

Statusrapport for ArkN4

pr. 21. desember 2011

Hallstein Bakken
Seksjon for digitalt depot
Riksarkivet

0. Innledning

0.1 Formål med rapporten

Når Riksarkivet foretar en versjonsoppgradering av ArkN4, som er det offisielle testverktøyet for Noark 4-arkivuttrekk, blir ny versjon lagt ut på Riksarkivets Web. Samtidig legges ut ”Statusrapport for ArkN4”. Denne inneholder en detaljert oversikt over det gjenstående kjente forbedringspotensialet. For de som ønsker en mer detaljert oversikt over punkter som er løst siden forrige versjon, henvises til dokumentet ”Endringslogg for ArkN4” på samme forum.

Kravspesifikasjonen for ArkN4 som opprinnelig lå i dette dokumentet ble fjernet herfra i forbindelse med utleggingen av versjon 1.3. Den finnes nå som eget dokument ”Utvidet kravspesifikasjon for ArkN4” på samme forum.

Årsaken til at Web-versjonen har hoppet fra versjonsnr. 1.3 til 1.8, er at Riksarkivet under den siste oppgraderingsfasen har operert med flere testversjoner internt.

De punkter som fremgår av denne statusrapporten er gjenstående kjente problemer.

0.2 Vesentligste endringer i versjon 1.3

- a. Overgang til ny parser (**XML::Twig**). Et uttrekk av størrelsesorden 1 GB går sju ganger raskere enn før. Et på 3,7 GB går 28 ganger raskere.
- b. Flere forbedringer i fremmednøkkelkontrollen.
- c. Flere feilmeldinger har fått nyanserte tekster.
- d. Fjerning av unødvendige repetisjoner av feilmeldinger.
- e. Rettet testing for attributter i *kursiv* i Noark 4-standard.

0.3 Vesentligste endringer i versjon 1.8

- a. Forbedret filreferansekontrollen i arkivuttrekk med elektroniske dokumenter. Valideringen av filer som inneholder dokumentreferanser er blitt 500 – 800 ganger raskere.
- b. Fremmednøkkelkontrollen ytterligere forbedret, bl.a. ved at rader som inneholder ugyldige verdier eller mangler obligatorisk innhold blir importert i MySQL-databasen.
- c. Implementert test på riktig dokumentstruktur iht. Avleveringsbestemmelsene, § 8-33.
- d. Forbedret test og behandling av blanke tegn i nøkkelfelt.
- e. Forbedret kontroll ved forekomster av attributt med manglende innhold.
- f. Forbedret kontroll av antall journalposter i NOARKSAK og JOURNPOST og analyse mot SAKDOK.
- g. Implementert kontroll på at arkivuttrekket minimum inneholder tabellene iht. O-kravet.
- h. Forbedret kontroll av attributter med datatypen BOOL.
- i. Forbedret kontroll av attributter med datatypen INTEGER.
- j. Korrigert utskrift av ”æ”, ”ø” og ”å” der de var feil i testrapporten.
- k. Forbedret feilmeldingshåndteringen, bl.a. fjernet overflødig informasjon i testrapporten i de tilfelle ingen avvik er funnet.
- l. Tekstlige rettelser/forbedringer i testrapporten.

0.4 Bruk av fargekoder

For å kunne benytte den opprinnelige statusrapporten fra 5. mai 2009 samtidig som historikken ivaretas, er det valgt å innføre fargekoder i teksten. Følgende fargekoder er tatt i bruk:

Svart.....: Punkter som er samlet opp (RA) i versjoner t.o.m. 1.1. **Ikke utført.**

Lysblå.....: Punkter oppdaget under tester (eksternt) foretatt med versjon 1.1. **Ikke utført.**

Brun.....: Punkter oppdaget under utviklingen av versjon 1.3. **Ikke utført.**

Blå.....: Punkter oppdaget under tester (RA) foretatt med versjon 1.3. **Ikke utført.**

Grønn.....: Punkter oppdaget under tester (eksternt) foretatt med versjon 1.3. **Ikke utført.**

I de fleste tabellene er det en kolonne for status. Teksten i denne er en beskrivelse som gjelder arbeidet i forbindelse med 1.8-oppgaderingen. Under konklusjoner fremkommer det en blanding av vurderinger foretatt av programskaper og Riksarkivet. Merk at det ikke er konkludert for alle punkter.

0.5 Nummerering

Underpunkt-benevningen på nivå to (f.eks. 3.3) og tre (f.eks. 3.6.1) er beholdt, selv om dette kan synes forvirrende i forhold til hovednummereringen etter at Kravspesifikasjonen ble skilt ut som eget dokument. Det er gjort for å beholde referansen til nummereringen i den opprinnelige statusrapporten fra mai 2009. Punkt som senere er kommet til følger også denne nummereringen.

I forbindelse med oppgradering til versjon 1.8 er det også innført betegnelser av type 1.A.1, 1.A.2 osv. Disse gjenspeiler prioritert arbeidsrekkefølge for hvert av de tidligere punkt.

Underpunkt-benevnelser i mindre font (10) henviser til meldingsnr. i testrapporten fra ArkN4.

1. Forbedringer som må gjennomføres

1A.4	Det er feil i punkt «4.0 Analyse av hull i saksnummerrekkefølge.» (I versjon 1. 8 er dette blitt pkt. 5.0).	Status																																																																								
Beskrivelse	<p>3.6.3 Det er en feil i punkt ”4.0 Analyse av hull i saksnummerrekkefølge.” at antall hull oppgis til 0.</p> <p>Nedenstående eksempel er hentet fra uttrekk fra Sjøfartsdir.</p> <p>4.0. ANALYSE AV HULL I SAKSNUMMERREKKEFØLGE</p> <table border="1"> <thead> <tr> <th>År</th> <th>Antall</th> <th>Første ID</th> <th>Siste ID</th> <th>ID/sak</th> <th>Ant.Hull</th> </tr> </thead> <tbody> <tr> <td>----</td> <td>-----</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>1994</td> <td>26094</td> <td>199400001</td> <td>199426241</td> <td>1.01</td> <td>0</td> </tr> <tr> <td>1995</td> <td>28671</td> <td>199500001</td> <td>199528820</td> <td>1.01</td> <td>0</td> </tr> <tr> <td>1996</td> <td>28038</td> <td>199600002</td> <td>199628165</td> <td>1.00</td> <td>0</td> </tr> <tr> <td>1997</td> <td>31642</td> <td>199700001</td> <td>199731812</td> <td>1.01</td> <td>0</td> </tr> <tr> <td>1998</td> <td>26517</td> <td>199800002</td> <td>199826634</td> <td>1.00</td> <td>0</td> </tr> <tr> <td>1999</td> <td>23621</td> <td>199900002</td> <td>199923768</td> <td>1.01</td> <td>0</td> </tr> <tr> <td>2000</td> <td>23090</td> <td>200000002</td> <td>200023252</td> <td>1.01</td> <td>0</td> </tr> <tr> <td>2001</td> <td>38306</td> <td>200100002</td> <td>200138448</td> <td>1.00</td> <td>0</td> </tr> <tr> <td>2002</td> <td>28772</td> <td>200200002</td> <td>200229078</td> <td>1.01</td> <td>0</td> </tr> <tr> <td>2003</td> <td>21441</td> <td>200300002</td> <td>200321844</td> <td>1.02</td> <td>357</td> </tr> </tbody> </table> <p>ID Hopp SA.ID Fra SA.ID Til</p>	År	Antall	Første ID	Siste ID	ID/sak	Ant.Hull	----	-----					1994	26094	199400001	199426241	1.01	0	1995	28671	199500001	199528820	1.01	0	1996	28038	199600002	199628165	1.00	0	1997	31642	199700001	199731812	1.01	0	1998	26517	199800002	199826634	1.00	0	1999	23621	199900002	199923768	1.01	0	2000	23090	200000002	200023252	1.01	0	2001	38306	200100002	200138448	1.00	0	2002	28772	200200002	200229078	1.01	0	2003	21441	200300002	200321844	1.02	357	Vurdert, Ikke utført
År	Antall	Første ID	Siste ID	ID/sak	Ant.Hull																																																																					
----	-----																																																																									
1994	26094	199400001	199426241	1.01	0																																																																					
1995	28671	199500001	199528820	1.01	0																																																																					
1996	28038	199600002	199628165	1.00	0																																																																					
1997	31642	199700001	199731812	1.01	0																																																																					
1998	26517	199800002	199826634	1.00	0																																																																					
1999	23621	199900002	199923768	1.01	0																																																																					
2000	23090	200000002	200023252	1.01	0																																																																					
2001	38306	200100002	200138448	1.00	0																																																																					
2002	28772	200200002	200229078	1.01	0																																																																					
2003	21441	200300002	200321844	1.02	357																																																																					

Her angis fremmednøkkelfeil for JP.SAKSDEL, ArkN4 ser ut til kun å sjekke omverdien i JP.SAKSDEL finnes igjen i feltet OV.ORDNVERD. Det sjekkes ikke om verdien til feltet SA.OPSAKSDEL i den tilhørende saken har verdi, og dermed sjekkes det heller ikke på om kombinasjonen av SA.OPSAKSDEL+JP.SAKSDEL finnes igjen OV.ORDNPRI+OV.ORDNVERD. Det er disse nøklene som er fremmednøkkel og kandidatnøkkel, og som det skal sjekkes på.

Konklusjon Slik den er beskrevet her lar det seg ikke gjøre å implementere en generell sjekk

JP.SAKSDEL peker til. Attributtet har i alle tilfeller kravtype A, og siden definisjonen av referansen er uklar har jeg valgt å fjerne denne sjekken. Alle data blir importert, og hvis JP.SAKSDEL er tatt med blir det ikke sjekket om innholdet i raden stemmer med data i andre tabeller.

1A.11	4.1.15 Ny feil i kontroll av postnummer	Status																						
Beskrivelse	<p>Eksempel fra Sjøfartsdir.:</p> <p>6.8.4. Fremmednøkkelfeil</p> <p>Feltet 'AM.POSTNR' som refererer til 'PO.POSTNR' i 'POSTNR' manglet, eller inneholdt, en verdi, som gjorde raden ugyldig i henhold til NOARK4. Antall feil totalt: 513537. Antall feil som blir skrevet ut: 10.</p> <table> <thead> <tr> <th>Attributt = verdi</th> <th>Kravtype</th> </tr> </thead> <tbody> <tr> <td>AM.POSTNR = ''</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '4045'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '5001'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '6002'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '1701'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '4601'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = ''</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '7053'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '0130'</td> <td>O1 => O1</td> </tr> <tr> <td>AM.POSTNR = '4800'</td> <td>O1 => O1</td> </tr> </tbody> </table> <p>Her kommer ut feilmelding både på postnummer som ikke finnes i postnr.tabellen (blank, 4800), og postnummer som finnes (de andre i eksemplet). I fremmednøkkelkontrollen for AK.POSTNR i ADRESSEKP kommer også på tilsvarende måte feilmeldinger for postnr. som finnes.</p>	Attributt = verdi	Kravtype	AM.POSTNR = ''	O1 => O1	AM.POSTNR = '4045'	O1 => O1	AM.POSTNR = '5001'	O1 => O1	AM.POSTNR = '6002'	O1 => O1	AM.POSTNR = '1701'	O1 => O1	AM.POSTNR = '4601'	O1 => O1	AM.POSTNR = ''	O1 => O1	AM.POSTNR = '7053'	O1 => O1	AM.POSTNR = '0130'	O1 => O1	AM.POSTNR = '4800'	O1 => O1	Vurdert, ikke aktuell
Attributt = verdi	Kravtype																							
AM.POSTNR = ''	O1 => O1																							
AM.POSTNR = '4045'	O1 => O1																							
AM.POSTNR = '5001'	O1 => O1																							
AM.POSTNR = '6002'	O1 => O1																							
AM.POSTNR = '1701'	O1 => O1																							
AM.POSTNR = '4601'	O1 => O1																							
AM.POSTNR = ''	O1 => O1																							
AM.POSTNR = '7053'	O1 => O1																							
AM.POSTNR = '0130'	O1 => O1																							
AM.POSTNR = '4800'	O1 => O1																							
Konklusjon	Denne feilen er fjernet som en konsekvens av at feilmeldinger vedr. Postnummer blir filtrert bort. Fremmednøkkel sjekken er uansett forbedret og skrevet om som en konsekvens av andre forbedringer av håndteringen av feilmeldinger.																							

1A.1 3	4.1.33 Manglende Case sensitivitet i MySql-basen	Status
<p>6</p> <p>Beskrivelse</p>	<p>Uttreksleverandør har meldt følgende: ...I tillegg er mysql-basen case insensitive</p> <p>Dette medfører at f eks følgende uoff-kolonner i praksis ses på som samme verdi</p> <pre> <TGHJEMMEL> <TH.TGKODE>P</TH.TGKODE> <TH.UOFF>Offl § 13</TH.UOFF> <TH.AGKODE>U</TH.AGKODE> <TH.ANVEND>Taushetsplikt</TH.ANVEND> </TGHJEMMEL> <TGHJEMMEL> <TH.TGKODE>P</TH.TGKODE> <TH.UOFF>Offl § 13</TH.UOFF> <TH.AGKODE>U</TH.AGKODE> <TH.ANVEND>Taushetsplikt</TH.ANVEND> </TGHJEMMEL> <TGHJEMMEL> <TH.TGKODE>P</TH.TGKODE> <TH.UOFF>offl § 13</TH.UOFF> <TH.AGKODE>U</TH.AGKODE> <TH.ANVEND>Taushetsplikt</TH.ANVEND> </TGHJEMMEL> </pre> <p>Arkn4 ville i dette tilfellet importert de to første til mysql-basen, men den tredje ville gitt sql'feilmelding under import fordi offl § 13 kræsjer med Offl § 13.</p> <p>Etter validering vil det bli rapportert ett tilfelle av at Offl § 13 allerede finnes, fordi valideringen ikke skiller mellom Offl § 13 med ett mellomrom og Offl § 13 med to mellomrom.</p>	<p>Vurdert, avvist</p>
<p>Konklusjon</p>	<p>Det stemmer at MySQL per default er satt opp slik at sammenlikning av tekststrenger ikke er case sensitive når de blir utført i databasen. Feilen beskrevet over inntreffer ved import av data når MySQL sjekker om nøkkelen for en rad eksisterer fra før.</p> <p>Det gjøres kontroll av data før insert i ArkN4 der det sjekkes om nøkler er unike. Siden ArkN4 er case sensitive blir raden godkjent av ArkN4, men senere avvist av MySQL.</p> <p>Før det kan implementeres en løsning må Arkivverket ta en prinsippavgjørelse om nøkkelfelter (i eksempelet over skrevet manuelt inn av saksbehandler) skal ta hensyn til case eller ikke.</p> <p>Når konklusjon er på plass må enten ArkN4 eller MySQL oppdateres slik at riktig prinsipp følges.</p>	

2. Forbedringer som bør gjennomføres

2.1	4.1.13 Kontroll om tabeller er tilstede iht. O2-krav	Status
Beskrivelse	For N 4-uttrekk med elektroniske dokumenter bør det legges inn kontroll iht. O2-kravet dvs. at de 39 definerte tabellene er med. Hvorvidt uttrekket er uten/med elektroniske dok. bør ArkN4 kunne finne ut selv, evt. kan gis inn som parameter?	Ikke utført
Konklusjon	Ikke realitetsbehandlet	
2.2	Endre overskriften på rapporten til: "TESTRAPPORT FOR ARKIVUTTREKK".	Status

2.3	4.1.5 I meldingen "Fremmednøkkelfeil": Her bør det legges ut linjenummer til posten i den tabellen meldingen refererer til.	Status
Beskrivelse	SSIA	Vurdert, avvist
Konklusjon	Ikke mulig å implementere uten større omskriving. Fremmednøkkelsjekken baserer seg på innholdet i databasen, ikke XML-flene.	

2.4	4.1.7 Rapporten bør ha sidetall.	Status
Beskrivelse	(Iflg. programskaperen tar dette noe tid. Hvor aktuelt er det siden vi normalt sender testrapporten på CD, og de forskjellige arkivskapere kanskje åpner denne i forskjellige lesere?).	Vurdert, avvist
Konklusjon	Sidetall er avhengig av printer, papirark og fontstørrelse. Rapportene vi genererer vil ikke ha faste sidetall før de blir etterarbeidet, eller skrives ut. Anbefalingen til programskaperen er å etterarbeide testrapportene i LibreOffice eller Microsoft Office og lagre de som PDF, eller tilsvarende. Eventuelt kan vi velge å implementere PDF som outputformat direkte.	

2.10	4.1.31 Sammenslåing av attributter under validering av TGMEDLEM	Status
Beskrivelse	<p style="text-align: center; color: green;">En uttrekksprodusent har meldt om følgende tilfelle:</p> <p style="text-align: center; color: green;">Når arkn4 validerer TGMEDLEM, slår den sammen PG.PEID og PG.GRUPPEID til en verdi F eks</p> <pre style="text-align: center; color: green;"> <TGM EDLEM> <PG.PEID>667</PG.PEID> <PG.GRUPPEID>0</PG.GRUPPEID> <PG.INNMAV>0</PG.INNMAV> <PG. FRADATO>20080122</PG. FRADATO> </TGMEDLEM> <TGM EDLEM> <PG.PEID>66</PG.PEID> <PG.GRUPPEID>70</PG.GRUPPEID> <PG.INNMAV>0</PG.INNMAV> <PG. FRADATO>20080218</PG. FRADATO> </TGMEDLEM> </pre> <p style="text-align: center; color: green;">Her blir resultatet av PEID + GRUPPEID 6670 for to personer, selv om de er to separate personer, og dette rapporteres som feil i arkn4</p>	Ikke startet
Konklusjon		

3. Forbedringer som kan gjennomføres

4.1.32 Forslag fra en uttrekksprodusent: Forbedret funksjonalitet

Hvis dere tar imot forslag til funksjonalitet i arkn4, har jeg ett forslag som ville vært hjelpsomt for meg i enkelte tilfeller, og det ville vært om arkn4 hadde hatt en mulighet til å avbryte en import, for så å fortsette med samme xml senere. For å konkretisere: Det ser ut som om arkn4 tar filene i en avlevering i en gitt rekkefølge hver gang, og jeg har også inntrykk av at import, reparering og validering er tre helt separate prosesser. Hvis arkn4 f eks allerede har importert sak.xml og jpost.xml, men jeg av en eller annen grunn må avbryte midt i avsmot.xml (dette skjer faktisk ganske ofte) så hadde det vært glimrende om man kunne lagt til ett parameter for hvilken fil man skal gjenoppta prosessen fra. F eks ved å gi avsmot.xml som ett innparameter, så hopper arkn4 over steget med å gjenopprette tabeller, hopper over import av xml inntil den kommer til filen som er gitt som parameter. Når den kommer til denne, kan den ta en delete from <avsmottabellen>, for så å fortsette import som normalt.

3.6.7 Feilmelding om uoverensstemmelser mellom filnavn

”2.1 .x. Forskjell på filnavn i NOARKIH og i avlevering
Filnavnet i NOARKIH stemte ikke med filnavnet i avleveringen for flere filer. Filnavnene er:

Linje	Navn i NOARKIH	Navn i avlevering
298	NOARKSAK.XML =>	NOARKSAK.xml
856	AVSMOT.XML =>	AVSMOT.xml

Osv.”

Disse tilfellene skulle ikke vært skrevet ut som feilmelding da det nå godtas kombinasjoner av store og små bokstaver i filnavn (jf. kommende kommentarer og veiledning til Avleveringsbestemmelsene § 8-32). Meldingene kan komme som info.

NB! Dette punktet er erstattet av pkt. 4.2.4 og 4.2.7

3.6.4 Unødvendig melding:

Dersom medfølgende DTD-er ligger i samme mappe som XML-filene, skrives følgende melding:

”2.1 .x. Ekstra filer i avleveringen

Følgende filer i avleveringen er ikke definert i NOARKIH: ’ARKIV.DTD’ osv.”

Denne meldingen burde ikke forekomme. Det er ikke noe krav i Avleveringsbestemmelsene om at DTD-ene skal følge med Noark 4-uttrekk. Riksarkivet mottar uttrekk både med og uten disse. Det forutsettes imidlertid at de offisielle DTD-ene fastsatt av Riksarkivaren benyttes ved uttrekksgenereringen (jf. Avleveringsbestemmelsene § 8-23). Uansett er det disse som brukes av ArkN4.

3.6.8 Skrivefeil i rapporten: ”6.xx.x. Mangler verdi i felt som krever verdi. **Forekomsten** mangler i feltet ’xx.xxxxxxxx’”. Er feilaktig registrert som ”NY” i estimatet nedenfor!

Tabell	Attributt	Beskrivelse	Status/ estimat
		Feilmelding for «Mangler verdi i felt som krever verdi. Her står det «forekomsten» istedenfor «forekomsten»	

3.6.9 Skrivefeil i rapporten: ”6.xx.x. Feil innhold i attributt

Attributtet ’PR. STDROLLE’ **innhold** en ugyldig verdi i forhold til datatypen ’BOOL’. ” Bør stå **inneholder**.

3.6.10 Forbedring av utseende for overskriftslinje: Ta vekk punktumet i linja ”4.0

Analyse av hull i saksnummerrekkefølge.” i tillegg til å endre til STORE bokstaver da dette er en overskriftslinje. Se forøvrig pkt. 4, underpkt. 4.1.4.

4.1.6 Forslag til nytt utseende for forsiden av rapporten: Standardisering av forside og innledning på testrapportene fra alle Arkivverkets testverktøy.

4.2.1 Kravspesifikasjonen pkt. 18 <...>:

ArkN4 kontrollerer saksposter, men mangler kontroll på om det finnes journalposter som ligger utenfor den oppgitte arkivperioden. Må forholde seg til problematikken rundt mykt/skarpt periodeskilte. (Hva m/saker i dagens ArkN4?)

4.2.3 Kravspesifikasjonen pkt. 14:

Opptellingen bør grupperes på årstallet i journaldato istedenfor eller i tillegg til årstallet i løpenummeret.

4.2.4 Kravspesifikasjonen pkt. 15 og 16:

Opptellingen bør grupperes på årstallet i journaldato istedenfor eller i tillegg til årstallet i løpenummeret.

4.1.3 Tydeligere overskrifter i testpunkt 3 må fortelle om man refererer til tabellfiler eller rapportfiler.

4.2.9 Rette feilmelding om attributter som mangler i XML-fil

Under pkt. 2.0 KONTROLL AV INFORMASJON I NOARKIH har følgende feilmeldinger forekommet (Dir. for Naturforvaltn.):

2.21. POSTNR

2.21.1. Attributter mangler i XML-fil

Attributtene som manglet hadde alle kravtype K=A, eller så kan de

være tomme iht. NOARK4 (O=NULL). Det er ikke kritisk at attributtene mangler i tabellene når de er nevnt i NOARKIH, men det anbefales at kun attributter som er brukt i tabellene blir tatt med i NOARKIH.

Følgende attributter er definert i NOARKIH, men mangler i avleveringen

Ekstra Attributter	Krav type
--------------------	-----------

2.21.2. Fant attributter som ikke skal avleveres

Følgende attributter som ikke skal være med forekom i avleveringen, enten i NOARK.IH, eller i tabellfila. I kravspesifikasjonen del 2 er disse markert med kursiv i K-feltet

Attributt	Hvor
<i>PO.KOMMUNE</i>	<i>NOARK.IH POSTNR</i>
<i>PO.KOMNR</i>	<i>NOARK.IH POSTNR</i>
<i>PO.NEDLDATO</i>	<i>NOARK.IH</i>
<i>PO.OPRDATO</i>	<i>NOARK.IH</i>
<i>PO.POSTNR</i>	<i>NOARK.IH POSTNR</i>
<i>PO.POSTSTED</i>	<i>NOARK.IH POSTNR</i>

Dette er et spesielt tilfelle idet alle attributtene i en tabell (POSTNR) er merket med kursiv, dvs. at filen ikke skal avleveres. I dette tilfelle kan det være en løsning å la pkt. 2.21.1 (som også mangler attributtene PO.NEDLDATO og PO.OPRDATO i dette tilfelle) med alle meldingstekstene utgå. Et annet alternativ er at attributtene skrives ut selv om de ikke skal avleveres. Det siste er kanskje å foretrekke da det også er forekommet tilsvarende eksempel der noen attributter ikke skal avleveres i en tabell som forøvrig er avleveringspliktig (eksempel fra Senter mot etnisk diskriminering).

I tillegg er meldingstekstene under pkt. 2.21.2 mangelfulle. Det som mangler er avmerket i rødt. I de tilfelle alle attributter i en tabell er merket i kursiv betyr det tabellen ikke er avleveringspliktig, og da kunne kanskje dette også fremgå av meldingsteksten.

4.1.18 Forbedre feilmelding under 1.0 GYLDIGHETSSJEKK AV XML-FILENE

(Eksempel fra Oslo Byarkiv - Doculive-uttrekk)

1.1.2. Elementer i feil rekkefølge

Elementer kommer i feil rekkefølge i henhold til NOARK-4. 1. Dette gjør

at valideringen feiler. Feilen er ikke kritisk, men genererer en

feilmelding. Rekkefølgen var:

I avlevering	Riktig rekkefølge
EL.ARKSKAPER	EL.ARKSKAPER
EL.ORGNR	+
EL.SYSTEMNAVN	EL.SYSTEMNAVN

+ EI. ORGNR

EI.KOMMUNE		EI.KOMMUNE
EI.FRADATO	+	
EI.TILDATO	+	
EI.PRODDATO	+	
EI.MERKNAD		EI.MERKNAD

+ EI.FRADATO

I kolonnene ”I avlevering” og ”Riktig rekkefølge” bør meldingene gjøres litt ryddigere.

4.1.24 Manglende meldinger for attributter som ikke er definert i Noark-4

ArkN4 skriver ikke ut melding av typen
 x.x.x. Fant attributter som ikke finnes i Noark-4,
 se. 2.7.2 i feilrapp

for AD.RAPUTV i ARKIVDEL

4.1.27 Forbedre meldinger hvor attributter som ikke er definert i Noark-4 forekommer i flere tabeller

Eksempel fra Sjøfartsdir:

2.2.1. Fant attributter som ikke finnes i Noark-4

Attributter som ikke finnes i NOARK-4 ble funnet i avleveringen. Disse var:

Attributt	Hvor	
-----	-----	
AI.AKTIV	tabell	Her bør listes den/de tabell(er) hvor attributtet forekommer! NOARKIH (i dette arkivuttrekket) ADMINDEL (i dette arkivuttrekket)

I samme uttrekk, pkt. 2.20.1, forekom tilsvarende problemstilling for ett attributt (som takles tilfredsstillende):

Attributt	Hvor
-----	-----
JP.AVSREF	NOARK.IH

4.1.33 Manglende utskrift av attributter ved meldingstype «Attributter mangler i XML-fil»

Eksempel fra prøveuttrekk:

2.5. ARKIVDEL

2.5.1. Attributter mangler i XML-fil

Attributtene som manglet hadde alle kravtype K=A, eller så kan de

være tomme iht. NOARK4 (O=NULL). Det er ikke kritisk at attributtene mangler i tabellene når de er nevnt i NOARKIH, men det anbefales at kun attributter som er brukt i tabellene blir tatt med i NOARKIH.

Følgende attributter er definert i NOARKIH, men mangler i avleveringen

Ekstra Attributter	Krav type
-----	-----

De attributtene som ArkN4 finner manglende i ARKIVDEL skrives ikke ut i meldingen. Tilsynelatende er årsaken at de på filen er tomme, men representert kun som slutt-tag (eks.: <AD.BSPRINS/>). Her kunne ArkN4 skrive meldingen f.eks. på følgende format:

Ekstra attributter	Krav type
AD.BSPRINS	K=O, O= (Forekommer kun som som slutt-tag)

ArkN4 må her holde styr på om attributtet er representert som slutt-tag på hele filen. På filer der et attributt noen steder forekommer med verdi, andre steder med kun slutt-tag skal ingen melding skrives ut.

4.1.34 Manglende utskrift av attributter ved meldingstype «Attributter mangler i XML-fil»

Eksempel fra prøveuttrekk fra Senter mot etnisk diskriminering (i siste gangs uttrekk):

2.21. JOURNPOST

2.21.1. Attributter mangler i XML-fil

Attributtene som manglet hadde alle kravtype K=A, eller så kan de være tomme iht. NOARK4 (O=NULL). Det er ikke kritisk at attributtene mangler i tabellene når de er nevnt i NOARKIH, men det anbefales at kun attributter som er brukt i tabellene blir tatt med i NOARKIH.

Følgende attributter er definert i NOARKIH, men mangler i avleveringen

Ekstra Attributter	Krav type
-----	-----

2.21.2. Fant attributter som ikke skal avleveres

Følgende attributter som ikke skal være med forekom i avleveringen, enten i NOARK.IH, eller i tabellfila. I kravspesifikasjonen del 2 er disse markert med kursiv i K-feltet

Attributt	Hvor
<i>JP.UILDATO</i>	<i>NOARK.IH</i>
<i>JP.UILTIL</i>	<i>NOARK.IH</i>

Her kunne også ArkN4 i pkt. 2.21.1 skrevet ut hvilke attributter som er definert i NOARKIH. I dette tilfelle kommer forklaringen forsåvidt i pkt. 2.21.2, men siden man normalt leser rapporten i kronologisk rekkefølge ville dette ha gjort meldingen klarere. Mulig teksten da burde endres til: «Følgende attributter er definert i NOARKIH, men skal ikke avleveres».

Dette punktet bør muligens ses på sammen med 4.1.33.

Tabell	Attributt	Beskrivelse	Status
		I filen AVSMOT.XML i et arkivuttrekk er det et ulovlig tegn på linje 16 113 888, 0x1F Hvordan forholder vi oss til ulovlige/ugyldige tegn i XML?	AVK

Tabell	Attributt	Beskrivelse	Status
		Skal dette løses i ArkN4 eller f.eks. ved preprosessering?	

	Håndtere XML-filer som er delt i mindre filer bedre	Status
Beskrivelse	<p>Hva: Når filer i avleveringen, rapporter eller tabeller, er delt i flere deler avdekker vi for få potensielle feilsituasjoner. Vi bør håndtere dette skikkelig.</p> <p>Hvor: Mottatt arkivuttrekk</p> <p>Lite viktig</p>	NY

4. Forbedringer som er kjekt å ha

4.1.23 Antallsangivelse ved melding ”Feil innhold i attributt”

Ved meldinger av typene

6.X.X. Feil innhold i attributt
Attributtet 'AA..AAAAA' inneholdt en ugyldig verdi i forhold til datatypen 'BOOL'/TIME'/INT'
kan totalt antall forekomster skrives i teksten i de tilfelle de overskrider 10.

4.1.25 Skrivefeil i rapporten: Rette opp forskjellig skrivemåte for NOARKIH.

I meldingen ”2.x.x. Fant attributter som ikke skal avleveres” brukes betegnelsen NOARK.IH istedenfor NOARKIH.

3.6.5 Noen steder skrives det ut for mange repetisjoner av samme type feilmelding. (Rettet noen steder – se pkt.5).

3.6.6 I feilmelding for manglende JOURNAL- og SAKDOK-oversikt brekkes linj ene feil.

4.1.1 Planer om å nyansere punktet ”8.0 OPPSUMMERING” bedre og flytte det til begynnelsen av rapporten.

(I versjon 1. 8 er dette blitt pkt. 9.0).

4.1.4 Benytte samme skrifttype på samme nivå.

I dag forekommer for eksempel:
5.0 Test...
6.0 IMPORT

4.2.6 Rette feilmelding ved feil benevning av tabellnavn

Under pkt. 2.0 KONTROLL AV INFORMASJON I NOARKIH har følgende feilvarianter forekommet:

2.1.1. Ugyldig tabellnavn i TABELLINFO

Det ble funnet et eller flere ugyldige tabellnavn. Disse var:

'AdminDel' i TABELLINFO på linje 12.
o.s.v.

.....

2.1.2 Ekstra filer i avleveringen:

Følgende filer i avleveringen er ikke definert i NOARKIH:

'ADMINDEL.XML'

O.S.V.

.....

Når TABELLINFO som her inneholder tabellnavnet **AdminDel** bestående av en kombinasjon av store og små bokstaver (som ikke er godkjent iht. Noark 4-standarden), er det korrekt at ArkN4 skriver ut feilmeldingene som vist i under pkt. 2.1.1. I dette tilfelle medfølger det uttrekket i virkeligheten en XML-fil benevnt ADMINDEL.xml. Derfor er det korrekt at ArkN4 skriver ut feilmeldingene som vist i pkt. 2.1.2, men for dette punktet må ArkN4 rettes slik at filendelsen xml blir skrevet ut med små bokstaver.

4.2.7 Styrke robusthet ved feil benevning av tabellnavn

Her refereres til de viste tilfellene i punkt 4.2.6.

Dersom navnene i TABELLINFO korresponderer med de medfølgende filnavn (sett bort fra Case-problematikken både for navn og filendelse) slik som i dette tilfelle, og (hvis vi unntar nevnte problematikk) er i henhold til Noark-4. 1, bør ArkN4 fortsette den normale behandling og ikke som nå avslutte kun med å etablere en tom MySQL-base.

Feil med tegn i XML som ikke er gyldige iht. XML-standarden

Beskrivelse	<p>Type feil: Velformet</p> <p>Antall: Flere feil av samme type</p> <p>SystemID: D:\uttrekk\Test\JOURNPST.XML</p> <p>Engine name: Xerces</p> <p>Severity: fatal</p> <p>Description: An invalid XML character (Unicode: 0x1a) was found in the element content of the document.</p> <p>Start location: 23692:20</p> <p>Vurdering;</p> <p>Feilen oppstår i produksjonsbase og ikke ved konvertering, altså bruk av uttrekksmodul. Etter Google søk ser jeg av flere rapporterer at denne feilen oppstår når man kopierer inn tekst direkte fra MS Word. Tegnet "ø" blir liggende som</p> <p>Feilen er alvorlig i forhold til produksjonsbasen. I forhold til uttrekk er feilen uheldig, men ikke alvorlig da den lett kan fikses manuelt i uttrekk / xml filer. (søk og erstatt eksempelvis)</p>	NY
--------------------	---	-----------

Tiltak	<ul style="list-style-type: none"> •Implementere preprosessering av XML-filene som vasker ut alle ulovlige tegn iht. XML-standarden. •Genererer feilmelding i rapporten som forteller hvor de ulovlige tegnene ble funnet. 	
Konklusjon		

-		
Beskrivelse	<p>Type feil: velformet</p> <p>Antall: Flere feil av samme type</p> <p>SystemID: D:\uttrekk\Test\TILLEGG.XML</p> <p>Engine name: Xerces</p> <p>Severity: fatal</p> <p>Description: Element type "Ragnhild.x" must be followed by either attribute specifications, ">" or ">".</p> <p>Start location: 132147:145</p> <p>End location: 132147:162</p> <p>Vurdering:</p> <p>I produksjonsbasen eSak benyttes <> tegnene for å markere en epost adresse i aktuelt felt - <ragnhild.x@hotmail.com></p> <p>Feilen skyldes at produksjonsbasen benytter tegn som er reservert i henhold til XML standarden. Ved uttrekk blir ikke disse tegnene konvertert til html.entities slik Noark 4 standarden anbefaler, og resulterer i at XML dokumentet ikke validerer i forhold til syntaks. (dette skal meldes inn til leverandør)</p>	NY
Tiltak	<ul style="list-style-type: none"> •Oppdag og rett feilen i preprosessering av XML-filene •Generer feilmelding i rapporten som forteller hvor feilen var og at den kunne rettes. 	
Konklusjon		

4.2.5 Postnummer-tabellen.: Hvordan bruker ArkN4 postnr-tabellen? (Alle attributtene her står i *kursiv* under K-kravene, d.v.s. at det ikke kreves at arkivuttrekket omfatter denne tabellen).

En fullgod løsning her ville være at RA fra Postverket får en postnummer-tabell med fra-til gyldighet for alle postnr. Det må da lages et eget opplegg (DB-løsning?) med ajourhold ved endringer fra Postverket.

4.2.10 Mer funksjonalitet i pkt. 5.0 Test av periodisering.

Her bør det testes hvordan første og siste sak i NOARKSAK stemmer med det som står i 0.2 EKSPORTINFO under: Periode fra – til. Evt. avvik skrives ut. Uansett resultat er dette et egnet sted å legge inn en fast melding som forteller at det skal foretas en manuell kontroll av INFO.TXT for å se om de funn ArkN4 har avdekket samsvarer med opplysninger om start og slutt av arkivperioden i INFO.TXT.

4.2.11 Ytterligere styrking av robusthet ved feil benevning av tabellnavn

Under pkt. 2.0 KONTROLL AV INFORMASJON I NOARKIH har følgende feilvarianter forekommet (Dir. for Naturforvaltn.):

2.1.1. Ugyldig tabellnavn i TABELLINFO

Det ble funnet et eller flere ugyldige tabellnavn. Disse var:

'AdresseKP' i TABELLINFO på linje 31.

O.S.V.

.....

2.1.2 Ekstra filer i avleveringen:

Følgende filer i avleveringen er ikke definert i NOARKIH:

'ADRESSEK.XML'

O.S.V.

.....

I dette tilfellet som er en annen variant av pkt. 4.2.6, er også meldingene korrekte da navnene ikke samsvarer selv om en ser bort fra av Case-problematikken. For slike varianter kunne robustheten til ArkN4 styrkes ved å få den til å fortsette behandlingen. Kunne muligens funnet makker blant XML-filene ved å gå inn og hente <!DOCTYPE i hver fil og så sjekket mot tabellnavnet i TABELLINFO. Punktet er lavt prioritert og det bør lages et tidsestimat før evt. arbeid igangsettes.

4.1.26 Melding om fremmednøkkelfeil kompletteres/nyanseres

I eksemplet under er tekstene i blått hentet fra et testet uttrekk. Forslag til endringer er uthevet i rødt. Forbedringene bør gjelde alle meldinger av denne type.

6.31 NOARKSAK

6.31.1. Fremmednøkkelfeil

Feltet 'SA.STATUS' som refererer til 'SS.STATUS' i 'SAKSTATUS' manglet,

eller inneholdt, en verdi, som gjorde raden ugyldig i henhold til

NOARK4. Antall feil totalt: 17. Antall feil som blir skrevet ut: 10.

Attributt = verdi	Linje Meldingstekst	Kravtype
SA.STATUS = "U"	9999 Verdi finnes ikke i 'SAKSTATUS'	O => O
SA.STATUS = "	9999 Felt manglet	O => O

Se Noark-4.1 Del II, pkt. 14.2..5 Saksstatus (s. 20): Følgende faste verdier inngår (i SS.STATUS): R, B, A, X, U, KU. I tillegg kan det legges inn egne verdier.

4.1.29 Fremmednøkkelkontroll: 'AM.KORTNAVN' i AVSMOT mot 'AK.KORTNAVN' i ADRESSEKP

Dette punktet omfatter en antatt nyansering av det ArkN4 tester i dag, men før evt. utførelse må det sjekkes nærmere. Se pkt. 4.1.28.

I Noark-4. 1, Del II, side 51 står følgende under «Merknader» for AK.KORTNAVN: «Må være utfyllt dersom ikke ID benyttes som unik nøkkel. Det tillates ikke to poster med samme kortnavn».

Ut fra den gjengitte formulering vil mulige tolkninger her være:

1. ID trenger ikke å bli benyttet som unik nøkkel. I dette tilfelle må AK.KORTNAVN være utfylt og bestå av unike verdier.
2. ID benyttes som unik nøkkel. En implisitte konklusjon vil da være at AK.KORTNAVN kan, men ikke trenger å være utfylt. I dette tilfelle vil det heller ikke kunne kreves at eventuelle utfylte verdier er unike. Men er forutsetningen i dette tilfelle at ID består av KORTNAVN?

Hvis disse forutsetninger legges til grunn vil følgende ArkN4-testmetodikk kunne anvendes:

1. Test om ID er utfylt og unik.
 - a) Hvis NEI
test at AK.KORTNAVN er utfylt og unike
hvis JA OK
hvis NEI: Feilmeld
Fremmednøkkelkontroll utføres som i dag?
 - b) hvis JA
Fremmednøkkelkontroll utføres som i dag?
Mulig mindre streng tolkning enn i dag (i alle fall dersom ID består av KORTNAVN).

4.1.30 Identifikasjon av arkivskaper i uttrekk

Tabell	Attributt	Beskrivelse	Status/ estimat
ARKIV	AR.ABASEID	<p>K=O,O=b. I NOARK4, del 2, side 332, står det: «Ved avlevering til arkivdepot må attributtet være utfylt med en ID som er avtalt med mottakeren for å skille denne basen fra alle andre baser i depotet»</p> <p>Her må altså Arkivverket definere en unik ID til hver arkivskaper, og kreve at arkivsystemet merker avleveringen med denne ID i dette feltet.</p> <p>Tiltak: Implementere test som sjekker AR.ABASEID mot arkivverkets ID og genererer feilmelding.</p> <p>Alternativ: Man kan implementere funksjonalitet for å reparere denne feilen ved at arkivverket bruker databasenavnet man oppretter som ID og setter inn dette isteden.</p> <p>Feilmelding bør uansett genereres.</p>	Vurdert, ikke utført

Testers betraktninger vedrørende problemstillingen:

«Jeg er ikke sikker på hva de som har skapt NOARK 4-kravspesifikasjonen har tenkt her. Tilsynelatende er det en selvmotsigelse mellom definisjonen og det som står i teksten. Så lenge de ikke har definert attributtet som obligatorisk, har vi ikke hjemmel for å skrive feilmelding når det mangler i et uttrekk (vi har mottatt uttrekk både med og uten AR.ABASEID). Hvis det så forekommer må det jo iht. teksten være avtalt med "mottakeren", dvs. oss.

Da kan man stille spørsmål om hva slags ID dette er ment å være. Er det en ID som skal gjenspeile arkivskaperens benevnelse skulle det ikke være nødvendig å avtale denne med oss.

Teksten tyder imidlertid på at det er vår ID som menes. Det krever i tilfelle at vi gir arkivskaper denne opplysningen før ethvert uttrekk genereres. Neste spørsmål blir da hvilken ID som skal brukes. Man kunne jo skjele til Asta-benevnelse, men hvis jeg har oppfattet det rett, er Elmag styrende overfor Asta. Da blir dette en muligens en diskusjon om arkivpakkeinnhold og hører hjemme i en kontekst utover Noark-4. Spørsmålet blir jo også om å tildele en slik ID i forkant vil være i tråd med vår prosessløype i Elmag?»

Betraktninger fra Elark seksjonen:

«Attributtet AR.ABASEID ble ikke gjort obligatorisk i Noark 4 nettopp fordi det ikke var avklart hva slags verdi som skulle registreres her. Dersom attributtet skulle ha vært obligatorisk, måtte vi ha hatt rutiner som informerte arkivskaper hvilken verdi de skulle legge inn når uttrekket produseres. Det har vi ikke i dag, og jeg tror kanskje det vil bli litt omstendelig å innføre det i praksis. Jeg er enig i at dette også er noe en må se i sammenheng med definisjonen av arkivpakker. Sannsynligvis kan det også være like greit at depotet legger inn denne verdien etter at uttrekket er mottatt. Attributtet bør derfor fremdeles være valgfritt i uttrekk.

Tilsvarende metadataelement i Noark 5 er *M006 arkivskaperID*. I metadatakatalogen står det at denne IDen kan være organisasjonsnummeret som tildeles av Brønnøysundregistrene, men at det også er åpent for andre typer Ider.»

4.3.1 Utvidelse av minne.

Uttrekk av størrelse >500 MB har vist seg å medføre lange kjøretider. Etter utvidelse av minne med 1 GB pr. CPU (2X2 GB) har kjøretiden for et uttrekk fra Sjøfartsdirektoratet på 3,7 GB blitt redusert fra 13,5 til 12 døgn som fremdeles er uakseptabelt.

(I Riksarkivet skal ArkN4 nå testes ut på en server med betydelig mer minne og flere/raskere prosessorer.)

Etter parserbytte er kjøretiden kommet ned i 10,5 timer! Derfor et punktet lagt på is inntil nytt behov evt. dukker opp!