Cologne
- Instituto Superior Técnico - Lisbon Technical University
- The National Archives of Hungary
- Rahvusarhiiv: The National Archives of Estonia
- Arkivverket: National Archival Services of Norway
- ES Solutions (Sweden)
- Magenta (Denmark)
- KEEP Solutions LDA (Portugal)
- Agência para a Modernização Administrativa IP - Agency for Public Services Reform, IP (Portugal)
- Ministerio de Hacienda y Administraciones Pública - The Ministry of Finances and Public Administrations (Spain)
- University of Portsmouth
- I tillegg: Swiss Federal Archives

Prosjektet har ulike typer deltagere

- Programvare-leverandører
  - KEEP Solutions LDA (Portugal), Magenta (Danmark), ES Solutions (Sverige), AIT Austrian Institute of Technology (Østerrike)
- Nasjonalarkiv o.l.
  - Arkivverket/Riksarkivet (Norge), Statens arkiver/Rigsarkivet (Danmark), The National Archives of Hungary (Ungarn), The National Archives of Estonia (Estland), Swiss Federal Archives (Sveits), Archives of the Republic of Slovenia (Slovenia), The Ministry of Finances and Public Administrations (Spania), Agency for Public Services Reform, IP (Portugal)

Prosjektet har ulike typer deltagere

- Interesseorganisasjoner
  - The Digital Preservation Coalition, The DLM Forum Foundation
- Forskningsinstitusjoner, universiteter
  - University of Brighton (Storbritannia), University of Portsmouth (Storbritannia), Lisbon Technical University (Potrugal), The University of Cologne (Tyskland), AIT Austrian Institute of Technology (Østerrike)

Hva som utvikles i prosjektet

- Guidelines og studier
- Metadata spesifikasjoner, eksempelvis:
    - SIARD 2.0, standard for databasebevaring
    - Forsøk på å standardisere pakkestrukturer
- Krav til verktøy (programvare) og verktøyleverandører som deltar i prosjektet
- Programvareendringer basert på disse kravene
- Pilotering, uttesting av verktøy og metadata-spesifikasjoner

Hva som utvikles i prosjektet: Riksarkivets hovedaktiviteter

- Metadata spesifikasjoner:
  - SIARD 2.0, standard for databasebevaring
  - Forsøk på å standardisere pakkestrukturer
- Pilotering, uttesting av verktøy og metadata-spesifikasjoner:
  - Uttesting av ESSArch Tools og ESSArch Preservation Platform (nye versjoner)

TAKK FOR MEG