Depot N5 validering v1.2\_utkast IKAMR

|  |  |
| --- | --- |
| Navn | Depot N5 validering |
| Depot | IKAMR |
| Versjon | 1.2\_utkast |
| Dato | 2020-04-13 |

Dette dokumentet er en fortsettelse og fordypning av "Depot mottak IKAMR".

* Type uttrekk: Noark 5

**# Del V: A4 Test og validering**

* All test og validering i Arbeidsområde på en kopi av mottatte originalfiler
* Liste over mulige testverktøy i avsnittene under
* Rapporter og loggfiler bevares sammen med skjermdumper
* Mal for test-logg og verktøy med logger er allerede kopiert i pkt. *A3-1 Data innhenting*
  + \repository\_operations\analyse\..
  + \repository\_operations\.. (se de ulike filer vi har her)

A4-1 Oppretting av test-logg

* ..\repository\_operations\FFKK\_nnn\_test-log.txt
* Alle punkter i testfasen skal loggføres i test-logg med status:
  + Godkjent
  + Godkjent med anmerkning
  + Avvises

A4-2 Teknisk konsistens, korrekt og komplett

A4-3 Struktur og formatkrav

A4-4 Teknisk- og bevarings-metadata

A4-5 Sjekksummer i enkeltfiler

\*N5

* Noark 5 system
  + Sjekksum til arkivuttrekk.xml (ADDML) skal følge med i uttrekket (typisk i info.txt)
  + Sjekksummer på alle andre .xml og .xsd fra arkivuttrekk.xml
  + Korrekte .xsd som publisert av Arkivverket
  + Velformet .xml og validere .xml mot .xsd
  + Innholdsanalyse (antall registreringer pr. år i arkivstruktur, offentlig- og løpende journal,   
    endringslogg, dato- og datotid-felter, avskrivningsmåte)
  + Generelle Noark 5-tester
  + Spesifikke Noark 5-tester
  + Validere arkivversjon av dokumenter
  + Kryssreferanser
  + Dokumentasjon
* ..\repository\_operations\FFKK\_nnn\_Noark5-rapport.docx | .pdf (PDF/A)

**# Del VI: Opprette AIP**

A5; 2018-mm-ddThh:nn; Oppretting AIP; ;

**# Del VII: Lagre AIP**

A6; 2018-mm-ddThh:nn; Lagre AIP; ;

**## INNHOLD: A4-1 til A4-5 Test og Validering 0 til 13 ##**

1. Notater med resultater og avvik
2. Innledende kopiering oppsett
3. N5 VALIDATOR 1 (Arkade 5)
4. KDRS Query
5. Noark 5 XML og XSD med Diff Directories
6. XQuery/Xpath
7. N5 VALIDATOR 2 (Documaster Noark Extractor Validator)
8. DOK PRONOM (DROID)
9. DOK VALIDATOR 1 (veraPDF)
10. DOK VALIDATOR 2 (KOST-Val)
11. DOK ANALYSE (KDRS Analyse)
12. Noark 5 Word rapport
13. Noark 5 Excel logg
14. Elarkiv sjekkliste

**## ANALYSE: A4-1 til A4-5 Test og Validering 0 til 13 ##**

1. **Notater med resultater og avvik**
2. **Innledende kopiering oppsett**
   1. Analyse
      1. Katalog ..\repository\_operations\analyse\.. (kopiert under punkt a. over)
      2. Fast arbeidsdokument: 15KK\_010\_analyse.txt

Original ligger i OneNote - IKAMR produksjonslinje, kanal arkiv-work: [arkiv-work N5 Innholdsanalyse v0.2](onenote:#arkiv-work%20N5%20Innholdsanalyse%20v0.2&section-id={D9E2B82B-F989-47B1-A33D-589330041DB1}&page-id={224CC44C-18B5-4C3A-8B02-B5EA237A7373}&end&base-path=https://ikamoreogromsdal.sharepoint.com/sites/IKAMRELARKIV/Delte%20dokumenter/IKAMR%20produksjonslinje/OneNote%20-%20IKAMR%20produksjonslinje/arkiv-work.one)

Arbeidskopi som tekstfil ligger i mappestruktur (sjekk for oppdateringer jevnlig).

1. Arkade 5 info.xml fra Produsent
   1. Katalog ..\repository\_operations\innhold\.. (kopiert under punkt a. over)
   2. Lastet ned fra orkidenett.no med zip-fil (valgfri): 15KK\_nnn\_info.xml\_v1.5.1.0\_orkidenett.xml
   3. Pakket ut fra zip-fil sammen med tar-fil (valgfri): 15KK\_nnn\_info.xml\_v1.5.1.0\_zip.xml
   4. Mottatt via epost (obligatorisk, journalføres): 15KK\_nnn\_info.xml\_v1.5.1.0\_epost.xml

1. Arkade 5 rapport og/eller andre analyser, rapporter eller systemdokumentasjon fra Produsent
   * Levert fra Produsent i DIAS SIP/AIP pakken:
   * Katalog for rapporter og analyser: ..\content\\_report\..
   * Katalog for systemdokumentasjon: ..\content\sysdoc\..
   * Levert fra Arkivskaper/Produsent på annen måte lagres midlertidig:
   * Katalog for rapporter og analyser: ..\repository\_operations\\_report\..
   * Katalog for systemdokumentasjon: ..\repository\_operations\\_sysdoc\..
   * Arkade 5 rapporter gis følgende navn:   
     15KK\_nnn\_SIP\_Arkade5\_v1.6.1.0\_rapport.xml | .html | .pdf  
     xml og html følger med i Arkade 5 skapt .tar fil hvis pakketype AIP ble valgt  
     Hvis ikke må filene kopieres manuelt fra Arkade 5 prosjektmappe før program lukkes og sletter dem

1. Noark 5 skjema
   1. Katalog: ..\content\schema\.. (kopiert under punkt a. over - ev. slette andre N5 skjemaer enn i dette uttrekket )
   2. Kun en Noark 5-versjonsmappe beholdes, resten fjernes når versjon er bekreftet
      1. Mappe \arkivverket\.. inneholder xsd-skjema publisert fra Arkivverket
      2. xsd skjema levert med mottatt uttrekk   
         f.eks \ephorte-submission\_v2.5.0\.. uttrekk eksempel under, tilsvarende for WebSak og Documaster  
         1. Katalog for Noark 5 v3.1: ..\content\schema\noark5\_v3.1\..
            1. \arkivverket\..
            2. \ephorte-submission\_v2.5.0\..
            3. \mod-1\_ephorte\..
         2. Katalog for Noark 5 v4.0: ..\content\schema\noark5\_v4.0\..
            1. \arkivverket\..
            2. \ephorte-submission\_v2.7.1\..
            3. \mod-1\_ephorte\..
         3. Katalog for Noark 5 v5.0: ..\content\schema\noark5\_v5.0\..
            1. \arkivverket\..
            2. \documaster\_v2.21.0\..
            3. \mod-1\_documaster\..

1. **Arkade 5**
   * [**https://github.com/arkivverket/arkade5/releases**](https://github.com/arkivverket/arkade5/releases)
   * Katalog: ..\repository\_operations\arkade5\_v1.6.1.0\..
   * Last inn uttrekk
   * Knapp "Start testing"
   * Skjermdump når test ferdig lagres som 15KK\_nnn\_Arkade5\_v1.6.1.0.png
   * Knapp "Vis rapport" åpner rapport i nettleser, lagres som 15KK\_nnn\_Arkade5\_v1.6.1.0\_rapport.pdf med Adobe Acrobat  
     Pr. dato fungerer det ikke å lagre som PDF/A av ukjent årsak (test om MSWord utskrift til PDF/A fungerer?)
   * Kopier midlertidige rapportfiler fra Arkade 5 fra den valgte arbeidskatalogen som ble valgt for programmet, eks.:  
     D:\uttrekk-arkade\arkade-tmp\work\<uuid>\administrative\_metadata\repository\_operations\..  
     arkade-log.xml lagres som 15KK\_nnn\_Arkade5\_v1.6.1.0\_rapport.xml  
     report.html lagres som 15KK\_nnn\_Arkade5\_v1.6.1.0\_rapport.html  
     NB! Arkade sletter disse når program lukkes, så MÅ gjøres først!
   * Se over Arkade 5 testrapport og skrive notat i Analyse-dokument

1. **KDRS Query**
   * **3.1 KDRS Query app:** [**https://github.com/IKAMR/KDRS\_Query/releases**](https://github.com/IKAMR/KDRS_Query/releases)
   * Katalog: ..\repository\_operations\analyse\..
   * Kjør siste oppdaterte IKAMR standardliste (ev. for gjeldende Noark 5 versjon og uttrekksleverandør)  
     Orginal ligger i OneNote - IKAMR produksjonslinje, kanal arkiv-work: [xml-queries\_noark5\_2020-01-23](onenote:#arkiv-work%20N5%20arbeidsliste%20IKT%20ORKidé&section-id={D9E2B82B-F989-47B1-A33D-589330041DB1}&page-id={69768826-CC58-470E-B06D-518D53CA4CE2}&object-id={369F527E-8739-012E-177F-C7BD45D52CDF}&8D&base-path=https://ikamoreogromsdal.sharepoint.com/sites/IKAMRELARKIV/Delte%20dokumenter/IKAMR%20produksjonslinje/OneNote%20-%20IKAMR%20produksjonslinje/arkiv-work.one)  
     Arbeidskopi finnes i ..\repository\_operations\analyse\xml-queries\_noark5\_2020-01-23.xml
   * KDRS Query result (komplett fra standardliste b) over) : 15KK\_nnn\_kdrs\_query\_results.txt
   * IKAMR Xpath Runner (hardkodet data for kopiering til rapport): 15KK\_nnn\_ikamr\_xpath\_results.txt
   * Generer valideringsrapport med data fra KDRS XQuery klar for videre redigering  
     Mal finnes i ..\repository\_operations\maler\IKAMR-Noark5-C-rapportmal\_v1.2.0\_2020-01-22.docx  
     Resultat dukker opp i ..\repository\_operations\analyse\.. , men skal flyttes ut til ..\repository\_operations\..
   * **3.2 CheckSumControl app (fra IKAMR, kommer på GitHub)**
   * **Inkl. sjekksum kontroll arkivuttrekk.xml**
   * **For enkel kontroll og klipp og lim i logg/rapporter**
   * Start program
   * Dra og slipp arkivuttrekk.xml inn i applikasjonen, f. eks. fra plassering:  
     ..\arkiv-work\15KK\15KK\_nnn\_E-15KK\_YYYY\_nnnn\content\sip\content\arkivuttrekk.xml
   * Målmappe beregnes, f. eks. til plassering:  
     ..\arkiv-work\15KK\15KK\_nnn\_E-15KK\_YYYY\_nnnn\repository\_operations\content\
   * Klikk knapp "Run check sum control"
   * Kontroll av app dialog loggfelt, f. eks.:  
     Kontrollerer sjekksum  
     Writing to file A:\arkiv-work\15KK\15KK\_nnn\_E-15KK-YYYY-nnnn\repository\_operations\analyse\sjekksum\_kontroll.txt
   * Endre filnavn til: ..\15KK\_nnn\_sjekksum\_kontroll.txt
   * Kontroller at alle sjekksummer er identiske (mellom arkivuttrekk.xml og app beregnede verdier)
   * Fri gjenbruk, klipp og lim av listen med sjekksummer i fila

1. **XML Diff Directories**
   * **oXygen XML Editor med Diff Directories eller andre verktøy**
   * Katalog: ..\repository\_operations\analyse\..
   * Kopier uttrekkets xsd skjema til (avhengig av Noark 5 versjon og leverandør/uttrekksprogram):  
     ..\content\schema\noark5\_v3.1\ephorte-submission\_v2.5.0\..  
     ..\content\schema\noark5\_v4.0\ephorte-submission\_v2.7.1\..  
     ..\content\schema\noark5\_v5.0\documaster\_v2.21.0\..
   * Kjør Diff Directories mot (avhengig av Noark 5 versjon og leverandør/uttrekksprogram):  
     ..\content\schema\noark5\_v3.1\arkivverket\..  
     ..\content\schema\noark5\_v4.0\arkivverket\..  
     ..\content\schema\noark5\_v5.0\arkivverket\..
   * Dokumenter avvik i analyse arbeidsdokument iht. maltekster der for xsd skjema Diff Directories
   * Vurdere å ta skjermdump av viktige avvik (vurderes mot teksten i analyse tekstfila)
2. **XQuery/XPath**  
   Dokumenter avvik i analyse arbeidsdokument iht. maltekster der for de kjente problemområder
   * oXygen XML Editor, BaseX eller andre verktøy
   * Utgangspunkt
     1. Katalog: ..\repository\_operations\analyse\..
     2. Standardspørringer  
        Orginal ligger i OneNote - IKAMR produksjonslinje, kanal arkiv-work: [arkiv-work N5 Innholdsanalyse v0.2](onenote:#arkiv-work%20N5%20Innholdsanalyse%20v0.2&section-id={D9E2B82B-F989-47B1-A33D-589330041DB1}&page-id={224CC44C-18B5-4C3A-8B02-B5EA237A7373}&end&base-path=https://ikamoreogromsdal.sharepoint.com/sites/IKAMRELARKIV/Delte%20dokumenter/IKAMR%20produksjonslinje/OneNote%20-%20IKAMR%20produksjonslinje/arkiv-work.one)

Arbeidskopi som tekstfil ligger i mappestruktur (sjekk for oppdateringer jevnlig).

1. arkivstruktur.xml
   1. XQuery for antall oppføringer pr. år
   2. XQuery for avskrivningsmaate
2. Endringslogg.xml
   1. XQuery for antall oppføringer pr. år
3. loependeJournal.xml
   1. XQuery for antall oppføringer pr. år
4. offentligJournal.xml
   1. XQuery for antall oppføringer pr. år
5. **Documaster Noark Extractor Validator**
   * [**https://github.com/KDRS-SA/noark-extraction-validator/releases**](https://github.com/KDRS-SA/noark-extraction-validator/releases)
   * Katalog: ..\repository\_operations\noark-extractor-validator\..
   * Vurdere behov for ..\mod-1\_ephorte\.. Ut i fra oXygen XML Editor XQueries over
     1. F. eks. metdatakatalog.xsd hvis der er verdier av avskrivningsmaate som ikke er blant de 7 gyldige
     2. F. eks. arkivsturktur.xsd hvis fortsatt Arkade5 og Noark Extractor Validator bruker feil (konverteringsverktoy)
   * Startes fra C:\prog\noark-extractor-validator-0.7.0\_2020-01-14\..  
     alternativt ligge på disk med uttrekket eks. I:\prog\..
   * Java kompilert file har navn noark-extraction-validator-0.7.0-IKAMR\_2020-01-14.jar  
     Y:\elarkiv\software\documaster\ikamr\noark-extraction-validator-0.7.0-IKAMR\_2020-01-14\  
     noark-extraction-validator-0.7\_IKAMR\_2020-01-14.jar
   * Kjøres fra .bat fil med navn  
     15KK\_nnn\_n5-validate-IKAMR\_2020-01-14\_I.bat <<< stasjonsnavn \_I for katalog med uttrekk I:\  
     15KK\_nnn\_n5-validate-IKAMR\_2020-01-14\_I\_schema.bat <<< tar med \_schema hvis CUSTOM brukt  
     Merk at mal/tidligere .bat fil settes 'REM exit' før kjøring, og tilbake til 'exit" når ferdig
   * Kjøres til Excel (ikke med XML no fordi den får feil, ikke komplett og feil tegnset og krasjer)
   * Aktiver 'exit' øverst i .bat fil og kopier denne .bat fil til dette punktets hovedkatalog
   * Åpne Excel regneark, redigere første fane til gode marger og ev. ekstra versjonsinfo program
   * Sjekke avvik hovedpunkter (opp mot avvik oXygen XML Editor XQueries over og for øvrig)
   * Dokumenteres i analyse tekstdokument

1. **DROID**
   * [**http://digital-preservation.github.io/droid/**](http://digital-preservation.github.io/droid/)
   * [**https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/**](https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/)
   * Katalog: ..\repository\_operations\droid\_v6.4\_V96\.. av ..\dokumenter\.. mappa  
     Sjekk jevnlig for nyere versjoner av DROID, samt Binary and Container Signature files pr. 03.02.2020:
     1. DROID v6.4
     2. Binary Signature File: DROID\_SignatureFile\_V96
     3. Container Signature File: container-signature-20200121

1. Lagre filer
   1. 15KK\_nnn\_droid\_v6.4\_V96.droid
   2. 15KK\_nnn\_droid\_v6.4\_V96.csv
   3. 15KK\_nnn\_ephorte-SAK3\_report-planets.xml
   4. 15KK\_nnn\_ephorte-SAK3\_report-droid.xml
   5. 15KK\_nnn\_ephorte-SAK3\_report-droid.pdf
   6. Dokumentere i analyse tekstdokument
2. Regneark med PIVOT-tabell forekomster filtyper
   1. 15KK\_nnn\_droid\_v6.4\_V96.xlsx
      1. Høyreklikk .csv fil og åpne i Excel
      2. Lagre som 15KK\_nnn\_droid\_v6.4\_V96.xlsx
      3. Merk hele kolonne A, velg meny "Data\Tekst til kolonner",   
         "Datamed skilletegn" Neste, "Komma" skilletegn Neste,   
         Hold inne SHIFT scroll til høyre og merk alle kolonner og  
         velg "Tekst" som kolonneformat, Klikk "Fullfør"
      4. Meny "Visning \ Frys ruter \ Frys øverster rad"
      5. Uthev rad 1 kolonnetitler
      6. Minimer en del kolonne-bredder
      7. Utvid kolonnebredder:   
         Name, MIME\_TYPE, FORMAT\_NAME, FORMAT\_VERSION
      8. Kopier til nytt ark med navn "15KK\_nnn\_droid\_v6.4\_V96\_PUID"
      9. Meny "Data \ Filtrer"
      10. Filtrere "TYPE = Folder" og slett alle rader
      11. Filtrere "TYPE = File og ev. pakkede filer"
      12. Kopier til nytt ark med navn "15KK\_nnn\_droid\_v6.4\_V96\_ZIP"  
          Og gjenta samme (PUID kun filer, ZIP kun de pakket innhold)  
          {beskrive hvordan her seinere}
      13. For ark … PUID  
          Slett innhold kolonne C, og sett inn tittel igjen URI  
          Søk - Erstatt fast path i FILE\_PATH fram til dokumenter  
          Da beholdes kun relativ sti under N5 ..\dokumenter\..
      14. Sortering meny "Data + Sorter"  
          Nivå 1: MIME\_TYPE (A til Å)  
          Nivå 2: FORMAT\_NAME (A til Å)  
          Nivå 3: FORMAT\_VERSION (A til Å)  
          Nivå 4: PUID (A til Å)  
          Velg "Sorter som tall alt som se ut som tall"
      15. Merk hele ark PUID, velg meny "Sett inn \ Pivottabell"  
          La stå plassering "Nytt regneark", klikk "OK"
      16. Gi nytt ark navn "15KK\_nnn\_droid\_v6.4\_V96\_PIVOT"
      17. Velg kolonner i avkrysningsboks "Pivottabellfelt":  
          TYPE, EXT, PUID, MIME\_TYPE, FORMAT\_NAME, FORMAT\_VERSION
      18. Dra EXT til kolonner og TYPE til Verdier
      19. Sorter Rader slik, fra topp til bunn:  
          MIME\_TYPE, FORMAT\_NAME, FORMAT\_VERSION, PUID
      20. PIVOT topprad navn fra "Antall av TYPE" til  
          "Antall pr. MIME \ FORMAT navn \ FORMAT ver \ PUID"
      21. Merk alle celler i PIVOT tabell og "Alle kantlinjer"
      22. Slett feltverdi kolonnenavn "Kolonneetiketter"
      23. Sett kolonnebredde 6 og midtstill alle tellere
      24. Sett kolonnebredde 10 og midtstill "Totalsum"
      25. Sett annethvert MIME-område bakgrunnsvarge hvit / lys grå  
          I tillegg godkjente arkivformat lys grønn / lys blå hvis under grønn  
          Gjelder: PDF/A der 1a og 1b kan være lys grønn og lys blå osv  
          Samt: JPG, TIFF, XML, PNG, TXT (utvid denne liste)
      26. Slett 2 øverste tomme rader i ark
      27. Klikk celle B3 og velg meny "Visning \ Frys ruter \ Frys ruter"
   2. 15KK\_nnn\_droid\_v6.4\_V96\_pivot.png
      1. Ta en skjermdump av PIVOT-tabell (zoome hele eller 2 bilder)
3. **veraPDF**
   * [**https://verapdf.org/software/**](https://verapdf.org/software/)
   * Katalog: ..\repository\_operations\verapdf\_v1.14.2\..
   * Alternativ hvis v1.14.2 krasjer, Katalog: ..\repository\_operations\verapdf\_v1.12.1\..na
   * Lagre filer
     1. 15KK\_nnn\_verapdf\_v1.14.2\_htmlReport.html
     2. 15KK\_nnn\_verapdf\_v1.14.2\_xmlReport.xml
     3. 15KK\_nn\_verapdf\_v1.14.2.png
     4. Åpne html-fil i nettleser og hvis mulig skriv ut til PDF  
        15KK\_nnn\_verapdf\_v1.14.2\_report.pdf
     5. Dokumentere i analyse tekstdokument
4. **KOST-Val**
   * [**https://kost-ceco.ch/cms/index.php?kost\_val\_de**](https://kost-ceco.ch/cms/index.php?kost_val_de)
   * [**https://github.com/KOST-CECO/KOST-Val/releases**](https://github.com/KOST-CECO/KOST-Val/releases)
   * Katalog: ..\repository\_operations\kost-val\_v1.9.6\..
   * Alternativ hvis v1.9.6 krasjer, Katalog: ..\repository\_operations\kost-val\_v1.8.4\..
   * Lagre filer (kopieres fra temp-katalog til KOST-Val, lage link til denne mappe på skrivebordet)
     1. 15KK\_nnn\_kost-val\_v1.9.6\_log.xml
     2. 15KK\_nnn\_kost-val\_v1.9.6\_a.png ta skjermdump ved start kjøring hoveddialog
     3. 15KK\_nnn\_kost-val\_v1.9.6\_b.png ta skjermdump dialog for ferdig kjøring, som oftest med avvik
     4. Åpne xml-fil i nettleser og hvis mulig skriv ut til PDF (NB! Store XML-filer åpner seg kanskje aldri…)  
        15KK\_nnn\_kost-val\_v1.9.6\_log.pdf
     5. kost-val.xls (med i standardkatalog kopiert over)
     6. Dokumentere i analyse tekstdokument
     7. Takler ikke 1000-vis av filer, så må mest sannsynlig kjøre pr. år  
        Depot må vurdere om kost-val skal kjøres avhengig av tidspress og kvalitet
     8. Pr år noteres slik for både xml og bildefiler: 15KK\_nnn\_YYYY\_kost-val\_v1.9.6\_log.xml
5. **KDRS Analyse**
   * [**https://github.com/IKAMR/KDRS\_Analyse**](https://github.com/IKAMR/KDRS_Analyse) **<< ikke offisielt released**
   * Katalog: ..\repository\_operations\analyse\..
   * Last inn: info.xml
   * Last inn: DROID
   * Last inn: veraPDF
   * Last inn: KOST-VAL
   * (Last inn: Arkade 5.xml < ikke ferdig?)
   * (Last inn: Documaster Noark Extractor.xml < ikke ferdig? + problemer xml generering)
   * 15KK\_nnn\_kdrs-analyse\_v.9.6\_log.xml
   * Dokumentere i analyse tekstdokument
6. **Noark 5 Word rapport**
   * Mal: IKAMR-Noark5-C-rapportmal\_v1.1.2\_2018-11-30.docx
   * Rapport (opprettes av KDRS XQuery):   
     ..\repository\_operations\15KK\_nnn\_Noark5\_rapport.docx
7. **Noark 5 Excel logg**
   * Mal: IKAMR-Noark5-D-validator-logg\_v1.0.1\_2018-06-21.xlsx
   * Logg (ønskes automatisert):  
     ..\repository\_operations\15KK\_nnn\_Noark5\_logg.xlsx
8. **Elarkiv sjekkliste**
   * Mal: IKAMR-21\_Sjekkliste\_earkiv\_godkjenning\_v2.2.0\_2020-01-10.docx <<< Ny versjon med tabell anker
   * Rapport (ønskes automatisert):  
     ..\repository\_operations\15KK\_nnn\_Elarkiv\_sjekkliste.xlsx