

# Vejledning i aflevering af forskningsdata bestående af statistikdata og dokumenter

---

*Vejledning i hvordan du afleverer forskningsdata, som både indeholder data fra statistikfiler og dokumenter i form af tekst, lyd, video eller billeder*

*Rigsarkivet juli 2020*

*Version 1.0*

## Indhold

<b>0. Læsevejledning .....</b>	<b>2</b>
A. Vejledningens målgruppe og anvendelse.....	2
B. Henvisning til øvrig vejledning.....	2
C. Lovgivning og retsfor skrifter.....	2
D. Definitioner.....	2
<b>1. Format for aflevering .....</b>	<b>4</b>
A. Statistikfiler afleveres som en afleveringspakke .....	4
B. Digitale dokumenter konverteres til arkivformat og afleveres i en mappestruktur .....	5
C. Oplysninger om digitale dokumenter registreres i et regneark .....	5

## 0. Læsevejledning

Offentlige myndigheder er forpligtet til at aflevere bevaringsværdige data og dokumenter til et offentligt arkiv.

**Denne vejledning beskriver hvordan forskningsdata afleveres, hvis der er tale om et studie, som både indeholder data fra statistikfiler og dokumenter i form af tekst, lyd, video eller billeder. Det kan være et forskningsprojekt der har indsamlet data via en spørgeskemaundersøgelse og samtidigt udført kvalitative interviews med udvalgte respondenter.**

Arkivet træffer afgørelse om, hvorvidt data fra et forskningsprojekt kan afleveres samlet i henhold til denne vejledning eller om data skal afleveres i to separate afleveringer.

### A. Vejledningens målgruppe og anvendelse

Denne vejledning henvender sig til dem, som afleverer forskningsdata bestående af både statistikfiler og dokumenter.

### B. Henvisning til øvrig vejledning

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

- Eksempelafleveringspakke med statistikdata FD.18005
- Quickguide – til produktion og test af en afleveringspakke med ASTA
- ASTA brugervejledning
- Vejledning til produktion af afleveringspakke med data fra regneark eller csv-filer
- Vejledning til bekendtgørelse om arkiveringsversioner nr. 128 (se bilag 9 om afleveringspakker)
- Vejledning til Skab archiveIndex
- Vejledning til Skab contextDocumentationIndex
- Vejledning om konvertering af dokumenter til TIFF
- Vejledning om UTF-8

Følgende vejledning har betydning for aflevering af dokumenterne:

- Regneark til oplysninger om de digitale dokumenter (Skabelon)

Alt vejledningsmateriale kan tilgås fra Rigsarkivets hjemmeside [www.sa.dk](http://www.sa.dk).

### C. Lovgivning og retsfor skrifter

Information om lovgivning m.v. findes på Rigsarkivets hjemmeside [www.sa.dk](http://www.sa.dk).

### D. Definitioner

#### Afleveringsbestemmelse

Før afleveringen af arkiveringsversionen sættes i gang, udarbejder det modtagende arkiv en bestemmelse, som fastsætter indholdet af afleveringen. Afleveringsbestemmelsen er en kravspecifikation for **indholdet** af afleveringen, mens bekendtgørelse om arkiveringsversionen fastsætter **formatet** for afleveringen.

**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.

**Dokumenter** i afleveringen er data, fx dagbøger, kvalitative interviews, transskriberede interviews, videooptagelser til feltstudier osv.

**Kontekstdokumenter** er dokumenter, der beskriver de afleverede data/dokumenter.

## 1. Format for aflevering

Et forskningsprojekt bestående af både statistikfiler og dokumenter skal afleveres på følgende måde:

1. Statistikfiler afleveres som en afleveringspakke jf. bestemmelserne i bilag 9 i bekendtgørelse nr. 128 om arkiveringsversioner
2. Digitale dokumenter konverteres til arkivformat og afleveres som en kopi af en mappestruktur indeholdende dokumenterne
3. Registreringer af oplysninger om alle digitale dokumenter placeret i mappestrukturen registreres i et regneark, der overholder denne vejlednings metadatastandard for dokumentssamlinger

Arkivet konverterer afleveringspakken, de digitale dokumenter og dokumentregistreringerne i regnearket til en samlet arkiveringsversion jf. bestemmelserne i bekendtgørelse nr. 128 om arkiveringsversioner, bilag 1-8. Afleveringsformatet specificeres nærmere herunder.

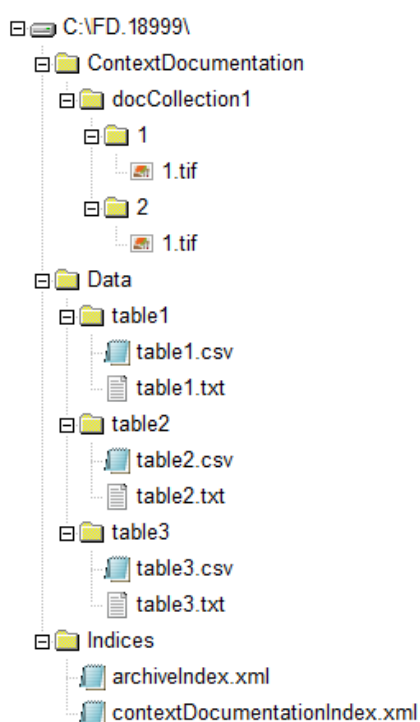
### A. Statistikfiler afleveres som en afleveringspakke

Denne vejledning omhandler data som oprindeligt er skabt og bearbejdet i statistikprogrammer som f.eks. SAS, Stata, SPSS eller R og lagret i et af disse statistikfilsformater eller indsamlede data registreret i et regneark og anvendt til statistisk analyse i f.eks. Excel.

Rigsarkivet har fastsat en række bestemmelser for aflevering af statistikfiler i form af en afleveringspakke af hensyn til bevaring og fremtidig brug af data, som alle myndigheder skal overholde. Disse bestemmelser er beskrevet i Rigsarkivets bekendtgørelse nr. 128 om arkiveringsversioner, bilag 9.

Afleveringspakkens struktur og indhold fremgår af figur 1 og kan skabes med programmet ASTA. Indeksfilerne kan skabes med Rigsarkivets indtastningsprogrammer *Skab archiveIndex.exe* og *Skab contextDocumentationIndex.exe*. Både bekendtgørelsen, vejledninger samt programmerne kan hentes fra Rigsarkivets hjemmeside [www.sa.dk](http://www.sa.dk).

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



Mappen **ContextDocumentation** indeholder dokumenter konverteret til bevaringsformat, f.eks. TIFF. Det kan være en metoderapport, et spørgeskema eller andre dokumenter, der beskriver de data, afleveringspakken indeholder.

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, f.eks. variabelbeskrivelser (variable labels), svarkategorier (value labels) og koder for manglende værdier (missing values), 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 f.eks. oplysninger som navn på datasættet, der afleveres, navn på forsker der har indsamlet data, perioden data dækker, adgangsbegrænsninger til data osv. Filen **contextDocumentationIndex.xml** indeholder oplysninger om kontekstdokumenterne placeret i mappen ContextDocumentation, f.eks. dokumentets titel, forfatter samt emnekategorisering af dokumentet.

## B. Digitale dokumenter konverteres til arkivformat og afleveres i en mappestruktur

En kopi af en mappestruktur indeholdende de digitale dokumenter i arkivformat afleveres sammen med afleveringspakken. Mappedstrukturen placeres på afleveringsmediet sammen med afleveringspakken med statistikdata i en mappe navngivet lbnr\_Dokumenter, hvor lbnr er løbenummeret for afleveringen udleveret af arkivet.

Rigsarkivet skelner mellem kontekstdokumenter og digitale dokumenter. Kontekstdokumenter betragtes som metadata der beskriver de afleverede data, fx en projektbeskrivelse, en metoderapport, et spørgeskema eller variabelbeskrivelser. Kontekstdokumenterne placeres i afleveringspakken jf. punkt 1. Andre dokumenter som kvalitative interviews, lydoptagelser af interviews eller videooptagelser til feltstudier betragtes derimod som data. Disse skal ligesom kontekstdokumenterne konverteres til et af følgende seks arkivformater inden aflevering til arkiv:

1. TIFF (.tif) til almindelige kontordokumenter i form af tekst, regneark og billeder
2. JPEG2000 (.jp2) til store tegninger, kort og fotos
3. MP3 (.mp3) til lydfile
4. WAV (.wav) til lydfile hvor høj kvalitet er vigtig
5. MPEG2 (.mpg) til videofile
6. MPEG4 (.mpg) til videofile

De digitale dokumenter skal overholde de specifikationer, som fremgår af bekendtgørelse nr. 128 om arkiveringsversioner, 5 E-F samt H-J.

De digitale dokumenter kan konverteres og placeres inden for den oprindelige mappestruktur, placeres i enkelt mappe eller placeres i en nyoprettet mappestruktur, der eventuelt ordner og emneinddeler dokumenterne.

Vejledning til konvertering af dokumenter, fx "Vejledning i konvertering til TIFF-format" med PDFCreator kan hentes fra Rigsarkivets hjemmeside [www.sa.dk](http://www.sa.dk).

## C. Oplysninger om digitale dokumenter registreres i et regneark

Oplysninger om de digitale dokumenter skal registreres i et regneark. En fast skabelon for dette regneark skal anvendes til registreringerne og kan hentes fra Rigsarkivets hjemmeside [www.sa.dk](http://www.sa.dk)

Regnearket indeholder kolonner til registrering af de 15 oplysninger fra metadatastandarden Dublin Core. Hertil er tilføjet oplysningen SourcePath. Metadatastandarden anvendt i skabelonen samt beskrivelser af hvordan oplysningerne udfyldes fremgår af figur 2.

Bemærk ved udfyldelse af regnearket:

- Der udfyldes en række i regnearket pr. dokument i mappestrukturen.
- Kolonner skal bibeholdes i regnearket i den angivne rækkefølge selvom de ikke er udfyldt.
- Kolonnenavne fremgår af række 1 i regnearket.
- Kolonnebeskrivelserne fra metadatastandarden i figur 2 fremgår af række 2 i regnearket.
- Det er tilladt at tilføje egne eller redigerede kolonnebeskrivelser i række 3 i regnearket, så de mere præcist beskriver indholdet af kolonnen.
- Oplysningerne **SourcePath** og **Titel** er obligatoriske at udfylde for alle dokumenter registreret i regnearket.

- **Dato** for dokumentet skal udfyldes, hvis den findes og skal overholde formatet ÅÅÅÅ-MM-DD, fx 2019-08-12.
- Oplysningen **Relation** skal udfyldes, hvis der findes en relation mellem det registrerede dokument og en række i en statistikfil som også findes i afleveringspakken (fra punkt 1 A). Hvis denne oplysning angives (fx med et cpr-nummer eller et ID) skal kolonnebeskrivelsen for kolonnen tilpasses (i række 3 i regnearket), så det fremgår af denne beskrivelse hvilken variabel i hvilken statistikfil i afleveringspakken denne reference knyttes til.
- Øvrige oplysninger i regnearket er frivillige at udfylde. De registrerede oplysninger skal anvendes af fremtidige brugere af materialet til fremsøgning af et relevant dokument.
- Det er tilladt at tilføje yderligere oplysninger om dokumenterne i ekstra kolonner indsat efter sidste kolonne i regnearket. Kolonnetitel i række 1 samt kolonnebeskrivelse i række 3 er obligatorisk at udfylde for de ekstra tilføjede kolonner.

## Figur 2. Vejledningens metadatastandard for dokumentssamlinger

Oplysninger markeret med \* er obligatoriske at udfylde

Kolonnenavn	Kolonnebeskrivelse
<b>SourcePath*</b>	Angivelse af den fulde sti til dokumentet i mappestrukturen inkl. filnavn og ekstension, fx S:\ Skilsmisseprojekt\Børn og skilsmisse\Interview A – barn.tif
<b>Title*</b>	( <i>Titel</i> ) Angivelse af original titel eller titel registreret efterfølgende på dokumentet. Dvs. billedets, filmens, videoens, lydoptagelsens eller tekstdokumentets titel.
<b>Creator</b>	( <i>Skaber</i> ) Angivelse af navn på den der har taget initiativ til eller skabt dokumentets indhold. Fx navn på forfatter, interviewer, forsker, fotograf, forsøgsperson eller andre personer der har produceret materiale, der indgår som empiri i forskningen.
<b>Subject</b>	( <i>Emne</i> ) Angivelse af emne eller kategori knyttet til dokumentet. Fx emneord der beskriver indholdet i dokumentet, navn på den mappe hvor dokumentet er placeret eller andre kategoriseringer, der grupperer dokumenterne.
<b>Description</b>	( <i>Indholdsbeskrivelse</i> ) Fritekstbeskrivelse af dokumentets indhold. Fx hvad dokumentet indeholder, hvad filmen handler om eller hvad billedet viser.
<b>Publisher</b>	( <i>Udgiver</i> ) Angivelse af udgiver af dokumentet. Fx filmselskab, producent, forlag. Sjældent relevant at udfylde for forskningsdata.
<b>Contributor</b>	( <i>Bidragyder</i> ) Angivelse af navne på personer eller andre enheder, der har bidraget til skabelsen af indholdet af dokumentet. Fx medvirkende i en video, folk der taler på en lydoptagelse, interviewpersoner eller respondenter.
<b>Date</b>	( <i>Produktionsdato</i> ) Angivelse af hvornår dokumentet er skabt. Fx oprettelsesdato for et dokument, dato for hvornår en video er optaget eller et interview er foretaget. Den angivne dato skal overholde formatet ÅÅÅÅ-MM-DD.
<b>Type</b>	( <i>Objekttype</i> ) Angivelse af dokumentets type valgt fra følgende faste liste: Lyd, Video, Tekst, Billede.

<b>Format</b>	<b>(Originalformat)</b> Angivelse af dokumentets originale format. Her kan fx angives originalt medie, såsom tegning, fotografi, VHS-bånd eller kassettebånd. Her kan også angives det digitale format dokumentet havde før konvertering til bevaringsformat, fx mov, wma, regneark, word, pdf, docx eller jpeg.
<b>Identifiser</b>	<b>(Identifikation)</b> Angivelse af et originalt ID for dokumentet anvendt i en given kontekst.
<b>Source</b>	<b>(Oprindelig kilde)</b> Angivelse af en reference til et andet materiale, som dokumentet er udledt af/digitaliseret fra. Fx angivelse af strekkode, navn, ID eller anden identifikation af materialet der refereres til.
<b>Language</b>	<b>(Sprog)</b> Angivelse af sprog anvendt i dokumentet.
<b>Relation</b>	<b>(Relation)</b> Angivelse af en oplysning der udgør en reference fra dette dokument til en række i en statistikfil som findes i afleveringspakken. Her kan fx angives et CPR-nummer eller et ID som også findes i statistikfilen og som udgør denne reference.
<b>Coverage</b>	<b>(Størrelse)</b> Angivelse af dokumentets mål. Fx størrelse på et billede hvor højde gange bredde er angivet i centimeter (cm) eller meter (m), et dokumentets størrelse angivet i Kilobytes (KB), Megabytes (MB) eller Gigabytes (GB) eller en videoptagelses spilletid angivet i minutter.
<b>Rights</b>	<b>(Rettigheder)</b> Angivelse af copyright ejer, dvs. hvem der skal give tilladelse til brug af dokumentet.