



ARKIVVERKET
RIKSARKIVET

Sector-specific, Standardized Long-term Archiving

Arne-Kristian Groven,
Senior Advisor,
The National Archives of Norway,
NorDig,
Stockholm, 25.-26.10 2017

Presentation overview

- The first step should be database preservation (Siard)
- But where lies the challenges?
- Motivations for standardization
- Sector specific challenges
- An outline of our approach and results
- Concluding remarks

But first a few words about the project resulting in this presentation

- Within a cooperative archival programme context called SAMDOK, involving all sorts of Norwegian archival stakeholders (<https://samdok.com/>)
- The final report, in Norwegian, is available here: https://samdokdotcom.files.wordpress.com/2017/09/sluttrapport-2017-samdok-komm_2-metodikk-for-bevaring-fra-kommunale-fagsystemer-v1-0-pr-11-09-2017.pdf
- The project team: Sigve Espeland (project leader), Arne-Kristian Groven (me), Pål Mjørlund, Geir Utmo, and Jørgen Vik-Strandli

Digital preservation (archiving) from relational databases is a solved problem

- Siard (1.0 and 2.0) are preservation formats for relational databases
- At least two different Siard-tools are now available (Siard Suite and DB preservation Toolkit)
- The standard is open, with well-defined semantics, providing proofs of data authenticity and direct access into the data sets

BUT: How to find the most valuable pieces of information contained in Siard files?

- *If a data dictionary is provided* together with the Siard file, use that to identify valuable information
 - But often such data dictionaries doesn't exist or is of poor quality
- *If user documentation with pictures and screenshots of the GUI is provided*, use that to identify valuable information
 - But that is also often of poor quality or lacking
- These «but's» are often the case for older systems and databases

Other important issues include

- How to secure tracability from (presumptive) valuable information into legal requirements for that domain?
- How to secure the same level of quality (of long-term digital archiving) across a domain or sector?
- How to make efficiency gains? We have 426 municipalities in Norway

The sector we used as a case

- The (municipal) health and social services (Pleie- og omsorgssektoren)
- Because it was urgent to move ahead here, in order for archives not to lose valuable information!
- The sector has three different systems, from three different vendors
- The information in the systems is a mix of EPJ (Electronic Patient Journals) and records management (Noark)

Our approach towards sector-specific standardization

- Identify the information types of value (w.r.t. long-term archiving) within the sector
 - derived from legal requirements, standards, previous work etc.

(Top-down approach)

- Create a mapping between data (-base) representation and the identified (valuable) information types

(Bottom-up approach)

Our starting point: The legal requirements, laws, and regulations

1. Lov om helsepersonell m.v.
2. Lov om pasient- og brukerrettigheter (pasient- og brukerrettighetsloven)
3. Lov om behandling av helseopplysninger ved ytelse av helsehjelp (pasientjournalloven)
4. Lov om arkiv
5. Lov om behandlingsmåten i forvaltningssaker (forvaltningsloven)
6. Lov om rett til innsyn i dokument i offentlig verksemd (offentleglova)
7. Lov om behandling av personopplysninger (personopplysningsloven)
8. Forskrift om pasientjournal

Another starting point

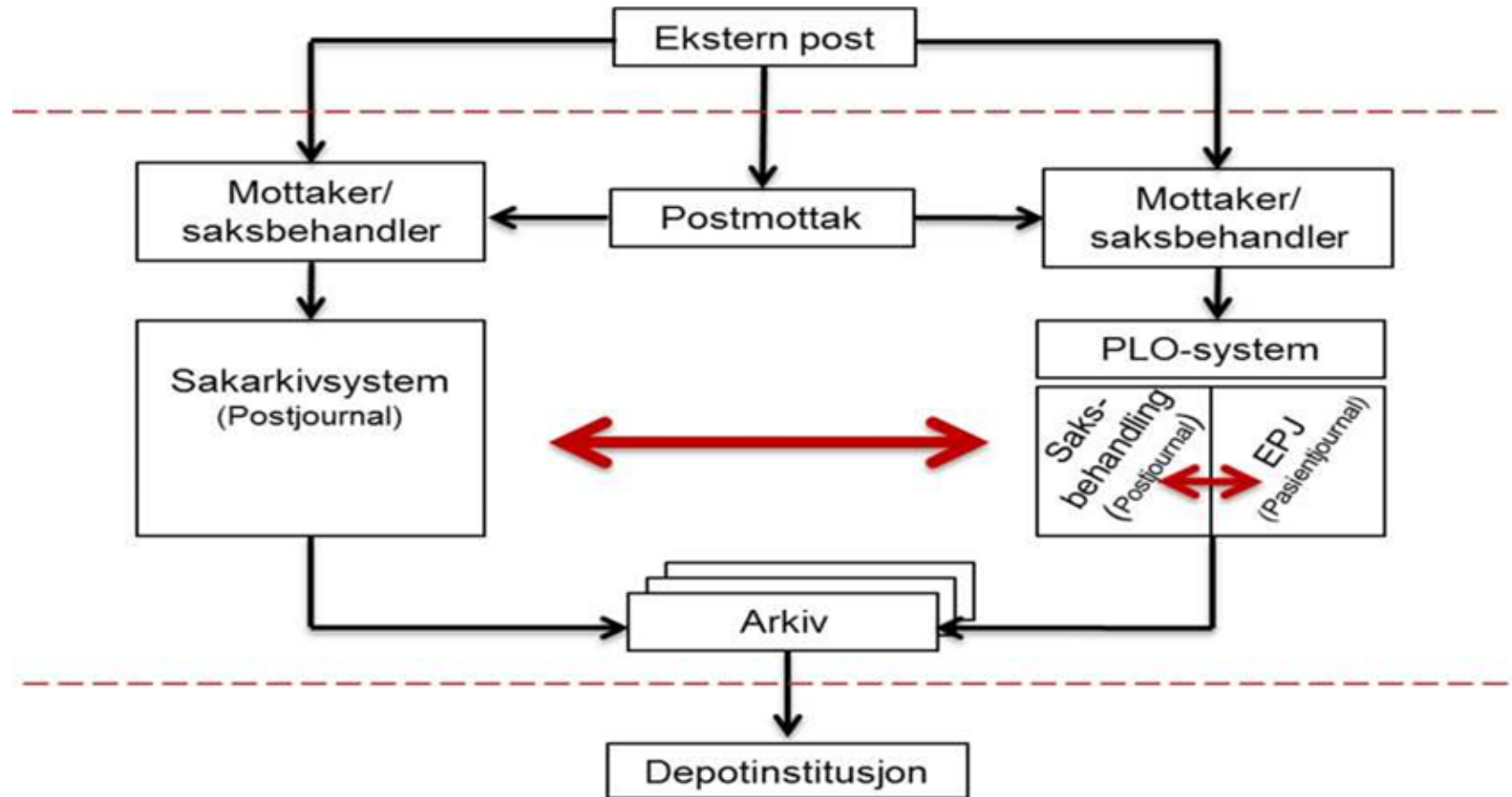
- High quality analytical work already done by inter-municipal archival institutions in cooperations with selected municipalities
- Siard-files from two of the system types and in addition a reversed engineered database model from the third
- Relatively scarce documentation
- Very little help from the vendors

Pre-existing analysis:

Identified information types

- Hjelpemiddelskjema
- Pasientforhold - prosedyrelister
- Pasientforhold - rutinepreget behandling
- Medisinkort/ordinasjonskort
- Hovedkort (navn på bruker/pårørende mm.)
- Utskriving - sykehus/institusjon
- Fagrapporter – sykepleie, tannhelse, fysio/ergo, rehabilitering, optiker, bilder, individuell plan (IP/IOP), mm.
- Journalnotat
- Notat til hjemmesykepleie/sykehjem fra fastlege/allmennlege
- Epikrise
- Prøvesvar
- Røntgensvar
- Trygghetsalarm
- Henvisninger
- Ventebrev/midlertidig svar
- Innkallinger
- Søknad - omsorgstjenester, trygghetsalarm, mm.
- Vedtak - omsorgstjenester, trygghetsalarm, mm.
- ...

Health and social services systems are hybrid: Records management and EPJ



Source: <http://www.ks.no/contentassets/a7aafcc06f034beaa46cab9603183535/veileder.pdf>

Decisions made on how to archive/preserve

- Two options, either to preserve Noark-related information (records management) and Electronic Patient Journal (EPJ) information separately, or to save it together
 - *We strongly concluded to do the latter*
- The next question was then, should we base the resulting digital preservation template on the Noark-standard or the EPJ-standard, or both
 - *We concluded to base it on the EPJ standard*

Examples of identified information types...

- EPJ overviews (10 in number)
 - Viktige opplysninger (ID: 50133)
 - Kartlegginger, vurderinger mv (ID: 50111)
 - Planer og tiltak (ID: 50126)
 - Tverrfaglig dokumentasjon (ID: 50131)
 - Medisinske opplysninger (ID: 50120)
 - Ekstern korrespondanse (ID: 50104)
 - Kontaktoversikt (ID: 50010)
 - Problembasertoversikt (ID: 50128)
 - Dokumentasjon av annet enn helsehjelp (ID: 50102)
 - Egenjournal (ID: 50103)
- EPJ cases (37 in number)
 - Cave (ID: 50001)
 - Underinndeling Cave (ID: 50002)
 - NB! (ID: 50005)
 - Underinndeling NB! (ID: 50006)
 - Reservasjoner og ønsker (ID: 50003)
 - Kartlegginger, vurderinger mv (ID: 50111)
 - Praktiske forhold (ID: 50127)
 - Hjelpemidler (ID: 50107)
 - Informasjon og tilbud til Pasient (ID: 50109)
 - Individuell plan (ID: 50108)
 - ...

Examples of identified information types

- EPJ documents (52 in number)
 - Allerginotat (ID: 11036)
 - Blodtype (ID: 11031)
 - Tidligere sykdom (ID: 11034)
 - Sykdommer/-tilstander i familien (ID: 10027)
 - NB-notat (ID: 10007)
 - Reservasjon/ønske fra pasient (ID: 11018)
 - Krav om reservasjon mot vaksinasjon (ID: 10001)
 - Generelt journalnotat (ID: 11009)
 - Kartlegging av funksjonsnivå (ID: 11013)
 - Omsorgssituasjon (ID: 10002)
- EPJ Fragments (49 in number)
 - Ansvarlig (ID: 2085)
 - Kartlagt funksjon (ID: 2056)
 - Samlet vurdering funksjonsnivå (ID: 2093)
 - Strukturelement (ID: 2097)
 - Journaltekst (ID: 5035)
 - Blodtrykk (ID: 2036)
 - Høyde og vekt (ID: 2046)
 - Tannstatus (ID: 2080)
 - Puls (ID: 2069)
 - Kroppstemperatur (ID: 2059)
 - ...

Comments about the work

- Pre-existing high quality analysis from the project leader, representing an inter-municipal archival institution
 - Identification of (valuable) information types was based on this work
- Studying the standards, laws and regulations involved was quite time consuming
 - Due to the large number involved
- Reverse engineering work to locate the data representation of the identified information elements:
 - A two-day workshop for each of the three systems types databases, with all five project members participating plus an (external) IT-professional at one workshop
 - Systems were based on generated Siard files
 - Reverse engineered visualization of one of the databases
 - Outcome: Identified data representation down to database tables, not fields (columns)

Concluding remarks: Do we need sector specific standardization?

- By standardizing we are able, in this case, to manage database content from three different IT-systems that look very different, in a consistent way
- Siard as a first step, will also strengthen trust in the material, regarding integrity and authenticity
- It supports effectiveness, since it is easier to do the right things
- The investments will be rewarded, since there are (at least) 426 deployments
- Improved traceability, from the information types into laws, regulations, and standards, and into the databases
- Efficiency gains, more work done using less time

Concluding remarks: Do we really need sector specific standardization?

- In some cases there might be reasons (from certain perspectives, maybe?!) for not standardizing on descriptive (meta-) data types within certain domains:
 - E.g. few deployed systems etc.
- BUT, the reason for doing it should be to improve usability for the end-user
 - Improved information searches
 - Improved vocabulary etc.

Concluding remarks: Do we really need sector specific standardization?

- If you only have a Siard file representing a database, you probably will be able to find (some) information of value
- If you in addition have:
 - (Up to date) data dictionaries are present, end/or
 - (Good) end-user manuals and technical documentation are present, and/or
 - Screenshots of various system views from an end-user perspective is present, hopefully annotated (by hand?) with database information
- Then you don't need descriptive information types, but it will be helpful from an end-user perspective

Concluding remarks:

Next step in this work

- This work has prepared the ground for a standardized long-term archiving approach for health and social services in Norwegian municipal sector, but we haven't yet reached the finish line
- QA has to be done, and identification at DB field level, not only table level
- Templates capturing the standardized descriptive datatypes has to be made
- Hopefully, this will be a prioritized task for next year (within the MAVOOD context)