

Utenfor NOARK – håndtering av dokumentorienterte databaser

Prosjektrapport 26. mars 2025



Innholdsfortegnelse

Innledning	3
Begrunnelse for prosjektet	3
Standardisert beskrivelse.....	3
Standarder og ontologi	3
Konkrete funksjonsbeskrivelser	5
Caser – databaseuttrekk	6
Forsøk på case - Lærdommer	6
Case – Dokumentorientert database og metadata i filer	7
Fra internt skjema til EAD - Program	9
Hierarkiet – Mapper	18
Case fil- og mappearkiv.....	19
Fil-/mappearkiv og EAD	19
Records in Contexts (RiC)	20
RiC - Introduksjon.....	20
Fra EAD til RiC	20
RiC-O Converter	23
Konklusjon/sluttord	24
Vedlegg 1: Funksjoner – Malbeskrivelser fagforbund	26
Funksjonsbeskrivelse – aktør ISAAR(CPF).....	26

Innledning

Denne sluttrapporten tar for seg prosjektet «Utenfor NOARK – håndtering av dokumentorienterte databaser». Prosjektet ble utført av Arbeiderbevegelsens arkiv og bibliotek (Arbark). Prosjektleder var Martin Ellingsrud. Prosjektdeltagere var Odd-Rune Hansen og Dag Jostein Juvkam.

Rapporten redegjør for prosjektgruppens arbeid, med et hovedfokus på beskrivelsesdelen av prosjektet, herunder både *Encoded Archival Description* (EAD) og *Records in Contexts* (RiC). Prosjektrapporten vil også beskrive et dataprogram som Arbark har utviklet som del av prosjektet, og som er spesifikt designet for å løse tilgjengeliggjøringsutfordringer Arbark har hatt over en lengre tidsperiode.

Prosjektperioden var januar 2023 – desember 2024 (inkludert rapportskriving til mars 2025). Prosjektet fikk støtte fra Arkivverkets utviklingsmidler både i 2023 og 2024.

Begrunnelse for prosjektet

Arbark har gjennom flere prosjekter støttet av Arkivverkets utviklingsmidler utviklet en god kompetanse på elektroniske arkiver. Det siste store prosjektet, «Fra grasrota til Stortinget – å dokumentere politiske partier i en digital tid», ble avsluttet i 2022. Det hadde høsting/avlevering av politiske partiers arkivmateriale som fokus. Rent overordnet har Arbark lenge konsentrert seg om avlevering og bevaring når det kommer til elektroniske arkiver. Prosjektet «Utenfor NOARK – håndtering av dokumentorienterte databaser» er et nødvendig ledd i å tilgjengeliggjøre elektroniske arkiver, ikke bare bevare dem.

Prosjektmålet var dermed å finne fram til en god og adekvat metodikk for håndtering av data fra private organisasjoners arkivsystemer, spesifikt arkivsystemer uten NOARK-godkjenning.

Standardisert beskrivelse

Standarder og ontologi

Ulike arkivbeskrivelsesstandarder har stått sentralt i prosjektet. OAIS – Open Archival Information System - er rammeverket i arbeidet med bevaring av digitale arkiver. *Encoded Archival Description* (EAD) er en standard for logisk arkivbeskrivelse i OAIS-

baserte arkivpakker. EAD er en XML-implementering av den generelle arkivbeskrivelsesstandarden *General International Standard Archival Description* (ISAD(G)). Begge standarder tar utgangspunktet i det klassiske arkivhierarkiet.¹

I prosjektet jobbet vi også med beskrivelse av aktører. ISAD(G)/EAD beskriver i utgangspunktet bare arkivmateriale, slik at konteksten ikke nødvendigvis kommer med for brukeren (i tilstrekkelig grad). I digitale arkiver kan derfor standardene *Encoded Archival Context – Corporate bodies, Persons and Families* (EAC-CPF) og *International Standard for Describing Functions* (ISDF) være nyttige.² EAC-CPF bygger på ISAAR(CPF), standarden for beskrivelsen av aktører, mens ISDF er et hjelpemiddel for å beskrive funksjoner, aktiviteter og transaksjoner.

EAD og EAC-CPF er begge standarder som brukes i Arkivverkets DIAS-pakkestruktur. Dette gjorde dem til åpenbare valg i prosjektet. Arkade 5 benytter seg også av disse standardene. Arkade 5 kan produsere EAD- og EAC-filer, men er etter prosjektgruppens syn ikke spesielt nyttige. Filene vi fikk ut, hadde kun en ID for selve filen.

Vi har også sett på arkivontologien *Records in Contexts* (RiC). RiC er relativt ny. Utviklingen har pågått noen år, men det var først i november 2023 at det kom en endelig versjon (1.0). En grunntanke bak RiC er å kombinere alle de eldre arkivstandardene til ICA, altså å ha én standard istedenfor fire.³ Semantisk web spiller en stor rolle i RiC. Semantisk web kan defineres slik:

Den semantiske veven er en del av internett med koblede datasett som er lesbare for datamaskiner [...]. Datasettene, som er organiserte samlinger av data, består av enkle påstander som er forståelige for maskiner. Den semantiske veven

¹ International Council on Archives [ICA], *ISAD(G): General International Standard Archival Description*, 2. utgave, 2000, (https://www.ica.org/app/uploads/2024/01/CBPS_2000_Guidelines_ISADG_Second-edition_EN.pdf) ; Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists [TS-EAS], *Encoded Archival Description Tag Library Version EAD3 1.1.2*, 2023, (<https://www.loc.gov/ead/EAD3taglib/EAD3-TL-eng.html#attr-id>).

² Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists og the Staatsbibliothek zu Berlin [TS-EAS & SB], *Encoded Archival Context - Corporate Bodies, Persons, and Families (EAC-CPF) Tag Library*, 2023, (<https://eac.staatsbibliothek-berlin.de/schema/v2/eac.pdf>); ICA, *International Standard for Describing Functions*, 2007, (https://www.ica.org/app/uploads/2023/12/CBPS_2007_Guidelines_ISDF_First-edition_EN.pdf).

³ International Council on Archives Expert Group on Archival Description [EGAD], *Records in Contexts conceptual model*, 2023, (https://www.ica.org/app/uploads/2024/01/ric-cm-1.0_0.pdf).

består ikke bare av datasett, men også av lenker mellom disse datasettene, slik at man får en samling koblede datasett.⁴

Beskrivelsene laget med RiC som mal skal kunne lenkes sammen på kryss og tvers. Primært gjelder dette arkivdata, men en vil også kunne ha muligheten til å kombinere arkivdatasett med andre datasett, for eksempel fra museum- eller bibliotekdomenet. Forutsetningen blir at alle datasettene kan nyttiggjøre seg av semantisk web.

RiC har spilt en sekundær rolle i dette prosjektet. Prosjektgruppen har sett det som fordelaktig bare å utprøve (RiC) for slik å legge et bedre grunnlag for videre arbeid med arkivontologien.

Konkrete funksjonsbeskrivelser

Prosjektgruppen har sett på hvordan funksjonsbeskrivelser for fagforbund kan standardiseres. Fagforbund er en type arkivskaper med samme type funksjoner, for eksempel lønnsforhandlinger. Vi har brukt tidligere prosjekterfaringer og vår egen arkivkunnskap for å konstruere såkalte malbeskrivelser. Disse er funksjonsbeskrivelser som kan brukes *as is*, men kan også utvides etter behov. I prosjektet er det tenkt at disse kan inngå i EAC-CPF-baserte beskrivelser, spesifikt *functions*-delen i EAC-CPF-baserte beskrivelser.⁵ (Se vedlegg 1 for malene.) Men det skal være mulig å bruke dem i andre sammenhenger der funksjoner beskrives. Malene er ikke avhengig av en spesiell type teknologi eller framgangsmåte.

Malbeskrivelsene bærer preg av å være et slags minimum, det er ikke lagt vekt på å lage detaljerte beskrivelser. Her kan nemlig detaljene variere mellom ulike arkivskapere (fagforbund). Vi har dermed funnet det mest hensiktsmessig å lage en slags basis som heller kan utvides. Det er også åpent for at fagforbundene selv kan bruke malene i sine klassifikasjonssystemer/arkivnøkler ved behov i arkivdanningsøyemed.

Tanken er at ved å ta med mer om de konkrete arbeidsoppgavene i en organisasjon, får brukere av arkivene, både Arbarks arkivarer og andre, lettere innganger til arkivmaterialet. Andre arkivinstusjoner kan også bruke malbeskrivelsene for sine

⁴ Gjersdal, A. Den semantiske veven, *Store Norske Leksikon*, 2018, (https://snl.no/Den_semantiske_veven).

⁵ TS-EAS & SB, 2023, s. 82.

arkivbeskrivelser. Arbark har tradisjonelt laget arkiv- og aktørbeskrivelser ved hjelp av Asta. Men i arbeid med digitale arkiver kan vi også se til ISDF, standarden for beskrivelse av funksjoner. Beskrivelse av en organisasjons funksjoner tydeliggjør formålet med de forskjellige funksjonene: Ved å beskrive funksjoner plasserer man dokumentene tydeligere i sin kontekst og forklarer *hvorfor* arkivinformasjonen ble skapt. I tillegg beskriver den *hvordan* dokumentene ble skapt gjennom prosessbeskrivelsene (hva som skjer).⁶ Aktørbeskrivelsene sier da *hvem* som utfører prosessene, mens arkivbeskrivelsene sier noe om *hva* arkivet er.

Arbark legger til rette for at slike beskrivelser kan brukes av andre arkivdepoter og arkivskapere for videreutvikling innen andre kontekster.

Caser – databaseuttrekk

Forsøk på case - Lærdommer

Helt i starten av prosjektperioden arbeidet vi med mottak av Norsk Transportarbeiderforbunds databasearkiv, som ble skapt i tidsrommet 2000-2019. Forbundet gikk i 2019 inn i Fellesforbundet. Arkivet ble skapt i saksbehandlingssystemet Symfoni, som er bygd på Lotus Notes-teknologi. Til slutt klarte vi å få overført en kopi av dataene til Arbark, men det tok såpass lang tid at vi måtte finne et annet case for prosjektet. (Se vedlegg 2 for en grundigere beskrivelse av utfordringene med dette arkivet.)

Her er det noen lærdommer. Den aktuelle databasen var en såkalt dokumentorientert database, med Lotus Notes som teknologisk basis. Denne teknologien er utdatert, så her står man som arkivinstitusjon overfor en utfordring. Det finnes mer moderne dokumentorienterte løsninger i dag, men selve Lotus Notes er i alle fall i norsk kontekst sjelden vare. Gitt at det finnes lite kompetanse, er det vanskelig å lage uttrekk. Fellesforbundet, arvtakeren av materialet, var interessert i å få til et uttrekk. Men det var en prosess preget av stadige forsinkelser.

Dette er trolig mest prekært i privat sektor, der det ikke er noen bevaringsplikt. Fra et arkivfaglig standpunkt blir den gamle lærdommen om teknologiforgjengelighet aktuell.

⁶ ICA, 2007, s. 7.

Selve dataene vil ikke nødvendigvis forsvinne helt, binære data kan fortsatt eksistere på en harddisk eller et annet lagringsmedium. Men det å få dataene til å bli et arkiv som kan tilgjengeliggjøres er en helt annen sak. Status i dag er at vi har en kopi av en virtuell server som har installert databasen med innhold hos oss. Videre arbeid med konvertering/standardisering vil skje utenfor prosjektet. Vi har ikke hatt ressurser til å arbeide videre med basen i 2024.

Case – Dokumentorientert database og metadata i filer

Det endelige caset som til slutt ble valgt, var et uttrekk fra Norsk Arbeidsmandsforbunds forrige database (Lotus Notes-basert). Uttrekket dekker perioden 2007-2015. Dette uttrekket var dels et resultat av prosjektet «Farvel DBS. Utredning av metoder for datahøsting fra private databasesystemer (2016-2018)», et tidligere prosjekt også støttet av Arkiverkets utviklingsmidler. Strukturen er slik at alle mapper har årstall (tosifret) og løpenummer (plass til fem sifre). Sak 07-00005 er da sak fem i år 2007. Denne metadataen finnes i selve navnet på mappen, men det er også egne filer i txt-format som har metadata om selve saksmappene. Metadataene om en sak er i en egen tekstfil kalt sak.txt, en i hver saksmappe. Se tabell under for flere detaljer:

Tabell 1

Metadata saksmappe i sak.txt

Metadatafelt/-type	Kommentar
Saksnr	Samme som mappetittel, f.eks. 07-00004
Tittel	Navn på saken
Opprettet	Opprettelsesdato, format dd.mm.åååå.
Siste aktivitet	Sist noe er endret i saken (for eksempel avslutning av saken), format dd.mm.åååå.
Status	Informasjon om saken fortsatt er pågående eller avsluttet.
Møtedato	Ikke brukt.
Saksbehandler	Navn på saksbehandler

Alternativ saksbehandler	Navn på alternativ saksbehandler (midlertidig tatt over saken eller assisterer hovedsaksbehandler).
Arkivnøkkel	Tallkode og navn på arkivnøkkelkodepunktet, for eksempel '050 MEDLEMSBEVEGELSE'.
Type	Overordnet arkivkode (ensifret og navn på kodepunktet), for eksempel '(0) Administrative saker'.
Lesetilgang	Tilgang til saker, hvem som kan få innsyn i saken internt i forbundet. Eneste status i uttrekket: 'åpen' (alle i forbundet med systemtilgang kan få tilgang).

Disse metadatatypene ble senere i prosjektet *mapnet* til EAD-standarden og Records in Contexts (RiC). Det er flere metadatatyper også, da på dokumentnivå. Hvert dokument har også tilknyttet metadata. Det varierer litt hvordan de er oppsatt: Er dokumentet en e-post, er metadata sammen med selve innholdet i samme txt-fil. Hvis ikke, er de i en separat fil: notater.txt. Følgende tabell viser metadatatypene:

Tabell 2

Metadata for dokumenter i epost-txt og/eller notater.txt

Metadatafelt/-type	Kommentar
Saksnr	Se tabell 1
Journalnr	Løpenummer for dokument basert på år. For eksempel, '008-00215' er det dokument 215 som ble journalført i 2008.
Type	Kan tilsvare for eksempel utgående og inngående dokument i Noark 5-standarden. For eksempel, verdien 'brev (utgående)' viser til et utgående brev. E-

	post vil ha samme system, bare at det står 'epost', ikke 'brev'.
Fra	Hvem som har sendt brevet.
Emne	Tittel på dokumentet.
Innhold	Ser ut til å være et slags notatfelt, for eksempel beskjeder om tidsfrister for svar på høringer o.l.
Status	Status på dokumentet, for eksempel 'avsluttet'. Vil indikere hvor i behandlingsprosessen dokumentet er.
Saksbehandler	Navn på saksbehandler for dokumentet, ofte identisk med saksansvarlig.
Alternativ saksbehandler	Navn på alternativ saksbehandler (midlertidig tatt over behandlingen av dokumentet eller assisterer hovedsaksbehandler).
Vedlegg	Viser til dokumentet metadataen er knyttet til hvis i notater.txt. Hvis ikke, viser det til vedlegg til e-posten.
Fra e-post	Spesifikk type for e-post. Avsender.
Til e-post	Spesifikk type for e-post. Mottaker.
Kopi til e-post	Spesifikk type for e-post. Kopimottaker.

Både tabellen for sak og tabellen for dokument viser typisk postjournalinformasjon. Det er langt fra like omfattende som en fullt utbygd Noark 5-registrering, men det er beskrivende metadata en bruker kan utnytte i søk. Om det er tilstrekkelig, vil avhenge av bruks- og søkescenarioet.

Fra internt skjema til EAD - Program

Planen for prosjektet var å bruke *Access to Memory* (AtoM), et program Arbark har godt kjennskap til gjennom flere prosjekter støttet av Arkivverkets utviklingsmidler. I dette prosjektet har AtoM vært redskapet for å produsere filer etter EAD og EAC-CPF. AtoM er

et brukervennlig program der man fyller inn beskrivende informasjon i felter som samsvarer med ISAD(G), eksempelvis (oversatt til norsk) 'arkivperiode', 'dataomfang', 'innholdsbeskrivelse', 'framfinningsmidler'. Noen felter er obligatoriske, ellers velger man feltene man vil fylle ut. Når beskrivelsen er laget, eksporterer vi den fra AtoM etter EAD-standarden i XML-format og bevarer den sammen med arkivdataene. Eksporten fra systemet blir da EAD selv om en bruker en ISAD(G)-mal for utfylling i AtoMs brukergrensesnitt.

Prosjektgruppen har brukt mye tid på å utvikle software som er viktig for å gjøre denne jobben. AtoM har en mal i CSV som må fylles ut. Deretter blir det utfylte maldokumentet prosessert av AtoM for å skape en EAD-basert beskrivelse.⁷ Arbark har nå et program som gjør at mye av utfyllingen av malen kan gjøres automatisk. Dette er et stort fremskritt, ettersom det effektiviserer beskrivelsesarbeidet. Programmet har vi valgt å kalle MAST (*Metadata Analyzing and Standardizing Tool*). Se vedlegg 3 for en nærmere beskrivelse av programmet og arbeidsflyten det legger opp til ved prosessering av et uttrekk i fil-/mappe-struktur.

Her er det fristende å tenke videre utviklingsarbeid. AtoM har ikke helt «tatt av» i Norge som brukerplattform. Da kan en annen mulighet være å implementere XSDen (et skjema som kontrollerer at strukturen til en XML-fil er etter en gitt standard) til EAD direkte inn i programmet. Dette vil gjøre at vi kan lage beskrivelser direkte i programmet via utfylling av CSV-malen. Her kan en i og for seg se for seg rettelser i programmet i etterkant, altså etter at CSV-filen har blitt brukt som input til softwaren.

En annen mulighet er at programmet kan implementere beskrivelser basert på ontologien *Records in Contexts* (RiC) basert på en lignende metode.

Arbark er noe i tvil om hva som er beste løsning og ser for seg at valg av løsning diskuteres i relevante fora, både internt i Arbark og eksternt med andre relevante aktører, hovedsakelig da i Norge. Prosjektgruppens foreløpige konklusjon er at vi helst ikke vil «oppfinne hjulet» på nytt og lage enda en ny arkivsoftware som helt eller i stor grad overlapper med annen software. Arbark kan ha både MAST og AtoM. Arbark har

⁷ Artefactual Systems, *CSV templates*, uten dato, (https://wiki.accessmemory.org/Resources/CSV_templates#AtoM_2.8_CSV_templates).

også begrenset programmeringskapasitet, så i den grad vi skal gjøre noe mer innen dette området, er det aktuelt for oss å samarbeide med andre. Det hadde også vært interessant for oss å se hvordan ASTA kan benyttes her. Med spesielt den nyeste versjonen, ASTA 7, er det ikke utenkelig å ha andre maler enn den basert på allment arkivskjema.

For å få til en vellykket konvertering til EAD, måtte prosjektgruppen bruke tid på å konvertere den interne leverandørbaserte metadatamalen (eller skjema/*schema*) til EAD. Eller rett og slett gjøre en *crosswalk*: “Metadata crosswalks translate elements and values from one schema to those of another.”⁸

Det er flere problemer med *crosswalks*:⁹

- 1) *Fuzzy matches*. Det er felter i begge skjema som tilsynelatende ser like nok ut, men ikke har helt samme semantiske mening.
- 2) Når det kommer til mediekonvertering, skilles det ikke nødvendigvis godt nok mellom den digitaliserte kopien og det originale dokumentet, selv om det er to forskjellige objekter med ulike egenskaper og metadata. Alt beskrives samlet.
- 3) Metadata i et skjema kan egentlig konverteres eller *mappes* til flere felt i et annet.
- 4) Data i separate felt i det originale skjemaet er i et felt i skjemaet du *mapper* til.
- 5) Det finnes ikke noe passende felt i målskjemaet, så metadataene «dumpes» i et restfelt.
- 6) Den opprinnelige standarden er en miks av flere «standarder». Standarden og veiledninger kan gradvis forandre seg (komme i flere utgaver). Så det

⁸ University of Texas, *Crosswalks*, 2025, (<https://guides.lib.utexas.edu/metadata-basics/crosswalks>).

⁹ Punktene under bygger på Woodley, M. S. *Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information*. Murtha Baca (red.). *Introduction to Metadata*, 2. utgave, 2008, s. 7-9, (https://books.google.no/books?id=qOdVAgAAQBAJ&printsec=frontcover&redir_esc=y#v=onepage&q&f=false).

introduseres forskjeller mellom eldre og nyere registreringer. Disse er kanskje bare uvesentlige nyanser til å begynne med, men kan etter hvert bli større. Dette er ikke aktuelt kun for crosswalk, men også i langtidsbevaringssammenheng. Et depot må ha kontroll på de ulike standardversjonene som er aktuelle for materialet det bevarer.

- 7) *Mapping* fungerer kun den ene retningen. Det vil si at du klarer å mappe skjema 1 til skjema 2, men ikke andre veien, i alle fall ikke like godt, fordi skjema 2 har i tillegg andre typer metadata som ikke er relevante for skjema 1.
- 8) Et skjema er mer komplekst, for eksempel gjennom bruk av et hierarki, mens det andre er enklere, for eksempel ved være «flat» (fravær av hierarki).

Vår største utfordring var først og fremst punkt 5. Det interne skjemaet tok utgangspunkt i norsk saksbehandling og journalføringspraksis (innen fagbevegelsen). EAD er på sin side XML-implementeringen av ISAD(G), som kom i andre og nåværende utgave i 2000. Utgangspunktet for ISAD(G) blir da primært papirarkiv, i mindre grad digitalt materiale. Nå skal det sies at også det interne skjemaet har en viss arv fra papirarkivene: det er dokumentfiler (A4) som arkiveres i mapper, noe vi gjenkjenner fra papirarkiv. Likevel, forskjellen blir saksbehandling og arkivdanning i databasens skjema versus arkivdepot og mer generisk tenkning (jf. internasjonal standard, finne fellesnevner) og arkivdepot i ISAD(G)/ EAD.

Det er verdt å nevne at også aktørbeskrivelsesstandard EAC-CPF (XML-implementering av ISAAR(CPF)) er brukt i prosjektet. I stedet for å komme oppi den allerede nevnte problemstillingen («dumping» i restfelt), kan bruken av en spesifikk aktørstandard utligne dette noe.

Tabell 3

Crosswalk internt databaseskjema til EAD (mappe/file)

Internt skjema fra databasen	EAD-felt (enten attributt eller element i XML)	Kommentar
------------------------------	--	-----------

Saksnr	unitid	<p>Unitid = “an identifier for the materials being described”¹⁰. Dette skal være en unik ID innen samme arkiv (eller <i>fonds</i>). Det skal sies at EAD ikke har en metadatatype som matcher saksnr direkte, en kunne for eksempel tenke seg «case file number» istedenfor.</p> <p>I EAD er utgangspunktet mer generiske kategorier, som identifikator. En mulighet her er å bruke <descriptivenote> som da kan beskrive hva slags type identifikatoren er (saksnr).</p>
Tittel	unittitle	Selvforklarende.
Opprettet	unitdate	<p>Her kan dato for opprettet og siste aktivitet beskrives samlet ved å bruke attributt og attributtverdi slik:</p> <p>unitdatetype="inclusive".</p> <p>Da setter en datoene sammen slik at det blir start- og sluttdato. Et</p>

¹⁰ TS-EAS, 2023, s. 462.

		<p>annet alternativ er å bruke to separate datoer og markere via attributter hva slags datoer de er (start og slutt, from- og todate).</p>
Siste aktivitet	unitdate	Se over.
Status	odd	<p>Her klarte vi ikke å finne et passende element. 'Status' er nært tilknyttet saksbehandling og arkivdanning, og passer dermed ikke helt inn i en mer arkivdepotrettet og generisk standard. Elementet i XML blir da <odd>, definert som <i>other descriptive data</i> "[f]or recording additional information about the described materials that is not easily incorporated into one of the other named elements".¹¹ Her går det også an å sette inn <descriptivenote> som forklarer elementet mer i detalj.</p>
Saksbehandler	Aktør eller <i>agent</i> i EAC-CPF	Kan lage en aktør via EAC-CPF og så bruke relations og <realtionType> for å

¹¹ TS-EAS, 2023, s. 343.

		beskrive relasjonen (her: 'saksbehandler') i fritekst. ¹²
Alternativ saksbehandler	Se over, igjen aktør.	Se over. Dog, her må relasjonen være 'alternativ saksbehandler' eller 'medansvarlig'.
Arkivnøkkel	Selve arkivnøkkelens i sin helhet legges under elementet <fileplan>. ¹³ Deretter kan koden saken har legges under <odd>.	Igjen, litt uheldig at kode ikke kan være tilleggselement under et mer spesifikt og passende element enn restelementet <odd>. Igjen blir det tydelig at EAD ikke er helt 100% tilpasset arkivet som skal beskrives. Det er derimot et pluss at arkivnøkkelens kan legges under <fileplan> (klassifikasjonssystemet). Da er relasjonen noenlunde klar. Igjen, <descriptivenote> kan brukes her og. ID kan også brukes som XML-attributt der selve tallkoden står. ¹⁴ Dette gjelder både for <odd> og

¹² TS-EAS & SB, 2023, s. 163-64.

¹³ TS-EAS, 2023, s. 270.

¹⁴ TS-EAS, 2023, s. 58.

		spesifikke koder i <fileplan>.
Type	Se over og kommentar.	Egentlig det samme som arkivnøkkel, bare overordnet kode. Kan da lages på tilsvarende måte, bare at type (overordnet kode) vil brukes som i <odd>. Eventuelt vil type komme mer indirekte frem ved kun å sette arkivnøkkelkoden (tresifret) i <odd>. Da vil en ved å se på <fileplan> se sammenhengen.
Lesetilgang	odd	Er ingen spesifikk metadatatype for dette i EAD. Elementet er bundet til lesetilgang hos arkivskaper, mens EAD er mer rettet mot innsynsbestemmelser hos depot. Må bli restelement. Muligens heller ikke så relevant i depotsammenheng, men kan røpe noe om en slags klausulering internt (hvem fikk se hva).

EAD har, som tabellen viser, en viss fleksibilitet som må utnyttes for å beskrive et digitalt sakarkiv skapt i en norsk fagforeningskontekst. Likevel, det kommer frem at *odd*-elementet eller «restelementet» må brukes for å få til en tilstrekkelig god nok konvertering. Det er dessverre heller ikke slik at alle metadatatyper i det interne databaseskjemaet har en 100% match i EAD-skjemaet. Da blir en tvunget til for eksempel å legge til mer metadata. Et eksempel er <unitid>, som da for å inkludere alle nyanser må ha mer metadata for å få med at IDen er et saksnummer konstruert etter en bestemt mal. Her er det vel slik at generiske kategorier som ID i standarder kan ha mer spesifikke manifestasjoner eller typer i mer spesialiserte og til dels ustandardiserte IT-systemer.

Tabell 4

Crosswalk internt databaseskjema til EAD (dokument/item)

Internt databaseskjema	EAD-felt	Kommentar
Saksnr	Se tabell 3.	Se tabell 3. Videre utdypning: Saksnummer trenger kun å bli tilknyttet saken. Item/dokument knyttes hierarkisk under saksmappen gjennom XML i EAD-filen.
Journalnr	Unitid	Tilknyttet spesifikt til selve dokumentet eller <i>item level</i> .
Type	Odd	Samme løsning som før med <descriptivenote> som går mer i detalj.
Fra	Aktør/agent i EAC-CPF.	Lage aktør, deretter relasjon og relasjonstype (se tabell 3).
Emne	Unittitle	Samme prinsipp som i tabell 3.

Innhold	Odd	I praksis er dette et slags notatfelt, muligens et felt <odd> faktisk passer til, siden det er så mye som kan registreres her.
Status	Se tabell 3.	Se tabell 3.
Saksbehandler	Se tabell 3.	Se tabell 3.
Alternativ saksbehandler	Se tabell 3.	Se tabell 3.
Vedlegg	ref (reference)	Lenke til dokumentfilen.
Fra e-post	Agent/aktør i EAC-CPF.	Se tabell 3. Igjen, relasjon. Men her trenger man ikke lage nye aktører hvis det er de samme aktørene som er verdier i fra- og saksbehandlerfeltet.
Til e-post	Agent/aktør i EAC-CPF.	Se over.
Kopi til e-post	Agent/aktør i EAC-CPF.	Se over.

I tabell 4 er det store likhetstrekk mellom metadatatypene for mappe (tabell 3) og dokument (tabell 4). Mye går igjen. Dette gjør arbeidet med *mapping/crosswalking* enklere, men igjen ser man de samme utfordringene, for eksempel bruk av *odd*-feltet.

Konklusjonen er derimot entydig: det går an å få til en relativt god konvertering fra databaseskjemaet til EAD.

Hierarkiet – Mapper

Det understrekes at programmet som er utviklet gjennom prosjektet kun har funksjonalitet til å få med arkiv-, serie -og mappenivå. Det er ennå ikke lagt inn funksjonalitet for å få med dokumentnivået (*item level* i ISAD(G) / EAD). Dette er dels et

resultat av at strukturen på metadatafilene var så «uryddig» at vi ikke rakk å lage et program som håndterte det tilfredsstillende. Noen av filene inneholder bare metadata, noen inneholder metadata og peker til dokumentfiler (som er konvertert til PDF/A, og dermed ikke lenger har sitt opprinnelige filnavn), og noen inneholder dokumentdata (som brev.txt og i noen tilfeller notater.txt).

Den andre grunnen er at Arbark anser mottatt/avlevert digitalt materiale som for ferskt eller for nytt til at vi vil tilgjengeliggjøre dokumenter og dokumentbeskrivelser direkte, blant annet av personvern hensyn. Tanken er at brukeren må søke innsyn. Dermed er det viktig at beskrivelsene er gode nok til at brukeren kan vurdere relevans, om enn på et mer overordnet nivå enn hvis brukeren hadde hatt direkte tilgang til dokumentene. Aktørbeskrivelsen, som i mindre grad enn arkivbeskrivelsene kan automatiseres, må klart få frem hva slags oppgaver aktøren/arkivskaperen har hatt. Oppgavene skal da reflekteres i arkivmaterialet.

Brukeren vil kunne benytte arkiv-, aktørbeskrivelsen og fritekst-/fulltekstsøk for søking straks innsyn er innvilget. Arbark ønsker å få til fritekstsøk eller *information retrieval* (IR) i dokumentbaserte arkiver.¹⁵

Case fil- og mappearxiv

Fil-/mappearxiv og EAD

I utgangspunktet var ikke fil- og mappearxiv dette prosjektets hovedansvarlig. Men helt siden vi måtte se bort fra basen til Norsk transportarbeiderforbund, har prosjektet vært opptatt av beskrivelse av fil- og mappearxiv. I beskrivelsene er det strengt tatt som oftest ikke mer enn navn på filer og mapper, dog også noen datoer som kan ses i Windows utforsker. Selv om datoer fort kan endre seg selv om dokumentet ikke har gjort det, er det de metadatatypene vi har å forholde oss til, i tillegg til selve arkivskaperen (aktørbeskrivelse i EAC-CPF/ISSAR(CPF)-format).

¹⁵ For en inngående diskusjon om IR og arkiv, se Andresen, H., Fremfinning og bruk av digitalt skapt materiale i arkivdepotene, *Norsk Arkivforum*, 25, 2019, (<https://oda.oslomet.no/oda-xmlui/bitstream/handle/10642/7420/Fremfinning-bruk-arkivdepot-NA25.pdf?sequence=1&isAllowed=y>).

Hierarkiene i slike arkiver er ikke nødvendigvis veldig avanserte. Men der databasen vi brukte i prosjektet egentlig kun har arkivnivå og deretter mappenivå, kan fil-/mappearkiv ha et varierende antall mappenivåer og filer under hvert nivå.

Til tross for den ikke helt optimale metadatasituasjonen, har det vært en prioritet for Arbark å få slike arkiver beskrevet for å øke tilgjengeligheten. Også i slike arkiver ser vi ikke for oss at digitale dokumenter og dokumentbeskrivelser «legges ut» i første omgang. Med tiden kan vi nok ha en noe mer «friere» tilgjengeliggjøring. Foreløpig har vi kun funksjonalitet for å få med mappene. Mappene i selve arkivet kan «tagges» som arkiv, serie eller mappe avhengig av hva som passer. Det kan for eksempel tenkes at mapper med «øverste» arkivkode blir serier og underliggende blir mappe etter EADs standard. Dette er altså en lignende metodikk som er utviklet i forbindelse med databaseprosjektet.

Records in Contexts (RiC)

RiC - Introduksjon

Records in Contexts (RiC) ble også en del av prosjektet. Dette er *International Council on Archives* sin nye arkivontologi.¹⁶ Oppgaven i prosjektet har vært å konvertere fra EAD til RiC. Arbark tenker ennå ikke at EAD skal forlates uten videre, gitt at det er en DIAS-standard. RiC er også såpass ny, noe som gjør at det enn så lenge ikke er gode arbeidsverktøy som gjør nytte av ontologien. Likevel har det vært viktig for prosjektgruppen å få noen erfaringer med RiC som Arbark kan bruke videre. Hvis RiC virkelig blir stor i internasjonal sammenheng, kan det hende at ontologien blir mer aktuell i Norge. Foreløpig er Arbarks inntrykk at RiC har et visst europeisk nedslagsfelt, noe som igjen kan gjøre den interessant for norske arkivdepoter på sikt.

Fra EAD til RiC

Igjen har det vært nødvendig med en mapping eller crosswalk, denne gangen fra EAD/EAC til RiC. Her gjaldt det å finne riktig predikat. RiC er laget med en semantisk web-tankegang i bunn. Den grunnleggende byggesteinen i semantisk web er URler (Universal Resource Identifier). De skal identifisere ressurser. For å være meningsbærende, må URler ha konkrete verdier tilknyttet seg. For eksempel vil en URI

¹⁶ EGAD, 2023.

som representerer en dokumenttittel, ha en verdi koblet til seg (jf. predikat). Verdien viser hva dokumenttittelen faktisk er. Videre, URler kan tilhøre ulike klasser, for eksempel klassen 'dokumenttittel'. Klasser har i standarder osv. stor bokstav først, mens predikater har liten bokstav først.

Ved hjelp av URler, klasser og predikater kan en da lage tripler, grunnstenen i RDF (Resource Description Framework). Kort sagt vil det tilsi at alle beskrivelsesmomenter har tre eksplisitte momenter: subjekt, predikat/verbal og objekt. Et gjenkjennelig eksempel er at Fellesforbundet *er skaper av* Fellesforbundets arkiv. RiC krever altså en definisjon av relasjoner. Det skal sies at ASTA ikke er så langt fra denne tankegangen. I ASTA er det mange relasjoner som kan defineres som setninger med subjekt, predikat og objekt. Forskjellen er at ASTA er best tilrettelagt flere nivåer, altså den hierarkiske arkivbeskrivelsen. RiC skal i større grad prøve å være multirelasjonell, altså ha ulike beskrivelsesenheter koblet sammen på kryss og tvers, ikke bare hierarkisk. Hierarkier kan lages, men i RiC har de en mindre plass enn i tradisjonelle arkivbeskrivelsestradisjoner.

Tabell 5

Crosswalk EAD og RiC – Mappe

Det understrekes at mapper i RiC vil klassifiseres som rico:RecordSet. Et recordset (arkivaggregering) vil da mer spesifikt klassifiseres som for eksempel mappe eller enkeltarkiv.

EAD-felt (enten attributt eller element i XML)	RiC-O	Kommentar
Unitid	rico:Identifiser (S), (P), rico:hasOrHadIdentifiser(P)	ID. Kan være mulig å bruke rico:IdentifiserType for å slå fast at det for er en type ID ('saksnummer').
Unittitle	rico:Title	Rimelig selvforklarende.

Unitdate	rico:Date(S), rico:hasEndDate (P), rico:hasCreationDate(P)	Dato for start og slutt.
Odd	rico:generalDescription(P), leder til fritekst.	Her klarte vi ikke å finne et passende element. 'Status' er det elementet vi vil skal inn i en RiC-beskrivelse. Det samme gjelder 'lesestilgang' (hos arkivskaper). Da blir det fort at rico:generalDescription blir restelementet.
Aktør eller <i>agent</i> i EAC-CPF	rico:Agent, eventuelt rico:Person eller rico:CorporateBody avhengig av om det er person eller organisasjon (S). rico:hasCreator (P) for mappe for saksbehandler. Alternativ saksbehandler kan få rico:hasOrganicProvenance (P) som bindeledd mellom seg selv og mappe.	Her kan en få med saksbehandler og alternativ saksbehandler. Likevel, for alternativ saksbehandler må en bruke rico:hasOrganicProvenance fordi denne generelle forbindelsen rett og slett er 'ledig', klar til bruk. Det ville vært uheldig med to rico:hasCreator-forbindelser når saksbehandler tross alt er hovedansvarlig for saken (hovedskaper). Her kan en også kanskje se at det er brukt skaper vitner om en typisk depottankegang (arkivskaper osv.). Her er

		det egentlig snakk om saksbehandlere, så det hadde vært ideelt om det kunne fås fram i RiC. Men det er dessverre ingen andre gode nok erstatninger i RiC.
Fileplan, odd (arkivkoder)	rdfs:Literal (P), rico:classification	Her kan koden kobles direkte til mappe. Det er lite behov for at hele arkivnøkkelen beskrives samlet først, og så kobles til mappe i eget element, som i EAD. Overordnet kode (såkalt type i opprinnelig database) kan også kobles til mappe direkte gjennom rico:classification. Skal en være riktig avansert, kan rico:Identifiser brukes for tallkode til arkivnøkkelkode (som så kobles konkret til tekstverdi gjennom bruk av rico:hasOrHadIdentifiser Identifiser).

RiC-O Converter

Det franske nasjonalarkivet har utviklet et program som skal kunne konvertere EAD-filer til RiC. Av et noe sparsomt utvalg per dags dato valgte Arbark å benytte dette verktøyet. Sannsynligvis kommer det flere verktøy for RiC, både for konvertering og direkte

produksjon (danne RiC-filer direkte uten EAD-filer o.l.). Det forutsetter at RiC blir en ontologi som tas i bruk av mange nok arkivdepoter. RiC-O Converter er utviklet i programmeringsspråket Java. Programoperasjoner startes via kommandolinje, noe som trekker ned brukervennligheten for eksempel vis-a-vis Arkade 5.¹⁷

Prosjektgruppen er delt i vurderingen av softwaren. På den ene siden er det positivt og uten tvil en stor fordel at det eksisterer en løype fra EAD til RiC. På den andre siden, uten justeringer for å modifisere output, altså RiC-filen som er selve arkivbeskrivelsen, blir ikke resultatet helt tilfredsstillende. Prosjektgruppen ønsker flere muligheter eller valg underveis, ikke bare et skript som «går på autopilot». Det kan være at dette delvis kan løses ved å bruke mer tid på konfigurasjon og kildefiler til konverteren, men vi valgte å begrense hvor dypt vi dykket ned i disse mulighetene.

Likevel, verktøyet var instrumentelt for å oppnå prosjektmålet om konvertering fra EAD til RiC. Arbark har produsert en viss mengde EAD-beskrivelser. Prosjektgruppen ser per dags dato (februar 2025) ikke noe poeng å avvikle EAD fullstendig, den er mer veletablert enn RiC (ergo gode verktøy som AtoM). Den er enn så lenge også en DIAS-standard. Foreløpig er det dermed utelukket å bruke RiC i tilgjengeliggjøringsøyemed, men vi vil følge nøye med på den videre utviklingen av standarden. Foreløpig er RiC-O Converter det verktøyet vi bruker å gå fra EAD til RiC.

Konklusjon/sluttord

Prosjektet har gitt oss viktige lærdommer. I eventuelle fremtidige prosjekter der ren programmering er en stor del av prosjektet, må Arbark samarbeide med flere institusjoner, først og fremst i Norge, og/eller opparbeide mer kompetanse på området. MAST ble en større belastning på våre IT-ressurser enn forutsatt, da flere trinn i utviklingsarbeidet avdekket nye nivåer av kompleksitet som måtte håndteres programmatisk. Det har imidlertid gitt oss verdifulle erfaringer og en dypere forståelse av problemstillinger både ved uthenting og sammenstilling av metadata fra arkiver, og prosesser i et utviklingsarbeid som innebærer programmering. Likevel, nå som programmet foreligger, er det et uttrykt mål å videreutvikle det. Arbark behøver

¹⁷ Archives Nationale, *RiC-O Converter*, uten dato, (<https://github.com/ArchivesNationalesFR/rico-converter>).

arkivbeskrivelser for digitalt skapt materiale. Å ha slike verktøy er av stor nytte for det fremtidige tilgjengeliggjøringsarbeidet.

Forhåpentligvis kan andre arkivinstitusjoner bruke funksjonsbeskrivelsene uten for mye modifikasjoner. Gitt at det ikke er noen teknologisk terskel her, vil arkivdepoter trolig finne beskrivelser de kan gjenbruke og eventuelt videreutvikle, også for beskrivelsesarbeid i ASTA.

Utover publisering av prosjektrapporten ønsker Arbark å formidle prosjektets resultater på arkivkonferanser og workshops for metodeutvikling. Arbark vil sende inn forslag til kommende aktuelle konferanser. Det er også mulig med egenarrangerte workshops hvis andre bevaringsinstitusjoner er interessert i erfaringsdeling rundt dette temaet.

Vedlegg 1: Funksjoner – Malbeskrivelser fagforbund

Funksjonsbeskrivelse – aktør ISAAR(CPF)

Funksjonsbeskrivelsen bygger her på felter i ISAAR(CPF)-standarden, spesifikt 5.2.5 - *funksjoner, yrker og aktiviteter* og 5.2.6 – *mandat/autoritetskilder*. Mandatet vil innebære kilden/grunnlaget til funksjonsutøvelsen, ofte vedtekter. I noen tilfeller kan den også bygge på et eksplisitt samtykke (behandling av personopplysninger, jf. GDPR). Funksjonsbeskrivelsen beskriver på sin side selve funksjonen og hva denne er.

I en viss grad er malbeskrivelsen også inspirert av ISDF, siden den har et godt kontekstområde for funksjonsbeskrivelser, se spesielt punkt 5.2.2 - *beskrivelse* og 5.2.3 – *historikk*. Beskrivelsen gjenbraker også en del fra Grasrota-prosjektet (2020-2022), fullført med midler fra Arkivverkets utviklingsmidler. Her ble det ved utarbeidelse av bevaringsveiledningene tatt utgangspunkt i et funksjonsbasert syn. Siden både partier og fagforbund er medlemsorganisasjoner med betydelig innslag av politisk virksomhet, er det mye i Grasrota-veiledningene som kan gjenbrukes.

Beskrivelsene er ment å være maler, de kan gjøres om ved behov for å være tilpasset en spesifikk arkivskaper. De er med andre ord en slags minimumsbeskrivelse.

Begrunnelsen for funksjonene følger i stor grad Shepherds og Yeos tankegang om en styringsfunksjon.¹⁸ Men en fagforenings virksomhet dreier seg i stor grad om å forsvare medlemmenes interesser, se her funksjon 1 og 2. Disse vil kreve en egen funksjon, siden de er såpass sentrale og ikke minst særegne for arbeidstakerorganisasjoner. I tillegg vil forbundene alltid ha flere medlemmer og holde på dem, en slags rekrutteringsfunksjon (se funksjon 4). Et forbund vil også ha et informasjons- og påvirkningsarbeid som går utover selve medlemsmassen (funksjon 5). Det vil også skje politikktutvikling, her kan man si at det her er tenkt mest på de fagspesifikke utvalgenes virksomhet. Altså, et fagforbunds mening om for eksempel yrkesfaglig utdanning eller andre områder et fagforbund kan fatte interesse for som går utover selve forhandlingsfunksjonen.

¹⁸ Shepherd & Yeo, *Managing records*, Facet Publishing, 2003, s. 51.

Funksjonene er nok ikke helt perfekt fordelt, materiale kan alltid sies å tilhøre flere funksjoner. Arbark er på sin side klar på at funksjonsbeskrivelsene er tilfredsstillende når det kommer til et fugleperspektiv på et fagforbunds virksomhet.

Funksjon 1

Funksjonsnavn	Beskrivelse
Forhandle med arbeidsgiver	<p>Fagforbundet forhandler med arbeidsgiversiden for å oppnå nye overenskomster eller forlenge allerede eksisterende.</p> <p>Forhandlingene kan være sentrale eller lokale (hver arbeidsplass).</p> <p>Hovedoppgjør er tariffoppgjør der man reviderer tariffavtaler som har gyldighetstid over flere år. I motsetning til mellomoppgjør, kan krav og tilbud i hovedoppgjør omfatte alle tariffavtalens bestemmelser (snl.no).</p> <p>Mellomoppgjør er en betegnelse på revisjon av tariffavtaler med gyldighetstid over flere år. I motsetning til hovedoppgjør, omfatter krav og tilbud i mellomoppgjørene som regel bare lønnsjusteringer (snl.no).</p>

Funksjon 2

Funksjonsnavn	Beskrivelse
Løse arbeidslivsrelaterte (konflikt)saker	Fagforbundet fører saker på vegne av enkeltmedlemmer i forskjellige saker, for

	<p>eksempel yrkesskader eller konflikter med kolleger og/eller overordnede.</p> <p>Denne funksjonen gjelder altså saker som dreier seg om det enkelte medlemmet i ulike konfliktsituasjoner, ikke lønns-/avtaleforhandlinger.</p>
--	---

Funksjon 3

Funksjonsnavn	Beskrivelse
Styre/lede fagforbundet	<p>Denne funksjonen tar for seg ledelsen av fagforbundet. Det inkluderer avgjørelser fattet på landsmøter og styremøter, eventuelt også i arbeidsutvalg. Dette er altså virksomheten i de styrende organene i fagforbundet. Funksjonen omfatter også å vedta strategier, samt landsmøtets vedtaksprosess.</p> <p>Innenfor denne funksjonen skjer også selve nominasjonsprosessen til den valgte ledelsen. Altså, nominasjonskomiteens møter med kandidater og diskusjon av kandidatenes kvalifikasjoner. Ikke minst, det endelige forslaget. Selve valget foregår på landsmøtet, et organ tilknyttet «styre/lede fagforbundet».</p>

Funksjon 4

Funksjonsnavn	Beskrivelse
Forvalte medlemsmassen	<p data-bbox="824 310 1341 401">Denne funksjonen er bred og kan deles i flere deler eller underfunksjoner.</p> <p data-bbox="824 478 1354 999">Underfunksjon 1 dreier seg om å rekruttere medlemmer, men også holde på dem (medlemspleie). Det inkluderer informasjonskampanjer og -tiltak med det formål å rekruttere og holde på medlemmer, herunder også materiellproduksjon. Det inkluderer også utvikling av medlemsfordeler, for eksempel innen forsikrings- og banktjenester.</p> <p data-bbox="824 1077 1354 1272">Underfunksjon 2 går på fastsettelse av hvem som kan bli medlem (regelverk), og ekskludering og suspensjon av medlemmer.</p> <p data-bbox="824 1350 1321 1608">Underfunksjon 3 dreier seg om å øke kompetansen til tillitsvalgte (så de kan bidra til å forvalte medlemmene og ivareta deres interesser) og andre gjennom kursvirksomhet.</p> <p data-bbox="824 1629 1321 1881">Kursvirksomheten kan også ha et yrkesfaglig fokus, ikke nødvendigvis begrenset til fagforeningsvirksomhet. Altså, kurs som er rettet mer mot selve profesjonsvirksomheten og</p>

	<p>yrkesutøvelsen, i mindre grad forhandlinger og lignende.</p> <p>Underfunksjon 4 går på å informere medlemmer, for eksempel gjennom nettside eller medlemsblad. Dette handler også i en viss grad om å gjøre medlemmene oppmerksomme på saker som betyr noe for forbundet, men også skape sympati for forbundets synspunkter (se funksjon 5).</p> <p>Underfunksjon 5 dreier seg om å utføre medlemsundersøkelser av forskjellig grad, for å få vite hva de synes om forbundet eller medlemstilbudene, men også andre saker, for eksempel politiske.</p>
--	---