

Det danske Rigsarkivs del i E-ARK- projektet

Præsentation v. Phillip Tømmerholt

NorDig

26. oktober 2017



NorDig
26. oktober 2017

Outline

- Rigsarkivets ansvar i E-ARK.
- Historisk databasebevaring i Danmark
 - a) SIARD 1.0*
 - b) SIARDDK*
 - c) SIARD 2.0*
- Pilot 1 i E-ARK (test af værktøj til databaseudtræk)



DoW - Description of Work

WP1 - Project Co-Ordination

WP2 - Use Cases and Pilots

WP3 - Transfer of Records to Archives

WP4 - Archival Records Preservation

WP5 - Archival Records Access Services

WP6 - Archival Storage, Services, and Integration

WP7 - Evaluation and Assessment

WP8 - Dissemination and Exploitation

(Organisation)

Test af specifikationer og værktøjer

SIP specifikation + værktøjer

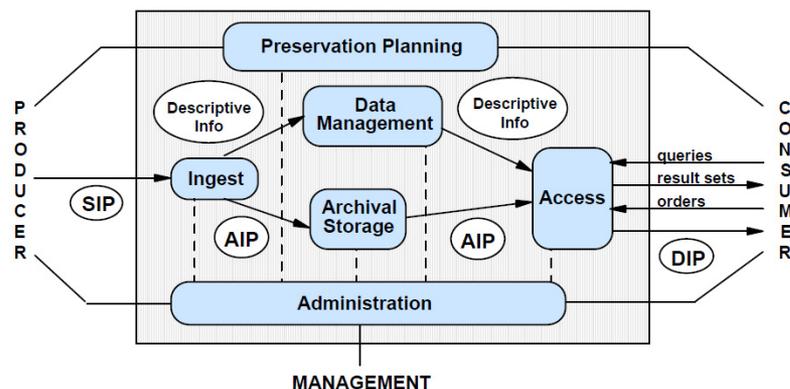
AIP specifikation + værktøjer

DIP specifikation + værktøjer

Software reference model (AIP)

(Organisation)

(Organisation)



Rigsarkivets del i projektet:

- Ansvarlig for arbejdspakke 5 (Archival Records Access Services)
- Ansvarlig for Pilot 1
- Bidragsydere til:
 - a) content information type om databaser (SIP, AIP, DIP)
 - b) Common Specification, METS, skalerbarhed

Jeg selv personlig var tilkøbet E-ARK siden november 2015 (1 år og 3 mdr)

Jeg har deltaget meget i alle tre dele, og har selv haft hovedansvaret for Pilot 1.



- Siden 1973 har Rigsarkivet arkiveret digitale arkivalier.
- Siden 1997 har vi arkiveret digitale arkivalier, hvor den relationelle databasemodel har været bærende.
(fagsystemer, dokumentssystemer, e-journaler, registre)
- Siden 2010 har arkivskaber afleveret efter den danske SIARD-variant = SIARDDK





Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

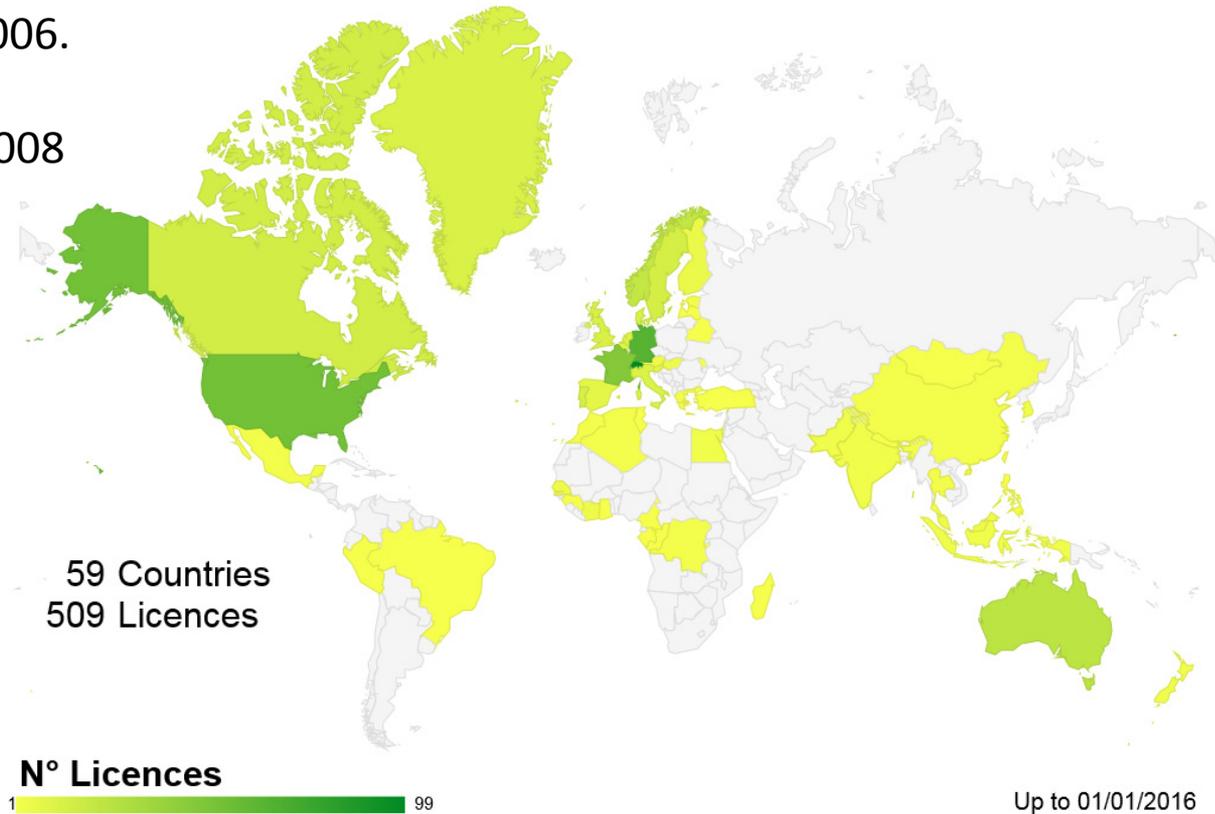
SIARD 1.0

Software Independent Archiving of Relational Databases

ARELDA project - 2006.

Gov. - 2008

PLANETS project -2008



NorDig
26. oktober 2017

SIARD 1.0 versus SIARD DK / 1007

SIARD 1.0 formatet tillader dokumenter, men slet ikke i en grad så det er skalerbart. SIARD1.0 formatet er ikke et informationspakkeformat, men det er 1007.

Dokumenter indgår i vores arkiveringsversioner, med et dokumentindeks, der peger 1:1 på tabeldata i relationel form. Derfor har vi også et sagsbegreb, dokumentmetadata osv. som ligger som en ekstra information i vores tabelindeks.

SIARD1.0 har dog mulighed for at bevare flere skemaer og informationer om kandidatnøgler, routines, triggers, og informationer om databasebrugeres roller og rettigheder.



SIARD 1.0 + SIARD DK / 1007 → SIARD 2.0

<http://www.eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20>



The screenshot shows the E-ARK website interface. At the top left is the E-ARK logo, a red boat with the text "E-ARK A European 'Showcase' Project". To the right is a search bar with "GO" and social media icons. Below the logo is a navigation menu with links: HOME, ABOUT, NEWS, EVENTS, COMMUNITY, PARTNERS, RESOURCES (highlighted), PILOTS, USE CASES, and STORIES. A secondary navigation bar shows: HOME, RESOURCES, SPECIFICATIONS, and SPECIFICATION FOR SIARD FORMAT V2.0. The main content area features a section titled "In This Section" with a list of links: PROJECT DELIVERABLES, E-ARK TOOLS, E-ARK GENERAL MODEL, E-ARK ARCHITECTURE, PRESENTATIONS, E-ARK PROJECT UPDATES, SPECIFICATIONS, POSTERS AND IMAGES, E-ARK ANNUAL SUMMARY REPORTS, and E-ARK BENEFITS. The central focus is the "Specification for SIARD Format v2.0 Popular" article, published on 31 July 2015 by Sharon McMeekin, with 323 downloads. A "Download (pdf, 2.02 MB)" link is provided, along with the filename "STAN_e_FINAL_2015-07-04_eCH-0165_V20_SIARD-Format.pdf". To the right, a "Latest News" section includes two items: "E-ARK Rated 'Excellent' by the EC - 'A European Showcase Project'" dated 03-04-2017, and "The European Commission's Year 3 Review of the E-ARK Project has rated it as 'Excellent' and have..." followed by "E-ARK Year 3 Summary Report Published" dated 16-03-2017. A paragraph below the second news item states: "The E-ARK Project have published their summary report to the EC on their project activities in the".



SIARD 1.0 + SIARD DK / 1007 → SIARD 2.0

What is new in SIARD 2.0?

The main new features are:

- 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
- Support for “deflate” as a compression mechanism
- 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



Pilot 1	SIP Creation of relational databases
Task leader	Danish National Archives
Supported by	Magenta
Scope	The scope of this Pilot is to test the E-ARK SIP Creation tool with not less than 4 databases of different sizes and complexities (one contains several million records)
Object	Creating SIPs for relational databases using the tool created in WP3, T3.3: SIP Creation Tools, for further evaluation
Short description	The goal of the pilot is to make four successful data extractions from live authentic databases into the SIARD 2.0 and SIARDDK format.



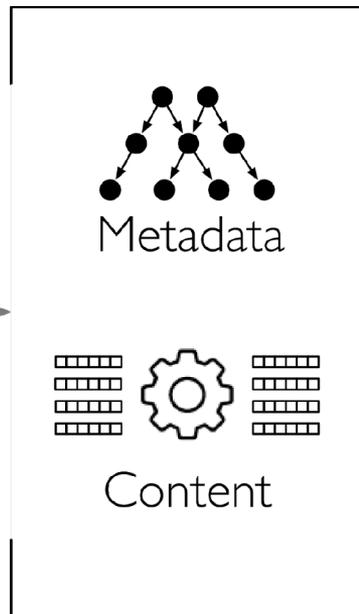
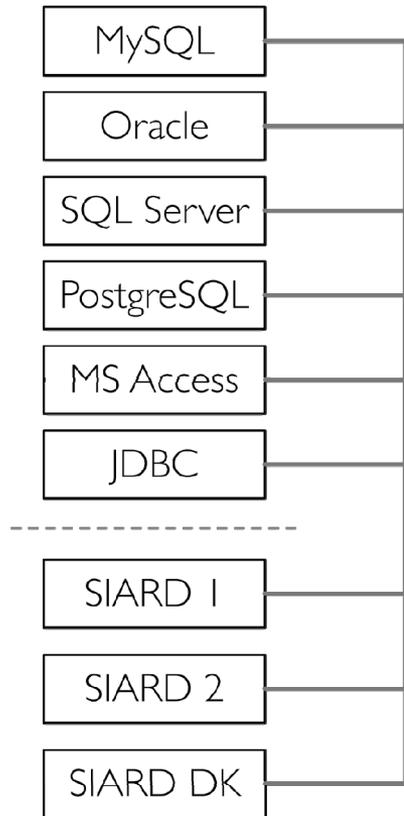


<http://www.database-preservation.com/>

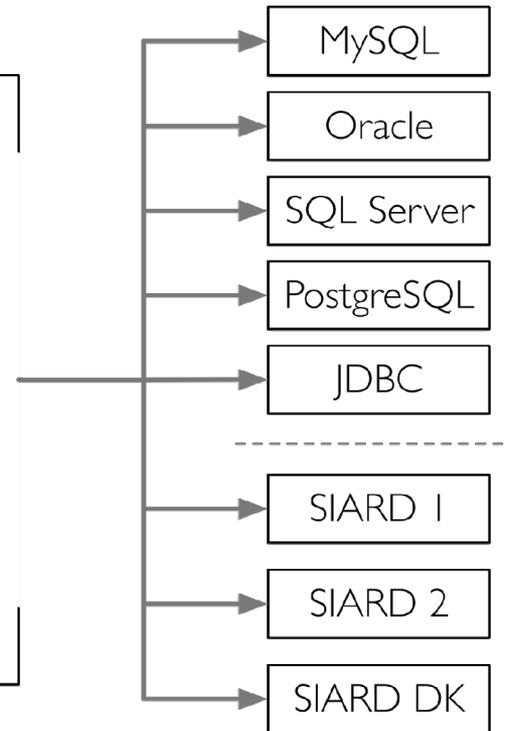


NorDig
26. oktober 2017

Import modules



Export modules



DBPTK can (via a user logon that have certain rights) attach to the relational database management system and export to SIARD format



Technical details	System
MS SQL Server 90 tables 90,000 records	Health system from the Danish Serum Institute Database containing information from reported infectious diseases at the national level. Infectious diseases for all Danish citizens.
MySQL 5 tables 33,000,000 rec.	System from a private sector business: Kultunaut Aps Harvesting and selling information about cultural events at national level from the smallest local gatherings to international exhibitions and events in Denmark.
MS SQL Server 180 tables 100,000 LOBs	Administrative and health records system from Ministry of Higher Education and Science Database containing information about social, psychological, and psychiatric counselling to students in their educational situation.
MS SQL Server 289 tables 1,300,000 LOBs	Administrative system from The Danish National Archives Database containing information about all incoming scientific research data, and public deliveries of research data.



Filters

is:issue author:PhillipTommerholt

Labels

Milestones

New issue

Clear current search query, filters, and sorts

5 Open 19 Closed

Author

Labels

Projects

Milestones

Assignee

Sort

- Beta6 does not create messageDigests in tableX.xml for BLOBs** meta: waiting for feedback  2
#295 opened on Jan 23 by PhillipTommerholt v2.1.0
- SIARDDK - include illegal datatypes** module: SIARD-DK priority: Pilot blocker type: bug  1
#290 opened on Dec 22, 2016 by PhillipTommerholt
- Cannot export SIARD-file to MS SQL Server due to naming of MySQL primary keys** module-db: MSSQLServer module-db: MySQL type: bug  2
#262 by PhillipTommerholt was closed on Feb 14 v2.1.0
- SIARDDK: documents/document folders not assigned unique numbers** module: SIARD-DK priority: Pilot blocker type: bug  3
#259 by PhillipTommerholt was closed on Jan 6 v2.1.0
- SIARDDK: wrong table namespaces when database has several schemas** module: SIARD-DK priority: Pilot blocker type: bug  4
#256 by PhillipTommerholt was closed on Jan 6 v2.1.0
- SIARDDK: fileIndex.xml only incorporates tables from first schema** module: SIARD-DK type: bug  2
#254 opened on Aug 22, 2016 by PhillipTommerholt v2.1.0
- SIARDDK can't handle bit varying** module: SIARD-DK priority: Pilot blocker  25
#250 by PhillipTommerholt was closed on Jan 6 v2.0.0-beta7
- SIARDDK handling of SQL identifiers** module: SIARD-DK priority: Pilot blocker type: bug  5
#244 by PhillipTommerholt was closed on Sep 1, 2016 v2.1.0
- SIARDDK: Schemas without tables causes fatal error** module: SIARD-DK type: bug  4
#242 by PhillipTommerholt was closed on Aug 16, 2016 v2.0.0-beta6
- SIARDDK Handling of missing primary keys** module: SIARD-DK  11
#241 by PhillipTommerholt was closed on Aug 16, 2016 v2.1.0

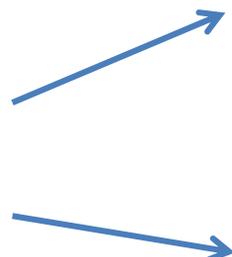




Applikationslag
- det brugeren
indtaster og
navigerer i



Databasen –
der informationerne
er gemt/ordnet



Name	SIARD Format Specification
Standard number	eCH-0165
Category	Standard
Maturity level	Tentative
Version	2.0
Status	Draft – Final E-ARK
Approval date	
Issue date	
Replaces standard	
Languages	English (Original)
Authors	Hedi Braggesser, Thurgau State Archives, hedi.braggesser@ts.ch Georg Büchler, KOST, georg.buechler@kost.admin.ch Alan Dubois, Valais State Archives, alan.dubois@admin.vs.ch Martin Kaiser, KOST, martin.kaiser@kost.admin.ch

