

Leverandørforum

18. maj 2017



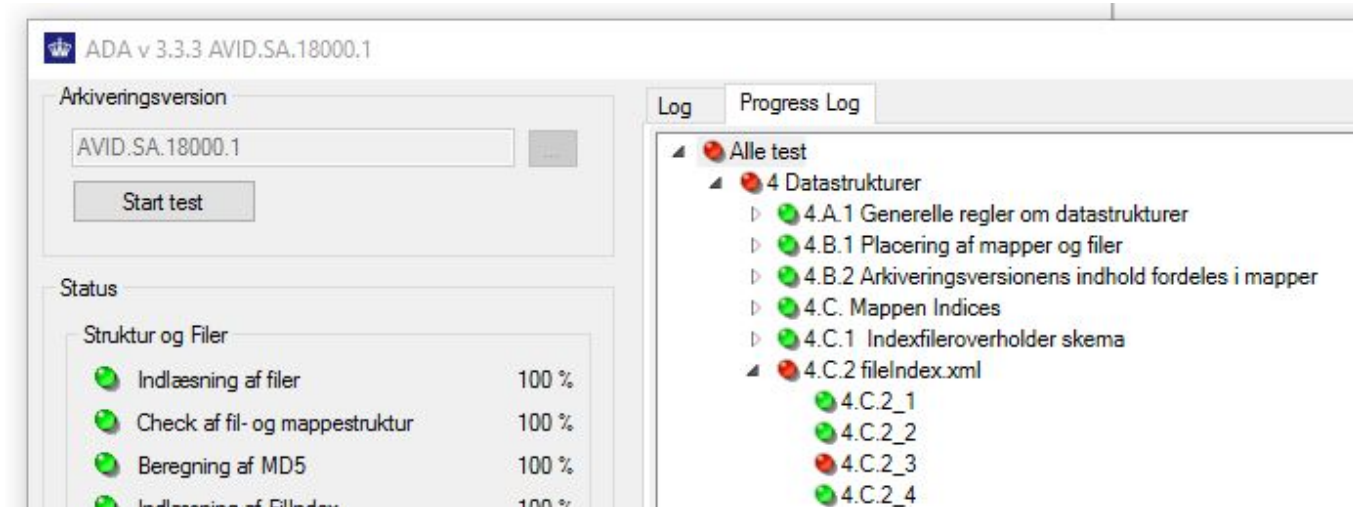
Dagsorden

1. Velkomst ~~og kort orientering~~
2. Status for det afsluttede ADA-projekt

- kaffepause -
3. Status for det afsluttede E-ARK-projekt
4. Eventuelt

ADA projektet

1. februar 2016- 15. marts 2017



The screenshot displays the ADA v 3.3.3 AVID.SA.18000.1 application window. The interface is divided into several sections:

- Arkiveringsversion:** A text box containing "AVID.SA.18000.1" and a "Start test" button.
- Status:** A section titled "Struktur og Filer" showing a list of tasks and their completion percentages:

Task	Percentage
Indlæsning af filer	100 %
Check af fil- og mappestruktur	100 %
Beregning af MD5	100 %
Indlæsning af FilIndex	100 %
- Log / Progress Log:** A tree view showing the test results:
 - Alle test (red icon)
 - 4 Datastrukturer (red icon)
 - 4.A.1 Generelle regler om datastrukturer (green icon)
 - 4.B.1 Placering af mapper og filer (green icon)
 - 4.B.2 Arkiveringsversionens indhold fordeles i mapper (green icon)
 - 4.C. Mappen Indices (green icon)
 - 4.C.1 Indexfileroverholder skema (green icon)
 - 4.C.2 fileIndex.xml (red icon)
 - 4.C.2_1 (green icon)
 - 4.C.2_2 (green icon)
 - 4.C.2_3 (red icon)
 - 4.C.2_4 (green icon)

ADA projektet

Outline

- Målene og resultaterne for projektforsløbet
- Kendte fejl og mangler - igangværende tilretninger
- Feedback fra jer

ADA projektet

Målene

- Performance
- Stabilitet
- Nyt testlogsystem
- Indførsel af manglende tests



Performance og stabilitet. Vi oplever stadig stigende datamængder, og de enkelte arkiveringsversioner bliver også større og større. Derfor er det vigtigt, at ADA generelt tester hurtigere, så vi kan få et større throughput. Det er også vigtigt at ADA ikke går ned, og hvis det sker, så skal alt ikke være tabt. Vi har før oplevet at flere dages maskintid er spildt fordi ADA har haft et nedbrud.

I forhold til ny testlogsystem, så har der tidligere været tvivl om, også fra vores egen side, hvad ADA testede og ikke testede for. Målet var, at enhver meddelelse som ADA kom ud med kunne henføres til et unikt id, således at alle ADAs deltests er samlet i et register, og med henvisning til bekendtgørelsens bestemmelser.

ADA projektet

Forberedelse

- Generel kravspecifikation
- Beskrivelse af alle tests (ADA dokumentation)
- Testsuite

ADA projektet

Hvad udgør en ADA-test?

Dokumentation af hvad ADA (version 3.3.3) tester for.

<https://www.sa.dk/aflevering-arkivet/statslige/ada-testprogram-arkiveringsversioner>

ADA projektet

Oversigt over testområder

	Ordlyd i bekendtgørelsen	Testområde
Bilag 4		
4.A	Generelle regler om datastruktur	
4.A.1	I en arkiveringsversion skal alle nøgler have en entydig identifikator. Der må ikke forekomme situationer, hvor det er nødvendigt at uddrage dele af nøglefelter for at forstå it-systemets indhold eller funktion.	4.A.1
4.B.	Placering af mapper og filer	
4.B.1	I roden af filsystemet på afleveringsmediet, jf. bilag 7, skal der være placeret en mappe navngivet med mediets navn. Mediets navn består af det unikke arkiveringsversionsID med tilføjelse af et suffiks »n«, der angiver rækkefølgen af medierne, hvor n er et fortløbende medieløbenummer begyndende med 1.	4.B.1
4.B.2	Arkiveringsversionens indhold fordeles i mapper, som angivet i figur 4.1.	4.B.2
4.B.3	Mapperne skal navngives som angivet i figur 4.1.	
4.B.4.a	Et arkiveringsversionsID består af præfikset AVID, en kode på 2-4 bogstaver, som angiver det modtagende arkiv, samt et arkiveringsversionsløbenummer. Elementerne adskilles med punktum.	
4.B.5.a	En arkiveringsversion, som ikke kan være på ét medie, jf. bilag 7, kan fordeles over flere medier.	
4.B.5.b	Mapperne <i>ContextDocumentation</i> , <i>Indices</i> og <i>Schemas</i> skal altid være placeret på afleveringens første medie.	
4.B.5.c	Mappen i roden på efterfølgende medier skal kun indeholde de mapper, hvis indhold kræver fordeling over flere medier. Disse mapper navngives altid som angivet i figur 4.1 uden brug af suffiks.	
4.B.6	Der må kun være én af hver af de i figur 4.1 angivne mapper på hvert medie.	
4.C	Mappen <i>Indices</i>	

ADA projektet

Testområde 4.C.6 (Mappen *Indices* – *docIndex.xml*)

4.C.6.a	docIndex.xml skal danne forbindelsen mellem hvert dokument og dets placering. »docIndex.xml« skal desuden indeholde oplysninger om dokumenternes oprindelige filnavne, filtype i arkiveringsversionen samt eventuelle overordnede dokumenter. »docIndex.xml« skal ikke indeholde oplysninger om dokumenterne i kontekstdokumentationen.
4.C.6.b	For hvert enkelt dokument i docIndex.xml angives de oplysninger, som fremgår af figur 4.4.

Testformulering:

Testnr.	Testnavn (internt referencenavn)	Testforløb beskrivelse	Implementeret
4.C.6_1	Dokumenttest, parentID	Der testes for, om der findes opmærkninger i docIndex.xml i elementet <pID>	V
4.C.6_2	Dokumenttest, GML	Hvis arkiveringsversionen har opmærket ekstension ".gml" i oFn samt gml i aFt testes der for, om opmærkningen i gmlXsd er udført korrekt.	V
4.C.6_3	Dokumenttest, parentID sammenhæng	Der testes for, om en docID opmærket i elementet <pID> har en overordnet reference til en docID opmærket i <dID>	V
4.C.6_4	Dokumenttest, parentID rekursiv sammenhæng	Der testes for, om alle "led" i indlejrede filers filer, har en endelig og afsluttende reference til en <dID>	V
4.C.6_5	Dokumentindekstest, refererer til tabeldata	Der testes for, om alle dokumenter i dokumentindekset refererer til tabeldata opmærket med functional description "Dokumentidentifikation".	V
4.C.6_6	Dokumentindekstest, dokumenter refereret i tabeldata	Der testes for, om alle dokumenter i tabeldata opmærket med functional description "Dokumentidentifikation" refererer til dokumenter i dokumentindekset.	V

ADA projektet

Deltest 4.C.6_5

Der testes for om alle dokumenter i dokumentindekset refererer til tabeldata opmærket med functional description
”Dokumentidentifikation”.

 AVID.SA.17171.1

Der 4 dokumenter i docIndex.xml, som ikke findes i tabeldata (doc 20, 21, 22 og 23)

ADA projektet

Resultaterne

- **Performance.** ADA tester cirka 10% hurtigere i vores miljø på eksempelarkiveringsversioner, filgennemløb sat til et minimum.
- **Stabilitet.** Der er kommet langt mere styr på afhængigheder - hvilke informationer skal være til stede før at en deltest kan køres. Alle problem-arkiveringsversioner i testsuiten kan køres igennem. Man kan teste videre fra sidst vellykkede test.
- **Nyt testlogsystem.** Alle fejlmeddelelser er bygget op på ny testlog. Samlet register over deltests, dokumentation ved hver release. Videre udvikling sikrer at deltests stadig tester efter hensigten.
- **Indførsel af manglende tests.** Kun ganske få er ikke indført.

ADA projektet

Resultaterne

- Progress log - alternativ fejllog.
- Tabelviewer - forbedret forespørgselsinterface.
- Tabelviewer - søgning i tabelindeks.
- Tabelviewer - E/R-diagram.

ADA projektet

8 fejl og mangler

1. 4.D_12 ADA skal komme med bedre lokalisation af foran- og efterstillede blanktegn og disse laves som hint.
2. Specifikke dokumenter i AVID får ADA til at smide "uventet fejl" for disse dokumenter. Sagsbehandlingen er stoppet pga. ADA-fejl.
3. Sagsbehandlerdefinerede SQL-forespørgsler i arkiveringsversioner skal tillade, at der dannes midlertidigt navngivne tabeller, selvom disse ikke "findes" i arkiveringsversionen.
4. 4.C.6_5 ADA skal tillade, at dokumenter i arkiveringsversionen ikke behøver at være repræsenteret direkte i tabeldata – det er ok, at det kun er muligt via pID-opmærkning i docIndex.xml at linke dette sammen.

ADA projektet

8 fejl og mangler

5. ADA skal omskrives i 6.C_25 så det er til at finde ud af, hvad fejlen er (maxlængde angivet i tableIndex.xml er overskredet i tabeldata) og testeren skal have at vide hvad den største længde er i tabeldata.
6. ADA skal fortælle, hvis der er GML i afleveringen.
7. Ændring af ordlyd i 4.A.1_11
8. 6.C_28 – status på nogle tests vises ikke korrekt

ADA projektet

8 fejl og mangler

Flere?

ADA projektet

Pause

E-ARK projektet

1. februar 2014 - 31. januar 2017



E-ARK præsentation
Harsdorfsalen
18. maj 2017



E-ARK projektet

1. februar 2014 - 31. januar 2017



Outline

- Ganske kort om projektet
- Ændret mappestruktur for arkiveringsversioner
- Bevaringsformat for databaser - SIARD 2.0
- Konkret fremvisning af udtræksværktøj
- Mangler og funktioner ved udtræksværktøj
- Umiddelbar feedback på udtræksværktøj

E-ARK præsentation
Harsdorfsalen
18. maj 2017



E-ARK projektets deltagere

Arkiver

-  Archives of the Republic of Slovenia
-  The Danish National Archives
-  The National Archives of Hungary
-  The National Archives of Estonia
-  National Archival Services of Norway

IT-firmaer

-  ES Solutions
-  Magenta
-  KEEP Solutions

Forskningsinstitutioner

-  University of Brighton
-  Austrian Institute of Technology GmbH
-  Universität zu Köln: The University of Cologne
-  University of Portsmouth
-  Lisbon Technical University

Andre

-  The DLM Forum Foundation
-  The Digital Preservation Coalition
-  Agency for Public Services Reform, IP
-  The Ministry of Finances and Public Administrations

E-ARK præsentation
Harsdorfsalen
18. maj 2017



E-ARK projektets finansiering



50 / 50



E-ARK is co-funded by the European Commission under its ICT Policy Support Programme (PSP) within its Competitiveness and Innovation Framework Programme (CIP)

E-ARK præsentation
Harsdorfsalen
18. maj 2017



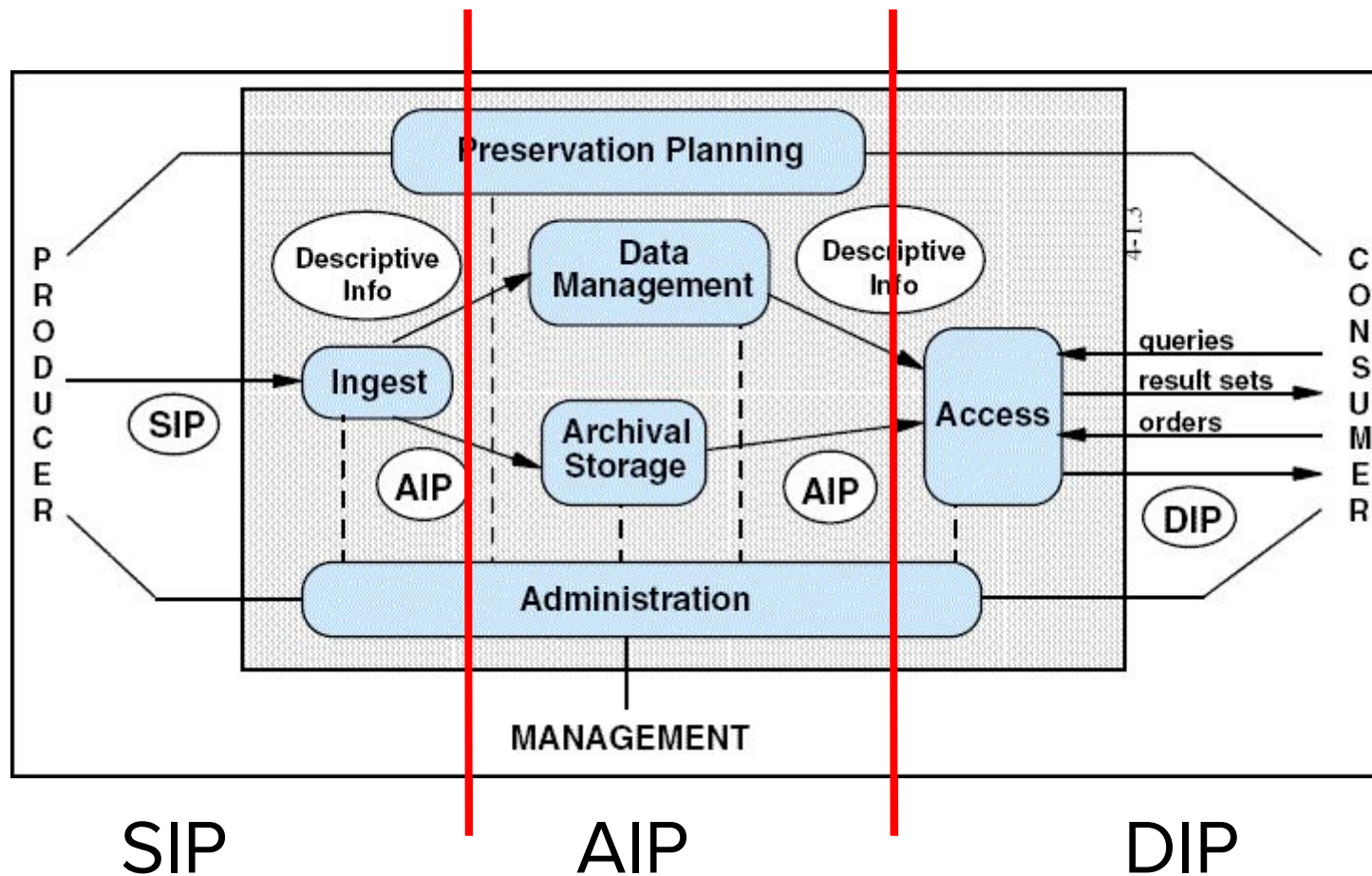
OAIS Open Archival Information System



E-ARK præsentation
Harsdorfsalen
18. maj 2017



OAIS Open Archival Information System



E-ARK præsentation
Harsdorfsalen
18. maj 2017



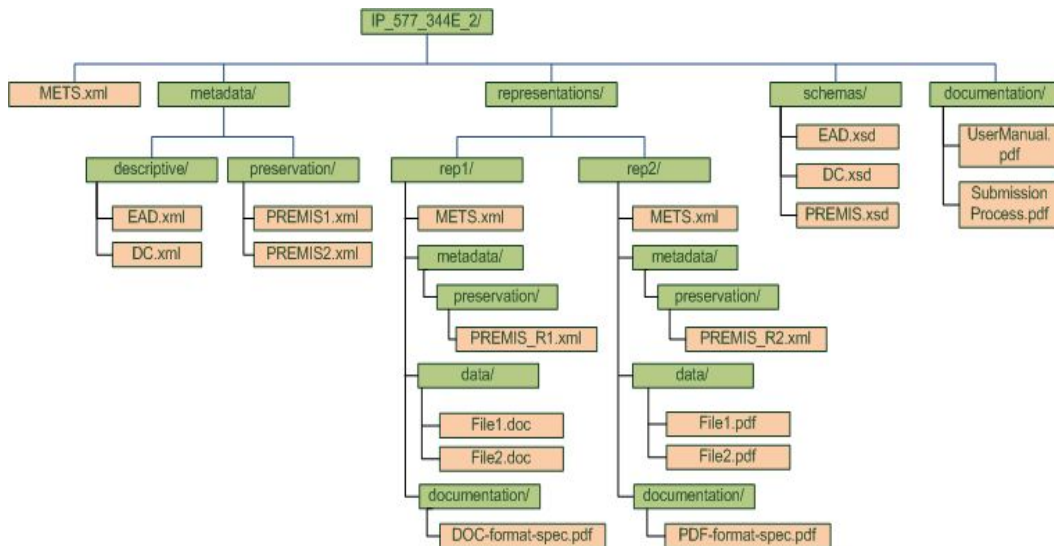
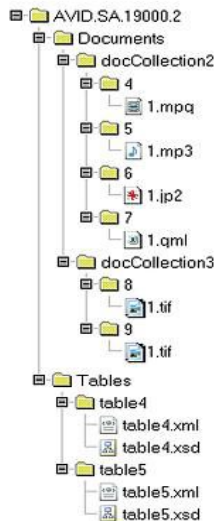
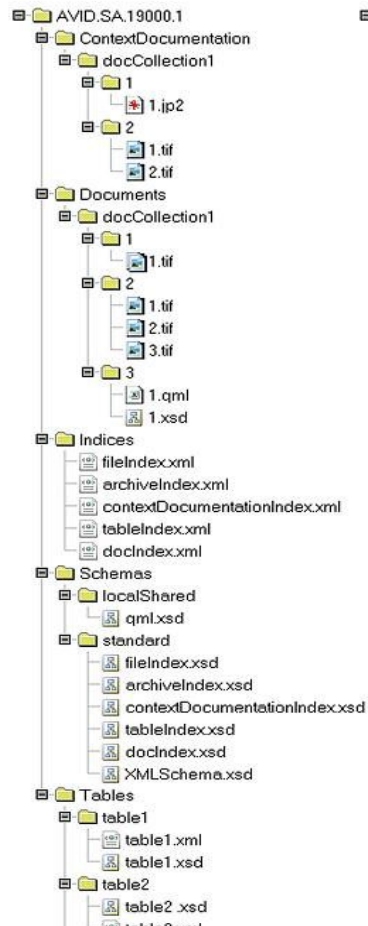
Common Specification - den vigtigste specifikation fra E-ARK



E-ARK præsentation
Harsdorfsalen
18. maj 2017



Common Specification - den vigtigste specifikation fra E-ARK



E-ARK præsentation
Harsdorfsalen
18. maj 2017



Common Specification - den vigtigste specifikation fra E-ARK

METS.xml <http://www.loc.gov/standards/mets/>

PREMIS.xml <https://www.loc.gov/standards/premis/index.html>

EAD.xml <https://www.loc.gov/ead/>



E-ARK præsentation
Harsdorfsalen
18. maj 2017

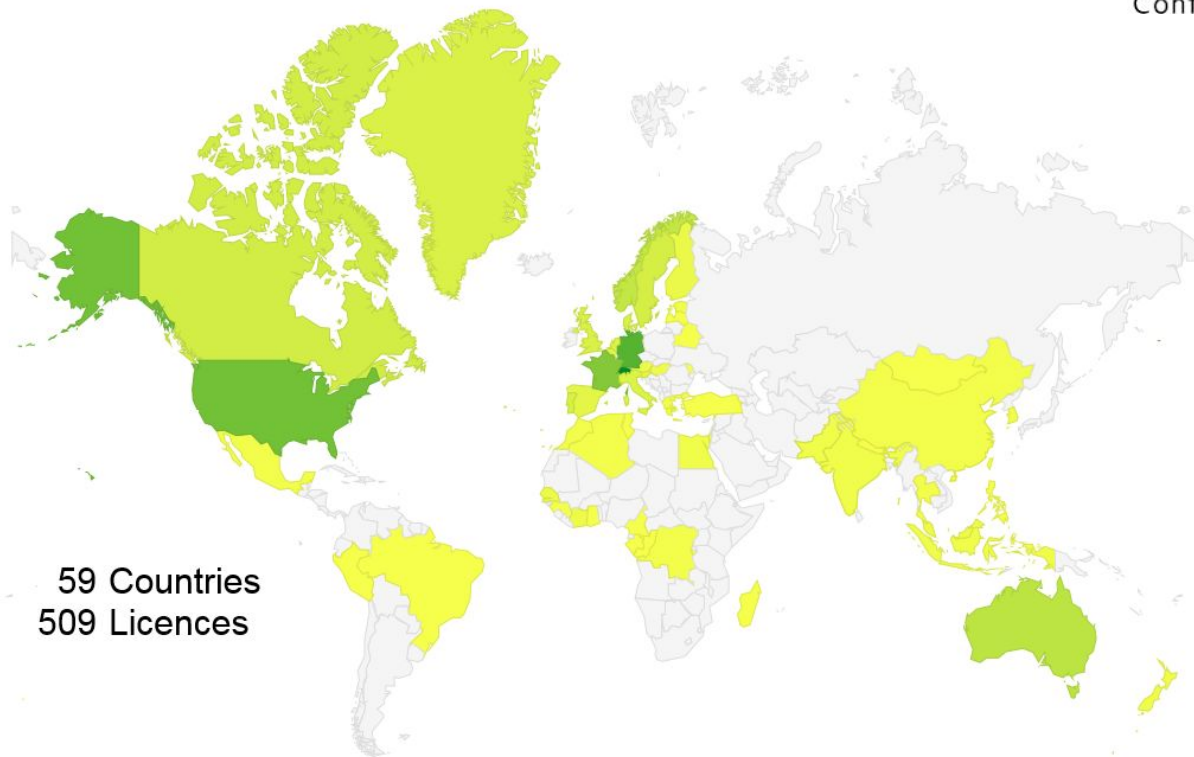


SIARD 1.0

Software Independent Archiving of Relational
Databases



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



59 Countries
509 Licences

N° Licences



Up to 01/01/2016

E-ARK præsentation
Harsdorfsalen
18. maj 2017





Lovtidende A

2010

Udgivet den 24. august 2010

20. august 2010.

Nr. 1007.

Bekendtgørelse om arkiveringsversioner

I medfør af § 5, stk. 1, § 13, stk. 1, og § 14 i bekendtgørelse nr. 591 af 26. juni 2003 om offentlige arkiver og offentlige arkivers virksomhed og efter drøftelse med de kommunale og regionale parter fastsættes:

Område

§ 1. Bestemmelserne i denne bekendtgørelse gælder for arkivalier skabt af den offentlige forvaltning og domstolene, og som af Statens Arkiver er bestemt til bevaring.

§ 2. Bevaring af data fra it-systemer og af lyd og video skal ske i form af arkiveringsversioner.

Stk. 2. Statens Arkiver kan af bevaringsmæssige grunde fastsætte, at bevaringen af andre arkivalier skal ske i form af arkiveringsversioner.

§ 3. En arkiveringsversion af bevaringsværdige data skal fremstilles efter de anvisninger, der udstedes af rigsarkivaren, jf. bilag 2-8.

Stk. 2. Rigsarkivaren kan anvisne anden bevaringsform end en arkiveringsversion, fremstillet efter anvisningerne i bilag 2-8, hvis bevaringsmæssige hensyn tilsiger det.

Produktion og aflevering

§ 4. Produktion og aflevering af arkiveringsversioner af data fra den statlige forvaltning og domstolenes it-systemer skal finde sted på tidspunkter, der fastsættes af Statens Arkiver.

§ 5. Produktion og aflevering af arkiveringsversioner af data fra kommunernes og regionernes it-systemer, som indeholder personoplysninger, skal finde sted, inden data skal slettes. Den pågældende kommune eller region kan træffe aftale om tidligere tidspunkter med det modtagende arkiv.

Stk. 2. Produktion af arkiveringsversioner af data fra øvrige it-systemer skal finde sted, inden data slettes, eller når it-systemet tages ud af brug.

Stk. 3. Statens Arkiver kan derudover fastsætte, at der skal produceres en arkiveringsversion af data fra et it-system, når det af bevaringsmæssige hensyn er nødvendigt.

§ 6. Arkiveringsversioner af statlige myndigheders data skal godkendes af Statens Arkiver. Arkiveringsversioner af kommunale og regionale myndigheders data skal godkendes af det modtagende arkiv. Arkiveringsversioner af bevaringsværdige kommunale og regionale data, som ikke er omfattet af afleveringspligt, skal godkendes af samme offentlige arkiv, som modtager myndighedens afleveringspligtige arkiveringsversioner.

Stk. 2. Data, som er overført til en arkiveringsversion, må ikke slettes hos myndigheden, for arkiveringsversionen er godkendt.

Ikrafttrædelse m.v.

§ 7. Bekendtgørelsen træder i kraft den 1. september 2010.

Stk. 2. Bekendtgørelse nr. 342 af 11. marts 2004 om arkiveringsversioner af bevaringsværdige data fra elektroniske arkivsystemer ophæves.

Stk. 3. Bekendtgørelse nr. 302 af 16. april 2009 om aflevering af lyd og video ophæves.

Stk. 4. Til og med 1. juli 2011 kan der indgås aftale med offentligt arkiv om, at arkiveringsversioner af data udarbejdes efter reglerne i den i stk. 2 nævnte bekendtgørelse.

Stk. 5. Arkiveringsversioner af data efter reglerne i den i stk. 2 nævnte bekendtgørelse skal være afleveret til det pågældende offentlige arkiv inden 31. december 2011.

Stk. 6. Rigsarkivaren kan efter ansøgning meddele dispensation for reglerne i stk. 4 og 5.

= SIARDDK

Rigsarkivet, den 20. august 2010

ASBJØRN HELLUM

/ Kirsten Villadsen Kristmar

E-ARK præsentation
Harsdorfsalen
18. maj 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.

E-ARK præsentation
Harsdorfsalen
18. maj 2017



SIARD 1.0 + SIARD DK / 1007 → SIARD 2.0

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



A European "Showcase" Project

Search ...

[Site Map](#) [Register](#) [Login](#) [SHARE](#)

HOME ABOUT NEWS EVENTS COMMUNITY PARTNERS **RESOURCES** PILOTS USE CASES STORIES

HOME RESOURCES SPECIFICATIONS SPECIFICATION FOR SIARD FORMAT V2.0

In This Section

- 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
- E-ARK BENEFITS

Specification for SIARD Format v2.0 Popular

Published on 31 July 2015 • By Sharon McMeekin • 323 downloads

Specification for SIARD Format v2.0 [Download \(pdf, 2.02 MB\)](#)

STAN_e_FINAL_2015-07-04_eCH-0165_V20_SIARD-Format.pdf

Latest News

E-ARK Rated "Excellent" by the EC - "A European Showcase Project" 03-04-2017

The European Commission's Year 3 Review of the E-ARK Project has rated it as "Excellent" and have...

E-ARK Year 3 Summary Report Published 16-03-2017

The E-ARK Project have published their summary report to the EC on their project activities in the

E-ARK præsentation
Harsdorfsalen
18. maj 2017



Udtræksværktøj



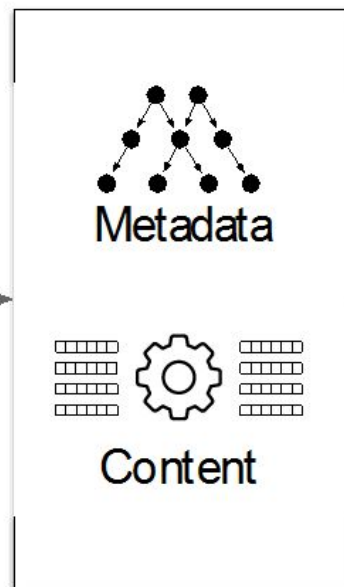
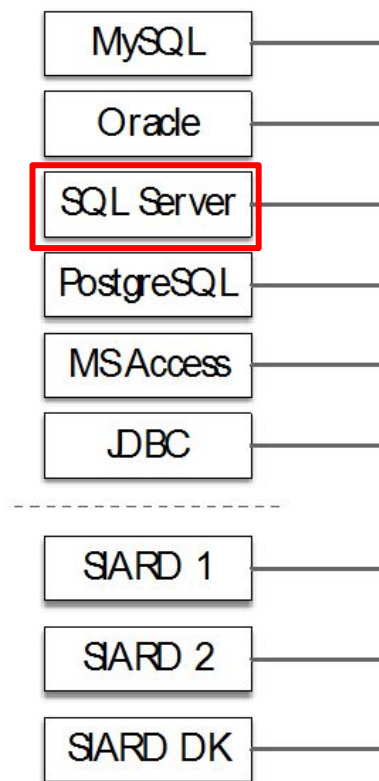
<http://www.database-preservation.com/>

E-ARK præsentation
Harsdorfsalen
18. maj 2017

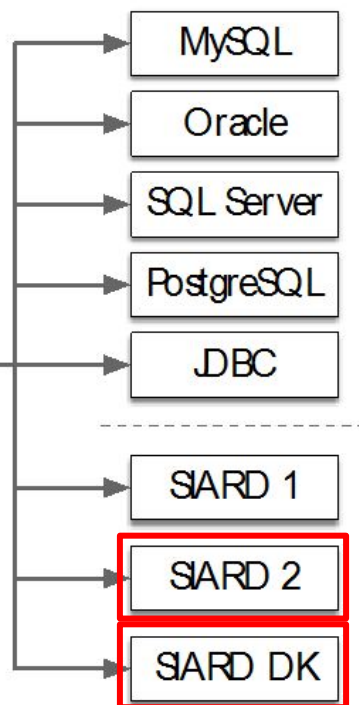


Udtræksværktøj

Import modules



Export modules



E-ARK præsentation
Harsdorfsalen
18. maj 2017



Fejl, mangler, funktioner, setting, etc.

- Trækker kun BLOBs/CLOBs (dokumenter etc) ud som .bin filer (derfor ingen filgenkendelse eller tiff-konvertering)
- Trækker kun det ud, som reelt findes i selve databasen (tabelbeskrivelser, kolonnebeskrivelser, forklaring af koder)
- Antagelse: alt data findes i en database (dokumenter uden for databasen skaber problemer med docIndex.xml)
- Man kan vælge hvilke tabeller, der skal med.
- Databaseadministratorer vil kunne bruge værktøjet.



Referencer - billeder

Welcome, slide 1	https://www.flickr.com/photos/90371939@N00/4344878104/in/photolist-7BWCTs-HHgKu-8UGZ7y-5JpSK9-bdQr9k-9gyKP4-8oRu6t-RSKsn5-M6Tip-qHKb6o-gjUxfC-9zCFMc-mbxz1R-bqSPUM-4PmjCC-iCoWHY-5hjWdy-cWP5z-enUzS1-9ynKoP-4ZUdep-JWQoH-oHgXdt-4qqi7k-4VtAGK-6q51Fg-cdJwd5-6RwtFw-6oTcWo-gWs4Q4-Nd5Nb-gc3RFD-h4tNb2-aFte4V-mBGdZG-srNFRu-iU7Ca-bs3mg5-7bgJh5-6wow15-qsi9pg-2g2M2-buxJDs-8eawNK-55jzxb-5jXpQS-9PsLAL-eknXEQ-9t4aDF-5FTHK9/
Slide 16, gruppebillede	http://www.eark-project.com/
Slide 9, split 50/50	https://pixabay.com/da/cupcake-v%C3%A6rtinde-split-fl%C3%B8de-640276/
Slide 20, NASA	https://www.nasa.gov/
Slide 20, CDCCS	www.ccsds.org/
Slide 24, Library of Congress	www.loc.gov
Slide 25, SFA SIARD Suite License	https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html
Slide 30, Udtræksværktøj	KEEPS https://www.keep.pt/

E-ARK præsentation
Harsdorfsalen
18. maj 2017



Referencer - relevante hjemmesider

1. Videreførsel af specifikation og oversigt over værktøjer til specifikationer: www.dashboard.eu
2. E-ARK projektsite: <http://www.eark-project.com/>
3. SA hjemmeside: <https://www.sa.dk/k/aflevering-arkivet/it-leverandoerer>
4. ADA: <https://www.sa.dk/aflevering-arkivet/it-leverandoerer/ada-testprogram-arkiveringsversioner>
5. OAIS: https://en.wikipedia.org/wiki/Open_Archival_Information_System
6. METS.xml <http://www.loc.gov/standards/mets/>
7. PREMIS.xml <https://www.loc.gov/standards/premis/index.html>
8. EAD.xml <https://www.loc.gov/ead/>
9. Udtræksværktøj <http://www.database-preservation.com/>

E-ARK præsentation
Harsdorfsalen
18. maj 2017

