



ika

Interkommunalt arkiv for Møre og Romsdal

Database Ingest, Preservation & Visualization

Using SIARD as method at IKAMR, a Norwegian Municipal Archive

Torbjørn Aasen, IKAMR

IT archivist

torbjorn.aasen@ikamr.no

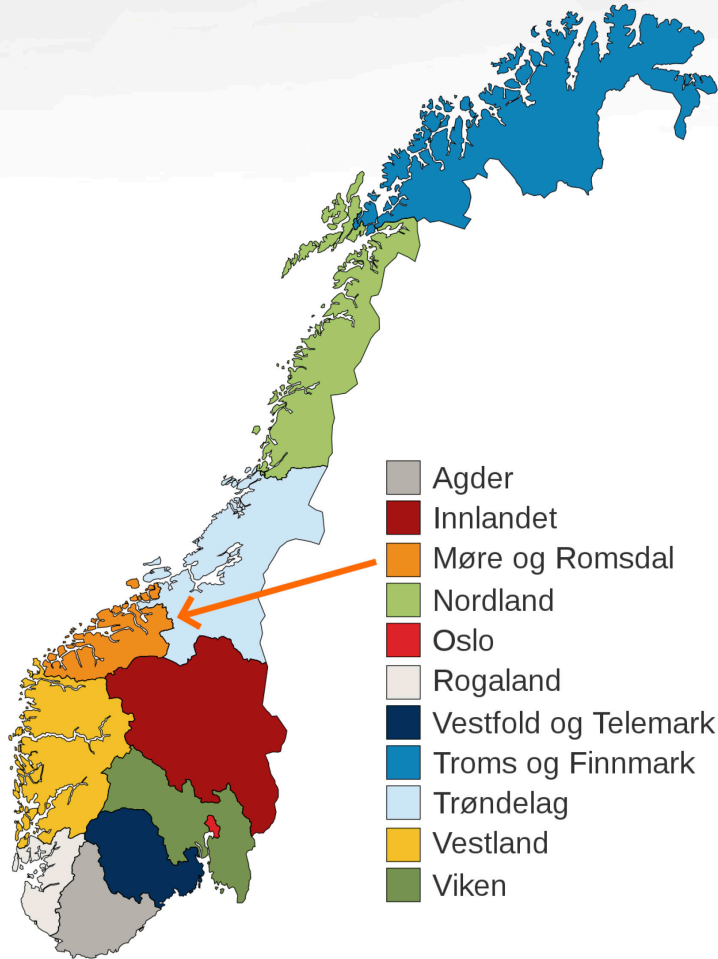
Databases for 2080, Landesarchiv Baden-Württemberg, 06.10.2021

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal



- Agder
- Innlandet
- Møre og Romsdal
- Nordland
- Oslo
- Rogaland
- Vestfold og Telemark
- Trøndelag
- Vestland
- Viken

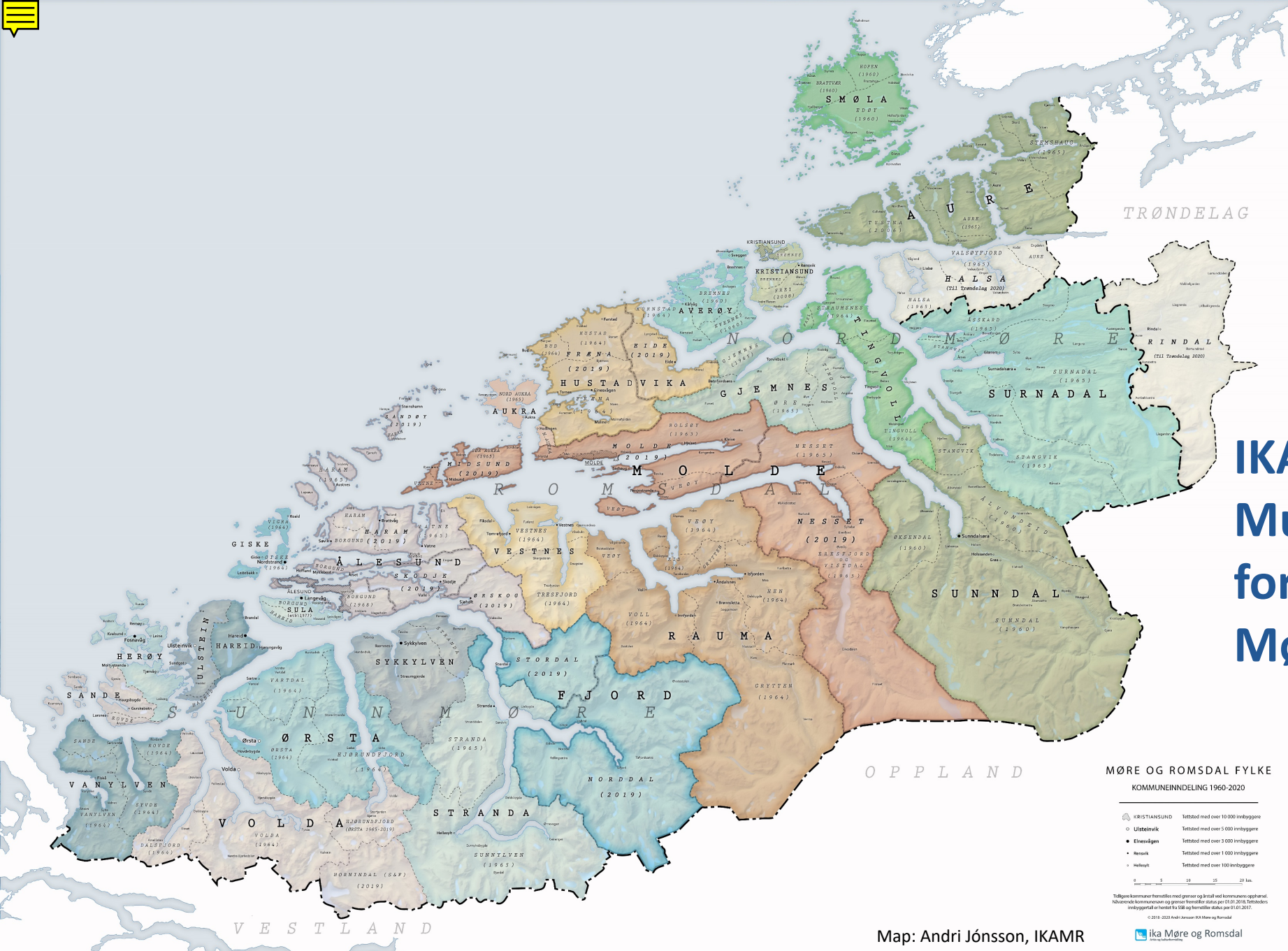
IKAMR – a Norwegian Municipal Archive for the County of Møre og Romsdal

Map: https://en.wikipedia.org/wiki/Counties_of_Norway

<https://www.ikamr.no>

Fortid
& notid
for framtid

Arkiv og kulturformidling



IKAMR – a Norwegian Municipal Archive for the County of Møre og Romsdal

MØRE OG ROMSDAL FYLKE
KOMMUNEINDELING 1960-2020

- KRISTIANSUND Tettsted med over 10 000 innbyggere
- Ulsteinvik Tettsted med over 5 000 innbyggere
- Eneåsen Tettsted med over 3 000 innbyggere
- Rensvik Tettsted med over 1 000 innbyggere
- Helseth Tettsted med over 100 innbyggere

Tallene kommuner fremmes med grenser og innval med kommunene og har et 1:500 000 kart over kommunene og grensene for 2018. Tettstedene er bygdertall og forlest fra SSB og kommunale data på 01.01.2017.

© 2019-2022 Andri Jónsson/IKAMR Møre og Romsdal

Map: Andri Jónsson, IKAMR



Fortid & notid for framtid

Arkiv og kulturformidling



ika

Interkommunalt arkiv for Møre og Romsdal

IKAMR – a Municipal Archive for the County of Møre og Romsdal

1. Located in Ålesund, a Jugendstil city at the west coast of Norway
2. Have 28 employees, including 2 IT archivists
3. Owned by 27 municipalities and the county municipality of Møre og Romsdal

Picture: Torbjørn Aasen, IKAMR

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

KDRS – Municipal Archive institutions Digital Resource Centre

1. Covers 81% of the municipalities and county municipalities
2. Have 18 Municipal Archives as members
3. Offers an OAIS based Digital Repository to its members (ESSArch)
4. Community for IT archivists forming the SIARD production line
5. Cooperation with The National Archives of Norway



KDRS

Kommunearkivinstitusjonenes
Digitale RessursSenter

Fortid
& notid
for framtid

<https://www.kdrs.no>

Arkiv og kulturformidling



ika

Interkommunalt arkiv for Møre og Romsdal

Preservation methods

1. Table Extractions (2005-2011)
2. Noark 5 Extractions (2015-2021)
3. SIARD Extractions (2014-2021)

Torbjørn Aasen, IKAMR

IT archivist

torbjorn.aasen@ikamr.no

Databases for 2080, Landesarchiv Baden-Württemberg, 06.10.2021

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Table Extractions

1. Each Database Table into its own table.xml
2. Metadata DB structure and info in ADDML.xml
3. From 2005 to 2011 preserved 50 Database Systems
4. Starting first system preserving 3 tables (tool [Pervasive](#))
5. Then 5 more tables, still directly from database
6. Finally ingest full database in archival work area

```
<addml version="7.3">
  <reference>
    <archives ar_id="molde.kommune.no">
      <ar_name>Molde kommune</ar_name>
    </archives>
    <system sy_id="MOL2006-001">
      <sy_name>Sak/-arkivsystem og elektroniske postjournaler</sy_name>
      <startdate>19990501</startdate>
      <enddate>20050501</enddate>
    </system>
  </reference>
  <structure>
    <dataset ds_id="molde.kommune.no/MOL2006-001">
      <ds_descr>Datasettet inneholder registeropplysninger fra Noark-3 systemet
      Forum Winsak. Registeret omfatter saks- og dokumentjournal, med tilhørend
      opplysninger (avdeling, ansatt, delarkiv, utvalgssaker, saksparter m.m.)
    </ds_descr>
    <charset>ISO-8859-1</charset>
    <format>XML</format>
    <recsep>NO</recsep>
    <nu_files>24</nu_files>
    <file name="SAKSDOKUMENT" path="saksdokument.xml">
      <fi_descr>SAKSDOKUMENT - informasjon om saksdokumenter</fi_descr>
      <nu_rectypes>1</nu_rectypes>
      <rectype name="DOKUMENT">
        <primkey>LOEPENR</primkey>
        <forkey>
          <fields>SAKSNR</fields>
          <reffilerec>SAKSSKJEMA:SAK</reffilerec>
          <reffields>SAKSNR</reffields>
          <reltype>n:1</reltype>
        </forkey>
        <nu_fields>65</nu_fields>
        <fieldtype name="LOEPENR">
          <ft_descr>Løpenummer</ft_descr>
          <datatype>INTEGER</datatype>
          <nonnull/>
          <unique/>
        </fieldtype>
      </rectype>
    </file>
  </structure>
</addml>
```

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Noark 5 Extractions

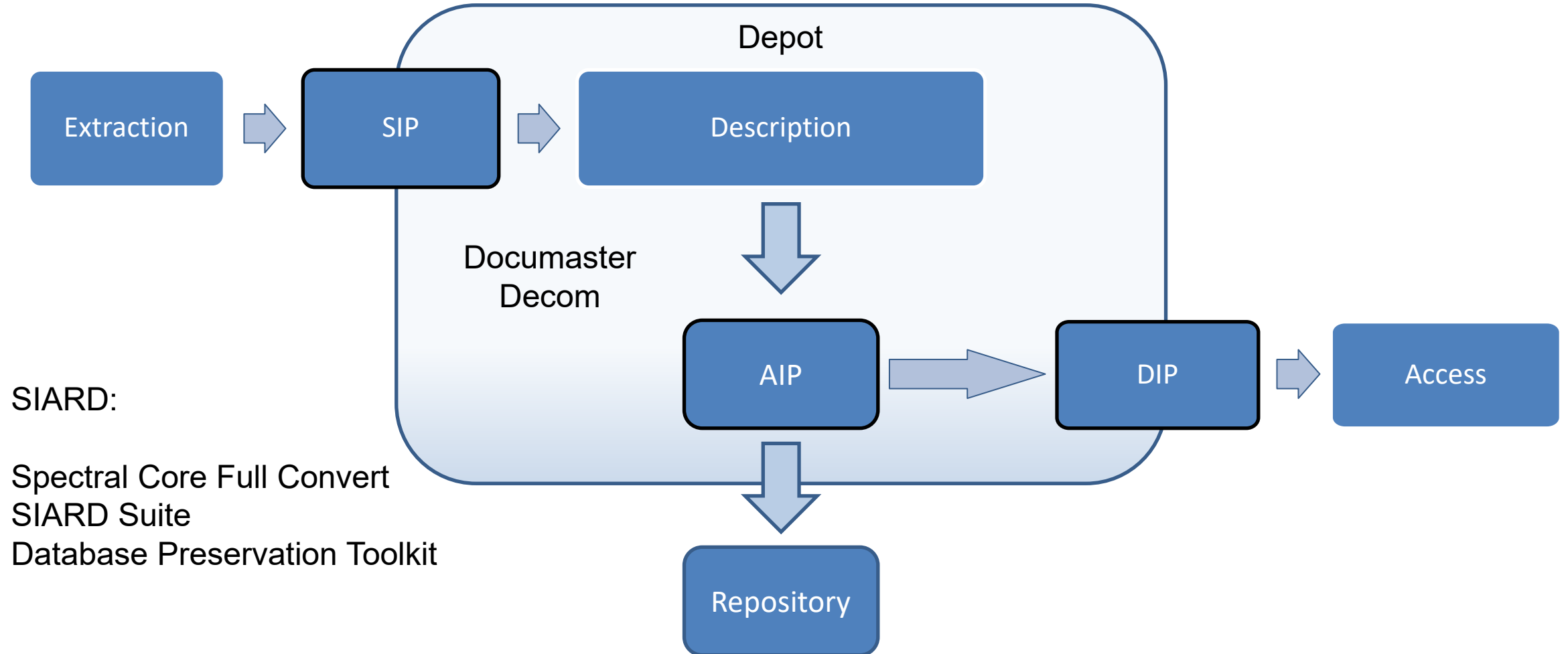
```
<arkivdel>
<systemID>arkivdel:SAK1</systemID>
<tittel>Saksarkiv basert på K-koder</tittel>
<arkivdelstatus>Avsluttet periode</arkivdelstatus>
<dokumentmedium>Elektronisk arkiv</dokumentmedium>
<opprettetDato>0001-01-01T00:00:00</opprettetDato>
<opprettetAv>NULL</opprettetAv>
<avsluttetDato>2017-05-10T14:25:25.613</avsluttetDato>
<avsluttetAv>Per Person</avsluttetAv>
<arkivperiodeStartDato>2008-01-01</arkivperiodeStartDato>
<arkivperiodeSluttDato>2011-12-31</arkivperiodeSluttDato>
<referanseArvtaker>arkivdel:SAK2</referanseArvtaker>
<klassifikasjonssystem>
<systemID>arkivdel:SAK1:klassifikasjonssystem:EMNE</systemID>
<klassifikasjonstype>EH</klassifikasjonstype>
<tittel>K-Kode fagklasse/fellesklasse bokmål</tittel>
<opprettetDato>0001-01-01T00:00:00</opprettetDato>
<opprettetAv>NULL</opprettetAv>
<klasse>
<systemID>arkivdel:SAK1:klassifikasjonssystem:EMNE:klasse:0</systemID>
<klasseID>0</klasseID>
<tittel>Organisasjon og administrasjon</tittel>
<opprettetDato>0001-01-01T00:00:00</opprettetDato>
<opprettetAv>NULL</opprettetAv>
<klasse>
<systemID>arkivdel:SAK1:klassifikasjonssystem:EMNE:klasse:00</systemID>
<klasseID>00</klasseID>
<tittel>Organisasjon og administrasjon - Felles</tittel>
<opprettetDato>2003-01-11T00:00:00</opprettetDato>
<opprettetAv>NULL</opprettetAv>
```

1. Noark 5 is the current “Norwegian Archive Standard” for recordkeeping systems used in public administration
2. Noark 5 extractions are well-ordered and should comply to the standard. The Archive will validate accordingly.
3. Approved or Rejected, the Archive will not edit the content
4. From 2015-2021 ingested 50 Noark 5 systems

Fortid
& notid
for framtid

Production Line – Electronic Archives

Ingest – Description/Storage – Access





ika

Interkommunalt arkiv for Møre og Romsdal

From SIARD Extraction to SIP delivery to the Archive

1. A free tool [Arkade 5](#) from The Norwegian National Archives puts the SIARD into a Norwegian DIAS SIP (.tar file) and metadata in a corresponding info.xml file
2. The coming Arkade 5 v2.5.0 can validate SIARD 2.1 using DBPTK Developer v2.9.9, and make a PRONOM filelist
3. Norway is not using the EARK SIP/AIP/DIP package formats

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

From SIARD Extraction to SIP delivery to the Archive

Arkade 5

Arkade 5

Testing av uttrekk

Filsti: Y:\uttrekk-dbptk-desktop\version\v2.5.9\t01bcl0b12_dbptk-desktop-2.5.9_int

UUID: bd3e5dde-b2be-4b30-a686-586bcb6f5ed2

Arkivtype: Siard

Framdrift: Utført

[Start testing](#) [Vis rapport](#) [Opprett pakke](#) [Ny kjøring](#)

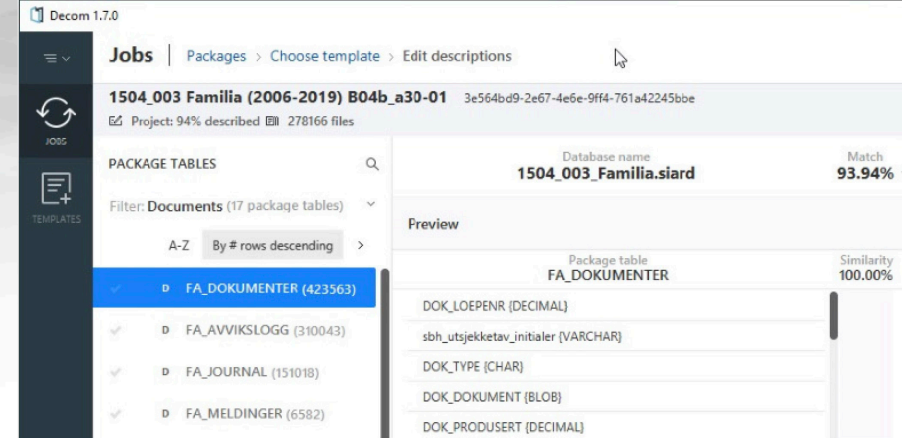
Meldinger:

✓ Ok	Testingen er fullført
✓ Ok	Lager rapport Rapport opprettet
⚠ Feil	WARN Could not create report file in current working directory. Attempting to use a temporary file
ⓘ NB!	Siard-validering t01bcl0b12_dbptk-desktop-2.5.9_int.siard blir validert med Database Preservation Toolkit Developer versjon 2.9.9
✓ Ok	Validerer uttrekket... Utført

Fortid
& notid
for framtid

Arkiv og kulturformidling

SIARD ingest and refining in our Archive



1. The SIARD extraction is imported into our tool [Decom](#), developed by the vendor [Documaster](#) for the KDRS production line.
2. A KDRS Decom Server holds over 60 system description templates that is used to handle the unordered extraction.
3. LOBs in production format is transformed from production to archival document formats (f. ex. doc/rtf to PDF/A)



ika

Interkommunalt arkiv for Møre og Romsdal

Access to preserved SIARD extractions

1. Systems for visualization of preserved SIARD extractions is under evaluation ([DBPTK Desktop/Enterprise](#), [Asta 7](#) from [Stiftelsen Asta](#))
2. Currently just a few requests for Access, but expecting increase
3. Archive handling Access today using a Database & tools like [DBeaver](#), recently testing DBPTK Desktop (and DBPTK Enterprise)

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Experienced challenges preserving Municipal public archives

1. Time is on my side, no it isn't.
2. Finding, prioritizing, extracting and deliver to the archive
3. The users/employees have left or not recall +3 years back
4. Describing the system, user and technical documentation
5. Front end logic lost or hard to do reverse engineering
6. Access fails over time, even administrator password missing

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

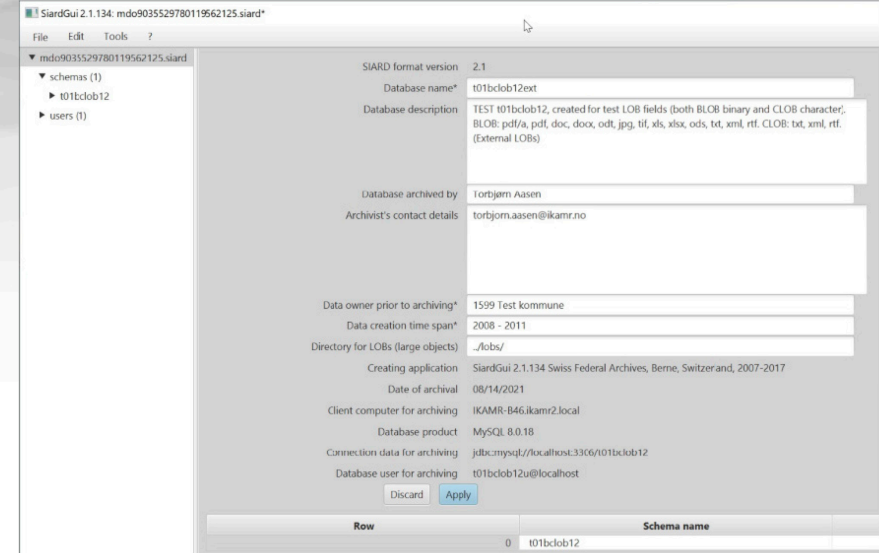
Experienced challenges using SIARD as method

1. Some old Oracle system slices files into 32 kB bits
Solution: Ad-hoc custom made C# code to DB copy with full files
2. Some systems have encrypted documents
Solution: Ad-hoc custom made C# code to decrypt documents
3. Som old systems store binary files into a CLOB text column
Solution: In a copy of DB, transform MSSQL ntext to varbinary

Fortid
& notid
for framtid



Feedback on tool [SIARD Suite](#)



1. Good progress on GUI and console version of 2.1 extractions
2. Good quality in SIARD 2.1 extractions
3. Produced SIARD 1.0 packages proven stable long time preservation
4. Sometimes problem connect to databases (MSSQL and Oracle)
5. Latest version [SiardGui](#) v2.1.134 dated 25.09.2020



ika

Interkommunalt arkiv for Møre og Romsdal

Feedback on tool Database Preservation Toolkit (DBPTK) - 1

1. Very good progress on GUI in [DBPTK Desktop](#) and [Enterprise](#), while keeping console based option in [DBPTK Developer](#)
2. Good quality in SIARD 2.1 extractions (reported [GitHub issue](#) LOBs)
3. Sometimes problem connect to databases (MSSQL)
4. Latest versions DBPTK Developer 2.9.9
(which Desktop 2.5.9 and Enterprise is using)

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Feedback on tool Database Preservation Toolkit (DBPTK) - 2

1. Excellent solution including a SIARD validator
2. SIARD validator need be tuned for Interoperability, approving full SIARD 2.1/2.2 spec (approving tools SIARD Suite and Full Convert)
3. A new SIARD extraction has none or empty Description elements:
The validator now reports every of those as error?
4. <https://github.com/keeps/dbptk-developer/wiki/Validation>

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Feedback on tool Database Preservation Toolkit (DBPTK) - 3

1. Excellent solution including a SIARD viewer (visualization)
2. Wish for an easier way to implement preferred views, to easily and automatically bring up a known system in the viewer (view templates separated from the system itself for reuse?)
3. Wish for easy reuse and display of element Descriptions
4. Our community looks at ways to do 2) & 3) above

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Feedback on tool Database Preservation Toolkit (DBPTK) - 4

DBPTK Desktop

Home Create Manage Preferences Help

Home > Databases > t01bcl0b12 > SIARD edit metadata

Filter sidebar

- Database
- Users & Roles
- t01bcl0b12
 - Tables
 - Views
 - Routines

Database Information

Global information at database level

Name *	t01bcl0b12ext
Archival date *	2021-08-14
Archivist	Torbjørn Aasen
Archivist contact	torbjorn.aasen@ikamr.no
Client machine	IKAMR-B46
Product	MySQL 8.0.18
User	
Data origin time span *	2008 - 2011
Data owner *	1599 Test kommune
Description	TEST t01bcl0b12, created for test LOB fields (both BLOB binary and CLOB character). BLOB: pdf/a, pdf, doc, docx, odt, jpg, tif, xls, xlsx, ods, txt, xml, rtf. CLOB: txt, xml, rtf. (External LOBs)
Producer application	DBPTK Developer (version 2.9.9)

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Feedback on tool Spectral Core Full Convert (SCFC) - 1

1. Very easy to use Windows high quality database migration tool
2. Good quality in SIARD 2.1 extractions
3. Very high successrate in connect and extract to SIARD 2.1
4. Easy create full transaction, warning and error logfiles
5. Latest version Full Convert v21.08.1657

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Feedback on tool Spectral Core Full Convert (SCFC) - 2

Full Convert Pro 21.08.1655 (64-bit) - [C:\uttrekk\scfc_v1655\1500_003_Extens\report\1500_003_Extens_scfc1655_int.fcpx]

COPYING THE DATABASE...

Oracle > SIARD
Admin@EXTENS > C:\uttrekk\scfc_v1655\1500_003_Extens\1500_003_Extens_scfc1655_int.siard

STOP

COPYING DATA

IST.VITNEMAL
○ 44.0 thousand records ○ 1.6 thousand/s

416
of 436 in total
Copying data

RECORDS
23.7 million of 0.0 ○ 1.6 thousand/s

ERRORS
2 errors in 1 table

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal


Feedback on tool Spectral Core Full Convert (SCFC) - 3

Full Convert Pro 21.08.1655 (64-bit) - [C:\uttrekk\sfc_v1655\1500_003_Extens\report\1500_003_Extens_sfc1655_int.fcpx]

← Back to project
Project overview screen.

Show details
See detailed conversion report

Browse target
Examine data in target database.



Your data is now delivered to the target database.

● 99% of tables processed successfully

We had some **errors**, **warnings** you should check out.
See details for more info.

436 tables processed in **en time** with just a single worker used.

Copied 26 million records (in fact, exactly 25,788,313) of data in **en time**.
On average, copied 4.7 thousand records each second.

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Experienced SIARD Interoperability Issues - 1

1. DBPTK doubles 2 x lobFolder elements
+ full table[n].xml file= attribute with full filepath
2. DBPTK will not accept empty schema name element
(this is valid towards metadata.xsd schema and not disallowed)

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Experienced SIARD Interoperability Issues - 2

1. Full Convert empty schema element at default schema SQL server
2. Full convert uses an extra subfolder for each 1000 files when storing Internal LOBs (stored inside the SIARD package)

Ex.: \content\schema0\lob9\seg0\rec2.bin

This is not strictly disallowed in the spec,

but a matter of interpretation => Interoperability

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Experienced SIARD Interoperability Issues - 3

1. <https://github.com/DILCISBoard/SIARD/issues>
2. <https://github.com/DILCISBoard/SIARD/issues/43> Interoperability
3. <https://github.com/keeps/dbptk-developer/issues>
4. <https://github.com/keeps/dbptk-developer/issues/476> LOB double

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

SIARD Interoperability Work in Progress

1. [SIARD-Interoperability 2021-08-14 – Google Disk](#)

Case study the tools DBPTK, SIARD Suite/Gui, Full Convert

2. [KDRS_open_SIARD – Google Disk](#)

Main folder for SIARD analysis (need update SIARD 2.1 and 2.2)

3. [KDRS_open_database – Google Disk](#)

Main folder for KDRS testdatabases and open work

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

SIARD Production Line Work in Progress - 1

1. [KDRS Metadata](#) Analyze SIARD metadata.xml and Decom to Excel
2. [KDRS Query](#) Predefined automated Queries (XML, SQL, JSON)
3. [KDRS SIARD Extdoc](#) Create SIARD extraction from file folders
4. Make note that all tools above are in betarelease, as first release v1.0 need be fully tested. Still useful for internal usage.

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

SIARD Production Line Work in Progress - 2

1. Add automation and pipelines for all steps and elements in the Production Line
2. Automated logging
3. Full integration of Ingest, Preservation and Access

Fortid
& notid
for framtid



ika

Interkommunalt arkiv for Møre og Romsdal

Questions?

Feel free to contact me on any subject for the IT archivist.

1. SIARD standard, tools, extraction, validation, preservation
2. IKAMR/KDRS production line of preservation and visualization
3. Community activities on events or netmeeting 1 on 1

Our common goal is full SIARD Interoperability among all the Tools in our line

Torbjørn Aasen

Mobil: +47 924 65 108

Email: torbjorn.Aasen@ikamr.no

Fortid
& notid
for framtid