



Rigsarkivet

Danmarks
hukommelse

Vejledning til produktion af afleveringspakke med statistikdata fra regneark og csv-filer

Vejledning til produktion af en afleveringspakke med
forskningsdata fra regneark og csv-filer, der overholder
krav i bilag 9 til Bekendtgørelse om arkiveringsversioner.





Indhold

0. Læsevejledning	2
a. Vejledningens målgruppe og anvendelse	2
b. Henvi sning til øvrig vejledning	2
c. Lovgivning og retsforskrifter	3
d. Definitioner	3
1. Hvad er en afleveringspakke?	4
2. Sådan skaber du manuelt en afleveringspakke	5
e. Tjekliste – trin før skabelse af afleveringspakken	5
f. Afleveringspakkens indhold	11
g. Opret afleveringspakkens mappestruktur	12
h. Placer indeksfiler i afleveringspakken	13
i. Placer kontekstdokumenter i afleveringspakken	13
j. Skab undermapper i Data-mappen.....	17
k. Udtræk data fra regnearket til datafilen (.csv)	18
l. Kontroller at udtræk af data til csv-fil er tabsfrit.....	19
m. Skab metadatafilen (.txt)	19
n. Placer data og metadatafilerne i afleveringspakken	20
3. Aflevering af csv-fil som afleveringspakke	27
4. Sådan tester du afleveringspakken	27
5. Support i Rigsarkivet	31





0. Læsevejledning

Offentlige myndigheder, herunder forskningsinstitutioner, er forpligtet til at aflevere en kopi af data og dokumenter af bevaringsværdige forskningsdata. Hvis de forskningsdata, der skal afleveres, stammer fra statistikfiler eller tilsvarende (fx regneark), kaldes afleveringsformatet en afleveringspakke. Statslige myndigheder skal aflevere til Rigsarkivet. Kommuner og regioner kan vælge, om de vil aflevere til Rigsarkivet eller oprette deres eget arkiv.

Rigsarkivet har fastsat en række bestemmelser for en afleveringspakke af hensyn til bevaring og fremtidig brug af data, som alle myndigheder skal overholde ved aflevering. Disse bestemmelser er beskrevet i Rigsarkivets bekendtgørelse om arkiveringsversioner, bilag 9: Afleveringspakke for visse typer af forskningsdata.

Rigsarkivet har udviklet værktøjet ASTA (Aflevering af Statistikdata Til Arkiv) til at skabe og teste en afleveringspakke. ASTA kan kun anvendes til at udtrække data fra statistikfiler som SAS, Stata og SPSS til en afleveringspakke. Data anvendt til statistisk analyse kan dog også findes i andre formater, fx regneark eller csv-filer udtrukket fra andre programmer.

Denne vejledning beskriver hvordan du kan skabe en afleveringspakke, hvis dine bevaringsværdige statistikdata findes i et regneark eller en csv-fil

a. Vejledningens målgruppe og anvendelse

Denne vejledning henvender sig til dem, som skal aflevere statistikdata fra regneark og csv-filer i form af en afleveringspakke. Det kan fx være myndighedens egen datamanager, it-afdeling, forskeren selv eller en ekstern it-leverandør.

b. Henvisning til øvrig vejledning

Foruden denne vejledning har Rigsarkivet udarbejdet andre vejledninger, der har betydning for produktion og aflevering af afleveringspakker:

- Quickguide – til produktion og test af en afleveringspakke med ASTA
- Vejledning til bilag 9 om afleveringspakker i bekendtgørelse om arkiveringsversioner nr. 128
- Vejledning til Skab archiveIndex
- Vejledning til Skab contextDocumentationIndex
- Vejledning om konvertering af dokumenter til TIFF
- Vejledning om UTF-8
- Eksempelafleveringspakke med statistikdata FD.18005

Alt vejledningsmateriale kan tilgås fra [Rigsarkivets hjemmeside](#).





Rigsarkivet

c. Lovgivning og retsforskrifter

Information om lovgivning m.v. findes på [Rigsarkivets hjemmeside](#).

d. Definitioner

Afleveringspakker med data fra statistikfiler består overordnet set af kontekstdokumenter, der skal afleveres i Rigsarkivets arkivformater, udtræk af data og metadata fra de statistikfiler som skal afleveres, samt to indeksfiler i xml-format, der indeholder overordnet metadata om de afleverede data og kontekstdokumenterne.

Arkivformater: Rigsarkivet benytter 6 arkivformater: TIFF, JPEG2000, MP3, WAV, MPEG2 og MPEG4.

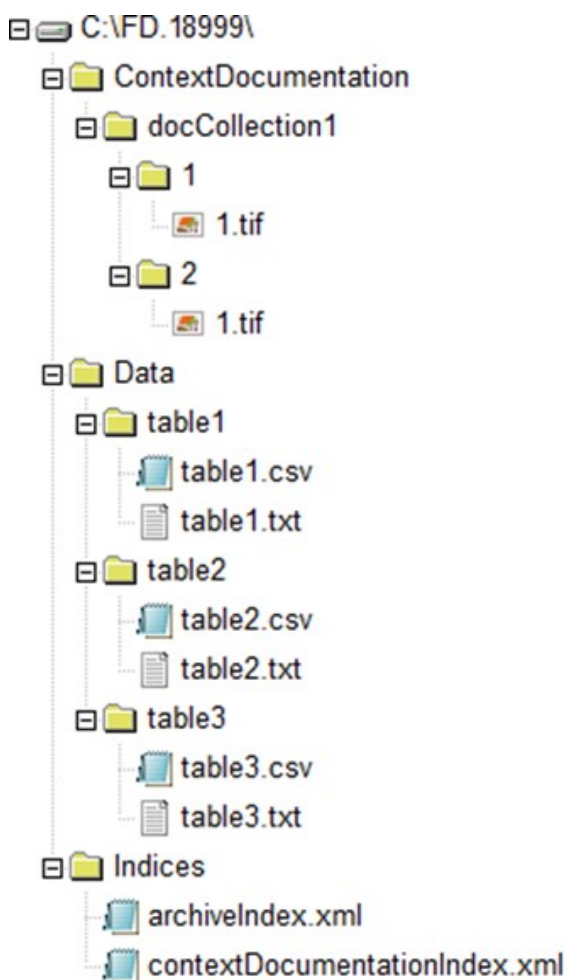


1. Hvad er en afleveringspakke?

Som en del af Rigsarkivets krav til aflevering af statistiske data, skal data fra de oprindelige statistikfiler udtrækkes til en afleveringspakke og testes inden aflevering til arkivet.

Afleveringspakken skal følge nogle specifikke krav vedrørende mappestruktur, navngivning af mapper, filer, datatyper, variabelnavne, variabellabels, formatnavne, manglende værdier osv. Alle krav til en afleveringspakke findes under bilag 9 i Bekendtgørelse om arkiveringsversioner.

Et eksempel på en afleveringspakke fremgår af figur 1.



Mappen **ContextDocumentation** indeholder dokumenter konverteret til bevaringsformat fx tif, der beskriver de data, afleveringspakken indeholder. Fx hvordan data er indsamlet, metoderapport eller et spørgeskema.

Mappen **Data** indeholder både en datafil og en metadatafil, der begge overholder kravene i bilag 9. Data udtrukket fra de originale statistikfiler skal afleveres i en semikolonsepareret csv-fil (**table1.csv**). Metadata udtrukket fra statistikfilen, fx variabelbeskrivelser, svarkategorier og koder for manglende værdier, skal afleveres som en metadatafil i txt-format (**table1.txt**)

Mappen **Indices** indeholder to indeksfiler med metadata på et mere overordnet niveau. **archiveIndex.xml** filen indeholder fx oplysninger som navn på datasættet, der afleveres, navn på forsker der har indsamlet data, perioden data dækker, adgangs begrænsninger til data osv. Filen **contextDocumentationIndex.xml** indeholder oplysninger om kontekstdokumenterne placeret i mappen **ContextDocumentation**, fx dokumentets titel, forfatter samt emne kategorisering af dokumentet.

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





2. Sådan skaber du manuelt en afleveringspakke

Herunder beskrives hvordan du skaber en afleveringspakke med udtræk fra et eller flere regneark. Se afsnit 3, hvis du skal skabe en afleveringspakke med data, som allerede er i csv-format.

e. Tjekliste – trin før skabelse af afleveringspakken

Du bør gennemføre følgende trin før du danner selve afleveringspakken.

NR.	OPGAVER	FÆRDIGT
GENERELT		
1. Ny mappe med kopi af materiale	Opret en ny mappe på computeren med en kopi af alle data og dokumenter, der skal afleveres. Dvs. statistikfilerne, der skal afleveres, kontekstdokumenterne, som fremgår af afleveringsbestemmelsen, og de to indeksfiler. Navngiv evt. mappen med afleveringspakkens løbenummer, der fremgår af arkivets afleveringsbestemmelse.	
DATAFILER		
2. Datafiler	Datafilerne skal være regneark (fx .xls eller .xlsx) eller csv-filer udtrukket fra et andet program. Dette afsnit i vejledningen omhandler primært udtræk fra regneark. Hvis du har en csv-fil udtrukket fra et andet program, kan du læse mere herom i afsnit 3.	
3. Variable i datafiler	Variabelnavne (kolonnenavne) skal placeres i den første række i regnearket. Variabelnavne må ikke være længere end 128 tegn, starte med et tal eller inkludere blanke mellemrum eller specialtegn såsom %.	





NR.	OPGAVER	FÆRDIGT
4. Variable Kategori/type	<p>Kolonnekategorier/datatyper skal tildeles i overensstemmelse med den type data, der vises i en celle/kolonne. Fx hvis en celle indeholder datoer, vælges kategorien "Dto" med formatet åååå-mm-dd.</p> <p>I Excel kan du ændre datatypen ved at markere alle celler i kolonnen og højreklikke. Vælg "Formater celler" > fanebladet "Tal" > Vælg den kategori/datatype, som stemmer overens med cellernes indhold. Her kan du også definere brugerdefinerede formater.</p> <p>Se figurerne 2 – 22.7 for en beskrivelse af tilladte dataformer/kategorier.</p>	
5. Value labels	<p>Hvis der findes koder blandt værdier i data skal disse forklares med kodeforklaringer i en metadatafil i afleveringspakken. Hav derfor disse kodebeskrivelser klar. Hvordan du angiver kodebeskrivelserne i metadatafilen beskrives under afsnit M. Skab metadatafilen (.txt).</p>	
6. Missing values	<p>Hvis du i dine data har anvendt koder for manglende værdier, såsom for eksempel 9 = uoplyst, 10 = irrelevant eller 11 = deltager ikke, skal disse værdier beskrives i en metadatafil i afleveringspakken. Hvordan du angiver koder for manglende værdier manuelt i metadatafilen beskrives i afsnit M. Skab metadatafilen (.txt).</p> <p>Der må kun anvendes koder for manglende værdier i numeriske og kategoriske variable. Kontakt arkivet, hvis du anvender koder for manglende værdier i variable med typerne tekst, tidspunkt eller dato.</p> <p>Alle koder for manglende værdier skal også tildeles en værdi i en kodeliste i metadatafilen.</p>	
7. Referencer (ved aflevering af flere datasæt)	<p>Når der afleveres mere end et datasæt i den samme afleveringspakke med en reference mellem disse statistiske filer (en fletningsnøgle) skal flettevariable have samme type/format og længde.</p>	
KONTEKSTDOKUMENTATION		





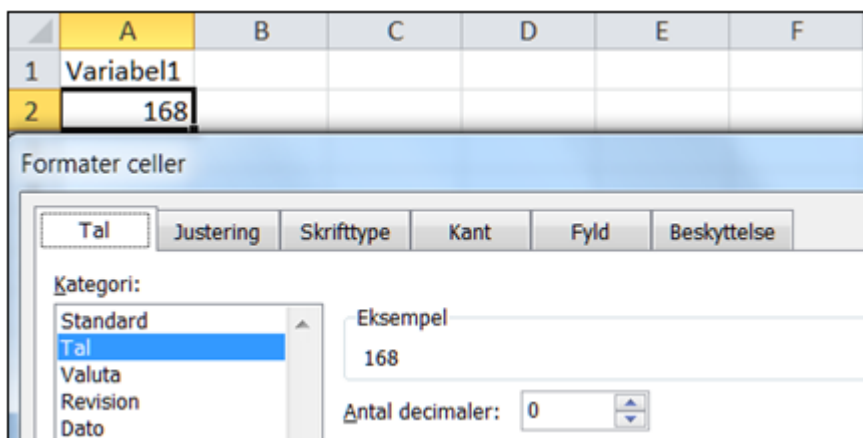
Rigsarkivet

NR.	OPGAVER	FÆRDIGT
8. Godkendelse fra arkivet	Kontekstdokumentationsfilen contextDocumentationIndex.xml skal godkendes af arkivet, inden den indgår i afleveringspakken. Filen specificerer alle de ekstra dokumenter, der er inkluderet i afleveringspakken. Se 'Vejledning til Skab contextDocumentationindex'.	
9. TIFF dokumenter	Hvert dokument registreret i kontekstdokumentationsfilen skal konverteres til TIFF eller andet arkivformat. Se 'Vejledning i konvertering af dokumentet til TIFF'.	
ARKIVBESKRIVELSESFIL		
10. Godkendelse fra arkivet	Arkivbeskrivelsesfilen skal godkendes af arkivet, inden den indgår i afleveringspakken. Filen indeholder overordnede metadata om de afleverede data. Se 'Vejledning til Skab archiveIndex'.	

Tilladte dataformater i datafilen

Jf. figur 9.3 i bilag 9 i Bekendtgørelse om arkiveringsversioner er følgende dataformater tilladte i en afleveringspakke med statistikdata

Heltal (int) jf. DS/ISO 6093:1985 (NR1) standard svarer til kategorien **Tal** med 0 decimaler i Excel. Bemærk at heltal er tal uden decimaler, så antal decimaler skal angives til 0.



Figur 2: Valg af kategori for heltal i Excel

Decimal (decimal) jf. DS/ISO 6093:1985 (NR2) standard svarer til **Tal** i Excel.

Bemærk at decimaltal er tal med decimaler, men datatypen må også gerne indeholde heltal uden



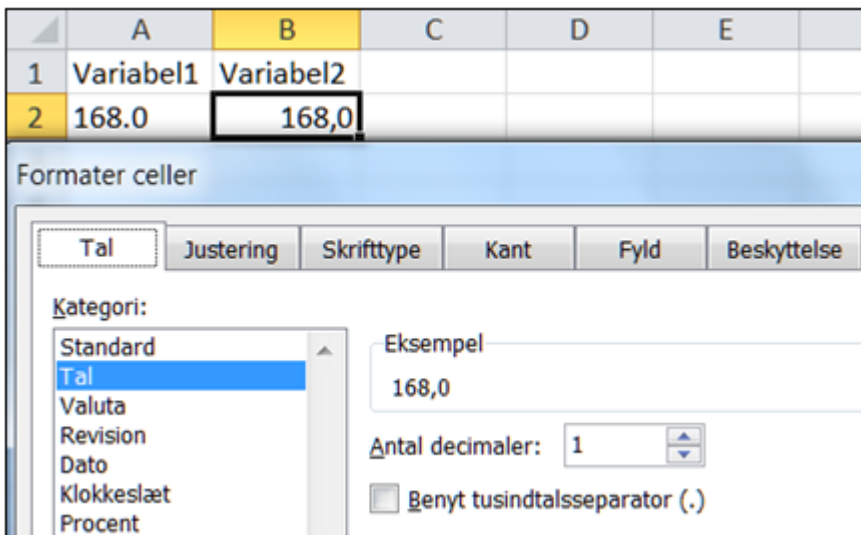


Rigsarkivet

decimaler.

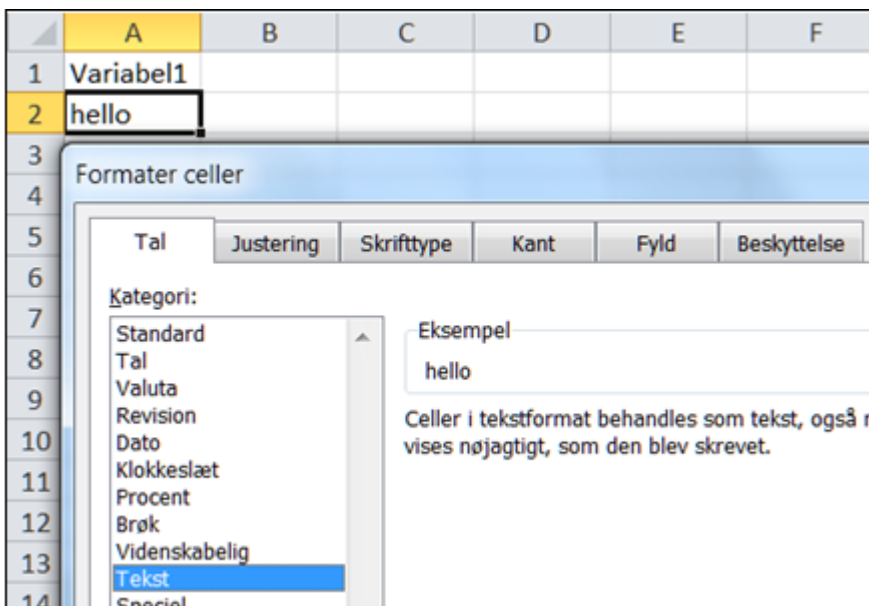
Bemærk, at der **ikke** må benyttes tusindstalsseparator.

Bemærk, at decimalseparatoren kan være enten '.' eller ','



Figur 3: Valg af kategori for decimaltal i Excel

Tekst (string), som svarer til **Tekst** i Excel



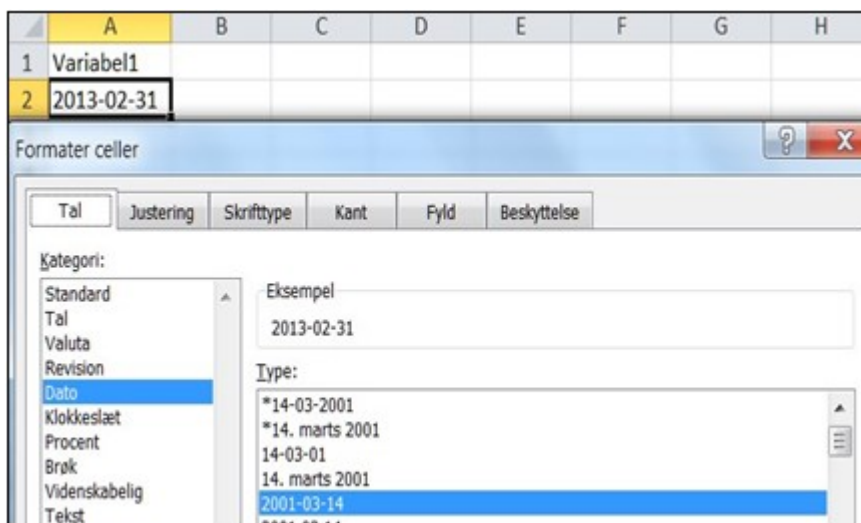
Figur 4: Valg af kategori for tekst i Excel





Rigsarkivet

Dato (date), DS/ISO8601:1993 standard, som svarer til **Dato** i Excel
Bemærk at formatet er CCYY-MM-DD, fx 2019-11-30



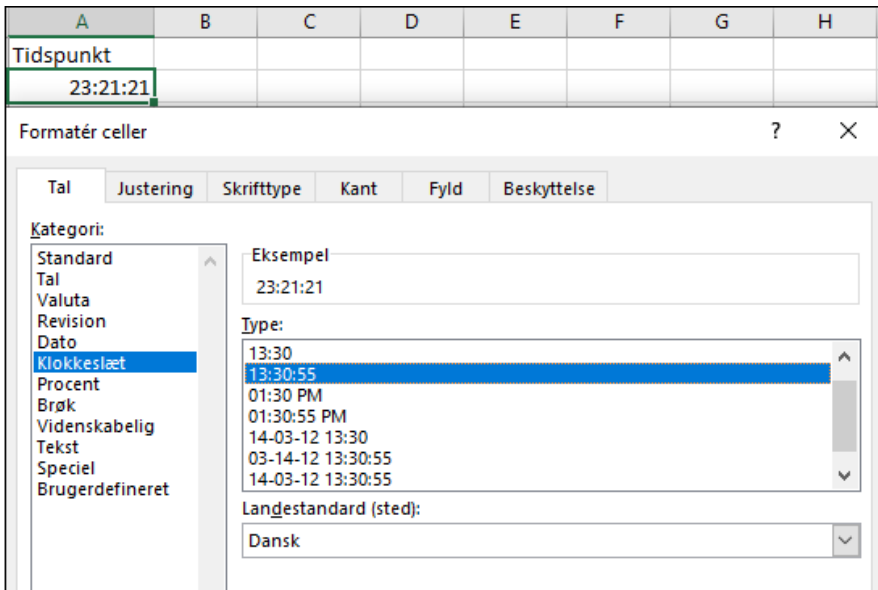
Figur 5: Valg af kategori for dato i Excel

Tidspunkt (time), DS/ISO8601:1993 standard, som svarer til **Klokkeslæt** i Excel
Bemærk at formatet er HH:MM:SS, fx 13:30:55





Rigsarkivet

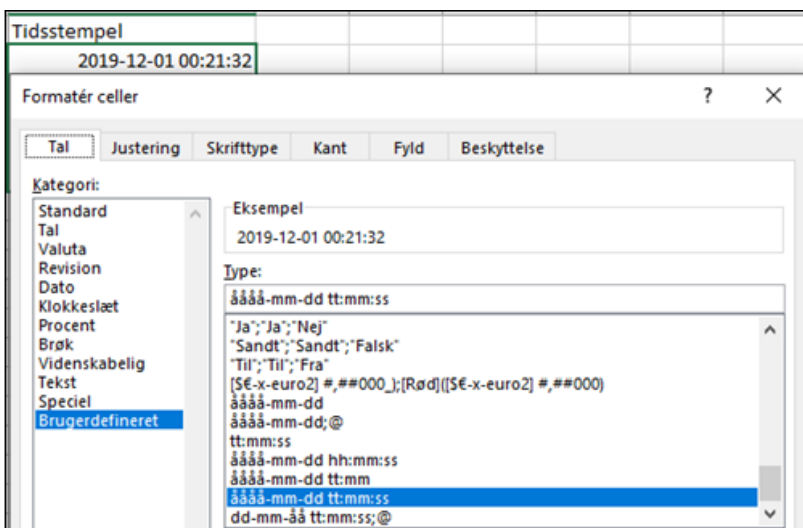


Figur 6: Valg af kategori for tidspunkt/klokkeslæt i Excel

Tidsstempel (datetime), DS/ISO8601:1993 standard

Bemærk at formatet er CCYY-MM-DD HH:MM:SS.ssssss eller CCYY-MM-DDTHH:MM:SS.ssssss
Bemærk at angivelse er millisekunder er valgfrit (.ssssss).

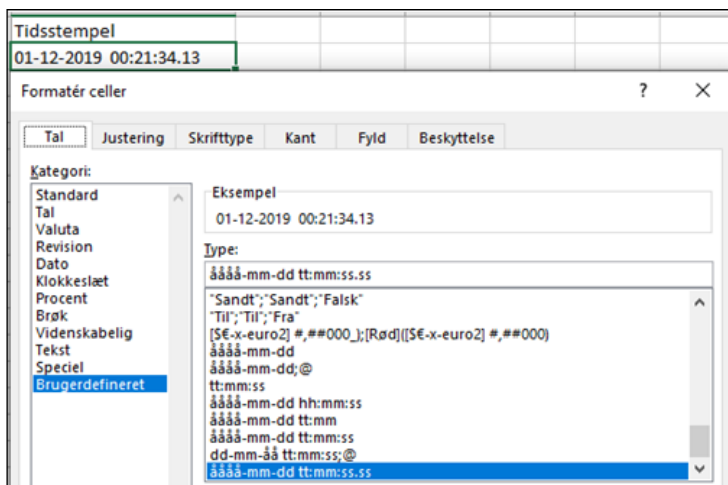
Bemærk at Excel ikke har en prædefineret kategori, der overholder dette tidsstempelformat. Du skal derfor selve definere en brugerdefineret kategori, der overholder dette format som vist herunder.



Rigsarkivet

Figur 7: Definition af brugerdefineret kategori for tidsstempel (både dato og tidspunkt) i Excel

Bemærk, at Excel maksimalt kan angive 2 cifre i millisekunder, åååå-mm-dd tt:mm:ss.ss, hvilket du angiver på følgende måde i Excel:



Figur 8: Definition af brugerdefineret kategori for tidsstempel med millisekunder (både dato og tidspunkt) i Excel

f. Afleveringspakkens indhold

En afleveringspakke består af følgende elementer, jf. figur 1:

- Mappedstruktur
- Indeksfilerne archiveIndex.xml og contextDocumentationIndex.xml placeret i mappen Indices
- Mappen ContextDocumentation med kontekstdokumenter konverteret til TIF
- Datafiler i form af csv-filer placeret i mappen Data
- Metadatafiler med fast struktur i form af txt-filer placeret i mappen Data

Du kan oprette afleveringspakkens mappedstruktur samt placere indeksfiler og kontekstdokumenter i afleveringspakken ved hjælp af programmet ASTA. Når du anvender ASTA, sørger dette program for at oprette mapper i den rette struktur og navngive mapper og filer korrekt i overensstemmelse med krav til en afleveringspakke. Du kan downloade ASTA fra [Rigsarkivets hjemmeside](#). Derefter skal du udtrække datafilerne og lave metadatafilerne og placere disse filer korrekt i afleveringspakken i Data- mappen.

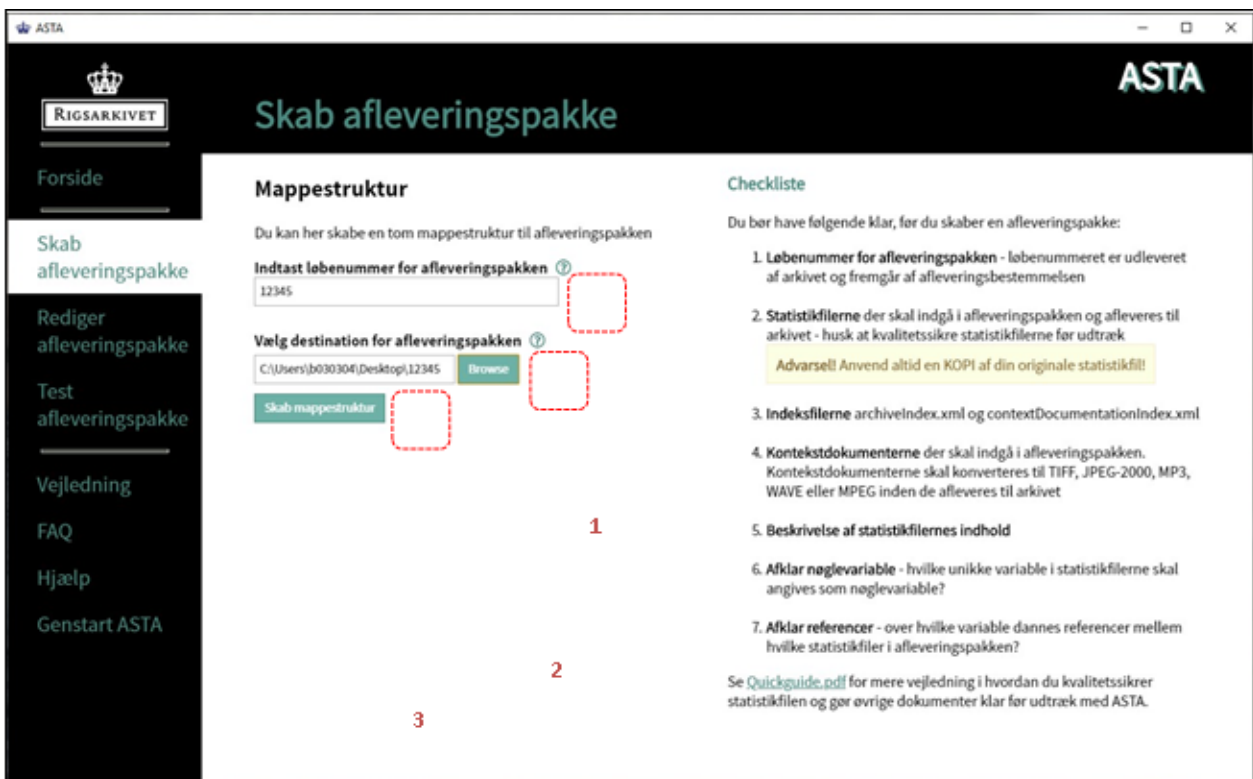
1. Download programmet ASTA fra [Rigsarkivets hjemmeside](#)
2. Klik på AstaInstaller.exe for at starte ASTA



I afsnittene herunder beskrives trin for trin, hvordan du skaber de forskellige dele af afleveringspakken.

g. Opret afleveringspakkens mappestruktur

Anvend ASTA til at oprette den ydre mappestruktur i afleveringspakken. Vælg menupunktet "Skab afleveringspakke og udfør **kun** skabelsen af mappestruktur, se figur 9.



Figur 9: Siden "Mappestruktur" under "Skab afleveringspakke" i ASTA

1. Indtast **løbenummeret** på afleveringspakken i overensstemmelse med det løbenummer, der er tildelt af arkivet og som fremgår af din afleveringsbestemmelse, fx. 12345
2. Vælg destinationen, hvor afleveringspakken skal gemmes, ved at klikke på knappen '**Browse**' og finde den rigtige mappe.
3. Klik på knappen 'Skab mappestruktur' for at oprette afleveringsmappestrukturen. Bemærk at du ikke skal fortsætte med de øvrige trin i programmet herefter.

Efter du har klikket på at knappen "Skab mappestruktur" (figur 9), vælger du menupunktet "**Rediger afleveringspakke**" for at placere indeksfiler og kontekstdokumenter i afleveringspakken. Siden i figur 10 fremkommer, hvor du vælger den afleveringspakke, du netop har skabt fx FD.12345.





Figur 10: Siden "Tilføj indeksfiler og kontekstdokumenter" under "Rediger afleveringspakke" i ASTA

1. Vælg den afleveringspakke (fx FD.12345), der skal redigeres, ved brug af '**Browse**' knappen
2. Klik på knappen '**Næste**' for at gå til næste side

h. Placer indeksfiler i afleveringspakken

Ved klik på 'Næste' i figur 10, fremkommer siden vist i figur 11. Du kan her placere indeksfilerne `archiveIndex.xml` og `contextDocumentationIndex.xml` i afleveringspakken.



Figur 11: Siden "Placer indeksfiler i afleveringspakken" under "Rediger afleveringspakke" i ASTA

1. Vælg filen `archiveIndex.xml` ved brug af '**Browse**' knappen.
2. Vælg filen `contextDocumentationIndex.xml` ved brug af '**Browse**' knappen.
3. Klik på knappen '**Næste**' for at placere de valgte filer i afleveringspakken.

i. Placer kontekstdokumenter i afleveringspakken

Ved klik på 'Næste' i figur 11, fremkommer siden vist i figur 12.





Rigsarkivet

Du kan her placere kontekstdokumenter i afleveringspakken eller erstatte eksisterende kontekstdokumenter med nye.

1. Alle kontekstdokumenter registreret i den valgte contextDocumentationIndex.xml vises i en dokumentliste. Klik på knappen '**Browse**' ud for hvert dokument og vælg det kontekstdokument, der svarer til dokumenttitlen i dokumentlisten.

Bemærk: Du kan godt placere dokumenterne, selvom de ikke er konverteret til arkivformat fx TIFF. Så skal du blot konvertere dem senere inde fra afleveringspakken. Du kan også forlade ASTA her og tilføje dem senere når du har konverteret dem.

2. Når du er færdig med at placere alle kontekstdokumenterne, anbefales det, at du udskriver dokumentlisten på knappen '**Udskriv dokumentliste**' og kontrollerer, at indholdet af de valgte dokumenter svarer til dokumenttitlen. Når du klikker på knappen 'Udskriv dokumentliste' vises dokumentlisten i din standardbrowser fx Edge (se figur 13). Anvend browserens printfunktion til at udskrive listen (se figur 14). Dokumentlisten gemmes også automatisk i en html-version samme sted, hvor den valgte afleveringspakke er placeret, når du klikker på knappen 'Udskriv dokumentliste'.

3. Klik på knappen '**Næste**' for at placere de valgte dokumenter i afleveringspakken. Et skærmbillede med bekræftelse af, at afleveringspakken er færdig fremkommer.

Bemærk: Din afleveringspakke er ikke færdig endnu. Du skal placere data- og metadatafiler i afleveringspakken, før du tester den med ASTA's funktion "Test afleveringspakke".





Rediger afleveringspakke

Indeksfilerne er placeret

Placer kontekstdokumenter i afleveringspakken

FD.12345

Du skal tilføje nedenstående kontekstdokumenter til afleveringspakken. Du kan kun placere en fil per dokumentmappe. Hvis et kontekstdokument ikke tilføjes, oprettes en tom dokumentmappe navngivet med mappennummeret. Hvis du ikke har kontekstdokumenterne klar, kan du udskrive dokumentlisten og tilføje dem manuelt senere.

Information! Kontekstdokumenter skal konverteres til bevaringsformat før de afleveres til arkiv, fx .tif eller .mp3

Dokumentliste

Mappennummer	Dokumenttitel	Vælg dokument
1	Afleveringsbestemmelse	<input type="text" value="C:\Users\b030304\Desktop\FD.12345\ContextDocumentation\docCollection1\1\1.tif"/> <input type="button" value="Browse"/>
2	System formål notat	<input type="text" value="C:\Users\b030304\Desktop\FD.12345\ContextDocumentation\docCollection1\2\1.tif"/> <input type="button" value="Browse"/>
3	Indsamlingsmetode	<input type="text" value="C:\Users\b030304\Desktop\FD.12345\ContextDocumentation\docCollection1\3\1.tif"/> <input type="button" value="Browse"/>
4	Forskningsresultater	<input type="text" value="C:\Users\b030304\Desktop\FD.12345\ContextDocumentation\docCollection1\4\1.tif"/> <input type="button" value="Browse"/>
5	Projektbeskrivelse	Vælg sti med knappen <input type="button" value="Browse"/>
6	Spørgeskema	Vælg sti med knappen <input type="button" value="Browse"/>
7	Protokol	Vælg sti med knappen <input type="button" value="Browse"/>

INFO: Filerne archiveIndex.xml, contextDocumentationIndex.xml er kopieret til mappen "Indices" i afleveringspakken

Figur 12: Siden "Placer kontekstdokumenter i afleveringspakken" under "Rediger afleveringspakke" i ASTA





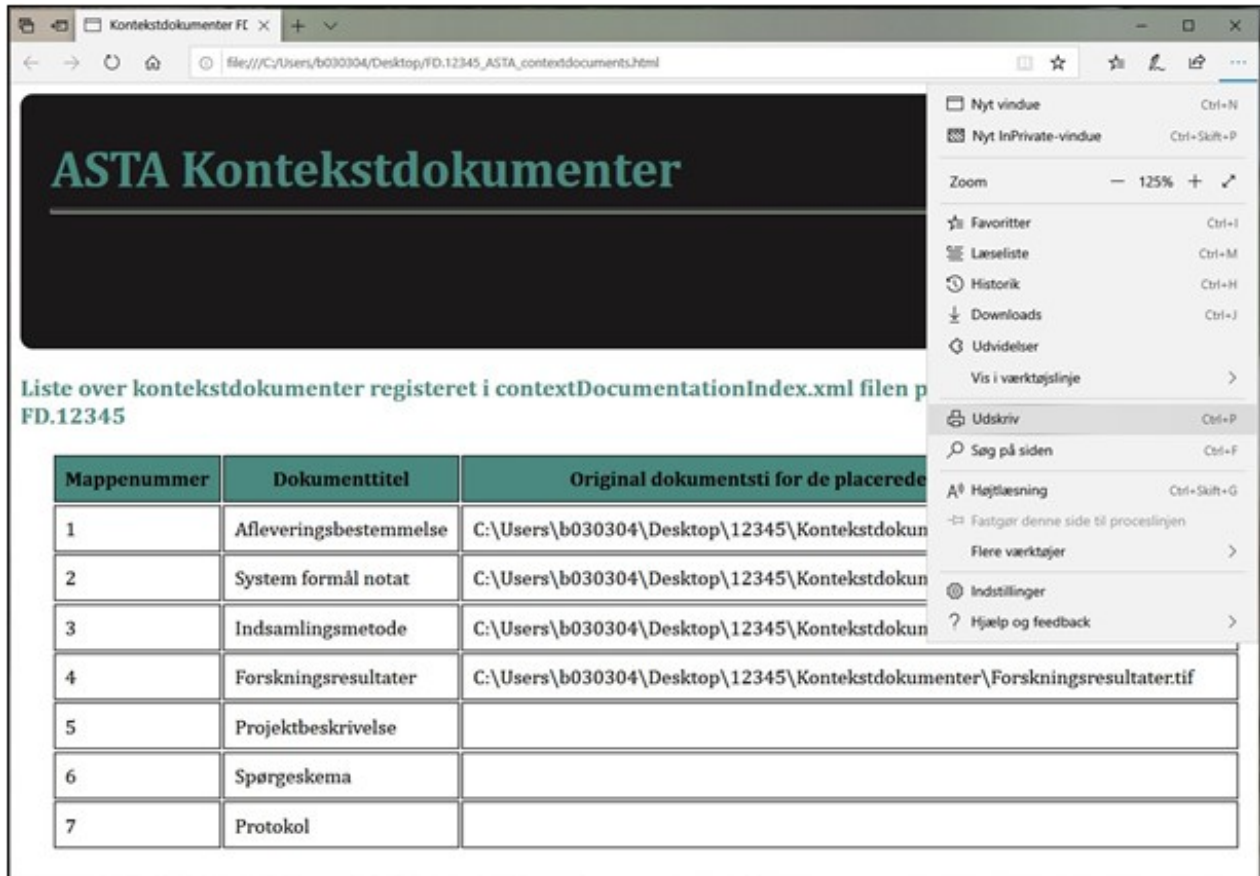
ASTA Kontekstdokumenter

Liste over kontekstdokumenter registeret i contextDocumentationIndex.xml filen placeret i afleveringspakke: FD.12345

Mappenummer	Dokumenttitel	Original dokumentsti for de placerede kontekstdokumenter
1	Afleveringsbestemmelse	C:\Users\b030304\Desktop\12345\Kontekstdokumenter\Afleveringsbestemmelse.tif
2	System formål notat	C:\Users\b030304\Desktop\12345\Kontekstdokumenter\System_formål_notat.tif
3	Indsamlingsmetode	C:\Users\b030304\Desktop\12345\Kontekstdokumenter\Indsamlingsmetode.tif
4	Forskningsresultater	C:\Users\b030304\Desktop\12345\Kontekstdokumenter\Forskningsresultater.tif
5	Projektbeskrivelse	
6	Spørgeskema	
7	Protokol	

Figur 13: "Dokumentliste" som fremkommer ved klik på "Udskriv dokumentliste" i ASTA

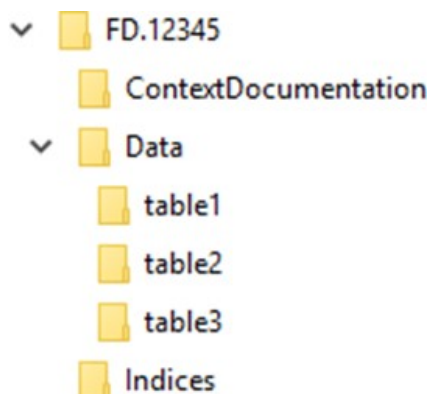




Figur 14: Print af "Dokumentliste" via Browserens printfunktion

Du mangler nu kun at skabe og placere data- og metadatafilerne i mappen "Data".

j. Skab undermapper i Data-mappen



Start med at skabe undermappestrukturen ved at oprette en undermappe med navnet 'table1' under mappen "Data". Der skal være en 'table'-mappe for hvert datasæt, som skal afleveres. Hvis der skal afleveres tre regneark, skal der oprettes mapperne 'table1', 'table2' og 'table3' under 'Data'-mappen. Hvis du kun skal aflevere ét datasæt, oprettes kun 'table1'.

Figur 15





k. Udtræk data fra regnearket til datafilen (.csv)

Før du udtrækker data til en .csv-fil fra Excel, skal du sikre dig, at alle dataformater (formater på indhold i celler) stemmer overens med de tilladte dataformater, jf. punkt 5 i tjeklisten i afsnit "E. Tjekliste – trin før dannelse af en afleveringspakke" oven for samt illustreret i figur 2.

For at udtrække data fra regnearket til en csv-fil, der overholder formatet for csv-filer i afleveringspakken, skal du gøre følgende:

1. Åben filen i Excel. Klik på 'Filer' > 'Åben' > og vælg det regneark du skal lave udtræk fra.
2. Tjek at kolonnenavne fremgår af første række i regnearket, jf. punkt 4 i tjeklisten i afsnit "E. Tjekliste – trin før dannelse af en afleveringspakke" ovenfor.
3. Tjek at alle kolonner har tilladte dataformater, jf. punkt 5 i tjeklisten i afsnit "E. Tjekliste – trin før dannelse af en afleveringspakke" oven for samt illustreret i figur 2. Tjek særligt celler med klokkeslæt, datoer og tidstempler (både tid og dato)
4. Gem filen som csv ved at klikke på 'Filer' > 'Gem som'
5. Vælg den placering, hvor du vil gemme csv-filen fx direkte i afleveringspakken i mappen 'Data' > 'table1'
6. Vælg filtypen "CSV (Semikolonsepareret) (*.csv)" i drop down, der fremkommer ved klik på lille pil ud for "Filtype"
7. Angiv filnavn på csv-filen fx "table1.csv", hvis udtrækket skal placeres i mappen 'table1'
8. Klik på 'Gem'

Data fra det valgte regneark er nu udtrukket til en semikolonsepareret fil og placeret i afleveringspakken.

Hvis du har flere regneark, der afleveres i samme afleveringspakke, skal du udføre trin 1-8 ovenfor for hver datafil og placere dem i mapperne table2, table3 osv.





	A	B	C	D	E	F	G	H	I
1	Lbnr	Koen	Afdeling	Antal_hceller	Antal_rceller	Beskrivelse	Laege_bes	Tidspunkt	Proeve_tidpkt
2	1	2	9	1,22	1.22	glad	2013-12-31	10:00:00	2013-12-31 10:01:32
3	2	2	2	10,3	10.3	Jeg er "blue"	1999-03-30	13:00:00	1999-03-30 13:02:12
4	3	3	1	8,14	8.14	rolig; glad	2012-11-21	14:00:00	2012-11-21 14:03:56
5	4	3	1	99,12	99.12	stresset	2001-01-15	11:00:00	2001-01-15 11:04:46
6	5	1	1	7,23	7.23	sur	1980-10-10	08:00:00	1980-10-11 08:05:05
7									

Figur 16: Eksempel på en Excel-fil med tilladte dataformater

```
Lbnr;Koen;Afdeling;Antal_hceller;Antal_rceller;Beskrivelse;Laege_bes;Tidspunkt;Proeve_tidpkt
1;2;9;1,22;1.22;glad;2013-12-31;10:00:00;2013-12-31 10:01:32
2;2;2;10,3;10.3;"Jeg er "blue"";1999-03-30;13:00:00;1999-03-30 13:02:12
3;3;1;8,14;8.14;"rolig; glad";2012-11-21;14:00:00;2012-11-21 14:03:56
4;3;1;99,12;99.12;stresset;2001-01-15;11:00:00;2001-01-15 11:04:46
5;1;1;7,23;7.23;sur;1980-10-10;08:00:00;1980-10-11 08:05:05
```

Figur 17: Eksempel på indhold af csv eksporteret fra regnearket i figur 16

I. Kontroller at udtræk af data til csv-fil er tabsfrit

Det er dit ansvar at sikre, at det udtræk, der laves fra originalfilen med Excel eller andre værktøjer, er tabsfrit. Du bør derfor foretage en visuel kontrol og sammenligne data i originalfilen med data i den udtrukne csv-fil.

m. Skab metadatafilen (.txt)

Du skal nu skabe metadatafilen (table1.txt), der beskriver indholdet i datafilen (fx table1.csv). Indholdet i metadatafilen skal overholde specifikke regler vedrørende struktur og indhold. Disse regler er defineret i figur 9.11 i bilag 9 til bekendtgørelse om arkiveringsversioner og forklares nærmere herunder.





Rigsarkivet

Regler for metadatafil:

1. Metadatafilen skal være i tekstformat med ekstension .txt
2. Tegnsættet i metadatafilens skal være indkodet som UTF-8 (se Vejledning om UTF-8)
3. Metadatafilen skal navngives 'table1.txt', hvis den placeres i mappen 'table1' i Data-mappen i afleveringspakken
4. Teksten i metadatafilen skal venstrestilles, og der må ikke laves tabuleringer
5. Metadatafilen skal indeholde følgende 9 overskrifter/etiketter angivet med upper-case: SYSTEMNAVN, DATAFILNAVN, DATAFILBESKRIVELSE, NØGLEVARIABLE, REFERENCE, VARIABLE, VARIABLEBESKRIVELSE, KODELISTE, BRUGERKODE
6. Regler og formater for indhold under overskrifter/etiketterne er beskrevet i figur 10

Se eksempel på indhold i en metadatafil i figur 10

Du kan også bruge eksemplet på en afleveringspakke, der er tilgængelig på [Rigsarkivets hjemmeside](#), som inspiration til hvordan du kan danne en metadatafil, der overholder kravene i figur 10.

Du skal navngive metadatafilen med samme navn som den datafil, den beskriver, fx table1.txt, og placere i samme mappe som denne datafil, jf. punkt 7 i afsnit K.

Du skal skabe en metadatafil til hver datafil, som indgår i afleveringspakken.

n. Placer data og metadatafilerne i afleveringspakken

Navngiv data- og metadatafilerne og kopier dem ind i table-mapperne i Data-mappen.

Table1 mappen skal indeholde datafilen navngivet table1.csv og metadatafilen navngivet table1.txt osv. Se figur 1.

Når du har navngivet og placeret alle udtrukne data- og metadatafiler korrekt i afleveringspakken er den færdig og klar til test.



Figur 10 Metadatafilens indhold

Etiket	Beskrivelse af etiketindhold	Udfaldsrum	Obligatorisk	Eksempel
SYSTEMNAVN	Navn på det program, som data udtrækkes fra, eller datas oprindelige format	SPSS SAS Stata Excel eller fritekst	Ja	SYSTEMNAVN Excel
DATAFILNAVN	Navnet på datafilen, som den benævnes i brugssammenhæng. Bemærk at datafilnavn ikke skal indeholde filens ekstension (fx .xls) Datafilnavne skal være unikke inden for samme afleveringspakke	ISO/IEC 9075:1999 – Database Language SQL (SQL-99). Dvs. at: Datafilnavn må ikke være længere end 128 tegn, starte med et tal eller inkludere blanke mellemrum eller	Ja	DATAFILNAVN Sundhedsundersøgelsen



		specialtegn såsom %		
DATAFILBESKRIVELSE	<p>Beskrivelse af datafilens indhold.</p> <p>Fx den fulde titel for statistikfilen, en eksisterende projektbeskrivelse eller et abstract, der præcist dækker statistikfilens indhold.</p> <p>Bemærk at datafilbeskrivelsen ikke må indeholde tegn for linjeskift</p>	Fritekst	Ja	DATAFILBESKRIVELSE Sundhedsundersøgelse af danskere som bor i Rostock, Tyskland
NØGLEVARIABLE	<p>Datafilens unikke nøglevariabel angivet med navne på den/de variable, nøglevariablen består af.</p> <p>En nøglevariabel kan være sammensat af flere variable. Flere variable i en nøglevariabel adskilles af mellemrum.</p> <p>Nøglevariablen skal entydigt identificere en række i datafilen.</p>	ISO/IEC 9057:1999 – Database Language SQL (SQL-99) Dvs. at: Variabelnavn, der indgår i en nøglevariabel, må ikke være længere end	Ja, hvis unik nøglevariabel findes	NØGLEVARIABLE Fødselsdato lbnr



	Dvs. to værdier i variabelen må ikke have samme værdi.	128 tegn, starte med et tal eller inkludere mellemrum eller specialtegn såsom %.		
REFERENCE	<p>Referencer til andre datafiler i afleveringspakken</p> <p>En reference til en anden datafil i afleveringspakken angives med navnet på den datafil, der refereres til (fremmeddatafil), efterfulgt af variabelnavnet for fremmeddatafilens nøglevariabel (fremmedvariabel) efterfulgt af variabelnavnet for den variabel (referencevariabel) i datafilen, der refererer til fremmedvariablen i fremmeddatafilen.</p> <p>Der skal være fuld overensstemmelse mellem</p>	<p>ISO/IEC 9057:1999 – Database Language SQL (SQL-99). Dvs. at:</p> <p>Datafilnavn og variabelnavne, der indgår i referencen, må ikke være længere end 128 tegn, starte med et tal eller inkludere blanke mellemrum eller specialtegn såsom %.</p>	Ja, hvis reference til anden datafil i afleveringspakken findes	REFERENCE Nøglefil 'nøglefil_id' 'hovedfil_id'



	<p>datatype og længde i nøglevariablene, der indgår i referencen.</p> <p>Hvis der findes flettevariable, som kan koble to datasæt (datafiler) i afleveringspakken sammen, skal disse angives. Der skal kun angives de relationer, hvor sammenfletningen resulterer i, at variable tillægges datasættet (horisontal sammenfletning). Dvs. i det tilfælde, hvor der er to datasæt med forskellige variable men samme individer, som over en fælles flettevariabel kan samles til et datasæt.</p> <p>Relationer til eventuelle nøglefiler skal også angives her. Hvis du fx afleverer et datasæt, hvor personfølsomme oplysninger fra datasættet er udtrukket til en nøglefil, skal relationen mellem datasættet og nøglefilen defineres i en</p>			
--	--	--	--	--



	reference.			
VARIABEL	<p>Her angives alle variable/kolonner i datafilen. Hver variabel/kolonne angives på separat linje.</p> <p>En variabel/kolonne angives som et sæt bestående af variabelnavn efterfulgt af notation for variabelens dataformat samt en eventuel kodelistereference.</p> <p>For hver variabel angives således 2-3 oplysninger:</p> <p>Første oplysning er angivelse af variabelnavn. Variabelnavne skal være unikke inden for samme metadatafil.</p> <p>Variabelnavne skal begynde med et bogstav og må bestå af en blanding af bogstaver, tal samt underscore på max 128 tegn.</p>	<p>Navngivning af variable skal overholde ISO/IEC 9075:1999 - Database Language SQL (SQL-99). Dvs. at:</p> <p>Variabelnavne, der indgår i referencen må ikke være længere end 128 tegn, starte med et tal eller inkludere blanke mellemrum eller specialtegn såsom %.</p>	Ja	<p>VARIABEL</p> <p>Lbnr int</p> <p>Koen int Koen.</p> <p>Afdeling int Afdeling.</p> <p>Antal_hceller decimal</p> <p>Antal_rceller decimal</p> <p>Beskrivelse string</p> <p>Laege_bes date</p> <p>Tidspunkt time</p> <p>Proeve_tidpkt datetime</p>



	<p>Anden oplysning er angivelse af variabelens datatype. Følgende xml-datatypeer er tilladte:</p> <ul style="list-style-type: none">• string (anvendes til tekst)• int (anvendes til heltal)• decimal (anvendes til decimaltal)• date (anvendes til datoer)• time (anvendes til tidspunkt/klokkeslæt)• datetime (anvendes til tidsstempel - både dato og klokkeslæt) <p>Tredje oplysning er angivelse af en eventuel kodelistereference. Hvis datatypen er numerisk heltal eller decimaltal, angives kodelistereferencen med det valgte kodelistenavn og et efterstillet punktum ».«</p> <p>Hvis datatypen er tekst, angives referencen med det valgte kodelistenavn med et foranstillet</p>	<p>Hvis variabelnavnet er et reserveret ord i SQL:1999, skal det enten omdøbes eller omkranses af dobbeltapostrof.</p>		
--	---	--	--	--



	dollartegn »\$« og et efterstillet punktum ».«.			
VARIABELBESKRIVELSE	<p>Beskrivelser af variabelenes/kolonnerens indhold med angivelse af variabelnavn efterfulgt af beskrivelsen omkranset af enkeltapostroffer.</p> <p>Hver beskrivelse angives på separat linje,</p> <p>Alle variable/kolonner skal have en beskrivelse, der udfyldes så udførligt som muligt, således at en fremtidig bruger har mulighed for at forstå hvilken oplysning, der er registreret i kolonnen. Alle forkortelser i beskrivelsen skal forklares, og eventuelle måleenheder skal angives (fx meter, centimeter, kilo). Findes datoer skal det angives, hvad datoen omhandler fx bryllupsdato eller fødselsdato.</p>	Fritekst	Ja	<p>VARIABELBESKRIVELSE</p> <p>Lbnr 'Løbenummer' Koen 'Patientens køn'</p> <p>Afdeling 'Undersøgelsesafdeling' Antal_hceller 'Antal af hvide celler' Antal_rceller 'Antal af røde celler' Beskrivelse 'Ord som bedst beskriver dit humør'</p> <p>Laege_bes 'Dato for lægebesøg (CCYYMMDD)'</p> <p>Tidspunkt 'Tidspunkt lægeundersøgelse' Proeve_tidpkt 'Tidspunkt blodprøve'</p>



	<p>Findes oplysninger om flere individer i et datasæt fx både mor og barn, skal det tydeligt fremgå af variabelbeskrivelsen hvilke variabeloplysninger, der er knyttet til hvilke individer fx barnets køn eller respondentens køn.</p> <p>For spørgeskemaundersøgelser anbefaler Rigsarkivet, at alle spørgsmålsteksterne indsættes i variabel-/kolonnebeskrivelserne. Hermed bliver det lettere at søge i og anvende data for fremtidige brugere. Spørgsmålsteksterne bør indsættes i deres oprindelige formulering og fulde længde.</p>			
KODELISTE	<p>Kodelister angivet med kodelistens navn efterfulgt af flere sæt bestående af kode og kodebeskrivelse. Begge oplysninger skal hver især omkranses af enkeltapostroffer og adskilles med mellemrum</p>	<p>Navngivning af kodelister skal overholde ISO/IEC 9075:1999 - Database Language SQL</p>	<p>Ja, hvis der findes kodelister</p>	<p>KODELISTE</p> <p>Koen</p> <p>'1' 'Mand'</p> <p>'2' 'Kvinde'</p> <p>Afdeling</p>



	<p>Hvis du anvender koder i datafilen, som ikke er forklaret i andre kolonner, kan du angive dem her. Husk også at angive en kodelistereferencenotation for kolonnen hvor koden anvendes under etiketten VARIABEL.</p> <p>I spørgeskemaundersøgelser svarer kodelister til svarkategorier til et spørgsmål.</p>	<p>(SQL-99). Dvs. at:</p> <p>Kodelistenavne ikke må være længere end 128 tegn, starte med et tal eller inkludere blanke mellemrum eller specialtegn såsom %.</p>		<p>'1' 'Celile Blodbank'</p> <p>'2' 'Marie Vang Blodbank'</p> <p>'9' 'Uoplyst'</p>
BRUGERKODE	<p>Brugerdefinerede koder for manglende værdier angivet med navnet på den variabel, hvor koderne anvendes, efterfulgt af de brugerdefinerede koder omkranset af enkeltapostroffer og adskilt med mellemrum.</p> <p>Eksempler på brugerdefinerede koder for manglende værdier:</p> <p>9 = Uoplyst</p> <p>10 = Irrelevant</p>	<p>Navngivning af variable skal overholde ISO/IEC 9075:1999 - Database Language SQL (SQL-99). Dvs. at</p> <p>variabelnavne ikke må være længere end 128 tegn, starte</p>	Ja, hvis der findes brugerdefinerede koder for manglende værdier	BRUGERKODE Afdeling '9'



	<p>11 = Deltager ikke</p> <p>Brugerdefinerede koder skal også fremgå af en kodeliste under etiketten KODELISTE knyttet til den kolonne, hvor de anvendes med en kodelistereference under etiketten VARIABEL.</p> <p>Brugerdefinerede koder for manglende variable må kun anvendes i kategoriske og numeriske variable med datatyperne heltal, decimaltal og tekst.</p>	<p>med et tal eller inkludere blanke mellemrum eller specialtegn såsom</p> <p>%.</p>		
--	--	--	--	--

Filer Rediger Formater Vis Hjælp

SYSTEMNAVN

Excel

DATAFILNAVN

Sundhedsundersøgelsen

DATAFILBESKRIVELSE

Sundhedsundersøgelse af danskere som bor i Rostock, Tyskland

NØGLEVARIABLE

Lbnr

REFERENCE

VARIABLE

Lbnr int

Koen int Koen.

Afdeling int Afdeling.

Antal_hceller decimal

Antal_rceller decimal

Beskrivelse string

Laege_bes date

Tidspunkt time

Proeve_tidpkt datetime

VARIABLEBESKRIVELSE

Lbnr 'Løbenummer'

Koen 'Patientens køn'

Afdeling 'Undersøgelsesafdeling'

Antal_hceller 'Antal af hvide celler i 1mm³ blod'Antal_rceller 'Antal af røde celler i 1mm³ blod'

Beskrivelse 'Ord som bedst beskriver dit humør'

Laege_bes 'Dato for lægebesøg (CCYYMMDD)'

Tidspunkt 'Tidspunkt for lægeundersøgelse (hh:mm:ss)'

Proeve_tidpkt 'Dato og tidspunkt for blodprøvetagning (CCYYMMDD hh:mm:ss)'

KODELISTE

Koen

'1' 'Mand'

'2' 'Kvinde'

Afdeling

'1' 'Cellule Blodbank'

'2' 'Marie Vang Blodbank'

'9' 'Uoplyst'

BRUGERKODE

Afdeling '9'

Figur 10 Eksempel på struktur og indhold i metadatafil



3. Aflevering af csv-fil som afleveringspakke

Hvis du har csv-filer udtrukket fra andre programmer end Excel, kan du også lave en afleveringspakke med disse på følgende måde:

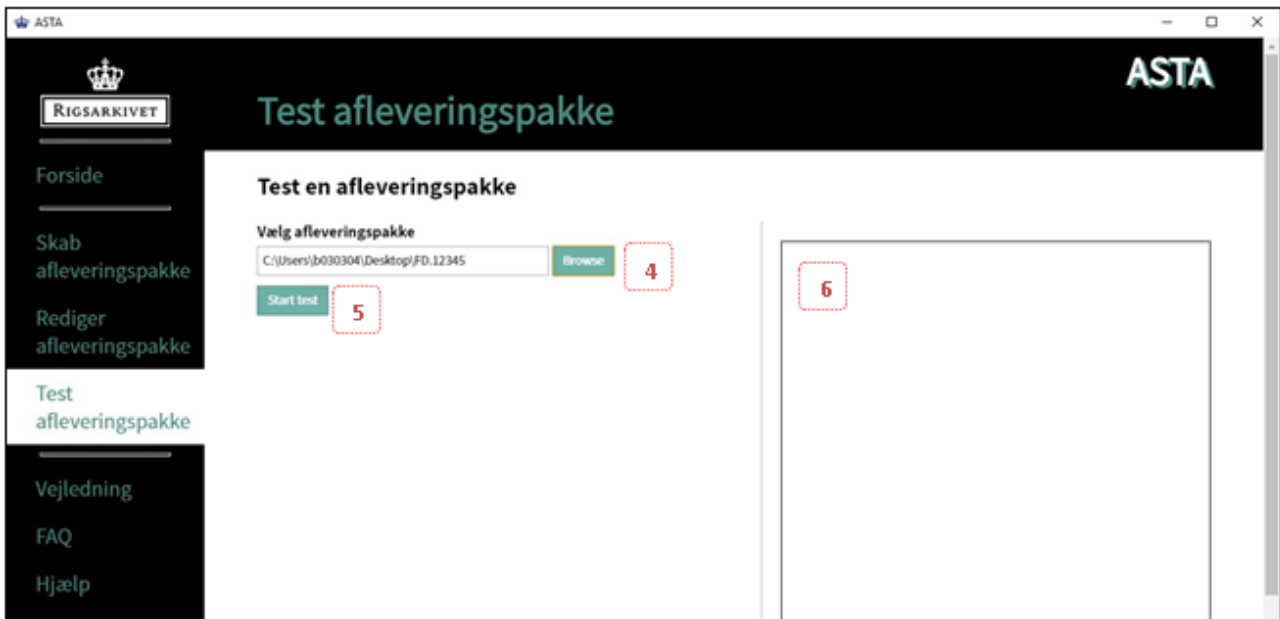
1. Skab afleveringspakkens mappestruktur og placer indeksfiler og kontekstdokumenter som beskrevet i afsnit G-I.
2. Opret undermapperne i Data-mappen som beskrevet i afsnit H.
3. Du skal sikre dig, at csv-filen overholder krav til indhold af datafiler jf. krav til datafiler i bilag 9 punkt 9.F-9.H i bekendtgørelse om arkiveringsversioner. Med værktøjet ASTA kan du teste om din csv-fil overholder regler for datafiler. Regler for datafiler omfatter, at
 - a. csv-filen skal være semikolonsepareret
 - b. første linje i datafilen skal angive kolonnenavne/variabelnavne
 - c. csv-filen skal overholde regler for escape og semikolon i tekstværdier. Dvs. hvis semikolon indgår i en værdi for en variabel, skal hele værdien omslutes med dobbelt apostrof »"« (U+0022). Hvis dobbelt apostrof indgår i en værdi for en variabel, skal dobbelt apostroffen foranstilles med en dobbelt apostrof, og hele værdien omslutes med dobbelt apostrof.
 - d. dataformaterne i csv-filen skal overholde de formater datatyper som er beskrevet i figur 2- 8
 - e. manglende værdier i datafilen skal angives som ingen værdi »« eller et mellemrum » «.
4. Skab metadatafilerne manuelt som beskrevet i afsnit M

4. Sådan tester de afleveringspakken

Det er dit ansvar, at struktur og indhold i afleveringspakken overholder kravene til afleveringspakken i bilag 9 i bekendtgørelse om arkiveringsversioner. Derfor er det vigtigt, at du tester afleveringspakken og retter eventuelle fejl, inden du afleverer afleveringspakken til arkivet. Programmet ASTA kan hjælpe dig med at teste dette.

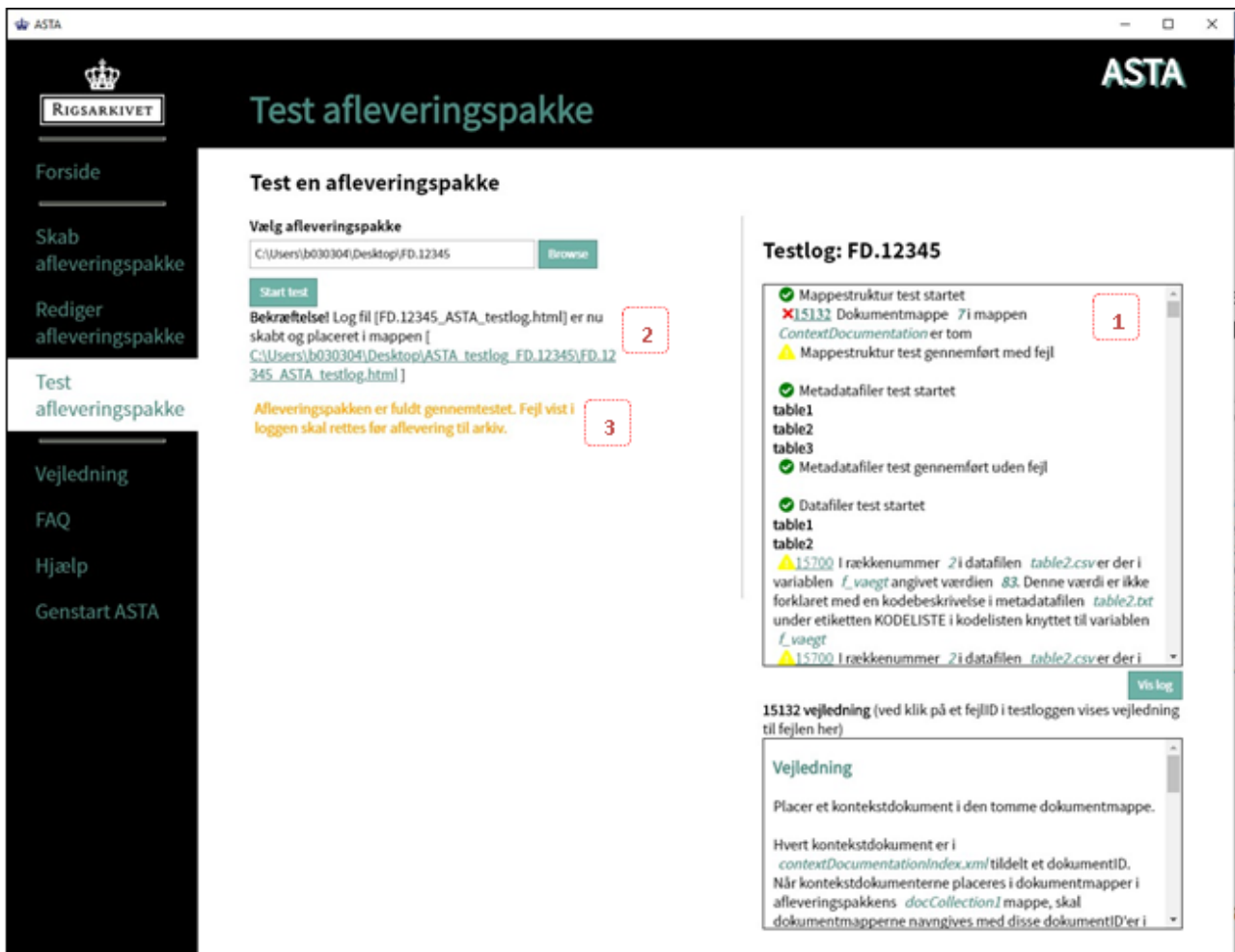
1. Download programmet ASTA fra [Rigsarkivets hjemmeside](#)
2. Klik på Astainstaller.exe for at starte ASTA
3. Vælg menupunktet 'Test afleveringspakke', se figur 18
4. Vælg den afleveringspakke, du vil teste (fx FD.12345) ved at klikke på knappen '**Browse**' i figur 18
5. Klik på knappen '**Start test**'
6. Testen er nu påbegyndt, og teststatus vises i **testloggen**, se også figur 19





Figur 18: Siden 'Test afleveringspakke' i ASTA før testen påbegyndes





Figur 19: Siden 'Test afleveringspakke' i ASTA når testen er afsluttet

Teststatus og rettelse af fejl i testlog

1. Gennemgå nøje de genererede **fejlmeldelser** markeret med rødt kryds i testloggen og ret altid disse inden aflevering til arkivet. **Hints** vises som en advarsel med gult udråbstegn. Hints er ikke altid fejl og skal kun rettes, hvis det vurderes, at det er en fejl eller mangel i den pågældende afleveringspakke.
2. Ud over den testlog, der vises på skærbilledet, genereres også en **html-version af testloggen**, som også gemmes automatisk på samme placering, som den valgte afleveringspakke efter endt test (se figur 20). Hvis du klikker på linket til denne testlog, åbnes den i din standard browser og kan printes ved hjælp af browserens printfunktion (se figur 14). Html-testloggen indeholder også vejledning til, hvordan fejlbeskeder i testloggen forstås og rettes (se figur 21).



Rigsarkivet

3. Når testen er kørt færdig, vises en af følgende tre **statusmeddelelser**:

a. **Afleveringspakken er gennemtestet uden fejl og kan nu afleveres til arkivet**

Dette indikerer, at afleveringspakken er klar til at blive afleveret til arkivet.

Bemærk: Da ASTA ikke tester for alle krav i bilag 9, kan arkivet vende tilbage med yderligere fejl eller mangler, der skal rettes.

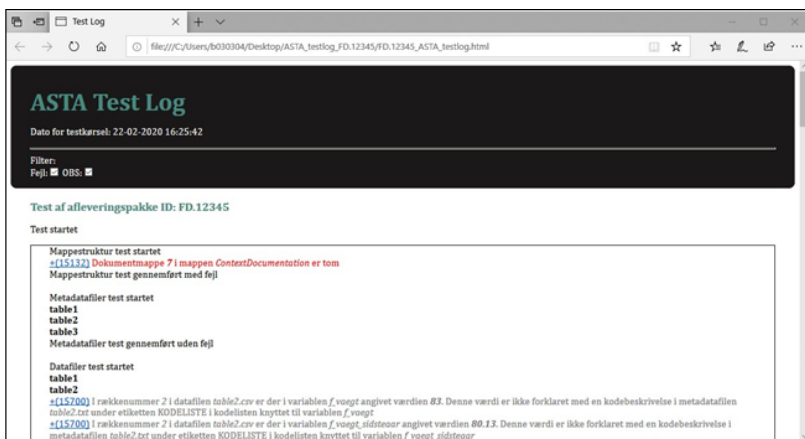
b. **Afleveringspakken er fuldtgennemtestet. Fejl vist i loggen skal rettes før aflevering til arkiv**

Dette indikerer, at testen er afsluttet, men der stadig findes fejl, der skal rettes. Når du har rettet de viste fejl i testloggen og kører en ny test, fremkommer der ikke flere nye fejl.

c. **Testen er afbrudt, og afleveringspakken er ikke testet færdig. Fejl vist i loggen skal rettes, før testen kan fortsætte.**

Dette indikerer, at der er fejl, der skal rettes, inden testen kan fortsætte. Når du har rettet de viste fejl i testloggen og kører en ny test, kan der fremkomme flere nye fejl. Ved flere tests vises max 40 af samme type fejl i testloggen, hvilket betyder, at når du har rettet de viste 40 fejl og kører en ny test, kan der forekomme flere af samme type.

Html

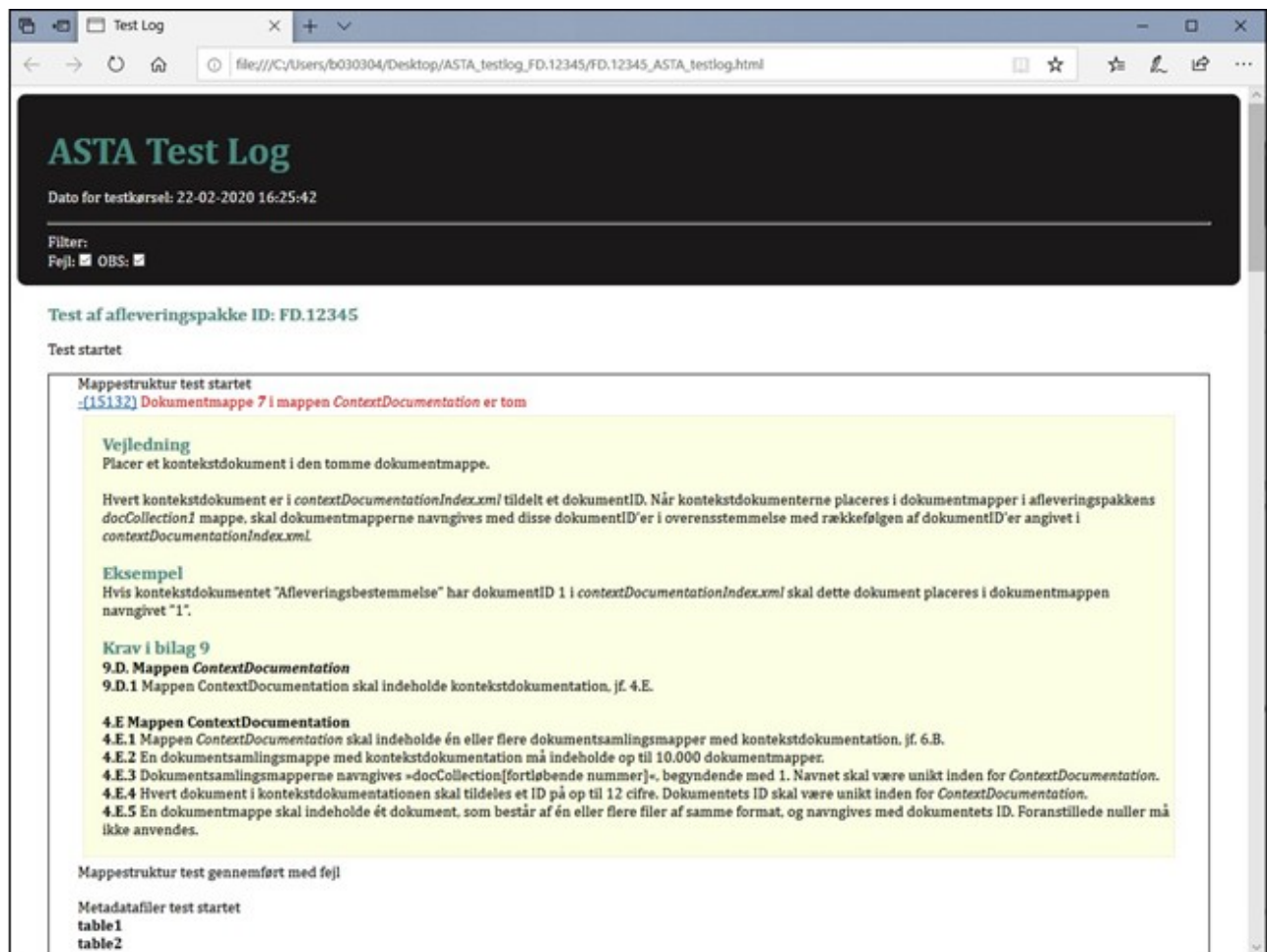


Figur 20: Html-version af testloggen fra 'Test afleveringspakke' i ASTA



Html-version af testlog med vejledning

1. I html-testloggen som fremkommer, når du klikker på linket fra ASTA til denne (se 2 i figur 19) vises et fejl-ID (fx 15132) ud fra hver funden fejl. Ved klik på dette fejl-ID fremkommer en vejledningstekst til, hvordan fejlen skal forstås og rettes. Der vises også et eksempel på data og metadata uden fejl samt angivelse af den konkrete ordlyd af kravene, som de står skrevet i bilag 9 i Bekendtgørelse om arkiveringsversioner (se figur 21)



Figur 21: Html-testlog med visning af vejledning ved klik på fejl-ID 15132

5. Support i Rigsarkivet

Hvis du oplever problemer med at skabe og teste afleveringspakken, bedes du kontakte datamanageren for forskningsdata i Rigsarkivet på følgende e-mail: mailbox@sa.dk.

