

Quickguide

- til produktion og test af en afleveringspakke med ASTA

Rigsarkivet marts 2020

Version 1.0

Hvad er ASTA?

Rigsarkivet har skabt værktøjet ASTA (Aflevering af Statistikfiler Til Arkiv), som du kan anvende til at skabe og teste en afleveringspakke, der indeholder data udtrukket fra statistikfiler, som overholder Rigsarkivets krav til en afleveringspakke. Disse krav fremgår af bilag 9 i bekendtgørelsen om arkiveringsversioner. Bekendtgørelsen og vejledningen til denne finder du på Rigsarkivets hjemmeside www.sa.dk.

Når du skaber en afleveringspakke med ASTA, sikrer programmet at afleveringspakken overholder de strukturelle krav til mappestruktur og data- og metadatafiler. Du skal derimod selv sikre at data og metadataindhold i statistikfilerne overholder krav i bilag 9 før udtræk med ASTA.

Sådan kvalitetssikrer du statistikfilen før du laver udtræk

Funktionen "Skab afleveringspakke" i ASTA udtrækker blot de data og metadata, der findes i statistikfilen. Det er derfor vigtigt, at du sikrer dig, at kvaliteten af data og metadata i statistikfilen lever op til arkivets krav i bilag 9 i bekendtgørelse om arkiveringsversioner, før du laver udtræk med ASTA.

Krav til data- og metadatakvalitet er listet herunder:

Datafil

- Tegnsættet for statistikfilen skal være indkodet som UTF-8. Se vejledning om UTF-8.

Variable

- Der må ikke være linjeskift i indhold i variable.
- Variabelnavne [Variable name] skal være angivet for alle variable.
- Variabelnavne må aldrig begynde med et tal, men må godt være en blanding af bogstaver og tal.
- Variabelnavne må maks. være 128 tegn lange.
- Variabelbeskrivelser [Variable labels] skal være angivet for alle variable.

Datatyper (Formats)

- Datatyper for variable i statistikfilen skal overholde de tilladte datatyper i en afleveringspakke angivet i figur 9.3 i bilag 9 i bekendtgørelse om arkiveringsversioner. Hvis dette ikke er tilfældet, skal du selv konvertere datatyperne til tilladte formater, før udtræk laves med ASTA

Value labels/Formats

- Alle kodeværdier i data [Value] skal være beskrevet med en kodeforklaring [Value labels]
- Kodelister [Value labels] må kun være tilknyttet numeriske og kategoriske variable med datatyperne numerisk heltal, decimal og tekst.
- Koder [Values] skal være unikke inden for samme kodeliste [Value labels knyttet til en variabel]

Koder for manglende værdier (missing values)

- Specialkoder for manglende værdier (SAS og Stata) må kun antage værdier A-Z eller .a-.z.
- Specialkoder (SAS og Stata) for manglende værdier må kun anvendes for numeriske og kategoriske variable med datatyperne numerisk heltal og decimal
- Brugerdefinerede koder (SPSS) for manglende værdier må kun anvendes for numeriske og kategoriske variable med datatyperne numerisk heltal, decimal og tekst
- Der må konsekvent kun anvendes enten specialkoder (.a-.z eller A-Z) eller brugerdefinerede koder for manglende værdier
- Hvis brugerdefinerede koder (SPSS) for manglende værdier er anvendt for variable med datatypen dato eller tidsstempel, skal dette beskrives i et kontekstdokument i afleveringspakken.
- Både specialkoder og brugerdefinerede koder for manglende værdier i en variabel skal altid fremgå af kodelisten [Value labels] for denne variabel
- Hvis der kun er angivet kodeforklaringer [value labels] for nogle af koderne [values], der fremgår af en skala, skal dette beskrives i variabelbeskrivelsen for variabelen.

Flettevariable (referencer)

- Der skal være fuld overensstemmelse mellem datatype og længde for de flettevariable, der indgår i en reference mellem flere statistikfiler i afleveringspakken

Læs bilag 9 i bekendtgørelsen om arkiveringsversioner for en mere detaljeret beskrivelse af, hvad en afleveringspakke er, og hvilke krav der stilles til data- og metadatafilen.

Dette skal du have klar før du skaber en afleveringspakke med ASTA

Da det ikke er muligt at fortryde eller gå tilbage i ASTA, bør du have nedenstående filer og informationer klar, før du skaber en afleveringspakke med ASTA.

Du kan tilføje eller erstatte eksisterende indeksfiler og kontekstdokumenter i en afleveringspakke efterfølgende med funktionen "Rediger afleveringspakke".

- **Fremfind de statistikfiler**, der skal afleveres til arkivet.

Information! Hvis personfølsomme oplysninger er adskilt fra statistikfilen i en nøglefil, skal udtræk både fra statistikfilen og nøglefilen indgå i afleveringspakken.

- **Fremfind løbenummer** for afleveringspakken. Løbenummeret er udleveret af arkivet og fremgår af afleveringsbestemmelsen.
- **Udfyld indeksfilerne** archiveIndex.xml og contextDocumentationIndex.xml. Arkivet har udviklet indtastningsværktøjer til udfyldning af indeksfilerne. Se vejledningerne til Skab archiveIndex og Skab contextDocumentationIndex på Rigsarkivets hjemmeside www.sa.dk.

Information! Arkivet skal godkende de udfyldte indeksfiler, før du placerer dem i afleveringspakken.

- **Fremfind de kontekstdokumenter**, der skal indgå i afleveringspakken. Disse fremgår af afleveringsbestemmelsen.
- **Konverter kontekstdokumenterne** til bevaringsformat, fx .tif (dokumenter og billeder), .mp3 (lyd) og .mpg (video). Se vejledning om TIFF-konvertering på Rigsarkivets hjemmeside www.sa.dk.

Information! Du kan også konvertere kontekstdokumenterne, efter afleveringspakken er skabt.

- **Afklar om der findes unikke variable** i statistikfilene, som skal angives som nøglevariable under udtræk med ASTA.
- **Afklar om der er flettevariable** på tværs af flere statistikfiler i afleveringspakken som skal angives i en opmærkning af en reference under udtræk med ASTA samt over, hvilke variable disse referencer skal danne.

Disse dele af afleveringspakken kan du skabe med ASTA

Anvend funktionen "Skab afleveringspakke" til at gøre følgende:

- Navngive afleveringspakken med korrekt løbenummer.
- Danne en tom mappestruktur til afleveringspakken med struktur og navngivning af mapper, der overholder krav i bilag 9.
- Lave udtræk af data og metadata fra dine originale statistikfiler til data- og metadatafiler, der overholder de strukturelle krav til disse filer i bilag 9.
- Tilføje archiveIndex.xml og contextDocumentationIndex.xml til afleveringspakken. Se vejledningerne til Skab archiveIndex og Skab contextDocumentationIndex.
- Tilføje kontekstdokumenter til afleveringspakken.

Disse dele af afleveringspakken skal du selv lave

I ASTA kan du ikke følgende:



- Udfylde indeksfilerne
- Konvertere kontekstdokumenterne til bevaringsformat, fx .tif eller .mp3

Sådan kontrollerer du om udtræk med ASTA er tabsfrit

Det er vigtigt, at de data, der afleveres til Rigsarkivet, er autentiske og svarer til indholdet i de originale statistikfiler. Det er den afleverende myndigheds ansvar at sikre, at det udtræk, der laves med ASTA eller andre værktøjer, er tabsfrit. Dette kan gøres på mange måder både automatisk og visuelt. Nedenfor er listet eksempler på, hvad der skal kontrolleres samt, hvordan det kan gøres.

Automatisk kontrol af dataudtræk i SPSS

Denne kvalitetskontrol kan i SPSS udføres ved at køre en compare-syntaks i statistikprogrammet, der sammenligner den originale statistikfil med den output statistikfil, som ASTA skaber. Compare-syntaksen til SPSS er indarbejdet i det eksport script til SPSS, som ASTA danner. Ved kørsel af dette eksport script udføres således automatisk en kvalitetskontrol af udtræk fra SPSS lavet med ASTA. Resultatet af denne sammenligning gemmes automatisk i en logfil, navngivet Exportscriptlog.spv, på samme destination, hvor den valgte originale statistikfil ligger. Du bør åbne denne logfil i SPSS og kontrollere, om der er forskelle mellem originalfilen og outputfilen. Resultatet af sammenligningen fremgår nederst i logfilen under overskriften "Compare datasets".

Compare-syntakser til SAS og Stata er ikke indarbejdet i eksport scriptene i ASTA.

Visuel kontrol

Ud over den automatiske kontrol bør du også foretage visuel kontrol af data før og efter konvertering samt tjekke, om metadata i metadatafilen (fx table1.txt) er udtrukket korrekt og tabsfrit fra datafilen.

Sådan tester du en afleveringspakke

Anvend funktionen "Test afleveringspakke":

- Vælg den afleveringspakke, du vil teste, og klik på knappen "Start test"
- Læs eventuelle fejlbeskeder vist i testloggen
- Klik på fejl-ID'et i html-versionen af testloggen, som fremkommer ved klik på knappen "Vis log" for at se vejledning i, hvordan fejlen rettes
- Ret fejlene vist i testloggen og test afleveringspakken på ny
- Når afleveringspakken er testet helt fejlfri kan den afleveres til arkivet efter de regler for aflevering, som fremgår på Rigsarkivets hjemmeside www.sa.dk.

Sådan retter du fejl i en eksisterende afleveringspakke

- **Fejl i data- og metadatafilerne** rettes direkte i statistikfilen, hvorefter der laves et nyt udtræk med ASTA.

Kan fejlen ikke rettes i statistikfilen, er det tilladt at rette direkte i datafilen (fx table1.csv) og metadatafilen (fx table1.txt) udtrukket fra statistikfilen. Disse fejl bør dog



rettes som noget af det allersidste, når man er sikker på, at der ikke er flere fejl, der skal rettes ved at lave nyt udtræk fra statistikfilen.

- **Fejl i indeksfilerne** kan rettes i Rigsarkivets indtastningsværktøjer til disse. Værktøjerne validerer om filerne overholder krav til indeksfiler. Ændringer i indeksfilerne skal godkendes af arkivet.
- **Fejl i kontekstdokumenter** rettes ved at erstatte det eksisterende kontekstdokument med et tilrettet.

Information! Tilføjes nye kontekstdokumenter til afleveringspakken, som ikke fremgår af afleveringsbestemmelsen, skal denne opdateres med disse dokumenter og godkendes af arkivet.