



Forvaltning av SIARD og E-ARK spesifikasjoner

KDRS samling, 5.-6. juni 2024

Arne-Kristian Groven

Arkivverket

arngro@arkivverket.no



Litt om meg selv

- ▶ Jeg er IKT-utdannet ved Institutt for Informatikk, Universitetet i Oslo
- ▶ Har drevet med innovasjon og FoU innen (anvendt) IKT i mange år
- ▶ Har vært ansatt i Arkivverket (Riksarkivet) siden 2011 hvor jeg også har arbeidet med FoU og innovasjon
- ▶ Har også vært tilknyttet OsloMet i perioder, som ansvarlig for kursene:
 - ▶ ARK2000, Digital Depot, ved Institutt for arkiv -, bibliotek og informasjonsfag.
 - ▶ ADSE2400, Informasjonsarkitektur, ved Institutt for informasjonsteknologi.
- ▶ Var en av de første i Norge benytte SIARD-metodikken for databasebevaring i et prosjekt i 2012/2013



“


E-ARK prosjektene:
Europeiske samarbeids-
prosjekt rundt digital
langtidsbevaring

”

Fra 2014...

Historien om E-ARK prosjektene

- Første prosjekt ble gjennomført i perioden 2014-2017, deretter fulgte flere
- Arkiverket var med fra starten sammen med nasjonalarkivene i Danmark, Estland, Slovenia, Ungarn, Sveits og etter hvert også Finland
- I tillegg til europeiske forskningsmiljøer, arkivnettverk, offentlige institusjoner og noen mindre programvare selskap
- Her ble det gjennomført kartlegginger og studier, OAIS spesifikasjoner ble produsert, SIARD 2.0 ble spesifisert og programvare prototyper ble utviklet



NOTICE: Historic website, not updated.

Search ...

E-ARK A European "Showcase" Project Site Map

HOME ABOUT NEWS EVENTS COMMUNITY PARTNERS RESOURCES PILOTS USE CASES STORIES

E-ARK FINAL CONFERENCE

In co-operation with the National Archives of Hungary, E-ARK held its closing conference in beautiful Budapest.

[Find out more](#)

NOTICE: This page is now static and will not have further updates. It's archived, available only for reference.

Welcome to the E-ARK Project

(European Archival Records and Knowledge Preservation)

E-ARK was a multinational CIP FP7 pilot project that improved 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 benefited 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 E-ARK was co-funded by the European Commission under its ICT Policy Support Programme (PSP) within its Competitiveness and Innovation Framework Programme (CIP).

@EARKProject on Twitter

For the latest EARK news follow us on Twitter [@EARKProject](#)

[Follow @EARKProject](#)

E-ARK Partners

REPUBLIC OF ESTONIA NATIONAL ARCHIVES

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 summary report of project activities in the final year of the project can be downloaded by...

Fram til i dag...

<https://www.e-ark-foundation.eu/>



E-ARK Foundation

Consortium members

Events

Webinars

E-ARK Foundation

The **E-ARK Foundation** is a **consortium of partners** working with the European Commission as part of the **eArchiving Initiative** to provide core specifications, software, training and knowledge with the aim to promote the interoperability of digital archives in Europe and to help organisations and people to preserve information for the long term.

Running from 2014 to 2017 and co-funded by the European Commission E-ARK was originally a multinational big data research project that improved the methods and technologies of digital archiving, in order to achieve consistency on a Europe-wide scale.



After the E-ARK project ended, a series of projects were undertaken to first become a Building Block of the Core Services Platform of the European Commission's Connecting Europe Facility, and now to establish the eArchiving Common Services Platform (eArchiving CSP) under the **European Commission's Digital Europe programme**.



E-ARK arrangemententer

Events

Future events

The eArchiving Initiative and DLM Geoforum announce the “Preservation of Digital Geospatial Records, the Cornerstone of Europe’s Digital Decade” workshop will take place on Tuesday May 28 2024 in Brussels.

- [Workshop: Preservation of Digital Geospatial Records, the Cornerstone of Europe’s Digital Decade](#)

The eArchiving Initiative announces a new event in Brussels on Wednesday 29 May 2024 to celebrate the tenth anniversary of the first E-ARK project funded by the European Commission and discuss the milestones over these ten years and the challenges of electronic archiving and digital preservation for the next decade.

- [Event 10 years of eArchiving: Reflections and insights for the next decade, 29 May 2024, Brussels](#)

Past events

Since the start of the eArchiving Initiative there was the eArchiving Initiative Summit in October 2023 in Salamanca, Spain; the online event about how the eArchiving Initiative can help in the implementation of eIDAS electronic archiving trust services in September 2023; and the eArchiving Initiative 2023-2024 online event.

- [eArchiving Initiative Summit – October 2023 in Salamanca, Spain](#) | October 25, 2023
- [How the eArchiving Initiative can help in the implementation of eIDAS electronic archiving trust services – September 2023](#) | September 22, 2023
- [The eArchiving Initiative 2023-2024](#) | April 25, 2023



E-ARK webinar



Webinars

2024

- [The eArchiving Reference Architecture](#) | March 21, 2024 | Anna Sancho
presenter : István Alföldi
- [The eArchiving validation framework: its rationale and how it is linked to validating against specifications](#) | February 15, 2024 | Diogo Proença
presenter : Sven Schlarb
- [The long-term accessibility of Cancer Registry data](#) | January 18, 2024 | Anna Sancho
presenter : Vesna Zadnik, Joanna Bartnicka, Jože Škofljanec, Tina Žagar

2023

- [Using eArchiving for FAIR access to long-term geospatial records in the digital economy](#) | October 24, 2023 | Diogo Proença
presenter : Gregor Završnik
- [Using open-source tools for eArchiving](#) | September 25, 2023 | Diogo Proença
presenter : Anssi Jääskeläinen
- [Creating E-ARK conformant Archivematica preservation workflows](#) | July 20, 2023 | Diogo Proença
presenter : Stephen Mackey
- [The Large object handling capabilities of SIARD](#) | June 15, 2023 | Diogo Proença
presenter : István Alföldi
- [Distributed Digital Preservation in practice](#) | May 18, 2023 | Diogo Proença
presenter : Luís Faria
- [Using eArchiving standards for system-independent storage of authorisation metadata](#) | April 27, 2023 | Diogo Proença
presenter : Sven Schlarb



E-ARK webinar

<https://youtu.be/VMDV0yqnsA4>

The Large object handling capabilities of SIARD

June 15, 2023 by [Diogo Proença](#)

June 2023 - Webinar

Learn about SIARD's large object handling (LOB, BLOB) capabilities in the third eArchiving Initiative training webinar.

The third eArchiving Initiative training webinar examines the large object (LOB) handling capabilities of SIARD (Software Independent Archiving of Relational Databases). Large objects are generally binary data in a database. Binary data is mainly referred to as a binary large object (BLOB), and large character-based data are named character large objects (CLOB). From version 2.0, the SIARD specification supports three possible archival scenarios for storing LOBs:

1. Internal BLOBs – BLOBs contained in the database records
2. External BLOBs – BLOBs stored outside the database records but within the SIARD file
3. External BLOBs – BLOBs referenced from SIARD but stored as files outside the SIARD file

This tutorial is based on the use case of the SQL database archiving pilot by the Hungarian Digital Heritage National Laboratory and the Digital Humanities Department of Eötvös Loránd University, Budapest. Three research databases about XVIII–XIX century Hungarian novels, popular publications and literary correspondence have been preserved as SIARD files. The pilot uses several E-ARK/eArchiving specifications such as SIARD, E-ARK SIP, and CITS SIARD, as well as the Database Preservation Toolkit, RODA-in and RODA software components.



E-ARK webinar

<https://youtu.be/4NZxrYMZ28g>

Creating E-ARK conformant Archivemata preservation workflows

July 20, 2023 by [Diogo Proença](#)

July 2023 - Webinar

The fourth training webinar examines how E-ARK conformant workflows (AIP, SIP, DIP) can be created and used with the open-source digital preservation platform Archivemata.

The fourth eArchiving Initiative training webinar examines how E-ARK conformant Archivemata preservation workflows can be created. The webinar explains what Archivemata is, who can benefit from implementing it, and what are its strengths and weaknesses. It looks at the open-source toolkit developed by the projects and how they work with the unaltered open-source Archivemata distribution and another related software product 'Enduro' to produce an E-ARK workflow. A demonstration shows the workflow in action from a technical perspective and discuss how it can be integrated into systems with SIP creation tools and storage systems or repositories.

The webinar is based on two recently completed Generic Services projects funded under the Connecting Europe Facility: OneClick and the Engineering Science Library. Both projects used E-ARK conformant workflows (SIP, AIP and DIP), which included the popular open-source digital preservation platform Archivemata.

Finally, the webinar considers recent developments, including a new lightweight Archivemata preservation engine from Artefactual Systems called A3M. This fully compatible digital preservation workflow engine is lightweight, scalable and ideal for applications like the E-ARK toolkit. We show how A3M could be integrated into the workflow in place of Archivemata and demonstrate a platform that already has A3M integrated within an E-ARK conformant workflow.



E-ARK/DILCIS Board

- Spredning av kunnskap om eArchiving
 - Arrangement
 - Webinar
 - Kurs osv.
- Vedlikehold og videreutvikling av spesifikasjoner og standarder
 - SIARD
 - CSIP (Common), SIP, AIP og DIP
 - CITS (Content) Geo, SIARD, Health, ERMS, PREMIS
 - Beslutninger rundt endring av spesifikasjoner tas av DILCIS Board



“

DILCIS Board: Digital Information LifeCycle Interoperability Standards Board

”

Ansvarlig for

SIARD (Software Independent Archiving of Relational Databases)

E-ARK OAIS spesifikasjoner

Home



We are the Digital Information LifeCycle Interoperability Standards Board (DILCIS Board).
We develop, publish and support standards which provide practical interoperability in digital archiving.

Common Specification

SIP

AIP

DIP

CS ERMS

CS Geospatial Data

SIARD

Latest news

[DILCIS Board in cooperation with E-ARK CSP publishes updated Common Specifications for the Information Packages and E-ARK SIP/AIP/DIP, Content Information Type Specifications \(CITS\) and guidelines](#)

2024-05-17 14:37:56

We are pleased to announce the publication of the Common Specifications for the Information...

[Read More ...](#)

[DILCIS eHealth1 specification review spring 2024](#)

2024-02-13 07:44:54

We are happy to announce the opening of the spring 2024 review of the eArchiving specification and...

[Read More ...](#)

[Welcomes and Goodbyes in the DILCIS Board](#)

2023-12-04 14:15:58

The DILCIS Board have some happy news and some sorry news. Our Board member Audun Lund will be...

[Read More ...](#)

[DILCIS Board says welcome to Stephen Mackey](#)

2023-06-12 08:02:07

The DILCIS Board want to say welcome to Stephen Mackey, Bonavero Limited

eArchiving Standards & Specifications

All transfers of information need to be built upon the use of standards and their transfer formats in order to secure reliable information storage and subsequent re-use. eArchiving is providing a common set of specifications for packaging digital information to meet this need. The specifications were originally developed in the [E-ARK project](#) and are maintained by the Digital Information LifeCycle Interoperability Standards, [DILCIS Board](#)

The eArchiving specifications are based on common, international standards for transmitting, describing and preserving digital data. The main standard is the Reference model for an Open Archival Information System (OAIS) ([OAIS Reference model](#)) which has Information Packages as its basis. The main standard for transmitting Information Packages is the Metadata Encoding and Transmission Standard ([METS](#)), and the main standard for preserving Information Packages is Preservation Metadata Implementation Strategies ([PREMIS](#)).

At the heart of the eArchiving specifications is the Common Specification for Information Packages, which delivers a basic core specification providing a necessary minimum for institutions across Europe to securely package their data, and then customise their data as required. The Submission Information Package (SIP) is for sending material to an archive, the Archival Information Package (AIP) is for storing material in an archive, and the Dissemination Information Package (DIP) is for accessing material from an archive. Use of these eArchiving specifications provides interoperability across borders; types of institution; and user communities. The specifications for Information Packages also currently covers Content Information types, a group of Content Information Types will be expanded to meet the needs of user communities!

who has accepted to joined the...

[Read More ...](#)

[DILCIS Board says thank you to Kuldar Aas](#)

2022-12-02 08:02:07

The DILCIS Board is want to say thank you to Kuldar Aas who has left the Board during 2022. We...

[Read More ...](#)

[E-ARK publishes updated Common Specifications for the Information Packages and E-ARK SIP/AIP/DIP, one new procedure and update to the eHealth2 guideline](#)

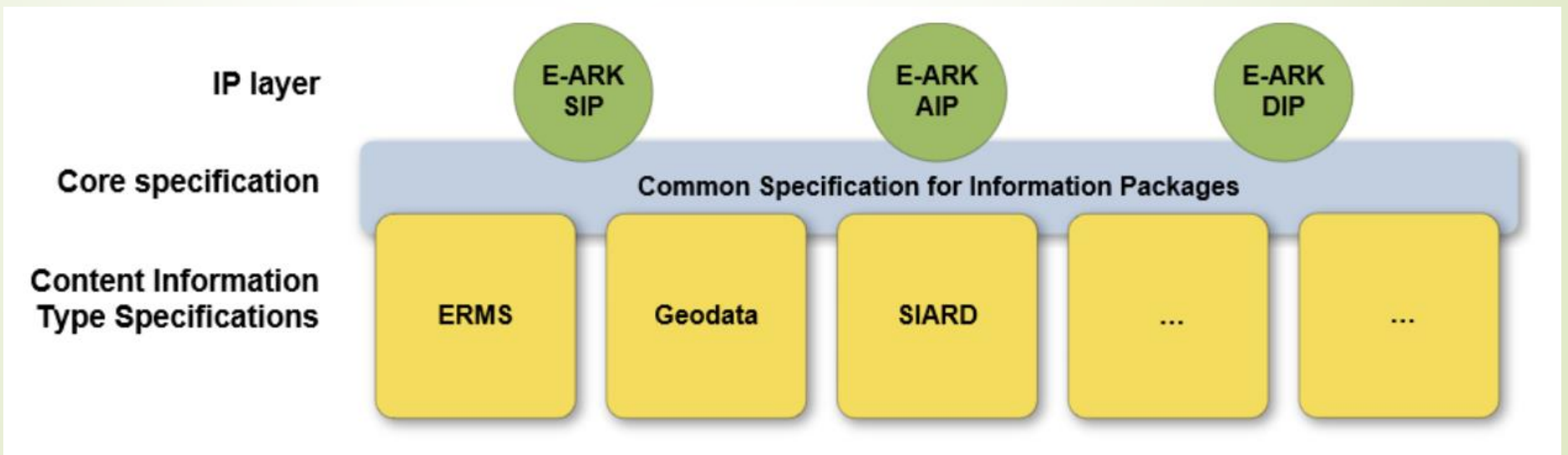
2021-11-01 09:54:07

We are pleased to announce the publication of the Common Specifications for the Information...

[Read More ...](#)

Ulike spesifikasjoner forvaltet av DILCIS Board

- ▶ SIARD 2.x spesifikasjonene
- ▶ OAIS generell pakkestruktur: Common Specification for Information Packages
- ▶ OAIS-pakkene: Submission (SIP), Archival storage (AIP), Dissemination (DIP)
- ▶ Innholdsspesifikke: Geodata, helse, database (SIARD), sak-arkiv (ERMS) osv.



SIARD (Software Independent Archiving of Relational Databases)

The SIARD specification is an open file format for the long-term archiving of relational databases in the form of text data based on XML that is packaged in a container file.

The link to SIARD here to the right have been updated to contain no spaces (2024/05/17).

If the structure and content of a relational database are translated into the SIARD format, it will subsequently be possible to access and exchange the data in the database in the future, even when the original database software is no longer available or can no longer be run. This long-term interpretability of the database content is based on the use of two widely supported international standards: XML and SQL:2008.

Please note that in addition to the SIARD specification the DILCIS Board maintains the [SIARD CITS](#) specification which describes how to package a SIARD file for long-term preservation along with additional representations of the data, metadata, and documentation.

We welcome all feedback in regard to the SIARD specification - if you have any comments or proposals, please contact us per [e-mail](#) or leave your comments on [GitHub!](#)

Download SIARD

Latest version (v2.2, August 31, 2021)

- [SIARD 2.2](#)
- [metadata.xsd](#)

Previous versions

[SIARD 2.1.1](#) (May 15, 2019)

[SIARD 2.0](#) (June 1, 2016)

[SIARD 1.0](#) (September 30, 2008)

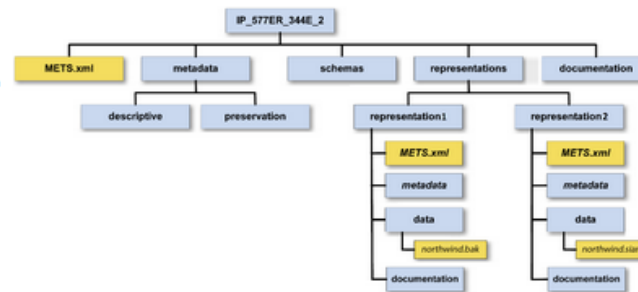
CITS SIARD

The CITS SIARD (Content Information Type Specification for Relational Databases using SIARD) is a specification that describes how to package and preserve relational database content. This is primarily done by packaging [SIARD](#) files into information packages that conform to the [Common Specification for Information Packages](#).

The specification helps you to apply a common way of storing multiple representations of a database (for example a proprietary backup and a SIARD snapshot) in a single package along with appropriate metadata and binary documentation of the dataset.

Used standards

- Software Independent Archiving of Relational Databases (SIARD)
- Data Entity Dictionary Specification Language (DEDSL) – XML/DTD: ISO 22643:2003 (CCSDS 647.3-B-1:2002)
- Common Specification for Information Packages (CS IP)
- E-ARK Submission Information Package (E-ARK SIP)

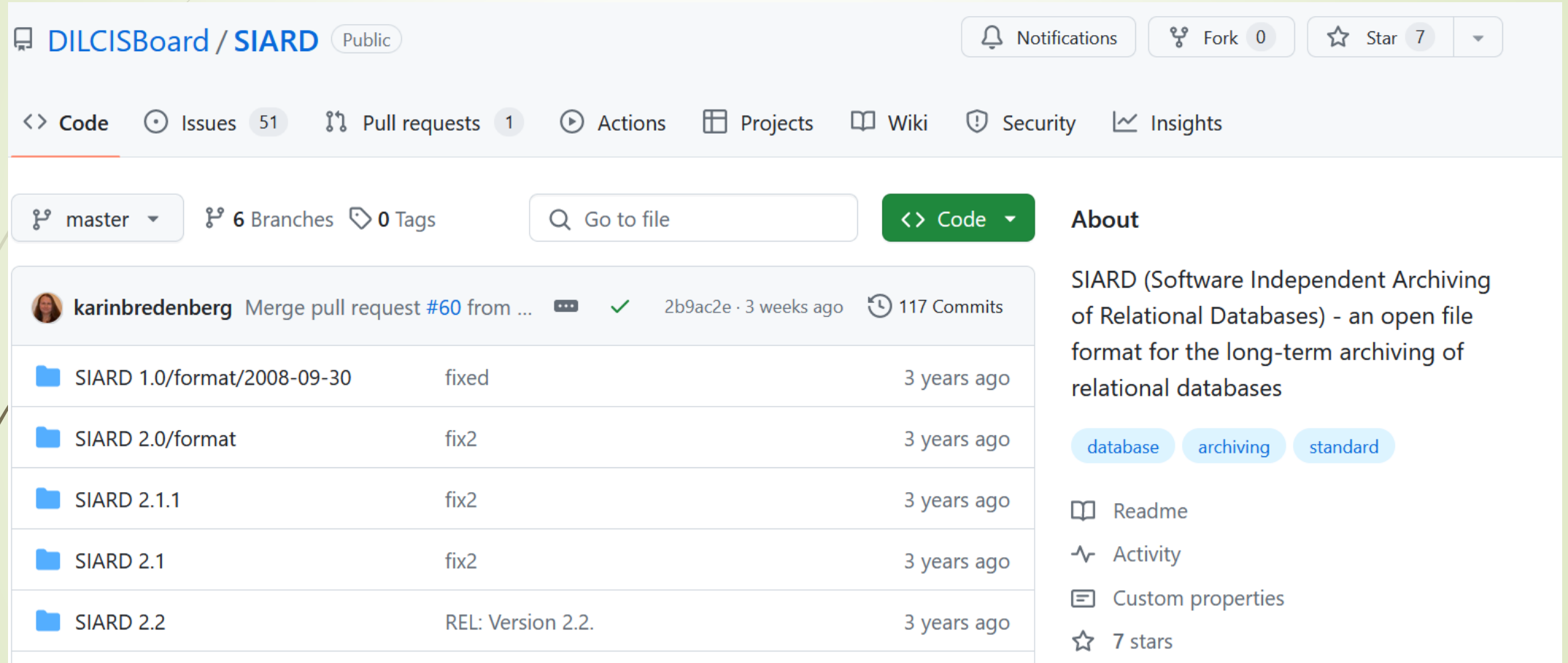


Download CITS SIARD

Latest version (v1.0.0, August 31, 2021)

- [CITS SIARD v1.0.0](#)
- [E-ARK-SIARD-ROOT.xml](#)
- [E-ARK-SIARD-REPRESENTATION.xml](#)
- [Guideline_CITS_SIARD_1_0_0.pdf](#)

SIARD på GitHub



DILCISBoard / SIARD Public

Notifications Fork 0 Star 7

Code Issues 51 Pull requests 1 Actions Projects Wiki Security Insights

master 6 Branches 0 Tags Go to file Code

karinbrenberg Merge pull request #60 from ... 2b9ac2e · 3 weeks ago 117 Commits

SIARD 1.0/format/2008-09-30	fixed	3 years ago
SIARD 2.0/format	fix2	3 years ago
SIARD 2.1.1	fix2	3 years ago
SIARD 2.1	fix2	3 years ago
SIARD 2.2	REL: Version 2.2.	3 years ago

About

SIARD (Software Independent Archiving of Relational Databases) - an open file format for the long-term archiving of relational databases

database archiving standard

Readme Activity Custom properties 7 stars

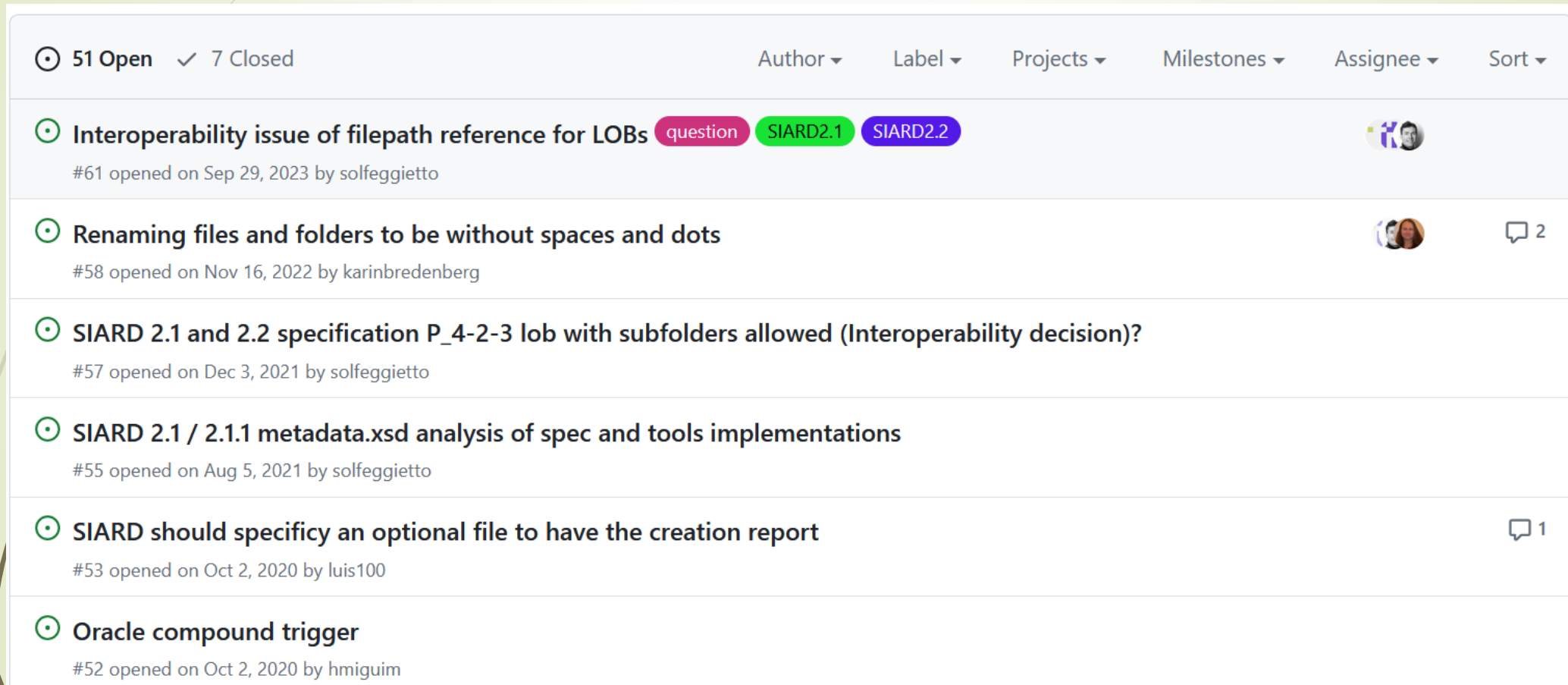
<https://github.com/DILCISBoard/SIARD>



Kort SIARD historikk




- ▶ SIARD 1.0
 - ▶ Normaliserer mot ANSI/ISO SQL:1999
 - ▶ Publisert i 2008, ble sveitsisk standard noen år senere
- ▶ SIARD 2.0
 - ▶ Normaliserer mot ANSI/ISO SQL:2008 (SQL:1999 er en delmengde)
 - ▶ Publisert i 2016
 - ▶ Motivasjon, egendefinerte datatyper i SQL:2008
- ▶ SIARD 2.1-2.1.1
 - ▶ Feilrettinger
- ▶ SIARD 2.2
 - ▶ Håndtering av store objekter, LOBs
 - ▶ Publisert i 2021

SIARD issue list på GitHub



51 Open ✓ 7 Closed

Author ▾ Label ▾ Projects ▾ Milestones ▾ Assignee ▾ Sort ▾

- 🕒 **Interoperability issue of filepath reference for LOBs** question SIARD2.1 SIARD2.2 
#61 opened on Sep 29, 2023 by solfeggietto
- 🕒 **Renaming files and folders to be without spaces and dots**  2
#58 opened on Nov 16, 2022 by karinbrendenberg
- 🕒 **SIARD 2.1 and 2.2 specification P_4-2-3 lob with subfolders allowed (Interoperability decision)?**
#57 opened on Dec 3, 2021 by solfeggietto
- 🕒 **SIARD 2.1 / 2.1.1 metadata.xsd analysis of spec and tools implementations**
#55 opened on Aug 5, 2021 by solfeggietto
- 🕒 **SIARD should specify an optional file to have the creation report**  1
#53 opened on Oct 2, 2020 by luis100
- 🕒 **Oracle compound trigger**
#52 opened on Oct 2, 2020 by hmiguim

<https://github.com/DILCISBoard/SIARD/issues>



SIARD oppsummert

- Siard standarden finnes på DILCIS Board nettsiden
 - <https://dilcis.eu/content-types/siard>
- Du kan rapportere inn kommentarer og endringsønsker på GitHub
 - <https://github.com/DILCISBoard/SIARD/issues>
- Informasjon E-ARK pakke-spesialisering mot SIARD kan finnes her
 - <https://dilcis.eu/content-types/cs-siard>



SIARD: Bak kulissene...

- ▶ Det arbeides både i E-ARK/eArchiving prosjektet og DILCIS Board med SIARD
- ▶ Kortsiktig
 - ▶ Reduksjon av issue-listen, det vil si
 - ▶ Lukke flest mulig saker som kan lukkes nå
- ▶ Langsiktig
 - ▶ Forberede neste SIARD-versjon
 - ▶ Spesielt gjennomgang av ulike elementer i SIARD spesifikasjonen
 - ▶ Kan noen elementer gjøres valgfrie og hvilke konsekvenser får det?
 - ▶ Hvilke krav må stilles til en verktøy-leverandør
 - ▶ for at de kan si at de produserer SIARD



Andre aktiviteter knyttet til E-ARK spesifikasjonene

- ▶ Flytting av mest mulig fra SIP, AIP og DIP spesifikasjonene til den generelle common specification
- ▶ Optimalisere AIP spesifikasjonen i forhold til endringshåndtering/versjonskontroll
 - ▶ Optimalisering, effektivitet, robusthet
 - ▶ Ser bl.a. på OCFL, Oxford Common Fileformat Layout
 - ▶ <https://ocfl.io/>



Avslutningsvis

- Håper jeg har gitt dere en oversikt over forvaltning av SIARD-spesifikasjonen
- At jeg også har gitt dere en oversikt over hvor dere kan finne mer informasjon og komme med tilbakemeldinger
- Pluss skissert hva som er i fokus i tiden framover
- I tillegg gitt et kort innblikk i andre aktiviteter rundt E-ARK spesifikasjonene

“

Takk for meg

”

arngro@arkivverket.no