



Rigsarkivet

Danmarks
hukommelse

Quickguide til produktion og test af en afleveringspakke med ASTA

Kort vejledning til Rigsarkivets afleveringskrav og konverterings- og testværktøjet ASTA til korrekt pakning af forskningsdata. Vejledningen dækker forberedelse, udtræk og kvalitetssikring af data fra de mest anvendte statistikprogrammer.

Rigsarkivet August 2025

Version 2.0





Indhold

Quickguide til produktion og test af en afleveringspakke med ASTA	3
Hvad er ASTA?	3
Hvad er en afleveringspakke?	3
Sådan kvalitetssikrer du statistikfilen før du laver udtræk	4
Datafil	4
Variable	4
Datatyper (Formats)	4
Value labels/Formats	4
Koder for manglende værdier (missing values)	5
Flettevariable (referencer)	5
Dette skal du have klar før du skaber en afleveringspakke med ASTA	5
Disse dele af afleveringspakken kan du skabe med ASTA	6
Disse dele af afleveringspakken skal du selv lave	7
Sådan kontrollerer du om udtræk med ASTA er tabsfrit	7
Automatisk kontrol af dataudtræk i SPSS	7
Visuel kontrol	7
Sådan tester du en afleveringspakke	8
Sådan retter du fejl i en eksisterende afleveringspakke	8
Figur 9.3 Tilladte datatyper	9





Quickguide til produktion og test af en afleveringspakke med ASTA

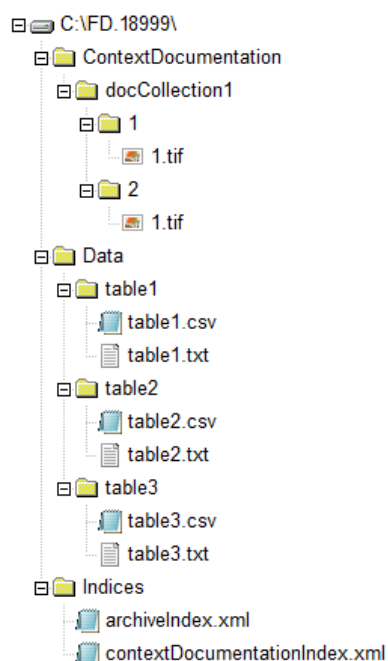
Hvad er ASTA?

ASTA (Aflevering af Statistikfiler Til Arkiv) er Rigsarkivets værktøj til at skabe og teste afleveringspakker fra statistikfiler. Værktøjet sikrer at din afleveringspakke overholder Rigsarkivets krav, som er beskrevet i bilag 9 i bekendtgørelse nr. 128 om arkiveringsversioner.

Du kan finde vejledningen til bekendtgørelsen og yderligere hjælp på Rigsarkivets hjemmeside under: <https://www.rigsarkivet.dk/aflever-data/for-dig-der-skaber-forskningsdata/aflever-forskningsdata/> efter at har valgt en datatype, i sektionen "Vælg dataformater".

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

Hvad er en afleveringspakke?



En afleveringspakke indeholder et udtræk af data, der har været anvendt til statistisk analyse fra statistikprogrammer som SAS, SPSS, Stata, R og regneark. Afleveringspakken består af både selve data og alt det materiale, der skal til for at forstå og genbruge data i fremtiden. Den skal overholde særlige krav til mappestruktur og navngivning:

- **Kontekstdokumentation:** Projektbeskrivelse, spørgeskemaer og metoderapporter (ContextDocumentation)
- **Datafil:** Selve data i CSV-format (f.eks. table1.csv)
- **Metadatafil:** Variabelbeskrivelser, datatyper og kodelister (f.eks. table1.txt)
- **Arkivbeskrivelsesfil:** Overordnede oplysninger om datasættet (archiveIndex.xml)
- **Kontekstdokumentationsfil:** Indeks over kontekstdokumenter (contextDocumentationIndex.xml)

Figur 1 - Grafisk oversigt over elementer og struktur i en afleveringspakke





Rigsarkivet

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

Statistikfilen skal være gemt med UTF-8 tegnsæt (sikrer at danske bogstaver som æ, ø, å vises korrekt i afleveringen).

Hjælp til UTF-8: Find vejledning på Rigsarkivets hjemmeside under <https://www.rigsarkivet.dk/aflever-data/for-dig-der-skaber-forskningsdata/aflever-forskningsdata/aflever-statistikfiler/> → sektionen "ASTA – udtræksprogram til statistikfiler".

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)

Dine variable skal bruge de tilladte datatyper for afleveringspakker. Hvis dine data har andre datatyper, skal du konvertere dem først, før du laver udtræk med ASTA.

Hvad skal du gøre?

- Tjek om dine datatyper er tilladte (se [Figur 9.3 Tilladte datatyper](#) i bunden af denne vejledning)
- Konverter om nødvendigt til tilladte formater i dit statistikprogram
- Lav derefter udtræk med ASTA

Value labels/Formats

- Alle kodeværdier i data [Value] skal være beskrevet med en kodeforklaring [Value labels]
- Nominale og kategoriske variable skal have kodebeskrivelser for alle kodeværdier
- Ordinale variable skal som udgangspunkt have kodebeskrivelser for alle kodeværdier. Undtagelse: Hvis kun yderværdierne i en skala har beskrivelse (f.eks. 1="Meget uenig",





Rigsarkivet

5="Meget enig"), kan variablen behandles som intervalvariabel, hvis dette afspejler hvordan data er anvendt. I så fald skal dette dokumenteres i variabelbeskrivelsen.

- Interval- og ratiovariable skal IKKE have tilknyttede kodelister. Dette gælder kontinuerlige målinger som f.eks. alder, vægt, indkomst, temperatur, hvor tallene repræsenterer faktiske målte værdier frem for koder
- 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

Flettevariable (referencer)

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

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.





Rigsarkivet

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.
- Fremfind afleveringsbestemmelsen, archiveIndex.xml og contextDocumentationIndex.xml som arkivaren har sendt til jer på mail.
- Fremfind de kontekstdokumenter, der skal indgå i afleveringspakken. Disse fremgår af afleveringsbestemmelsen.
- Konverter kontekstdokumenterne til bevaringsformatet:
 - Dokumenter og billeder: .tif (med LZW-kompression)
 - Lydfiler: .mp3
 - Videofiler: .mpg

Hjælp til konvertering: Find vejledning om TIFF-konvertering på Rigsarkivets hjemmeside under <https://www.rigsarkivet.dk/aflever-data/for-dig-der-skaber-forskningsdata/aflever-forskningsdata/aflever-statistikfiler/> → sektionen "Kontekstdokumenter".

Information! Du kan også konvertere kontekstdokumenter, 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.





Rigsarkivet

- 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.
- Tilføje kontekstdokumenter til afleveringspakken.

Disse dele af afleveringspakken skal du selv lave

I ASTA kan du ikke følgende:

- Konvertere kontekstdokumenterne til bevaringsformatet, fx .tif, .mp3 eller .mpg.

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, Stata og R 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 <https://www.rigsarkivet.dk/aflever-data-til-rigsarkivet/>

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, kan det blive nødvendigt 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 henvises til den ansvarlige arkivar på den pågældende sag hos arkivet.
- Fejl i kontekstdokumenter rettes ved at erstatte det eksisterende kontekstdokument med et tilrettet. Ellers kontakt Rigsarkivet på følgende mail: digitalbevaring@rigsarkivet.dk.

Information! Tilføjes nye kontekstdokumenter til afleveringspakken, som ikke fremgår af afleveringsbestemmelsen, skal du kontakte arkivaren, som fremsender en ny afleveringsbestemmelse og ny kontekstdokumentationsfil.





Figur 9.3 Tilladte datatyper

Datatype	Dataformat i datafil	Eksempler på data i datafil	Dataformatnotation i metadatafil		Eksempler på angivelse af dataformatnotation i metadatafil
Tekst	UTF-8 tegnsæt, jf. punkt 9.F.1.	"Antal hjorte; 6." "Hun sagde ""Hej""." Jeg spiller fodbold.	xml	string	string
			Stata	%ws	%20s
			SAS	\$w.	\$20.
			SPSS	aw	a20
Numerisk heltal	Repræsentation af et heltal med eller uden fortegn i henhold til DS/ISO 6093:1985 (NR1) standard, jf. syntaksregel i figur 9.6.	25 +45 -234 10000	xml	int	int
			Stata	%w.Of	%3.Of
			SAS	fw.	f3.
			SPSS	fw	f3
Numerisk decimaltal	Repræsentation af et decimaltal med eller uden fortegn i henhold til DS/ISO 6093:1985 (NR2) standard, jf. syntaksregel i figur 9.7.	23,75 .10 -123.76 150000.25	xml	decimal	decimal
			Stata	%w.df eller %w.dg	%9.2f eller %9.2g
			SAS	fw.d	f9.2
			SPSS	fw.d	f9.2
Dato	Angivelse af kalenderdato i henhold til DS/ISO 8601:1993 udvidet format: CCYY-MM-DD	2019-11-15	xml	date	date
		2019-11-15	Stata	%tdCCYY-NN-DD	%tdCCYY-NN-DD
		2019-11-15	SAS	yymmdd10.	yymmdd10.
	Alternativt kan følgende format anvendes, jf. syntaksregel i figur 9.8: CCYY/MM/DD	2019-11-15	SPSS	sdate10	sdate10
Tidspunkt	Angivelse af tidspunkt i henhold til DS/ISO 8601:1993 udvidet format, jf. syntaksregel i figur 9.9: hh:mm:ss	23:12:06 05:10.23	xml	time	time
			Stata	%tcHH:MM:SS	%tcHH:MM:SS
		4:22:59 05:10.23	SAS	time. eller time8.	time. eller time8.
			SPSS	time8	time8
Tidsstempel	Angivelse af dato og tidspunkt i henhold til DS/ISO 8601:1993 udvidet format:	2019-11-15T08:10:23 Med fraktioner af sekunder: 2019-11-15T08:10:23.123456	xml	datetime	datetime





Rigsarkivet

Datatype	Dataformat i datafil	Eksempler på data i datafil	Dataformatnotation i metadatafil		Eksempler på angivelse af dataformatnotation i metadatafil
	CCYY-MM-DDThh:mm:ss.sss	2019-11-15T08:10:23	Stata	%tcCCYY-NN-DD!THH:MM:SS	%tcCCYY-NN-DD!THH:MM:SS
	Eller CCYY-MM-DD hh:mm:ss.sss Fraktioner af sekunder er valgfrit og tilladt med en præcision på op til 6 cifre. Tidszone i tidsangivelser er ikke tilladt, jf. syntaksregel i figur 9.10.	2019-11-15T08:10:23.123	Stata	%tcCCYY-NN-DD!THH:MM:SS.sss	%tcCCYY-NN-DD!THH:MM:SS.sss
		2019-11-15T08:10:23.1			%tcCCYY-NN-DD!THH:MM:SS.s
		2019-11-15T08:10:23	SAS	e8601dt19.	e8601dt19.
		2019-11-15T08:10:23.123456	SAS	e8601dtw.d	e8601dt25.6
		2019-11-15T08:10:23.12345			e8601dt24.5
		2019-11-15 08:10:23	SPSS	ymdhms19	ymdhms19
		2019-11-15 08:10:23.123456	SPSS	ymdhmsw.d	ymdhms26.6
		2019-11-15 08:10:23.1234			ymdhms24.4
	Alternativt kan følgende format anvendes, jf. syntaksregel i figur 9.10: dd-mm-yyyy hh:mm:ss	15-NOV-2019 08:10:23	SPSS	datetime20	datetime20

