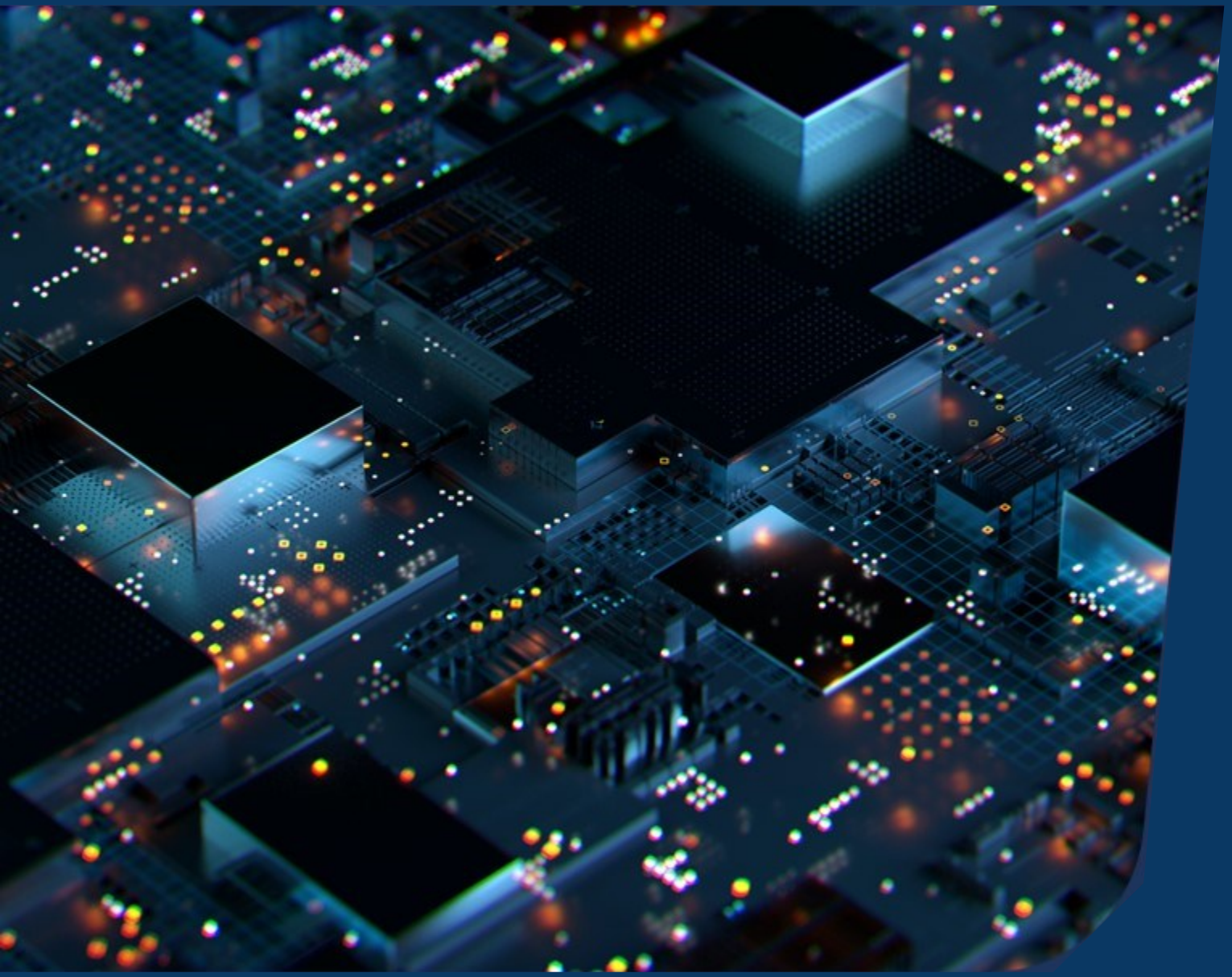


E-ARK standardisation for databases

Kuldar Aas
E-ARK3 Technical lead

What is E-ARK?

EC funded E-ARK projects

- 2014 – 2017: E-ARK
- 2018 – 2019: E-ARK4ALL
- 2019 – 2021: E-ARK3

eArchiving Building Block

- EC-owned building block that aims to support interoperability in digital archiving
- E-ARK3 as the current service provider

Digital Information LifeCycle Interoperability Standard (DILCIS) Board

- Maintains specifications developed in E-ARK projects ..
- .. and SIARD

eArchiving

Facilitate the preservation, migration, reuse and trust of your data

E-ARK vision (2014)

Vision: All digital preservation systems receive, store and provide access to information regardless of its size, type or format according to a set of agreed principles which allow systems to identify, verify and validate the information in a uniform way

Goal: Interoperability between data sources, archives and reuse environments is improved to a point where digital preservation tools can be reused across borders and institutions

E-ARK and databases

The idea of E-ARK was first developed among national archives

Focus still remains on ‘records’ in public sector and business

- Content in relational databases
- ERMS
- other

National Archives SIARD scenario

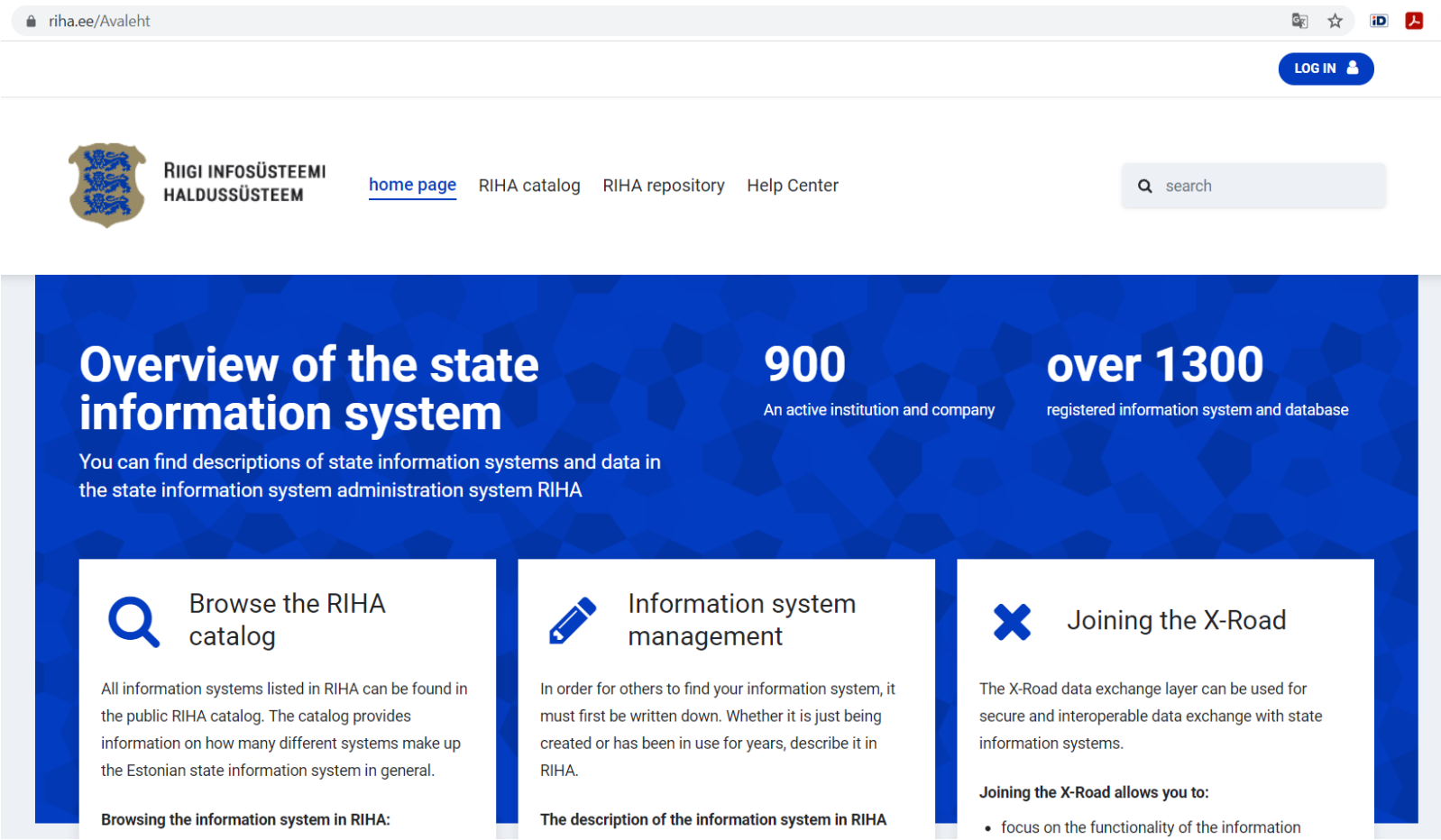
- Potentially 1000s on relational databases in the public sector

  *About 95% of public records in Estonia assumed to be relational data

- 100s of databases of archival value

- Need to preserve data and service definitions (i.e. views, queries)

- Need for a simple, standardised preservation format (i.e. SIARD)



DILCIS standardisation effort

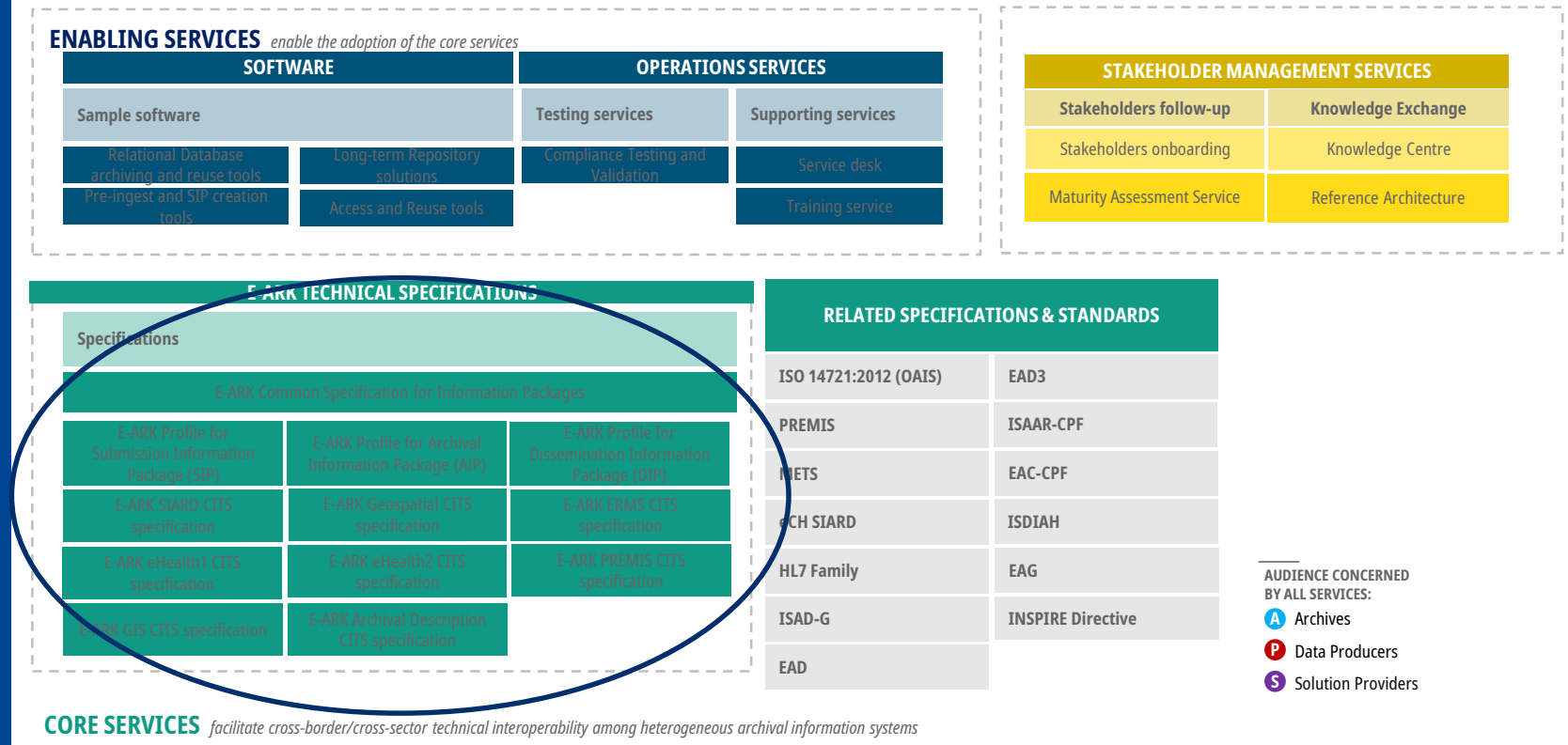
Generic Information Package Specifications

- *CSIP, SIP, AIP, DIP*

Content Information Type Specifications (CITS)

- SIARD, ERMS, Geodata, eHealth(x2)
- Archival description, PREMIS

Each specification has its own space and team



DILCIS standardisation support

Guidelines and procedures for CITS maintenance and development
https://www.dilcis.eu/guidelines

Dedicated landing pages for all specifications
https://www.dilcis.eu/content-types/siard

Open reviews
SIARD 2.2 RFC in 10.2020 – 01.2021

GitHub sites for spec development and issue handling
https://github.com/DILCISBoard/SIARD

dilcis.eu/about

DILCIS BOARD | HOME | ABOUT | SPECIFICATIONS | CONTENT TYPES | GUIDELINES | REVIEWS | SOFTWARE | CONTACT

The Digital Information LifeCycle Interoperability Standards Board

The Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) is an international group of experts committed to maintain and sustain maintain a set of **interoperability specifications** which allow for the **transfer, long-term preservation, and reuse of digital information** regardless of the origin or type of the information.

More specifically, the DILCIS Board maintains specifications initially developed within the E-ARK Project (02.2014 - 01.2017):

- Common Specification for Information Packages
- E-ARK Submission Information Package (SIP)
- E-ARK Archival Information Package (AIP)
- E-ARK Dissemination Information Package (DIP)
- SMURF Specification (Semantically Marked-Up Record Format)

The DILCIS Board collaborates closely with the Swiss Federal Archives in regard to the maintenance of the SIARD (Software Independent Archiving of Relational Databases) specification.

The DILCIS Board consists of up to ten international experts who act on a voluntary basis. You can read more about the setup, tasks and responsibilities of the Board here.

The DILCIS Board is supported by the European Commission through the CEF eArchiving Building Block (06.2018 -) and supervised by the DLM Forum.

Latest news

E-ARK publishes updated Content Information Type Specifications, procedures and supporting guidelines
2021-08-31 21:21:42
We are pleased to announce that the E-ARK eArchiving Content Information Type Specifications,...
Read More ...

Conclusion of 2020 - 2021 specification review
2021-07-23 07:39:10
The DILCIS Board wishes to thank all those who contributed to the reviews of the 25 eArchiving...
Read More ...

5th Call for Reviewers: eArchiving CITS for geospatial data
2021-06-15 07:52:55
We are happy to announce the opening of group 5 of the 2020 - 2021 review cycle of eArchiving

E-ARK and SIARD

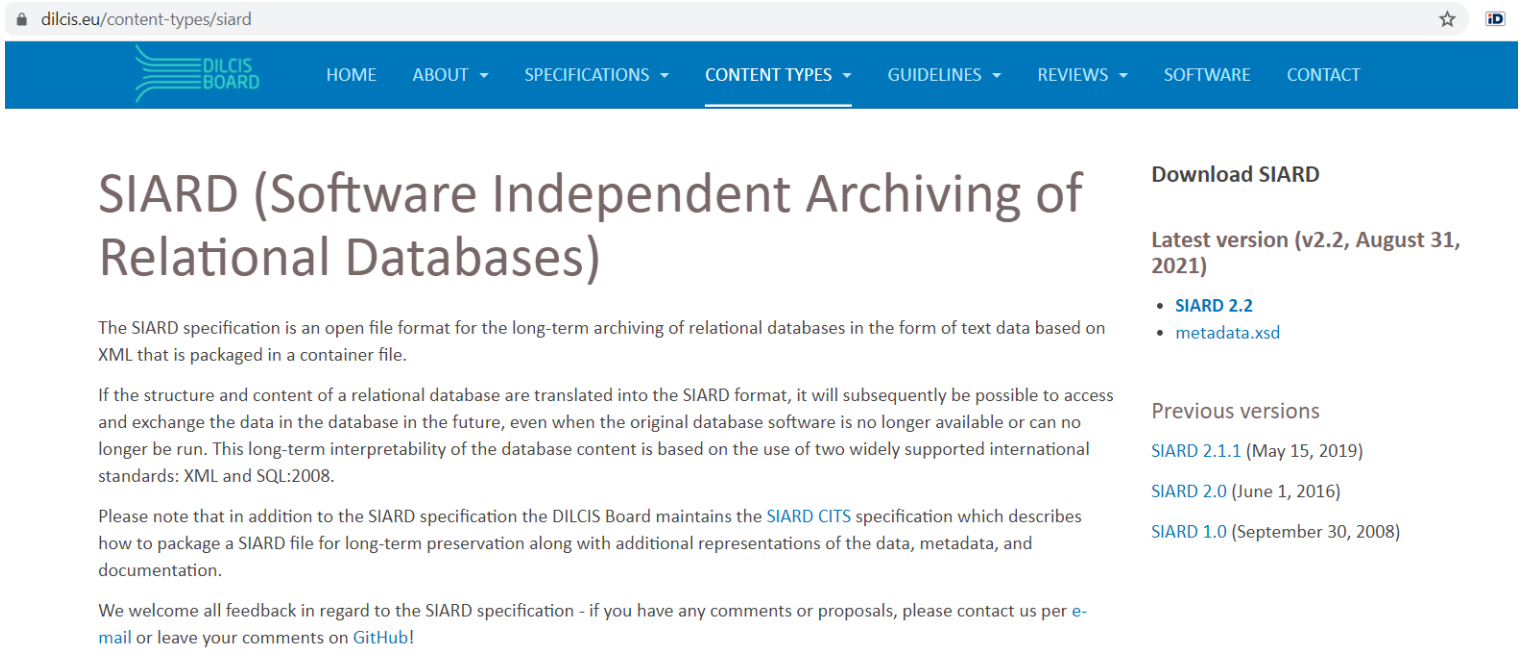
* SIARD is maintained in collaboration of DILCIS Board and Swiss Federal Archives

Common work on SIARD since 2017 (v2.0)

More people to work with SIARD = more errors and issues found

SIARD scalability as a special concern for E-ARK members (DNA, NAE)

- SIARD v2.2 includes support for LOBs outside SIARD



E-ARK and CITS SIARD

A relational database transfer might include more than just a SIARD snapshot:

- Original dump (and application)
- Additional metadata (Dublin Core, EAD, …)
- Documentation
- …

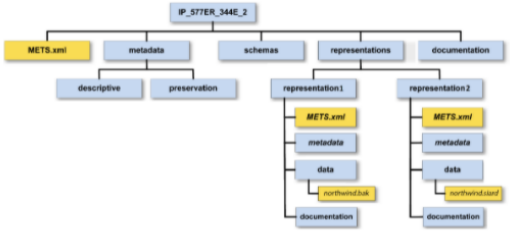
dilcis.eu/content-types/cs-siard

DILCIS BOARD

HOME | ABOUT | SPECIFICATIONS | CONTENT TYPES | GUIDELINES | REVIEWS | SOFTWARE | CONTACT

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.

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

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

How many ways are there to archive a database?

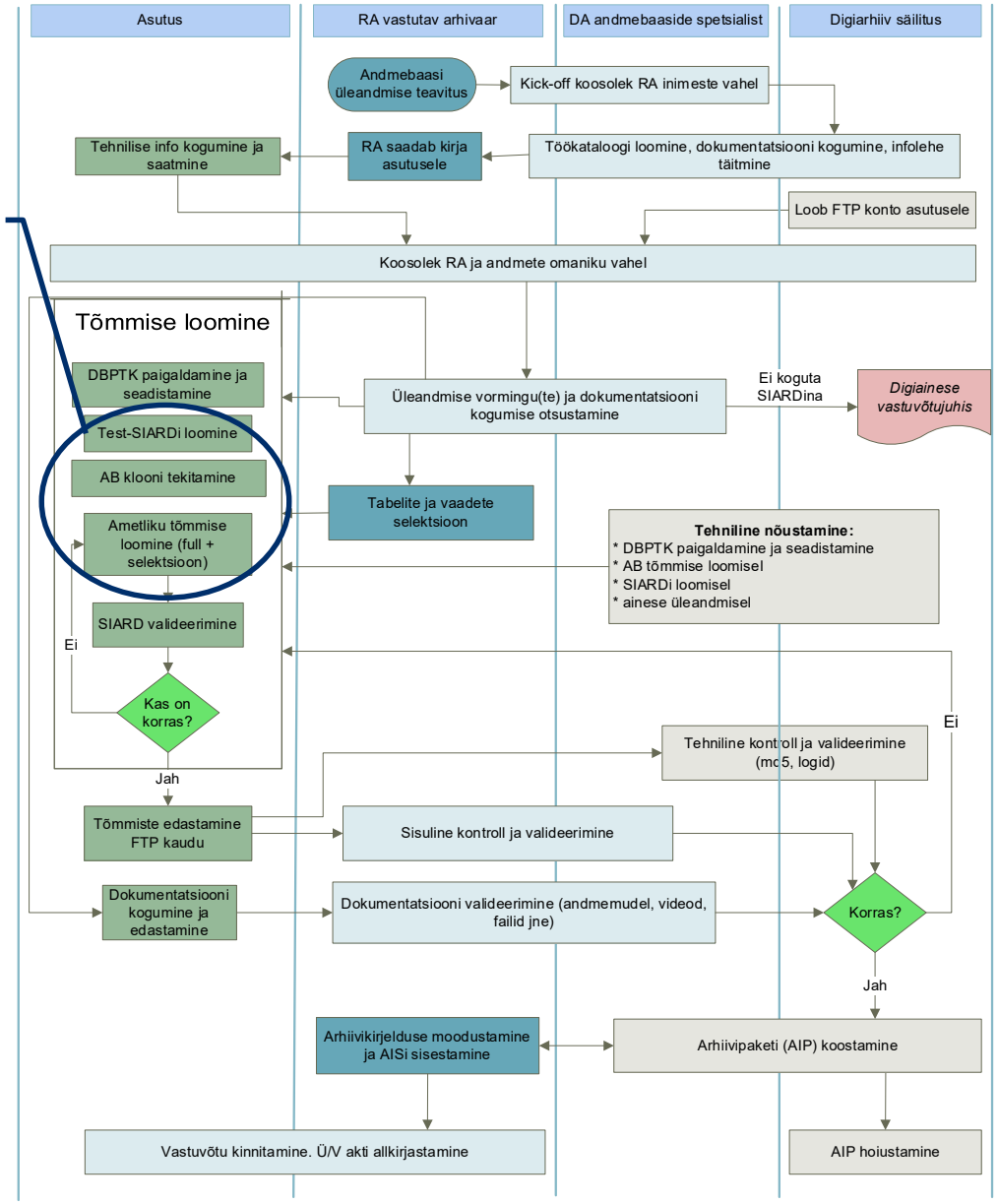
SIARD is just one of the options!

The whole archival process is much wider than just creating a snapshot

- Example: to archive the whole database or parts of it, relational data or materialised views / services?

SIARD software (DBPTK, SIARD Suite) behaves differently – which one to choose?

SIARD created here



How many ways are there to archive a database?

E-ARK thinks it's crucial to share our experiences

- Relational DataBase Archiving Interest Group (rdb-aig)
- Two case studies written in 2020

- NEW! rdb-aig mailing list
  https://listserv.dilcis.eu/info/rdb-aig

listserv.dilcis.eu/info/rdb-aig

DILCIS BOARD

Home | Search for List(s) | Support | Login

rdb-aig - Relational DataBase Archiving Interest Group Mailing List

List Options

Owners: Anders Bo Nielsen, achiggs, Kuldar Aas, Lauri Rätsep, Phillip Aasvang Tommerholt

Moderators: (same as owners)

Contact owners

List Home

Subscribe

Unsubscribe

Archive

Post

rdb-aig@lists.dilcis.eu

Subject: Relational DataBase Archiving Interest Group Mailing List

Description: Welcome to the rdb-aig mailing list!

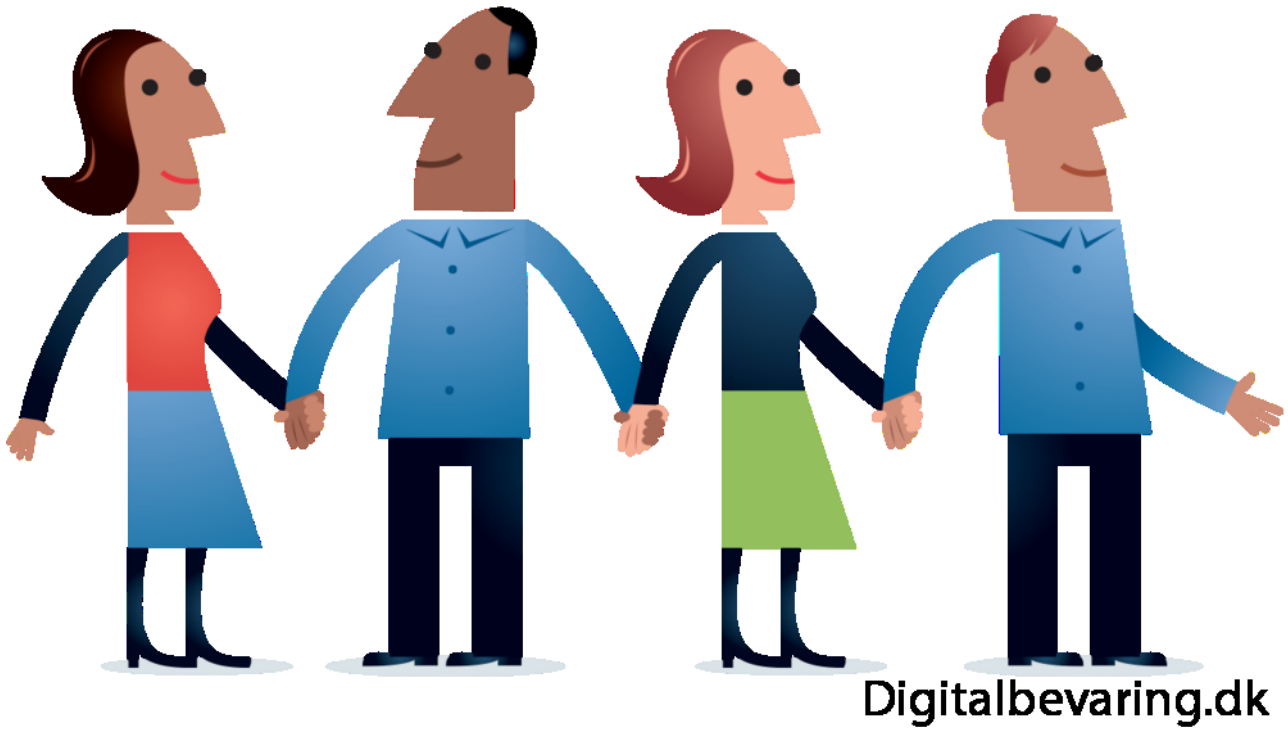
The Relational DataBase Archiving Interest Group mailing list is the official means of communication between the members of the Interest Group - the community where information about SIARD, CITS SIARD, database archiving tools, best practices, case studies, and related news can be shared.

As a member, you will receive the latest news and information about database archiving from the developers of the SIARD specification and associated tools. You are also welcome to ask any questions about and share your personal practices in relational database archiving!

The use of personal data. Please note that all posts posted to the list will be added to the list archive for future reference. The information added into the archive will therefore include the name of the author of the post as well as any additional details in the body (incl signature) of the message. The administrators of the list will make sure, that list archives are accessible only to registered members.

The future of ~~E-ARK and~~ SIARD and DILCIS

- DILCIS (and E-ARK) mindset is about being open and inclusive!
- DILCIS to support CITS development (as opposed to lead)
    - SIARD and CITS SIARD groups are currently led by DNA and SFA, with contributions by many others
    - Communication MUST be improved → 2021 SIARD v2.2 RFC got a total of three responses..
- SIARD validation and tool confrmance as a serious issue
- rdb-aig mailing list to serve as the birthplace for new ideas!
- E-ARK to lobby with EC for funding



Questions?

Kuldar Aas
kuldar.aas@ra.ee

Ready to join DILCIS?

Find out more at:
https://dilcis.eu/content-types/siard

Join the rdb-aig mailing list at:
https://listserv.dilcis.eu/info/rdb-aig