



Database ingest and use at public archives with the SIARD format

Torbjørn Aasen (Intercommunal Archive Møre og Romsdal, Ålesund, Norway)

Torbjørn Aasen is an IT specialist who has previously worked in systems development.

IKAMR is a Norwegian Municipal Archive in Middle Norway. It is a rural place with fjords and mountains. The organization has 28 employees including two IT archivists. IKAMR relies on the KDRS, a regional digital resource centre with 18 Municipal archives as members. They offer an OAIS based digital repository to members. Database archiving is done by routine. The IKAMR started XML table extractions in 2005, then was refined by using SIARD in 2014 and the national Noark standard in 2015.

- From 2005 onwards, table extractions were done by taking out each table and placing it into an individual XML file. Metadata on the DB structures were included as an ADDML.xml file (national standard). This way, IKAMR preserved 50 database systems between 2005 and 2011. They started by preserving three tables then added five more, finally ingesting full database in the archival work area.
- In 2014, IKAMR first ingested SIARD packages and has so far ingested 100 SIARD extractions. A few of them were extracted using SIARD Suite v1 and 2, some were using DBPTK 2, but most of them were using Spectral Core Full Convert, with SIARD 2.1 as the target format.
- Noark 5 extractions are the current Norwegian Archive Standard for recordkeeping systems used in public administration. The extractions are well-ordered, comply to the standard and can be validated against it. Fifty databases have been ingested since 2015 on this basis.

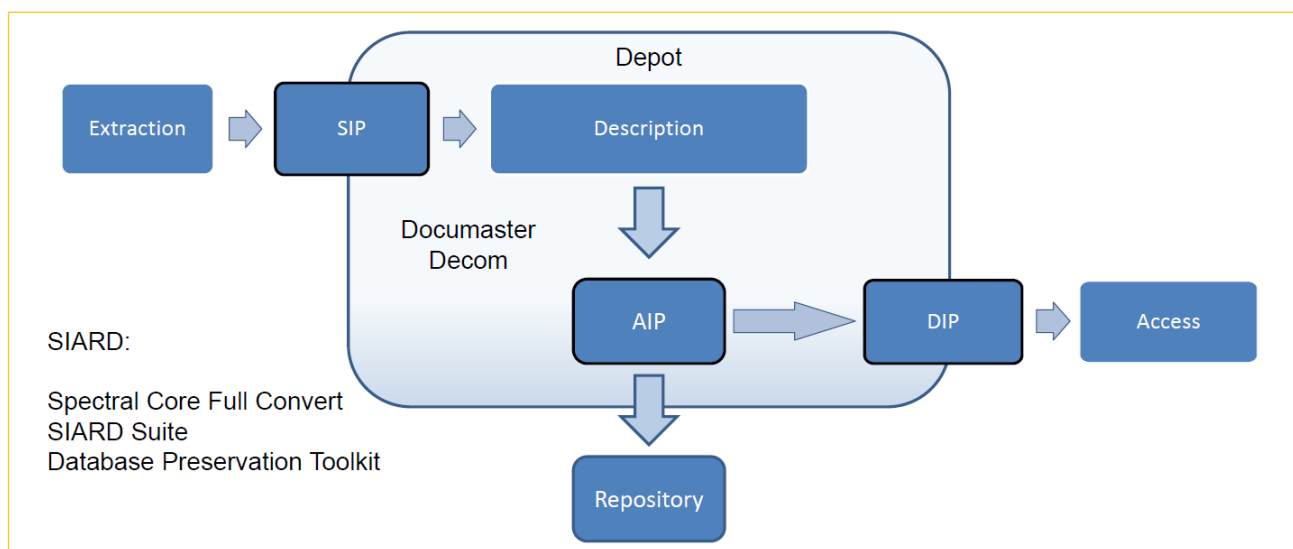


Figure 33: Production line for digital archives at Norwegian municipal archives



A SIARD extraction is imported into the tool Decom. The KDRS Decom server holds over 60 system description templates that are used to handle the unordered extractions (e.g. describing a particular childcare database system). This means that extractions from the same system over time are quicker as IKAMR can reuse the previous template. LOBS are transformed as part of the process, for example from office formats to PDF/A.

Regarding access to preserved SIARD extractions, the system for visualisation of preserved SIARD extractions is under evaluation. IKAMR is looking at DBPTK desktop enterprise and Asta 7 from Stiftelsen Asta. There are not many requests for access now, but this is likely to increase. Currently, Aasen migrates a SIARD file into a database for users and helps them use SQL queries to search for relevant content.

The challenges Torbjørn Aasen saw in the process are lack of time, especially for finding and prioritising the databases, loss of knowledge from staff who have left, scarce technical documentation, and in some cases a front-end logic that has not survived.

Challenges of using SIARD (but there are solutions and workarounds noted):

- Some old Oracle system slices files into 32 kb bits.
- Some systems have encrypted documents.
- Some old systems store binary files in CLOB text column and forces it into UTF-8.

Aasen gave positive feedback on DBPTK, an excellent solution including the SIARD validator. The validator needs to be tuned for interoperability. The viewer is also excellent. He also recommended Spectral Core Full Convert as a high-quality migration tool with high success rates. He has experienced some interoperability issues which are reported on GitHub (<https://github.com/DILCISBoard/SIARD/issues>). He is looking forward to exchanging experiences with the community for the common goal.