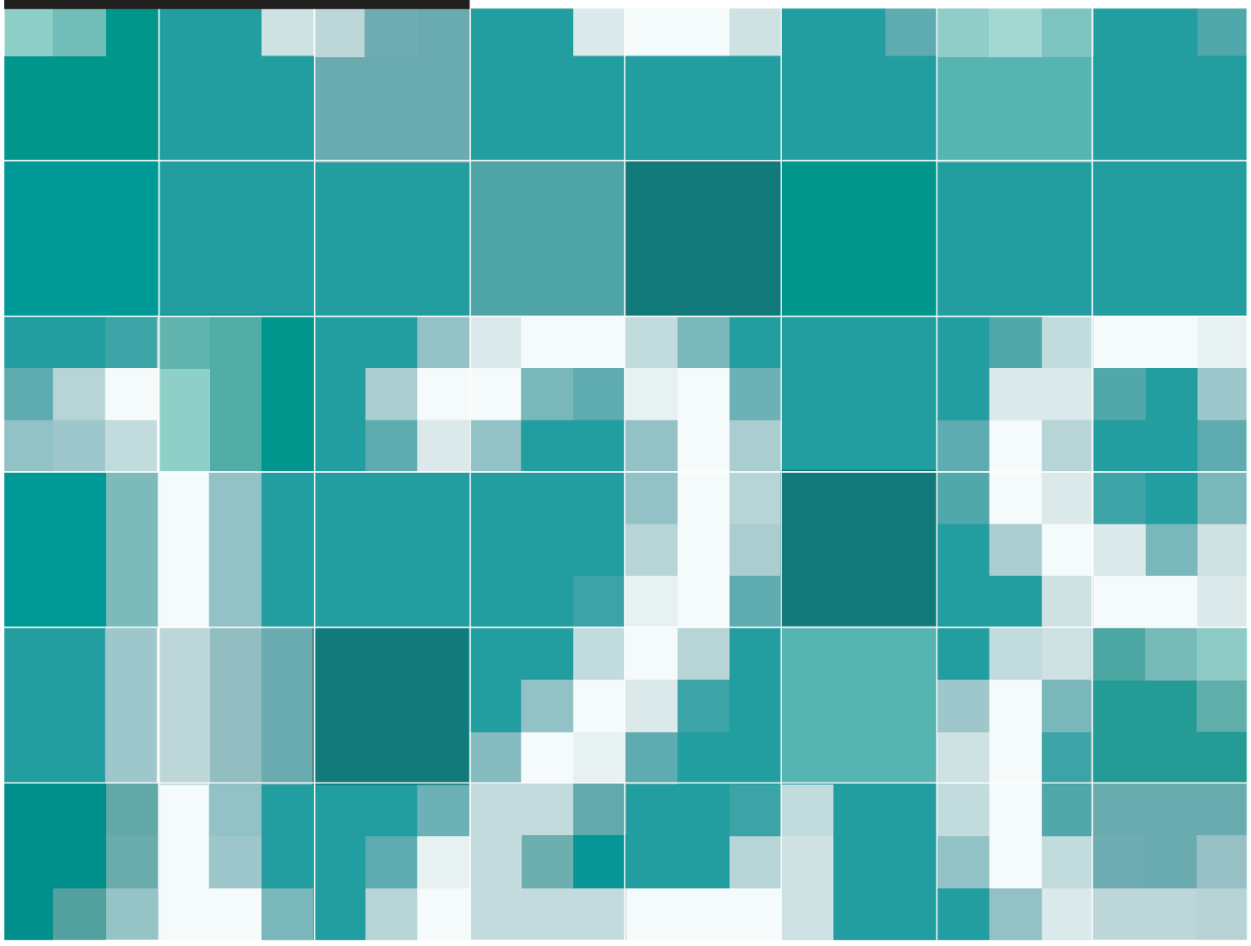
RIGSARKIVET

Vejledning til bekendtgørelse om arkiveringsversioner

Bekendtgørelse nr. 128 af 12. februar 2020



www.sa.dk

Indholdsfortegnelse

Paragraf 1-8 4
Bilag 1 – Bilagsfortegnelse 10
Bilag 2 – Strukturen i en arkiveringsversion 11
Bilag 3 – Arkiveringsversion af digitale data og eventuelle dokumenter 18
Bilag 4 – Datastruktur 24
Bilag 5 – Dataindhold 43
Bilag 6 – Information om arkiveringsversionen 65
Bilag 7 – Afleveringsmedie 111
Bilag 8 – Skemaer 111
Bilag 9 – Afleveringspakke for visse typer af forskningsdata 112
Bilag A – Ændringer i forhold til tidligere bekendtgørelse (bek. 1007) 161

For detaljeret fortegnelse af indholdet af bilagene se bilag 1, s. 10

Indledning

Dette er vejledning til bekendtgørelse nr. 128 om arkiveringsversioner, som fastsætter formatet for arkivering af bevaringsværdige data fra offentlige myndigheders it-systemer og forskningsdata samt dokumentationen af samme. Vejledningen uddyber og eksemplificerer bekendtgørelsesteksten.

Formålet med arkivering af bevaringsværdige data fra offentlige myndigheders it-systemer og forskningsdata mv. er at sikre, at data kan genanvendes til ny vidensproduktion både i nutiden og fremtiden.

Bevaringsværdige offentlige data skal arkiveres i det format, som er fastsat i bekendtgørelsen. Formålet er at sikre, at alle data ligger i et kendt format, så fremtidige konverteringer af data, som er nødvendige for at sikre bevaringen, kan ske efter en ensartet metode. Bevaringen skal sikre, at arkivbrugerne også om 100 eller 500 år kan få adgang til data skabt i de offentlige myndigheders it-systemer og forskningsdata i begyndelsen af det 21. århundrede.

Rigsarkivet afgør hvilke data fra offentlige it-systemer og hvilke forskningsdata, der skal bevares for fremtiden. Data fra statslige myndigheder skal afleveres til Rigsarkivet. Data fra kommuner og regioner skal afleveres til Rigsarkivet eller til et kommunalt eller regionalt arkiv, oprettet i henhold til arkivlovens § 7. Det modtagende arkiv bestemmer, hvordan bekendtgørelsen skal implementeres på data og tester, om arkiveringsversionen/afleveringspakken opfylder bekendtgørelsens krav.

Læsevejledning

Vejledningen henvender sig primært til personer, som skal producere en arkiveringsversion/afleveringspakke eller bidrage til dokumentationen af samme.

I vejledningen gennemgås og eksemplificeres bekendtgørelsesteksten paragraf for paragraf. Hele bekendtgørelsesteksten er indsat i vejledningen med fed skrift, og efter hver paragraf/afsnit følger vejledningsteksten i almindelig skrift, dog vil vejledningsteksten indgå i en særskilt kolonne i nogle skemaer.

I bilag A findes en oversigt over hvilke ændringer der er foretaget i bekendtgørelsen og vejledningen ved sidste revision i 2019.

På Rigsarkivets hjemmeside ligger flere kortere vejledninger, indtastningsprogrammer, udtrækprogrammer og eksempler på arkiveringsversioner, som kan understøtte produktionen af arkiveringsversioner. Endvidere ligger Rigsarkivets testprogram, ADA, også til fri download, så det er muligt at teste arkiveringsversioner inden aflevering til arkiv.

Begrebet ”data fra it-systemer” anvendes i bekendtgørelsen og dækker både data fra offentlige administrative IT-systemer såvel som forskningsdata og data fra private arkivskabere som findes i form af databaser eller dokumentsamlinger såsom tekst-, lyd- eller billedfiler placeret i en mappestruktur på et fildrev. Forskningsdata i disse formater er f.eks. kvalitative forskningsprojekter bestående af transskriberede interviews eller store indsamlede datasæt og måleresultater lagret i databaser. Alle disse typer af data skal afleveres som arkiveringsversioner efter reglerne i bekendtgørelsens bilag 1-8. Forskningsdata skabt eller bearbejdet i statistikprogrammer skal derimod afleveres som en afleveringspakke efter reglerne i bilag 9.

Paragraf 1-8

§ 1

Bekendtgørelsen gælder for arkivalier som hidrører fra den offentlige forvaltning og domstolene, og som af Rigsarkivet er bestemt til bevaring.

Bestemmelserne gælder for centrale og lokale statslige myndigheder, kommunale og regionale myndigheder samt domstolene. Bestemmelserne omfatter arkivalier i digital form. Digitale arkivalier kan være databærende it-systemer med og uden dokumenter, forskningsdata samt digitale lyd- og videofiler. Bestemmelserne omfatter også arkivalier på papir eller andre medier, såfremt disse af Rigsarkivet er bestemt til digitalisering f.eks. af bevaringsmæssige årsager.

It-systemer kan f.eks. være:

- journalsystem
- ESDH-system
- EDH-system
- fagsystem
- register
- system sammensat af data og eventuelt dokumenter fra flere forskellige systemer i en serviceorienteret arkitektur (SOA-system)

Lyd og video i analog form skal digitaliseres før aflevering til offentligt arkiv, jf. bilag 5.

Forskningsdata kan f.eks. være:

- Data fra spørgeskemaundersøgelser
- Lydfiler fra Interviewundersøgelser
- Databaser med opsamling af data til forskningsbrug
- Dokumentsamlinger som danner grundlag for forskning
- Observationsdata
- etc.

Bestemmelserne i denne bekendtgørelse omfatter kun data, der er bestemt til bevaring af Rigsarkivet:

- Statslige myndigheders data fra administrative it-systemer og forskningsdata er omfattet af en bevarings- eller kassationsbestemmelse, som Rigsarkivet udsteder i forbindelse med anmeldelsen af it-systemet/data til Rigsarkivet. Bestemmelsen for data kan slås op i Rigsarkivets arkivdatabase Daisy på www.sa.dk.

- Lyd og video kan enten indgå i et it-system eller være selvstændige filer. Hvis de er en del af et it-system, er de omfattet af bevarings- og kassationsbestemmelsen for dette it-system. Er lyd og video en særskilt enhed i en myndighedsarkiv, vil de være omfattet af bevarings- og kassationsbestemmelse for myndighedens øvrige arkivalier.

- Kommunale og regionale myndigheders it-systemer samt lyd og video er omfattet af bevarings- og kassationsbestemmelser i bekendtgørelsesform. Bestemmelserne kan ses på www.sa.dk på siderne for kommunale myndigheder.

§ 2

Bevaring af digitale data skal ske i form af arkiveringsversioner.

***Stk. 2*. Bevaringsværdig lyd og video digitaliseres inden aflevering til offentligt arkiv og bevares som arkiveringsversioner, jf. stk. 1.**

***Stk. 3*. Hvor det af bevaringsmæssige grunde er nødvendigt, kan Rigsarkivet anvise, at bevaringsværdige arkivalier digitaliseres inden aflevering til offentligt arkiv og bevares som arkiveringsversioner, jf. stk. 1.**

En arkiveringsversion består af:

- udtræk af datatabeller
- eventuelle dokumenter og
- information om arkiveringsversionen,

som er lagret i en fast struktur.

Information om arkiveringsversionen består af:

- data om arkiveringsversionens oprindelse og indhold (**Arkivbeskrivelsesfil**)
- dokumenter om arkiveringsversionens oprindelse, indhold og brug samt af det oprindelige it-systems administrative sammenhæng eller forskningsprojektets formål, forløb mv. (**Kontekstdokumentation**)
- data om arkiveringsversionens tabeller (**Tabelindeks**)

De nærmere anvisninger for arkiveringsversionen fremgår af bekendtgørelsens bilag 2-8.

Lyd og video i analog form
Lyd og video i analog form, som er bevaringsværdigt skal digitaliseres og afleveres i arkiveringsversioner til offentligt arkiv.

Andre bevaringsformer
Rigsarkivet kan beslutte, at statslige samt kommunale/regionale papirarkivalier eller arkivalier på et andet medium, som er i dårlig stand, skal afleveres i form af en arkiveringsversion. Det kan f.eks. dreje sig om mikrofilm, svampeskadede dokumenter eller billeder på papir, som skal digitaliseres.

Bemærk at forskningsdata skabt eller bearbejdet i et statistikprogram skal afleveres som en afleveringspakke som defineret i § 3, stk. 2.

§ 3

En arkiveringsversion af bevaringsværdige data skal fremstilles efter de retningslinjer, der fremgår af bilag 2-8 i denne bekendtgørelse.

***Stk. 2*. Data, som er skabt i forbindelse med forskning med anvendelse af videnskabelig metode, og som er skabt eller bearbejdet i statistikprogrammer eller tilsvarende, afleveres som en afleveringspakke efter reglerne i bilag 9.**

***Stk. 3*. Rigsarkivaren kan anvise andre bevaringsformer end arkiveringsversioner og afleveringspakker, fremstillet efter bilag 2-9, hvis bevaringsmæssige hensyn tilsiger det.**

Se vejledning til de enkelte bilag:

Bilag 1: Bilagsfortegnelse
Bilag 2: Grafisk oversigt over elementer og struktur i en arkiveringsversion
Bilag 3: Arkiveringsversion af digitale data og eventuelle dokumenter
Bilag 4: Datastruktur
Bilag 5: Dataindhold
Bilag 6: Information om arkiveringsversionen
Bilag 7: Afleveringsmedie
Bilag 8: Skemaer
Bilag 9: Afleveringspakke for visse typer af forskningsdata

På Rigsarkivets hjemmeside, www.sa.dk, findes den seneste version af skemaerne i bilag 8.

Bilag 9 omfatter aflevering af statistikformater fra statistikprogrammer som Stata, SPSS, R o.lign, men også forskningsdata, som f.eks.er lagret i regneark.

Rigsarkivaren kan anvise en anden bevaringsform end en arkiveringsversion f.eks. kan et lille elektronisk journalsystem eller forskningsdata i form af en lille dokumentsamling bevares i udskrift, hvis det skønnes, at omkostningerne til produktion af en arkiveringsversion ikke står mål med udbyttet af en digital bevaring.

§ 4

Produktion og aflevering af arkiveringsversioner og afleveringspakker af digitale data fra den statslige forvaltning og domstolene skal finde sted på tidspunkter, der fastsættes af Rigsarkivet.

Data fra administrative systemer
Det fremgår af Rigsarkivets bevaringsbestemmelse, hvornår der første gang skal afleveres en arkiveringsversion af et it-system, samt hvor ofte der efterfølgende skal afleveres. Bevaringsbestemmelsen udstedes i forbindelse med anmeldelsen af it-systemet. Rigsarkivet tager kontakt til den pågældende myndighed, når det er tid til aflevering af en arkiveringsversion.

Tidspunkt for aflevering samt efterfølgende afleveringsfrekvens fastsættes på baggrund af de oplysninger, Rigsarkivet har på anmeldelsestidspunktet. Det kan f.eks. være oplysninger om, hvornår slettes i henhold til GDPR/Databeskyttelsesloven. Data må ikke slettes, før de indgår i en arkiveringsversion, som er godkendt af Rigsarkivet.

Det er vigtigt, at myndighederne informerer Rigsarkivet, hvis der sker ændringer i it-systemet, f.eks.:

- Administrative ændringer i brugen af systemet
- Ændring i sletningsfristen for systemet
- Systemmæssige ændringer eller driftsmæssige forhold
- Nedlukning af systemet (gælder også, hvis der sker konvertering af data til nyt system)

Samtidig er det vigtigt, at myndigheden sikrer bevaring af information om systemet, f.eks. systemdokumentation, brugervejledninger og administrative beskrivelser, så informationsmaterialet kan understøtte arkiveringen af data og evt. dokumenter fra systemet.

Forskningsdata
På samme vis vil det fremgå af Rigsarkivets bevaringsbestemmelse, hvornår statslige forskningsdata skal afleveres. Fristen for aflevering aftales med myndigheden/den ansvarlige forsker. Der skal ved fristens fastsættelse tages højde for om der er persondata i datasættet og i givet fald hvornår disse skal slettes i henhold til GDPR/Persondataloven.

§ 5

Produktion og aflevering af arkiveringsversioner og afleveringspakker af digitale data fra kommuner og regioner, som indeholder personoplysninger, skal finde sted, inden data slettes. Den pågældende kommune eller region kan træffe aftale om tidligere tidspunkter med det modtagende arkiv.

Stk. 2. Produktion af arkiveringsversioner af øvrige data skal finde sted, inden data slettes, eller når det it-system, data er skabt i, tages ud af brug.

Stk. 3. Rigsarkivet kan derudover fastsætte, at der skal produceres en arkiveringsversion af data, når det af bevaringsmæssige hensyn er nødvendigt.

Data fra administrative systemer
Kommuner og regioner skal aflevere arkiveringsversioner, der indeholder data omfattet af persondataloven, og som skal bevares, til et offentligt arkiv.

Et offentligt arkiv kan enten være Rigsarkivet eller et offentligt arkiv, som er oprettet efter arkivlovens § 7 af én eller flere kommuner eller en eller flere regioner med henblik på at bevare og stille kommunale/-regionale arkivalier til rådighed for borgere og myndigheder.

Der skal produceres en arkiveringsversion af bevaringsværdige data og eventuelt dokumenter senest på det tidspunkt, hvor data skal slettes i henhold til GDPR/Databeskyttelsesloven. Data må ikke slettes, før de indgår i en arkiveringsversion, som er godkendt af det modtagende arkiv. Kommunen eller regionen og det modtagende arkiv kan aftale, at arkiveringsversionen afleveres før den fastsatte sletningsfrist, f.eks. ved et periodeskift eller når it-systemet lukkes.

For it-systemer uden personoplysninger skal der produceres en arkiveringsversion, hvis data slettes, eller it-systemet lukkes. Data må ikke slettes, før de indgår i en arkiveringsversion, som er godkendt af et offentligt arkiv.

Rigsarkivet kan fastsætte, at der skal produceres en arkiveringsversion af data fra et it-system for at sikre bevaring af data, som i henhold til bekendtgørelser om bevaring og kassation for kommunale og regionale arkivalier er bevaringsværdige. Det kan f.eks. være et it-system uden sletningsfrist eller med meget lang sletningsfrist, som er i administrativt brug i mange år.

Forskningsdata
På samme vis skal der produceres en arkiveringsversion/afleveringspakke af forskningsdata til det modtagende arkiv før der slette i data aht. GDPR/Databeskyttelsesloven. Data må ikke slettes før afleveringen er godkendt af modtagende arkiv.

§ 6

Såfremt den afleverende myndighed i forbindelse med produktion af arkiveringsversion konstaterer fejl eller mangler i dokumentkonverteringen eller i konverteringen af tabelindhold, skal der træffes aftale med det modtagende arkiv om udbedring heraf, inden arkiveringsversionen afleveres.

Ved produktion af en arkiveringsversioner kan det vise sig, at der er dokumenter, som ikke kan konverteres til bevaringsformaterne. Det kan være filen er korrupt, blank, beskadiget osv. I de tilfælde skal myndigheden indsende en liste over dokumenter, som ikke kan konverteres til det modtagende arkiv. Arkivet kan pålægge myndighederne at fremsøge de pågældende dokumenter andre steder i myndighedens virksomhed eller som minimum dokumentere de fejl og mangler, der måtte være i konverteringen, i arkiveringsversionen.

På samme vis skal myndigheden træffe aftale med det modtagende arkiv, hvis der sker ændringer i tabelindholdet i konverteringen fra driftssystemet til arkiveringsversionen.
Alle fejl og mangler i konverteringen, som ikke kan udbedres, dokumenteres i kontekstdokumentationen efter aftale med modtagende arkiv.

§ 7

Arkiveringsversioner af statslige myndigheders data skal godkendes af Rigsarkivet. Arkiveringsversioner af kommunale og regionale myndigheders data skal godkendes af det modtagende arkiv.

Stk. 2. Arkiveringsversioner af bevaringsværdige kommunale og regionale data, som ikke er omfattet af afleveringspligt, skal godkendes af samme offentlige arkiv, som modtager myndighedens afleveringspligtige arkiveringsversioner.

Stk. 3. Data, som er overført til en arkiveringsversion, må ikke slettes hos myndigheden, før arkiveringsversionen er godkendt.

Stk. 4. For forskningsdata afleveret som en afleveringspakke efter bilag 9, må data først slettes, når det modtagende arkiv har produceret og godkendt en arkiveringsversion af afleveringspakken.

Administrative data og forskningsdata efter bilag 2-8
Efter modtagelse af en arkiveringsversion tester det modtagende arkiv den for at kontrollere om:

- arkiveringsversionen stemmer overens med bestemmelserne i afleveringsbestemmelsen
- der er overensstemmelse mellem dokumentation og data
- de generelle krav som beskrevet i bilag 2-8 er opfyldt

Hvis en arkiveringsversion ikke opfylder bestemmelserne, returneres den til myndigheden med en beskrivelse af eventuelle fejl og mangler. Det modtagende arkiv sender en godkendelse af arkiveringsversionen, når alle bestemmelser er opfyldt. Data må ikke slettes, før arkiveringsversionen er godkendt af det modtagende arkiv.

Det modtagende arkiv kan enten være Rigsarkivet eller et offentligt arkiv, som er oprettet med hjemmel i arkivlovens § 7 af én eller flere kommuner/regioner med henblik på at bevare og stille kommunale arkivalier til rådighed for borgere og myndigheder.

Bemærk, at det modtagende arkiv først overtager ansvaret for arkiveringsversionen ved modtagelsen. Det anbefales derfor, at overdragelsen af arkiveringsversionen sker under betryggende forhold. Enten ved at sende arkiveringsversionen som anbefalet post eller ved personlig overdragelse. Medier med følsomme personoplysninger krypteres efter anvisning fra det modtagende arkiv jf. 7.A.5 i denne bekendtgørelse.

Ved aflevering til Rigsarkivet skal arkiveringsversionen **sendes anbefalet** eller **overbringes personligt** med følgende modtager og adresse (der må **ikke** angives navn på sagsbehandler):

Rigsarkivet,
DBO, Testværkstedet,
Rigsdagsgården 9,
1218 København K

Kommunen eller regionen kan vælge at undlade at *aflevere* arkiveringsversioner af it-systemer uden personoplysninger til arkiv. Men kommunen eller regionen skal stadig *producere* en arkiveringsversion, og arkiveringsversionen skal godkendes af det samme offentlige arkiv, som modtager myndighedens

afleveringspligtige arkiveringsversioner. Data må ikke slettes, før arkiveringsversionen er godkendt af et offentligt arkiv.

Forskningsdata, som afleveres efter bilag 9
For forskningsdata som afleveres bilag 9 gælder, at det modtagende arkiv skal have produceret en arkiveringsversion og godkendt samme af den afleveringspakke, som er afleveret til arkivet, før data må slettes hos myndigheden.
Der gælder samme fremsendelsesregler for afleveringspakker, som for arkiveringsversioner.

§ 8

Bekendtgørelsen træder i kraft den 1. marts 2020.

Stk. 2. Bekendtgørelse nr. 1007 af 20. august 2010 om arkiveringsversioner ophæves.

Bekendtgørelsen gælder for arkivering af data, for hvilke der er udstedt afleveringsbestemmelse for efter bekendtgørelsens ikrafttrædelse.

Bilag 1 – Bilagsfortegnelse

Bilag 2 Grafisk oversigt over elementer og struktur i en arkiveringsversion s. 12

Bilag 3 Arkiveringsversion af digitale data og eventuelle dokumenter **s. 18**
3.A. Generelle regler om arkiveringsversioner s. 18
3.B. Datastruktur s. 22
3.C. Dataindhold s. 23
3.D. Information om arkiveringsversionen s. 23

Bilag 4 Datastruktur **s. 24**
4.A. Generelle regler om datastruktur s. 24
4.B. Placering af mapper og filer s. 28
4.C. Mappen Indices s. 31
4.D. Mappen Tables s. 39
4.E. Mappen ContextDocumentation s. 40
4.F. Mappen Schemas s. 41
4.G. Mappen Documents s. 41

Bilag 5 Dataindhold **s. 43**
5.A. Tabelindhold s. 43
5.B. Datatyper s. 44
5.C. Konvertering af tabelindhold til digitale dokumenter, lyd, video eller geodata s. 48
5.D. Tekstformat s. 49
5.E. Digitale dokumenter s. 50
5.F. Lyd og video s. 54
5.G. Geodata s. 54
5.H. Komprimering s. 64
5.I. Optimering s. 64
5.J. Ingen forringelse s. 64

Bilag 6 Information om arkiveringsversionen **s. 65**
6.A. Arkivbeskrivelsesfil s. 65
6.B. Kontekstdokumentation s. 97
6.C. Data om arkiveringsversionens tabeller s. 103
6.D. SQL-forespørgsler s. 110

Bilag 7 Afleveringsmedie **s. 111**

Bilag 8 Skemaer **s. 111**

Bilag 9 Afleveringspakke for visse typer af forskningsdata **s. 112**
9.A. Aflevering af forskningsdata s. 115
9.B. Afleveringspakkens mappestruktur s. 116
9.C. Mappen Indices s. 116
9.D. Mappen ContextDocumentation s. 128
9.E. Mappen Data s. 131
9.F. Tekstformat s. 132
9.G. Datafil s. 132
9.H. Datatyper s. 137
9.I. Metadatafiler s. 142

Bilag A Tidligere opdateringer af vejledningen **s. 161**

Bilag 2 – Strukturen i en arkiveringsversion

I bilag 2 i bekendtgørelsen er en illustration, som viser alle mulige elementer i en arkiveringsversion (figur 2.1). De elementer, som ***ikke*** er gråmarkerede, er grundelementer, som altid skal være i en arkiveringsversion.

Der vil i de fleste arkiveringsversioner kun være behov for grundelementerne plus et udsnit af de øvrige viste elementer. Der er derfor tilføjet flere illustrationer, som viser arkiveringsversioner for henholdsvis systemer med og uden dokumenter (figur a og b), med geodata (figur c) og med forskningsdata afleveret efter bilag 9 (figur d). Strukturen for en afleveringspakke med visse typer af forskningsdata fremgår af bilag 9, s. 115.

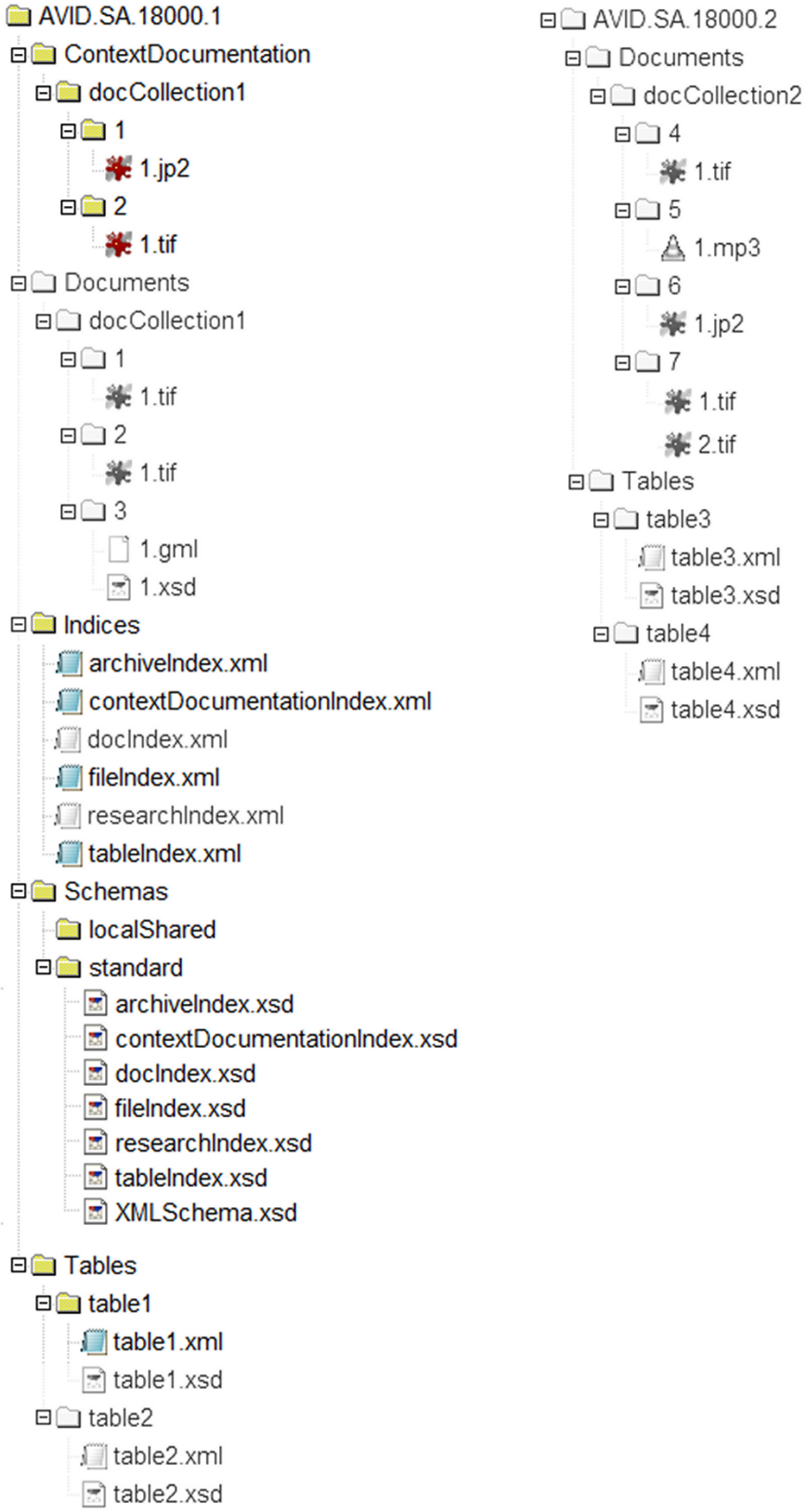
Bemærk, at skemaerne table1.xsd, table2.xsd, table3.xsd og table4.xsd der fremgår af figurerne ikke behøver at indgå i arkiveringsversionen, da Rigsarkivets testprogram ADA genererer dem ud fra tableIndex i forbindelse med test af arkiveringsversionen. localShared mappen skal altid oprettes, men kan være tom, med mindre arkiveringsversionen indeholder GML. Læs mere om ADA og download programmet på www.sa.dk.

Den anbefales at have illustrationen af den arkiveringsversion, som aktuelt skal afleveres, liggende ved siden af, når nærværende vejledning læses.

På Rigsarkivets hjemmeside, www.sa.dk, kan du endvidere download eksempler på arkiveringsversioner henholdsvis med og uden dokumenter, med geodata og med forskningsdata, som alle dels viser strukturen i en arkiveringsversion og dels selve indholdet i en arkiveringsversion for de pågældende typer af afleveringer.

Figur 2.1 Grafisk oversigt over mulige elementer og struktur i en arkiveringsversion

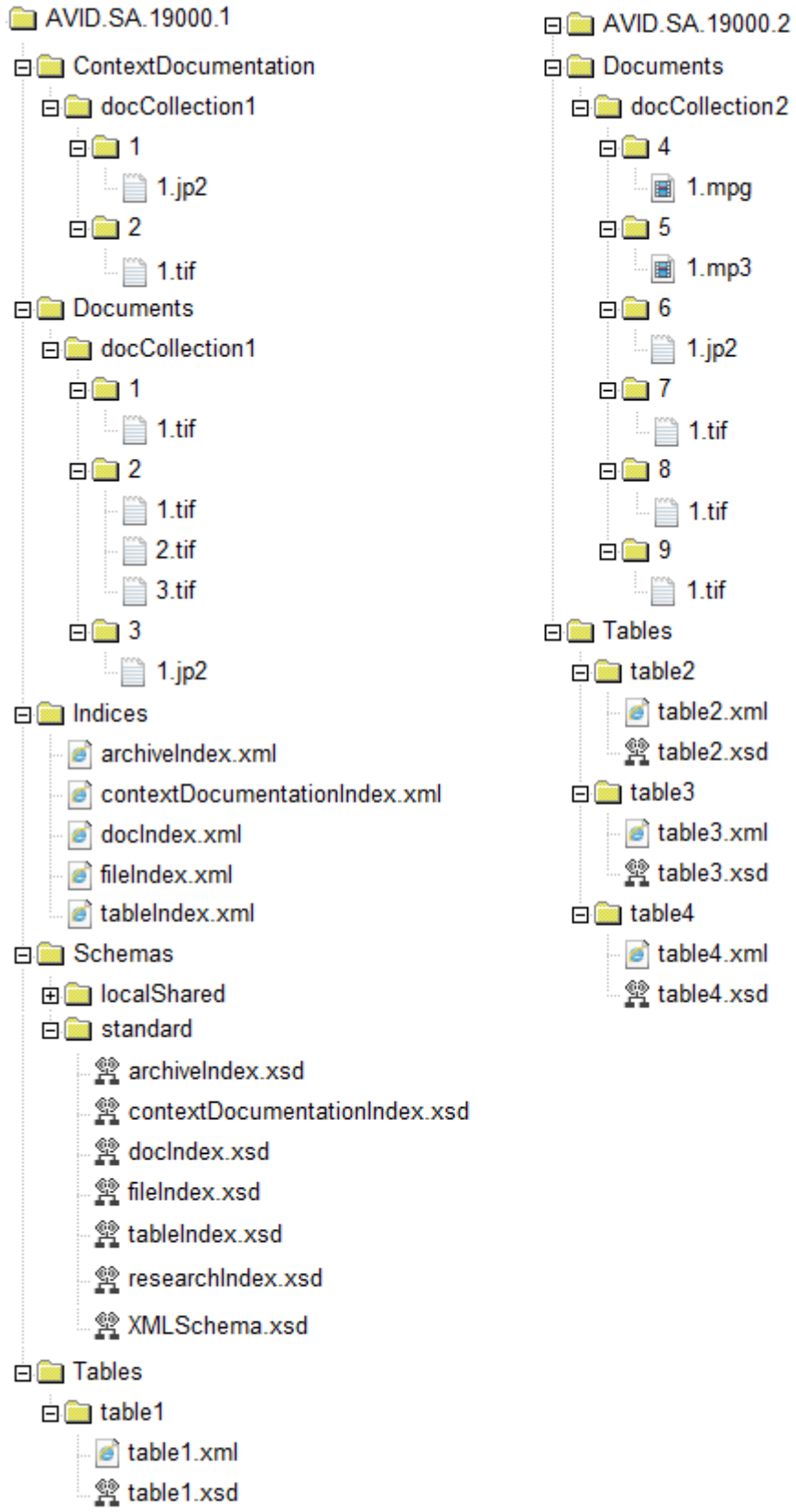
Elementer, som ikke er gråmarkerede, er obligatoriske i en arkiveringsversion.



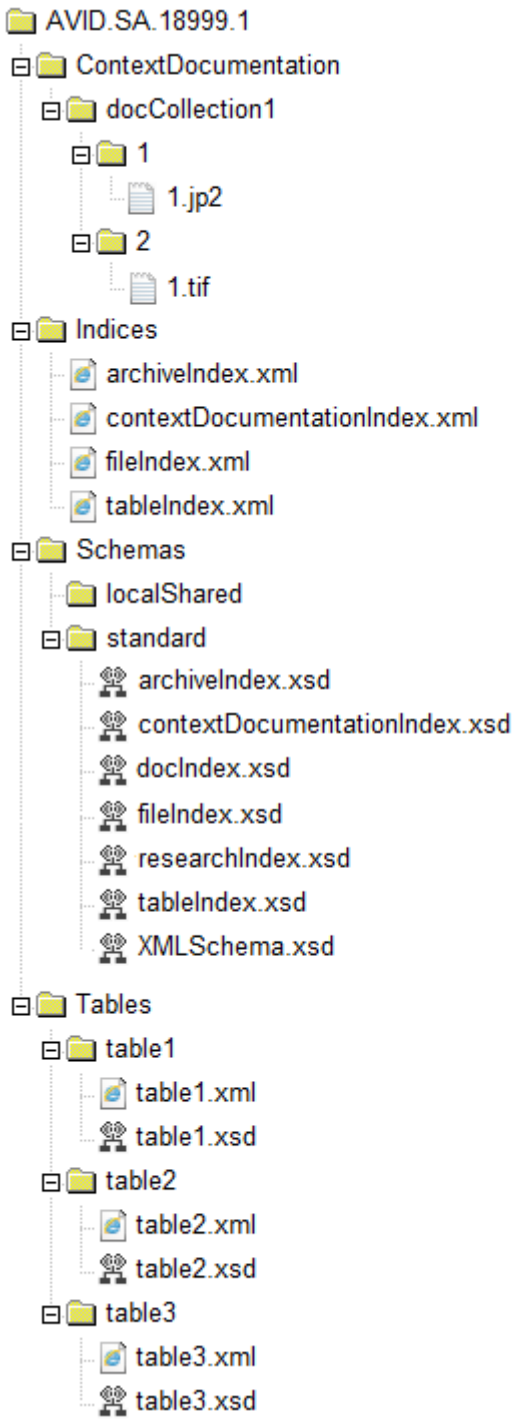
Figur 2.2 Placering af skemaer fra OGC til validering af GML version 3.1.1

Skemaer fra OGC til validering af GML version 3.1.1 placeres uændret i mappen *localShared* under mappen *Schemas*.

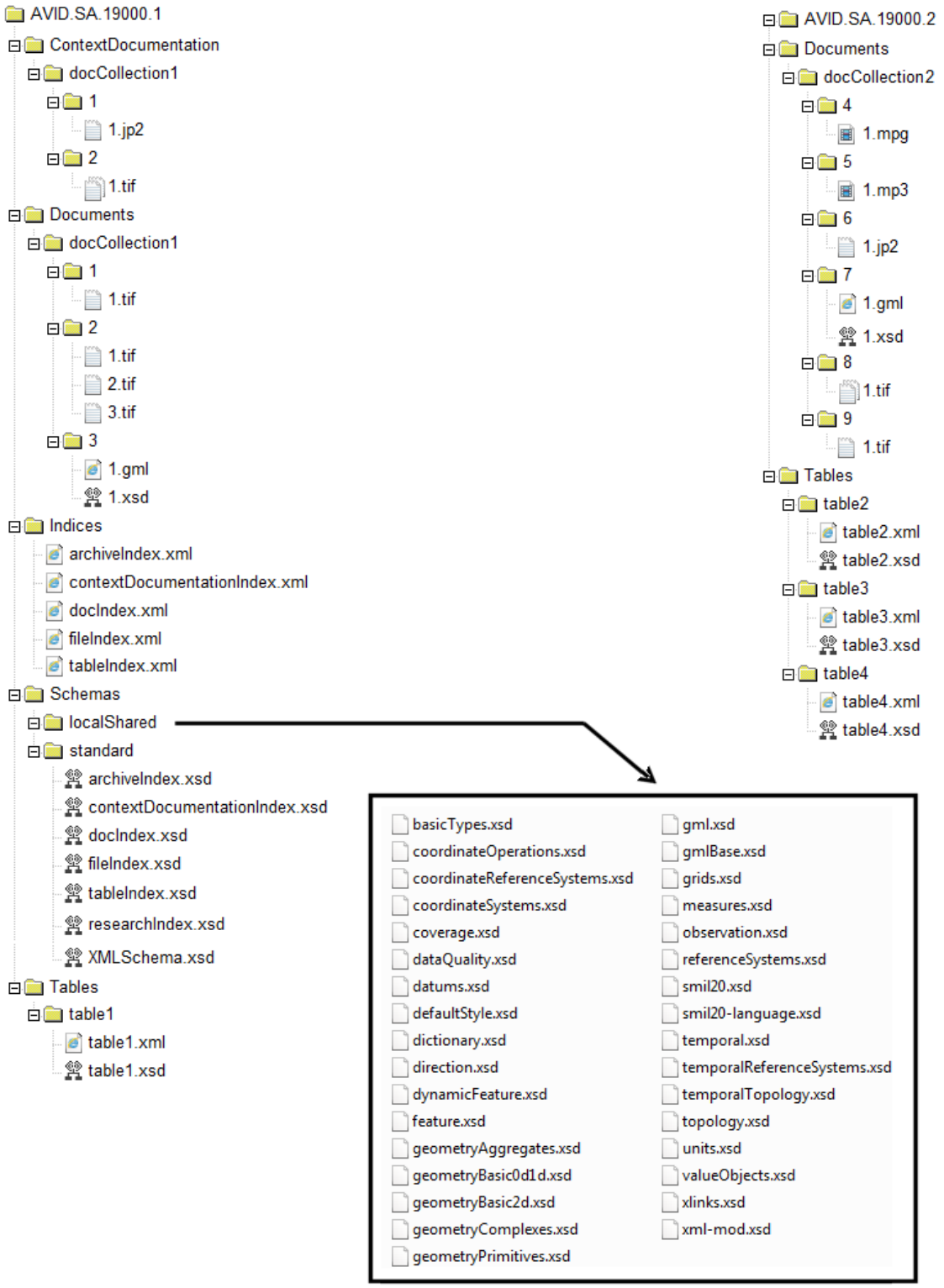
Figur a Arkiveringsversion af data fra et it-system med dokumenter



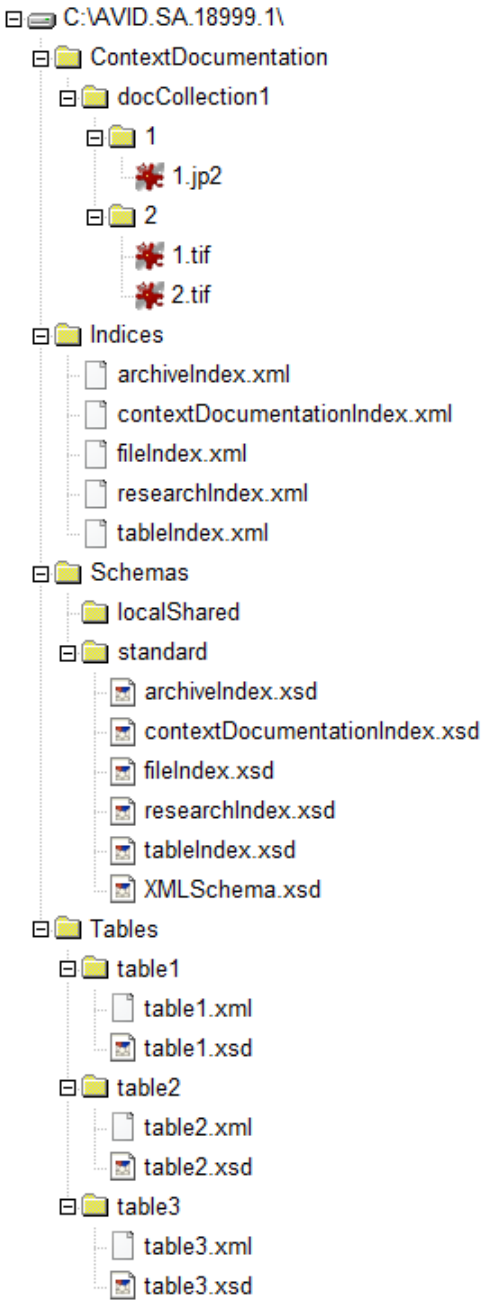
Figur b Arkiveringsversion af data fra et it-system uden dokumenter



Figur c Arkiveringsversion af data fra et IT-system med dokumenter og geodata (gml-filer)



Figur d Arkiveringsversion af forskningsdata afleveret efter bilag 9



Bilag 3 – Arkiveringsversion af digitale data og eventuelle dokumenter

3. A. Generelle regler om arkiveringsversioner

3.A.1 En arkiveringsversion består af:

- **datastruktur, jf. 3. B**
- **dataindhold, jf. 3. C**
- **information om arkiveringsversionen, jf. 3. D**

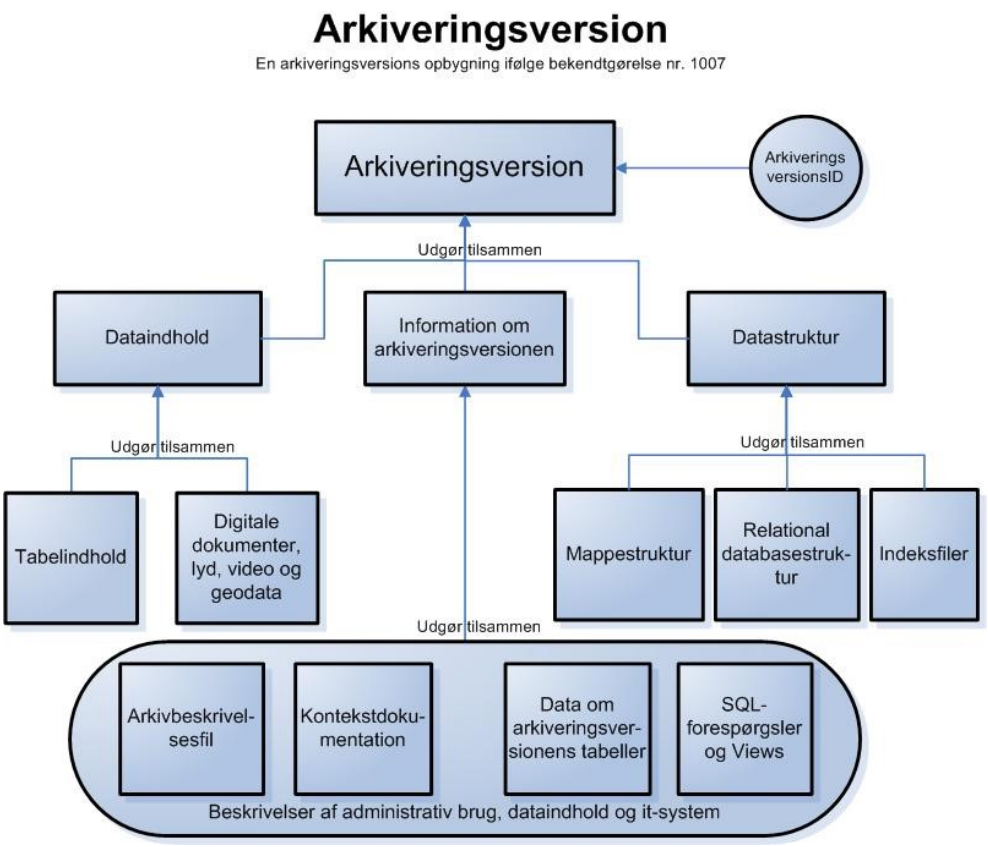
3.A.2 En arkiveringsversion må ikke indeholde krypteret information.

3.A.3.a En arkiveringsversion skal indeholde samtlige bevaringsværdige data og eventuelle dokumenter fra en afgrænset periode, hvor der ikke længere rettes i eller tilføjes data, eller udformes som et øjebliksbillede og indeholde samtlige bevaringsværdige data og eventuelle dokumenter på et bestemt tidspunkt, jf. dog punkt 3. A. 3.b.

3.A.3.b Under forudsætning af at der er gennemført kontrol med, at afsluttede sager er lukket korrekt, kan det modtagende arkiv godkende, at en arkiveringsversion kun indeholder dokumenter fra afsluttede sager. Metadata om samtlige sager og dokumenter, både afsluttede og uafsluttede, skal dog indgå i arkiveringsversionen.

3.A.4 Hvis arkiveringsversionen af et it-system med dokumenter udarbejdes uden skift af journalperiode, eller hvis der i forbindelse med skift af journalperiode sker overførsel af dokumenter til ny periode, skal de dokumenter, der indgår i arkiveringsversionen, markeres på en sådan måde i it-systemet, at de kan undtages fra efterfølgende arkiveringsversioner.

3.A.5 Mindre dokumentsamlinger, som er skabt i forbindelse med forskning med anvendelse af videnskabelig metode, kan afleveres som en del af institutionens ESDH-system efter aftale med det modtagende arkiv.



Kryptering

Krypteret information skal dekrypteres, inden den indgår i arkiveringsversionen. Det gælder f.eks. krypterede cpr-numre.

Ved forsendelse med post skal arkiveringsversionen dog krypteres i forbindelse med transporten fra myndigheden/leverandøren og til arkivet, såfremt den indeholder persondata eller andre typer følsom information. Ved aflevering til Rigsarkivet krypteres data med BitLocker eller VeraCrypt. Kode til dekryptering skal fremsendes separat.

Afgrænsning af data og dokumenter i arkiveringsversioner

Det modtagende arkiv fastsætter afgrænsningen af data og dokumenter i en arkiveringsversion efter drøftelse med den afleverende myndighed.

Arkiveringsversionen skal som udgangspunkt omfatte bevaringsværdige data og eventuelt dokumenter fra en afsluttet periode i et system, hvor der ikke længere rettes eller tilføjes data og eventuelt dokumenter.

En afsluttet periode kan f.eks. være:

- En afsluttet periode for et journalsystem, EDH- eller et ESDH-system
- Afslutning af en 5-årig periode for et fagsystem eller register
- Afslutning af en 1-årig periode fra f.eks. statistikregister
- Afslutning af et it-system, der ikke tidligere er afleveret, og som lukkes. F.eks. register med data fra en spørgeskemaundersøgelse

Nogle it-systemer har ikke afsluttede perioder, men tilføjes data og eventuelt dokumenter efter udtræk til arkiveringsversion. Arkiveringsversionen er således et 'øjebliksbillede' af it-systemet. I de tilfælde kan arkiveringsversionen omfatte data og dokumenter fra hele den periode, hvor systemet har været i brug. Den kan også være afgrænset til kun at indeholde data og dokumenter, der ikke er afleveret i en tidligere arkiveringsversion.

Det modtagende arkiv kan endvidere godkende, at arkiveringsversionen af et system med dokumenter alene indeholder de sager, som er lukket i systemet. Godkendelse kan gives under forudsætning af, at myndigheden har ført kontrol med at sager lukkes korrekt og i tilstrækkeligt omfang i systemet. Med korrekt menes at der er foretaget kvalitetskontrol af sagerne i forbindelse med lukning, så det er sikret, at de indeholder alle relevante dokumenter.

For statslige myndigheder finder godkendelsen af om denne model for arkivering må anvendes sted i forbindelse med Rigsarkivets to års tilsyn med systemet. For regionale og kommunale skal godkendelse finde sted senest i forbindelse med afleveringssagsbehandlingen.

For afleveringen af lukkede sager gælder i øvrigt at alle metadata for hele systemets levetid skal afleveres uanset om sagerne er lukkede eller ej, dog kan tidligere arkiverede metadata slettes, såfremt dette er nødvendigt i forhold til GDPR/persondataloven.

Afleverede dokumenter skal markeres i systemet (driftsversionen af systemet) på en sådan måde, at de kan undtages fra efterfølgende arkiveringsversioner. Det kan f.eks. gøres ved et felt i registreringsdelen, hvori det angives, om et registreret dokument er afleveret i en tidligere arkiveringsversion.

Denne information skal overføres til arkiveringsversionen med værdien 1 for dokumenter, som tidligere er afleveret og 2 for dokumenter, der ikke tidligere er afleveret jf. bilag 6, punkt C, figur 5. I arkivbeskrivelsesfilen skal det endvidere registreres, om der er tale om en afsluttet periode eller et øjebliksbillede.

Tre modeller for arkivering af systemer med dokumenter

Model A for arkivering af et system med dokumenter med periodeskift
(uden overlapningsperiode)

1. arkiveringsversion

2005 **2010** **2015** **2020**

Registreringsdata, 2005-2010

Dokumenter fra perioden 2005-2010

2. arkiveringsversion

Registreringsdata fra perioden 2010-2015

Dokumenter fra perioden 2010-2015

3. arkiveringsversion

Registreringsdata fra perioden 2015-2020

Dokumenter fra perioden 2015-2020

It-systemer med afsluttede perioder kan have overført udvalgte sager og dokumenter fra den forrige periode til den nye periode, f.eks. personalesager. Overførte dokumenter markeres i driftsversionen, så de kan undtages fra efterfølgende arkiveringsversioner jf. ovenstående retningslinjer. Hvis der er tale om overførelse af lukkede sager, skal dokumenterne markeres som afleveret i tidligere arkiveringsversion på samme måde.

Model B for arkivering af et system med dokumenter uden periodeskift

1. arkiveringsversion

2005 **2010** **2015** **2020**

Registreringsdata, 2005-2010

Dokumenter fra perioden 2005-2010

2. arkiveringsversion

Registreringsdata fra perioden 2005-2015

Dokumenter fra perioden 2010-2015

3. arkiveringsversion

Registreringsdata fra perioden 2005-2020

Dokumenter fra perioden 2015-2020

Model C for aflevering af lukkede sager

1. arkiveringsversion

2010 **2015** **2020** **2020** **2025**

Registreringsdata, 2010-2015

Sager (inkl. dokumenter) **lukket** i perioden 2010-2015

2. arkiveringsversion

Registreringsdata fra perioden 2010-2020

Sager (inkl. dokumenter) **lukket** i perioden 2015-2020

3. arkiveringsversion

Registreringsdata fra perioden 2005-2025

Sager (inkl. dokumenter) **lukke**t i perioden 2020-2025

Efter aflevering kan data og dokumenter, der indgår i arkiveringsversionen, slettes.
Læs mere om fordele og ulemper ved de forskellige modeller i vejledningen til bekendtgørelse 9290 om anmeldelse og godkendelse af statslige it-systemer. Vejledningen findes på www.sa.dk

Aflevering af forskningsdata i form af mindre dokumentsamlinger
Forskningsdata i form af mindre dokumentsamlinger kan efter aftale med modtagende arkiv indgå som en del af afleveringen af institutionens ESDH-system (system med dokumenter). Formålet med denne afleveringsmetode er at spare udgifterne til produktion af en arkiveringsversion for små mængder af dokumenter, lyd og video.

Det er en forudsætning for denne afleveringsform, at samlingen placeres i en sag, som kan identificeres via sagsnummer. Endvidere skal der i sagen indgå kontekstdokumentation af samlingen som svarer til bestemmelserne i bilag 9 i nærværende bekendtgørelse.

3.B. Datastruktur

3.B.1 Datastrukturen i arkiveringsversionen består af:

- **en mappestruktur, jf. 4. B, illustreret i figur 2.1.**
- **en relationel databasestruktur på 1. normalform eller højere, angivet i indeksfilen tableIndex.xml, jf. 4.C. 5.a.**
- **øvrige indeksfiler i XML, jf. 4. C, der strukturerer data om arkiveringsversionens overordnede indhold, samtlige filer i arkiveringsversionen, dens kontekstdokumentation og dens eventuelle digitale dokumenter.**

Arkiveringen af offentlige myndigheders data i it-systemer og forskningsdatasæt er baseret på den relationelle databasemodel. Ved aflevering til offentligt arkiv defineres en relationel databasestruktur på 1. normalform således:

1. Enhver tabel skal for alle rækker have samme antal kolonner (dvs. multivariable attributter er ikke tilladt).
2. Alle tabeller skal have et navn, og navnet skal være entydigt.
3. Alle kolonner skal have et navn, og navnet skal være entydigt indenfor den pågældende tabel.
4. Rækkefølgen af rækker må ikke være betydningsbærende i nogen tabel.
5. Rækkefølgen af kolonner må ikke være betydningsbærende i nogen tabel.
6. Enhver tabel skal have defineret et eller flere felter, som udgør en entydig primærnøgle (ingen rækker må være ens).
7. En fremmednøgle skal være relateret til en primærnøgle.
8. En fremmednøgle må ikke være relateret til dele af en sammensat primærnøgle. Der skal således være sammenfald mellem antal felter i primærnøgle og fremmednøgle.
9. Alle forbindelser mellem tabeller er udtrykkelige (direkte). Således kan der ikke forekomme forbindelser mellem tabeller på baggrund af fortolkning af deres feltindhold. Hierarkiske databaser, netværksdatabaser og objektrelationelle databaser er således per definition ikke relationelle databaser.

   Det følger heraf, at en fremmednøgle kun må relateres til en specifik tabel (ved dennes primærnøgle). Fremmednøglen må altså ikke angives at være relateret til flere tabeller med henblik på, at konkrete poster vil være relateret til én ud af flere forskellige tabeller alt efter ”henvisningens art”.
10. Der bør ikke være tabeller uden relation til en eller flere andre tabeller.

For it-systemer og forskningsdata, som ikke er relationelle, skal myndigheden udarbejde en relationel model for arkiveringen, som lever op til ovenstående definition. Modellen skal godkendes af det modtagende arkiv forud for produktion af arkiveringsversionen.

3.C. Dataindhold

3.C.1 Dataindholdet består af:

- **tabelindhold i standardiserede datatyper, jf. 5. B.**
- **eventuelle digitale dokumenter, lyd, video og geodata i standardiserede dataformater, jf. 5. E. - 5. G.**

Ved digitale dokumenter forstås dokumenter anvendt i myndighedens forvaltning af sit ressortområde eller som er produceret som forskningsdata. Det kan f.eks. være:

Administrative dokumenter i it-systemer

- Ind- og udgående breve
- Mødereferater, notater, indstillinger o. lign.
- Fotos
- Forskningsrapporter
- Regneark
- Præsentationsfiler f.eks. PowerPoints mv.

Lyd, video og geodata kan dels indgå i ESDH/EDH-systemer og dels være lagret i selvstændige systemer.

Dokumenter skabt vha. videnskabelige metode, forskningsdata, kan f.eks. være:

- Transskriberede interviews
- Fotos af forskningsobjekter
- Video af arkæologiske fundsteder
- Lydoptagelser
- GML-opmærkning af geologiske fænomener

3.D. Information om arkiveringsversionen

3.D.1 Information om arkiveringsversionen består af beskrivelser af administrativ brug, dataindhold og it-system, jf. bilag 6.

For at kommende brugere kan forstå og genbruge data i arkiveringsversionerne skal den administrative anvendelse af it-systemerne, den tekniske udformning af dem samt selve dataindholdet dokumenteres. På s. 65 beskrives hvilke oplysninger, der skal/kan indgå i informationen om arkiveringsversionen.

For forskningsdata skal der sammen med data leveres dokumentation, som muliggør forskningsmæssig genanvendelse af data.

Bilag 4 – Datastruktur

4.A. Generelle regler om datastruktur

4.A.1 I en arkiveringsversion skal alle nøgler have en entydig identifikator. Der må ikke forekomme situationer, hvor det er nødvendigt at uddrage dele af nøglefelter for at forstå it-systemets indhold eller funktion. Et felt, der indgår i en primærnøgle må ikke udelukkende bestå af blanktegn eller være NULL.

4.A.2 Når en værdi i et felt er en kode som repræsentation for en fast og entydig værdi, skal koderne forklares. Såfremt værdien ikke findes i en kode- eller opslagstabel i it-systemet, skal der i arkiveringsversionen genereres en eller flere tabeller med angivelse af værdierne i kodede felter. Alternativt skal der enten tilføjes et felt i data med den værdi, som koden dækker, eller kodeværdien i tabellen skal udskiftes med det egentlige indhold.

4.A.3 Såfremt dokumenterne i et it-system med dokumenter er lagret i en betydningsbærende struktur, skal denne struktur i arkiveringsversionen omdannes til en eller flere tabeller.

Nøglefelter
Det afhænger af it-systemets indretning og funktion, hvorvidt det er påkrævet at udskille/gentage dele af indholdet af et felt i it-systemet til flere felter i arkiveringsversionen. I nogle systemer forekommer der felter, der anvendes som en samlet nøgle, men som i realiteten indeholder flere forskellige informationer (se eksemplerne nedenfor). Hvis disse separate informationer har betydning for den fremtidige anvendelse, skal de anbringes i separate felter.

I nogle tilfælde kan det oprindelige felt uden problemer splittes op i flere felter i arkiveringsversionen. Hvis det oprindelige felt fungerer som nøglefelt eller kodet felt, vil det imidlertid være nødvendigt at beholde det i arkiveringsversionen, men samtidig tilføje felter indeholdende de separate dele af det oprindelige felt.

Eksempel, hvor CPR-nummer er brugt som nøgle

CPR-nummer kan nævnes som et eksempel på indhold, hvor man må vurdere it-systemets indretning og funktion for at afgøre, om det skal udskilles/gentages dele af i flere felter. I mange situationer er et CPR-nummer entydigt. Således kan man ofte i et it-system udmærket anvende CPR-nummer som nøgle. Imidlertid består CPR-nummer af to elementer, en fødselsdato og en kontrolsum, hvoraf køn kan udledes. Mens it-systemet er i brug hos myndigheden, er det tilladt, at it-systemet uddrager fødselsdato og køn af feltet indeholdende CPR-nummer, men i arkiveringsversionen er det **ikke** tilladt.

I arkiveringsversionen skal fødselsdato og køn være udskilt/gentaget i egne felter, når de således har været anvendt som egne oplysninger i it-systemet. Bemærk, at dette krav ikke forhindrer, at CPR-nummer kan anvendes som nøgle i arkiveringsversionen, heller ikke i sin oprindelige form. Grundet andre datoformater og kodede værdier (som f.eks. lige numre for kvinder, ulige for mænd i kontrolsum) er det ikke muligt rent at *udskille* fødselsdato og køn i egne felter fra feltet indeholdende CPR-nummer, men de skal i stedet gentages i anden form i egne felter.

I it-systemet:
CPR-nummer 140575-1306
(Fødselsdato og køn udledes heraf og anvendes i it-systemet)

I arkiveringsversionen:
Data gentages i egne felter:

CPR-nummer	Fødselsdato	Køn
140575-1306	19750514	K

K er en kodet værdi og skal således angives eller udlæses i sin egen tabel, eller værdien skal indsættes i tabellen og f.eks. skrives ”kvinde” i stedet for K.

Koder
Det er vigtigt, at **alle** kodede felter forklares, da denne viden er central for kommende brugeres genanvendelse af data. Koder, som i dag kan forekomme indlysende, kan reelt gøre data ubrugelige om 100 år, hvor brugerne ikke har de samme forudsætninger for at bruge data, som vi har i dag. Eksempelvis vil informationsindholdet af en kommunekode anvendt før strukturreformen i 2007 hurtigt gå tabt, hvis den ikke dokumenteres.

Koder kan arkiveres på tre måder:

- Angivelse i en separat kodetabel, som relaterer til de(n) tabel(ler), hvor den pågældende kode er anvendt, eller
- Ved i data at tilføje et felt, med den værdi, som koden dækker
- Ved at udskifte kodeværdien i data med den faktiske værdi

Kodetabeller, som oprettes til brug i arkiveringsversionen, skal registreres i *tableIndex.*

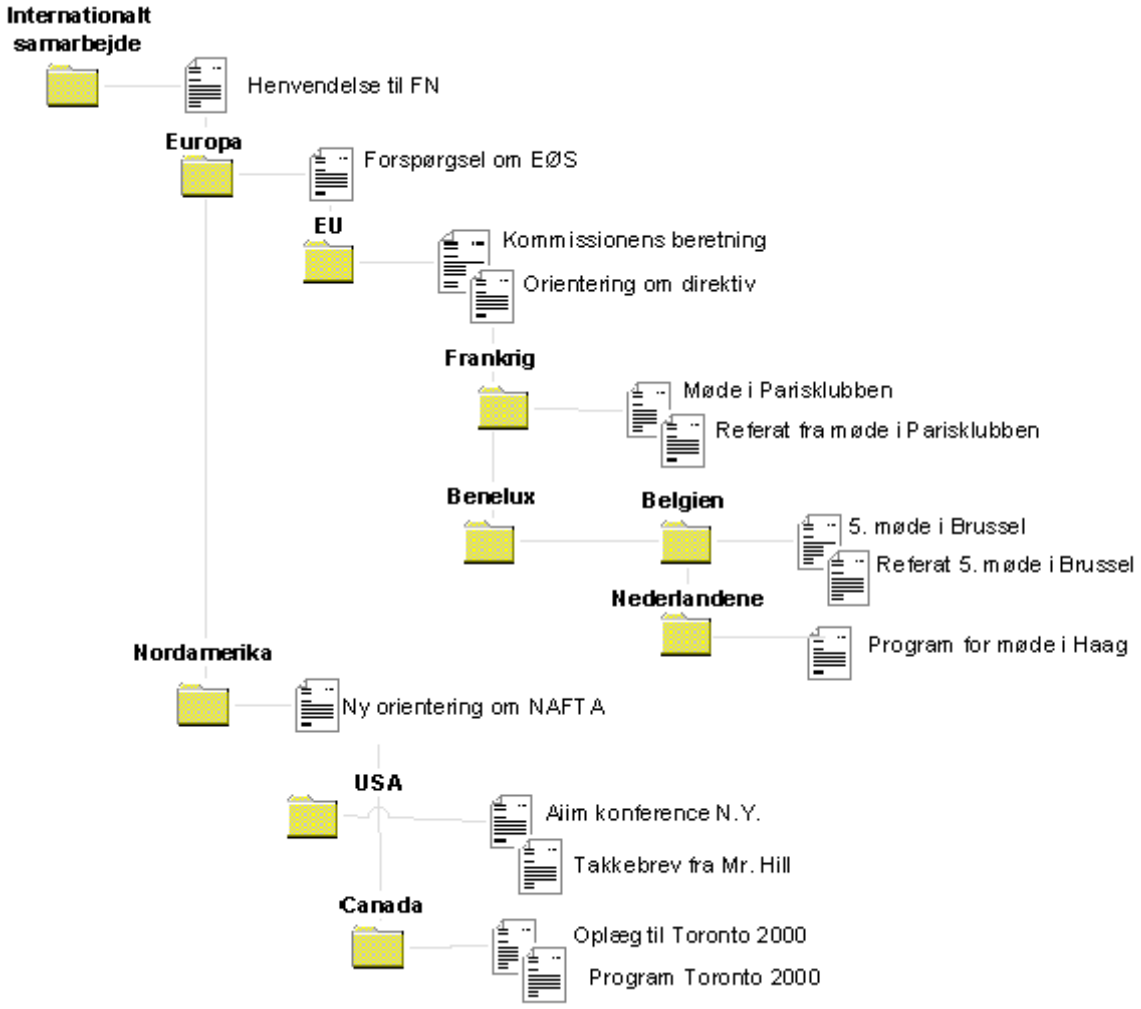
Eksempel på hvordan koden udskiftes med den værdi den dækker

I tabel X findes feltet Y, der indeholder en produktkode (M=mad, T=tøj, E=elektronik)

I stedet for at have en kodetabel, hvor der i felt Y optræder koderne M, M, T, T, E osv. kan man i data indsætte værdierne dvs. mad, mad, tøj, tøj, elektronik osv. Dette kan gøres ved enten direkte at udskifte koder med værdierne *eller* ved at indsætte et ekstra felt med værdierne efter det kodede felt.

Betydningsbærende mappestruktur
Hvis systemets dokumenter har været lagret i en betydningsbærende mappestruktur, skal denne struktur omsættes til en eller flere tabeller, som kan arkiveres i en relationel struktur. Det aftales nærmere med det modtagende arkiv, hvordan et konkret udtræk laves, så det afspejler mappestrukturen og systemets funktion.

Eksempel på betydningsbærende mappestruktur før den omsættes til en eller flere tabeller i en arkiveringsversion



Ovenstående eksempel på enkel mappestruktur omsat til én tabel
I dette eksempel er der tale om en mappestruktur, der er omsat til én tabel med en primærnøgle, en sti og et dokumentnavn, hvor hele mappestrukturen omsættes til en "String" f.eks.
Internationalt samarbejde\Europa\EU\Benelux\Nederlandene. Hvis der er mange mappeniveauer, frarådes det at benytte denne struktur. Der bør kun omformes til én tabel, hvis en mappestruktur har haft højst 5 niveauer, jf. eksempel nedenfor.

Dokument

ID	Sti	Dokument navn
1	Internationalt samarbejde	Henvendelse til FN
2	Internationalt samarbejde\Europa	Forespørgsel om EØS
3	Internationalt samarbejde\Europa\EU	Kommisionens beretning
4	Internationalt samarbejde\Europa\EU	Orientering om direktiv
5	Internationalt samarbejde\Europa\EU\Frankrig	Møde om Parisklubben
6	Internationalt samarbejde\Europa\EU\Frankrig	Referat fra møde i parisklubben
7	Internationalt samarbejde\Europa\EU\Benelux\Belgien	5. Møde i Brussel
8	Internationalt samarbejde\Europa\EU\Benelux\Belgien	Referat 5. Møde i Brussel
9	Internationalt samarbejde\Europa\EU\Benelux\Nederlandene	Program for møde i Haag
10	Internationalt samarbejde\Nordamerika	Ny orientering om NAFTA
11	Internationalt samarbejde\Nordamerika\USA	Aiim konference i NY
12	Internationalt samarbejde\Nordamerika\USA	Takkebrev fra Mr. Hill
13	Internationalt samarbejde\Nordamerika\Canada	Oplæg til Toronto 2000
14	Internationalt samarbejde\Nordamerika\Canada	Program Toronto 2000

Eksempel på kompleks mappestruktur omsat til flere tabeller

Hvis mappestrukturen er mere kompleks med et større og evt. varierende antal niveauer af undermapper, skal der laves en mere generel beskrivelse af mappestrukturen som i nedenstående eksempel. Mapperne er listet i en tabel med en primærnøgle, en fremmednøgle, der peger på denne primærnøgle (selvrefererende tabel) og en mappetitel. Hvis det har været af betydning i systemet, skal mappernes oprettelsesdato og evt. andre informationer medtages som separate felter. Beskrivelsen af dokumenterne er i en tabel for sig med en primærnøgle, en pegepind til den mappe, dokumentet lå i, og et dokumentnavn.

Mappe

ID	Over	Mappe Titel
1		Internationalt samarbejde
2	1	Europa
3	1	Nordamerika
4	2	EU
5	3	USA
6	3	Canada
7	4	Frankrig
8	4	Benelux
9	8	Belgien
10	8	Nederlandene

Dokument

ID	Dokument navn	I
1	Henvendelse til FN	1
2	Forespørgsel om ØS	2
3	Kommisionens beretning	4
4	Orientering om direktiv	4
5	Møde om Parisklubben	7
6	Referat fra møde i Parisklubben	7
7	5. Møde i Bryssel	9
8	Referat 5. Møde i Bryssel	9
9	Program for møde i Haag	10
10	Ny orientering om NAFTA	3
11	Aim konference i N.y.	5
12	Takkebrev fra Mr. Hill	5
13	Oplæg til Toronto 2000	6
14	Program Toronto 2000	6

4.B. Placering af mapper og filer

4.B.1 I roden af filsystemet på afleveringsmediet, jf. bilag 7, skal der være placeret en mappe navngivet med mediets navn. Mediets navn består af det unikke arkiveringsversionsID med tilføjelse af et suffiks ".n", der angiver rækkefølgen af medierne, hvor n er et fortløbende medieløbenummer begyndende med 1.

4.B.2 Arkiveringsversionens indhold fordeles i mapper, som angivet i figur 4.1.

Figur 4.1 Arkiveringsversionens mapper

Fordeling af arkiveringsversionens indhold på mapper

Navn på mappe	Beskrivelse	Obligatorisk	Placering
Indices	Indeksfiler, jf. 4.C.	Ja	Altid på første medie
Tables	Tabellindhold, jf. 4.D.	Ja	Kan placeres frit og fordeles over flere medier
ContextDocumentation	Kontekstdokumentation, jf. 4.E.	Ja	Altid på første medie.
Schemas	XML-skemaer, jf. 4.F. Skemamappen skal have to undermapper: *standard* og *localShared*.	Ja	Altid på første medie
Documents	Eventuelle dokumenter, lyd, video og geodata, jf. 4.G.	Nej	Kan placeres frit og fordeles over flere medier

4.B.3 Mapperne skal navngives som angivet i figur 4.1.

4.B.4.a Et arkiveringsversionsID består af præfikset AVID, en kode på 2-4 bogstaver, (som angiver det modtagende arkiv), samt et arkiveringsversionsløbenummer. Elementerne adskilles af et punktum.

4.B.4.b ArkiveringsversionsID udleveres af Rigsarkivet.

4.B.5.a En arkiveringsversion, som ikke kan være på ét medie, jf. bilag 7, kan fordeles over flere medier.

4.B.5.b Mapperne ContextDocumentation, Indices og Schemas skal altid være placeret på afleveringens første medie.

4.B.5.c Mappen i roden på efterfølgende medier skal kun indeholde de mapper, hvis indhold kræver fordeling over flere medier. Disse mapper navngives altid som angivet i figur 4.1 uden brug af suffiks.

4.B.6 Der må kun være én af hver af de i figur 4.1 angivne mapper på hvert medie.

ArkiveringsversionsID (AVID) er unikt og skal sikre en entydig identifikation af en arkiveringsversion. Nummeret udleveres af det modtagende arkiv.

Eksempel på arkiveringsversionsID

AVID.SA.19000, hvor

- AVID er et fast præfiks
- SA er en kodeværdi, der angiver, at Rigsarkivet er modtagende arkiv. Se nedenstående liste over bogstavkoder for andre offentlige arkiver
- 19000 er et løbenummer

Eksempel på medienavn

AVID.SA.19000.1, hvor

- AVID.SA.19000 er ArkiveringsversionsID'et og
- 1 er det fortløbende medieID, som angiver rækkefølge på de anvendte medier i arkiveringsversionen.

  Bemærk! Der vil ofte kun være anvendt et medie til hele arkiveringsversionen, men suffixet skal alligevel angives

ArkiveringsversionsID'er uddeles fra en landsdækkende liste over ID'er, som Rigsarkivet administrerer. Det modtagende arkiv skal således sørge for at rekvirere et arkiveringsversionsnummer hos Rigsarkivet i forbindelse med udarbejdelse af en afleveringsbestemmelse for et system. Medienavnet genereres løbende ved produktion af arkiveringsversionen.

Bogstavkoder for § 7-arkiver (listen opdateres jævnligt med nye arkiver):

Arkivnavn	**Bogstavkode**
Ballerup Stadsarkiv	BALK
Billund Stadsarkiv	BKA
Bornholms Ø-arkiv	BRK
Byhistorisk Samling for Lyngby-Taarbæk Kommune	LTK
Dragør Lokalarkiv	DRA
Egedal Kommunearkiv	EGE
Esbjerg Byhistoriske Arkiv	EBA
Faxe Kommunes Arkiver	FAXE
FredensborgArkiverne	FRED
Frederiksberg Stadsarkiv	FRB
Frederikshavn Stadsarkiv	FS
Furesø Museer	FURE
Gentofte eArkiv	GENT
Gladsaxe Byarkiv	GLXB
Grenaa Egnsarkiv	RA

Greve Stadsarkiv	GREV
Gribskov Stadsarkiv	GRIB
Guldborgsund Stadsarkiv	GBS
Hedensted Kommunearkiv	HK
Helsingør Stadsarkiv	HEL
Historisk Arkiv for Haderslev Kommune	HAH
Historisk Arkiv for Rudersdal Kommune	RUD
Hjørring Kommunearkiv	HJK
Holbæk Stadsarkiv	HOSA
Hørsholm Stadsarkiv	HØKO
Ishøj Stadsarkiv	IS
Kolding Stadarkiv	KKSA
KOMDA KMD Klyngeafleveringer	KMDA
Københavns Stadsarkiv	KSA
Lokalhistorisk Arkiv for Thisted Kommune	LATK
Lolland Kommunes Stadsarkiv	LOLL
Læsø Arkiv	LAES
Mariagerfjord Kommunearkiv	MFK
Middelfart Stadsarkiv	MS
NEA KMD Klyngeafleveringer	NKMD
NæstvedArkiverne	NÆK
Odense Stadsarkiv	ODSA
Odsherred Stadsarkiv	ODS
Randers Stadsarkiv	RDS
Ringsted Kommunearkiv	RIKO
Roskilde Stadsarkiv	ROS
Silkeborg Arkiv	SIA
Skive Byhistoriske Arkiv	RA
Slagelse Arkiverne	SLAG
Sorø Stadsarkiv	SOR
Stadsarkivet for Frederikssund Kommune	SFK
Struer Kommunearkiv	RA
Syddjurs Kommunearkiv	SYKO
Sønderborg Arkiv	SOND
Tønder arkiv	TOND
Tårnby Kommunes Stads- og Lokalarkiv	TÅRN
Vejen Kommunearkiv	VEJE
Vejle Stadsarkiv	VS
Vendsyssel Historiske Museum og Arkiv	VHMA
Aabenraa Stadsarkiv	AABK
Aalborg Stadsarkiv	AAS
Aarhus Stadsarkiv	AARS

Opdeling af arkiveringsversionen på flere medier

Arkiveringsversionen kan som oftest lagres på et medie, som størrelsesmæssigt passer til datamængden, f.eks. en CD-R til afleveringen af et lille fagsystem og en stor USB-harddisk til aflevering af et omfattende ESDH-system. Der kan dog opstå situationer, hvor det er nødvendigt at dele arkiveringsversionen over flere medier.

I de tilfælde skal de overordnede mapper have samme betegnelse på alle medier. En dokumentmappe, som er delt over flere medier, vil således hedde ”Documents” på såvel det første medie som de efterfølgende,

mens dens undermapper skal have en fortløbende nummerering inden for arkiveringsversionen, se eksemplet i bilag 2, s.10ff.

4.C. Mappen *Indices*

4.C.1.a Mappen *Indices* skal indeholde følgende indeksfiler med oplysninger om arkiveringsversionen og dens indhold:

- **fileIndex.xml**
- **archiveIndex.xml**
- **contextDocumentationIndex.xml**

4.C.1.b Hvis arkiveringsversionen indeholder digitale dokumenter, lyd, video eller geodata, skal mappen *Indices* endvidere indeholde følgende indeksfil:

- **docIndex.xml**

4.C.1.c Hvis arkiveringsversionen indeholder data, som er skabt i forbindelse med forskning med anvendelse af videnskabelig metode og er afleveret efter reglerne i bilag 9, skal mappen *Indices* endvidere indeholde følgende indeksfil:

- **researchIndex.xml**

4.C.1.d Alle indeksfiler skal overholde deres tilhørende skema, jf. bilag 8.

4.C.2.a fileIndex.xml skal indeholde en komplet liste over samtlige filer, der findes i arkiveringsversionen. fileIndex.xml er dog undtaget fra denne regel.

4.C.2.b For hver enkelt fil i arkiveringsversionen angives de oplysninger, som fremgår af figur 4.2.

Oversigt over Indeksfiler og deres indhold

Navn på indeksfil, som skal ligge i mappen Indices	Indhold
fileIndex	Komplet liste over samtlige filer, der findes i arkiveringsversionen (undtagen filen selv)
archiveIndex	Oplysninger om datas oprindelse, brug, videregivelse m.v. Indekset kaldes også for arkivbeskrivelsesfilen.
contextDocumentationIndex	Liste over alle dokumenter, som indgår i kontekstdokumentationen. Dokumenterne indeholder længere beskrivelser af datas oprindelse, brug, videregivelse og underbygger oplysningerne i arkivbeskrivelsesfilen
tableIndex	Liste over alle tabeller i arkiveringsversionen og relationerne mellem disse m.v. Indekset kaldes også tabelindeks.

docIndex	Liste over alle dokumenter i arkiveringsversionen med angivelse af deres placering i arkiveringsversionen
researchIndex	Oplysninger om hovedtabeller og koder for manglende værdier i forskningsdata afleveret efter bilag 9. Indekset genereres automatisk (f.eks. med programmet ASTA) i forbindelse med arkivets konvertering af en afleveringspakke iht. bilag 9 til en arkiveringsversion iht. bilag 1-8.

Indeksfilerne er filer, som skal udarbejdes i forbindelse med produktionen af arkiveringsversionen, og som skal anvendes til indeksering og beskrivelse af indholdet i arkiveringsversionen. Filerne skal altid placeres i mappen Indices på arkiveringsversionens første medie og skal overholde de skemaer, som indgår som bilag 8 i bekendtgørelsen.

Figur 4.2 Oplysninger i fileIndex.xml

Filindeks er en oversigt over alle filer i arkiveringsversionen på nær filen selv.

Element navn	Betegn else	Beskrivelse	Forekomst pr fil	Obliga-torisk	Eksempel
foN	**Sti**	**Angivelse af sti til den mappe i arkiveringsversionen, hvor filen findes**	**1**	**Ja**	AVID.SA.18000.1\ContextDocumentation\docCollection1\1
fiN	**Navn**	**Filens navn**	**1**	**Ja**	Notat.tif
md5	**Kontrolsum**	**Filens kontrolsum af typen MD5 iht. *IETF RFC1321 - The MD5 Message-Digest Algorithm*:** **128 bit (16 bytes) repræsenteret som 32 hexadecimale cifre, alle angivet med *enten* minuskler *eller* versaler (små eller store bogstaver). Krav om 32 hexadecimale cifre medfører således krav om, at der afhængigt at værdien anvendes foranstillede nuller.**	**1**	**Ja**	0624888DEE5389FCED97FD71562191B8

Formålet med angivelse af kontrolsum for alle filer er at sikre deres integritet og dermed kontrollere, at overførslen af f.eks. dokumenter er udført korrekt. Ved modtagelse af arkiveringsversioner, som afleveres til Rigsarkivet, kontrolleres det, om filernes kontrolsummer stemmer overens med den angivne kontrolsum i fileIndex. Ved uoverensstemmelse vil arkiveringsversionen blive returneret til den afleverende myndighed. Det

er derfor vigtigt, at md5 generes/opdateres, når arkiveringsversionen er færdigproduceret og ellers klar til afsendelse.

fileIndex valideres af skemaet fileIndex.xsd, som er udarbejdet af Rigsarkivet og kan downloades på hjemmesiden, www.sa.dk. fileIndex.xsd placeres i mappen ”standard” under mappen ”Schemas”, jf. illustration i bilag 2.

4.C.3 archiveIndex.xml skal indeholde de oplysninger, som fremgår af 6.A.

archiveIndex, også kaldt arkivbeskrivelsesfilen, skal indeholde en lang række oplysninger om datas tilvejebringelse, brug, videregivelse mv. Det er nærmere specificeret i bilag 6, punkt A, hvilke oplysninger, der skal indgå i filen.

archiveIndex valideres af skemaet archiveIndex.xsd, som er udarbejdet af Rigsarkivet og kan downloades på hjemmesiden, www.sa.dk. archiveIndex.xsd placeres i mappen ”standard” under mappen ”Schemas”, jf. illustrationen i bilag 2.

4.C.4.a contextDocumentationIndex.xml skal indeholde et indeks over de dokumenter, som findes i arkiveringsversionens kontekstdokumentation.

4.C.4.b For hvert dokument i kontekstdokumentationen angives de oplysninger, som fremgår af figur 4.3.

Figur 4.3 Oplysninger i contextDocumentationIndex.xml

Oversigt over de oplysninger, der skal angives for hvert dokument i kontekstdokumentationen.

Elementnavn	Betegnelse	Udfaldsrum	Forekomst pr. dokument	Obliga-torisk	Eksempel
documentID	**DokumentID**	**ID på op til 12 cifre**	**1**	**Ja**	170996
documentTitle	**Dokument titel**	**Fritekst**	**1**	**Ja**	F.eks. Journalvejledning for arkivstyrelsen *Eller* Projektansøgning for forskningsprojekt om slædehunde
documentDescription	**Dokument beskrivelse**	**Fritekst**	**0-1**	**Nej**	Beskrivelse af hvordan arkivstyrelsens ESDH-system har været brugt *Eller* Ansøgning om fondsmidler til projekt om slædehundes leve vilkår
documentDate	**Dato**	**År, år-måned, år-måned-dag, iht. XML-standarden**	**0-1**	**Nej**	2009-01-09
authorName	**Forfatternavn**	**Fritekst**	**0-m**	**Nej**	ESDH-koordinator Anne Østergård *Eller* Professor Niels Nielsen
authorInstitution	**Forfatter institution**	**Fritekst**	**0-m**	**Nej**	Arkivstyrelsen *Eller* xx-universitet, Grønlandsk institut
documentCategory	**Dokument kategori**	**Kategori iht. bilag 6, figur 6.2**	**1-m**	**Ja**	1 (et dokument kan registreres under flere kategorier)

contextDocumentationIndex er en indholdsfortegnelse for alle dokumenter, som indgår som kontekstdokumentation i arkiveringsversionen. Det er vigtigt for den fremtidige genanvendelse af data i arkiveringsversionen, at ophavssituation for data og dokumenter fremgår af kontekstdokumentationen.

contextDocumentationIndex valideres af skemaet documentationIndex.xsd, som er udarbejdet af Rigsarkivet og kan downloades på hjemmesiden www.sa.dk. documentationIndex.xsd placeres i mappen ”standard” under mappen ”Schemas”, jf. illustration i bilag 2.

4.C.5.a tableIndex.xml skal indeholde en angivelse af en relationel databasestruktur på 1. normalform eller højere. Samtlige tabeller i arkiveringsversionen skal angives.

4.C.5.b »tableIndex.xml« skal overholde det generelle XML-skema »tableIndex.xsd«, jf. 4.F.

4.C.5.c Hvis et felt må have værdien NULL, skal der i »tableIndex.xml« være angivet værdien »true« for kolonnens tilhørende element »nullable«.

Normalform

Arkiveringen af offentlige myndigheders it-systemer er strukturmæssigt baseret på den relationelle databasemodel. Ved aflevering til offentligt arkiv defineres en relationel databasestruktur på 1. normalform således:

1. Enhver tabel skal for alle rækker have samme antal kolonner (dvs. multivariable attributter er ikke tilladt).
2. Alle tabeller skal have et navn, og navnet skal være entydigt.
3. Alle kolonner skal have et navn, og navnet skal være entydigt inden for den pågældende tabel.
4. Rækkefølgen af rækker må ikke være betydningsbærende i nogen tabel.
5. Rækkefølgen af kolonner må ikke være betydningsbærende i nogen tabel.
6. Enhver tabel skal have defineret et eller flere felter, som udgør en entydig primærnøgle (ingen rækker må være ens).
7. En fremmednøgle skal være relateret til en primærnøgle.
8. En fremmednøgle må ikke være relateret til dele af en sammensat primærnøgle. Der skal således være sammenfald mellem antal felter i primærnøgle og fremmednøgle.
9. Alle forbindelser mellem tabeller er udtrykkelige (direkte). Således kan der ikke forekomme forbindelser mellem tabeller på baggrund af fortolkning af deres feltindhold. (Hierarkiske databaser, netværksdatabaser og objektrelationelle databaser er således per definition ikke relationelle databaser).

   Det følger heraf, at en fremmednøgle kun må relateres til en specifik tabel (ved dennes primærnøgle). Fremmednøglen må altså ikke angives at være relateret til flere tabeller med henblik på, at konkrete poster vil være relateret til én ud af flere forskellige tabeller alt efter ”henvisningens art”.
10. Der bør ikke være tabeller uden relation til en eller flere andre tabeller.

For data i it-systemer og forskningsdata, som ikke er relationelle, skal myndigheden udarbejde en relationel model for arkiveringen, som lever op til ovenstående definition. Modellen skal godkendes af det modtagende arkiv forud for produktion af arkiveringsversionen.

tableIndex

tableIndex er en oversigt over alle tabeller i arkiveringsversionen og deres indbyrdes relationer.

tableIndex valideres af skemaet tableIndex.xsd, som er udarbejdet af Rigsarkivet og skal downloades fra www.sa.dk. tableIndex.xsd placeres i mappen ”standard” under mappen ”Schemas” jf. illustration i bilag 2.

Null-værdier

Eksempel på opmærkning i tableIndex af et felt i en tabel, hvor data kan antage værdien ”null”:

Hvis feltet må indeholde null-værdier, skal opmærkningen være:

```
<column>
  <colId>c3</colId>
  <name>sagsdato</name>
  <type>DATE</type>
  <typeOriginal>datetime</typeOriginal>
   <defaultValue></defaultValue>
  <nullable>true</nullable>
  <descripton>Dato for sagsoprettelse </descripton>
  <functionalDescription></functionalDescription>
</column>
```

Hvis feltet *ikke* må indeholde null-værdier:

```
<column>
  <colId>c3</colId>
  <name>sagsdato</name>
  <type>DATE</type>
  <typeOriginal>datetime</typeOriginal>
   <defaultValue></defaultValue>
  <nullable>false</nullable>
  <descripton>Dato for sagsoprettelse </descripton>
  <functionalDescription></functionalDescription>
</column>
```

4.C.6.a docIndex.xml skal danne forbindelsen mellem hvert dokument og dets placering. »docIndex.xml« skal desuden indeholde oplysninger om dokumenternes oprindelige filnavne, filtype i arkiveringsversionen samt eventuelle overordnede dokumenter. »docIndex.xml« skal ikke indeholde oplysninger om dokumenterne i kontekstdokumentationen.

4.C.6.b For hvert enkelt dokument i docIndex.xml angives de oplysninger, som fremgår af figur 4.4.

Figur 4.4 Oplysninger i docIndex.xml

Oversigt over de oplysninger, der skal angives i docIndex.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Forekomst pr. dokument	Obligatorisk	Eksempel
dID	**DocID**	**ID, som entydigt udpeger det enkelte dokument**	**ID på op til 12 cifre**	**1**	**Ja**	6351747
pID	**Parent**	**ID på overordnet dokument**	**DocID**	**0-1**	**Nej**	6351746
mID	**MediaID**	**Navn på det lagringsmedie, som dokumentet ligger på**	**Jf. 4.B.1.**	**1**	**Ja**	AVID.SA.19005.2
dCf	**SubFolder**	**Dokumentsamlingsmappe, som dokumentet ligger i**	**Jf. 4.G.**	**1**	**Ja**	docCollection1500
oFn	**OrigFilename**	**Angivelse af filens navn (inkl. ekstension) i it-systemet**	**Fritekst**	**1**	**Ja**	Indstilling.doc
aFt	**Archival Format**	**Angivelse af filens format i arkiveringsversionen**	**Ekstension, jf. 4.G.**	**1**	**Ja**	tif
gmlXsd	**Schema**	**For GML filer angives reference til det skema, der skal bruges til validering af den pågældende GML-fil**	**Fritekst**	**0-1**	**Nej**	

Hvis dokumentID'er i driftsversionen af systemet består af tal og højst er på 12 tegn, kan disse anvendes som DocID i arkiveringsversionen og registreres i docIndex.xml. Hvis driftssystemets dokumentID er bygget anderledes op, eller hvis det af andre grunde findes hensigtsmæssigt, genereres i forbindelse med aflevering et dokumentID, som svarer til bekendtgørelsens krav.

Originalfilens navn (oFn) skal altid angives. Som filnavn anvendes den titel, som dokumentet er gemt med i driftsudgaven af it-systemet. Som filnavnets ekstension anvendes det dokumentformat, som er registreret i driftsudgaven, f.eks Word.Document.8 eller TIFImage.Document.

I de tilfælde, hvor der ikke umiddelbart er nogen titel, anvendes det objektID eller lignende identifikation af dokumentet, som findes i driftsudgaven.

Alle dokumenter, som i it-systemets driftsversion kan identificeres som separate dokumenter, skal i arkiveringsversionen lagres separat og have hvert sit entydige dokumentID. Det gælder eksempelvis indlejrede elementer som f.eks. billeder, vedhæftede bilag og e-mails med vedhæftede dokumenter.

Driftversionens basale funktionalitet til genfinding af sager skal bevares i arkiveringsversionen, hvorfor hierarkiet mellem dokumenterne skal bevares. Referencen mellem dokumenterne registreres i docIndex.xml, hvor dokumenternes indbyrdes relation registreres som en parent/child-relation.

4.C.7.a researchIndex.xml skal indeholde angivelse af hovedtabeller og koder for manglende værdier, jf. figur 4.5.

Figur 4.5 Oplysninger i researchIndex.xml

1. Oplysninger om arkiveringsversionen				
	Elementnavn	**Beskrivelse**	**Forekomst pr. aflevering**	**Obligatorisk**
1.a	**mainTables**	**Liste over hovedtabeller**	**1**	**Ja**
2. Oplysninger om liste over hovedtabeller				
	Elementnavn	**Beskrivelse**	**Forekomst pr. liste over hovedtabeller**	**Obligatorisk**
2.a	**table**	**Hovedtabel**	**1-m**	**Ja**
3. Oplysninger om tabel				
	Elementnavn	**Beskrivelse**	**Forekomst pr. tabel**	**Obligatorisk**
3.a	**tableID**	**Hovedtabellens ID, som angivet i elementet »folder« i tableIndex.xml, jf. regler for navngivning af tabeller i 5.A.**	**1**	**Ja**
3.b	**source**	**Datas oprindelige format eller navn på program hvorfra data stammer**	**1**	**Ja**
3.c	**specialNumeric**	**Boolsk værdi, som angiver, om der er specialkoder for manglende værdier**	**0-1**	**Ja, hvis der er anvendt koder for manglende værdier**
3.d	**columns**	**Liste over kolonner**	**0-1**	**Ja, hvis der er anvendt koder for manglende værdier**
4. Oplysninger om kolonneliste				
	Elementnavn	**Beskrivelse**	**Forekomst pr. liste over kolonner**	**Obligatorisk**
4.a	**column**	**Kolonne**	**1-m**	**Ja**
5. Oplysninger om kolonne				
	Elementnavn	**Beskrivelse**	**Forekomst pr. kolonne**	**Obligatorisk**
5.a	**columnID**	**Kolonnens ID som angivet i tableIndex.xml, jf. regler for navngivning af kolonner i figur 6.3, 4.b**	**1**	**Ja**

5.b	**missingValues**	**Liste over koder for manglende værdier**	**1**	**Ja**
6. Oplysninger om liste over koder for manglende værdier				
	Elementnavn	**Beskrivelse**	**Forekomst pr. liste over koder for manglende værdier**	**Obligatorisk**
6.a	**value**	**Kode for manglende værdi**	**1-m**	**Ja**

researchIndex indeholder oplysninger om hvilke tabeller i arkiveringsversionen, der er hovedtabeller, samt hvilke koder der er anvendt for manglende værdier (missing values). Læs mere om koder for manglende værdier i bilag 9.

researchIndex skal kun findes i arkiveringsversionen, hvis data stammer fra statistikfiler og afleveres som en afleveringspakke jf. bilag 9. Denne indeksfil skal således ikke indgå i arkiveringsversioner med andre typer forskningsdata, f.eks. kvalitative studier (dokumentsamlinger) eller databaser.

researchIndex genereres automatisk (f.eks. med programmet ASTA) i forbindelse med arkivets konvertering af en afleveringspakke iht. bilag 9 til en arkiveringsversion iht. bilag 1-8.

4.D. Mappen *Tables*

Det anbefales at have en figur fra bilag 2 ved siden af, når dette afsnit læses.

4.D.1 Mappen Tables skal indeholde én mappe for hver tabel i arkiveringsversionen.

4.D.2.a Mappen for en tabel navngives »table[fortløbende nummer]«.

4.D.2.b Den fortløbende nummerering begynder med 1. Foranstillede nuller må ikke anvendes.

4.D.3 Mappen for hver tabel skal indeholde en fil: table[fortløbende nummer].xml, jf. dog 4.D.5

4.D.4 »table[fortløbende nummer].xml« er en XML-instans, der indeholder data for den pågældende tabel, og dens struktur skal være i overensstemmelse med XML-instansen »tableIndex.xml«, således at datatyper i XML instansen er transformeret fra SQL:1999 til XML-datatyper i overensstemmelse med bilag 5 figur 5.1.

4.D.5 Det er tilladt i mappen for hver tabel at inkludere et XML-skema, der kun angiver strukturen for den pågældende tabel. Skemaet navngives »table[fortløbende nummer].xsd«. Skemaet skal være i overensstemmelse med XML-instansen »tableIndex.xml«, der angiver strukturen for hele den relationelle database herunder samtlige tabeller.

Jævnfør figuren i bilag 2

Skemaerne i 4.D.5 kan undlades i arkiveringsversionen, da de genereres i Rigsarkivets testprogram, ADA, på baggrund af beskrivelserne i tableIndex. ADA kan downloades fra Rigsarkivets hjemmeside www.sa.dk

4.D.6 Hvis et felt i en tabel har værdien NULL, skal XML-instansens (»table[fortløbende nummer].xml«) element have attributten xsi:nil="true" for dette felt, jf. W3C-standarden for håndtering af nil-værdier i XML.

Eksempel på skema med angivelse af NULL værdi

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema  xmlns:xs="http://www.w3.org/2001/XMLSchema"
                xmlns="http://www.sa.dk/xmlns/siard/1.0/schema0/table3.xsd"
                attributeFormDefault="unqualified"
                elementFormDefault="qualified"
                targetNamespace="http://www.sa.dk/xmlns/siard/1.0/schema0/table3.xsd">
                <xs:element name="table">
                <xs:complexType>
                                <xs:sequence>
                                <xs:element maxOccurs="unbounded" minOccurs="1" name="row"
type="rowType">
</xs:element>
                </xs:sequence>
</xs:complexType>
</xs:element>
<xs:complexType name="rowType">
                <xs:sequence>
<xs:element minOccurs="1" name="c1"  type="xs:integer" nillable="false"/>
<xs:element minOccurs="1" name="c2"  type="xs:string" nillable="false"/>
<xs:element minOccurs="0" name="c3"  type="xs:date" nillable="true"/>
<xs:element minOccurs="1" name="c4"  type="xs:string" nillable="false"/>
                </xs:sequence>
</xs:complexType>
</xs:schema>
```

4.E. Mappen *ContextDocumentation*

4.E.1 Mappen *ContextDocumentation* skal indeholde én eller flere dokumentsamlingsmapper med kontekstdokumentation, jf. 6.B.

4.E.2 En dokumentsamlingsmappe med kontekstdokumentation må indeholde op til 10.000 dokumentmapper.

4.E.3 Dokumentsamlingsmapperne navngives "docCollection[fortløbende nummer]", begyndende med 1. Navnet skal være unikt inden for *ContextDocumentation*.

4.E.4 Hvert dokument i kontekstdokumentationen skal tildeles et ID på op til 12 cifre. Dokumentets ID skal være unikt inden for *ContextDocumentation*.

4.E.5 En dokumentmappe skal indeholde ét dokument, som består af én eller flere filer af samme format, og navngives med dokumentets ID. Foranstillede nuller må

ikke anvendes.

4.E.6 Et dokuments fil (eller filer) navngives fortløbende med et nummer, begyndende med 1 samt formatets ekstension, jf. 4.G.6.

F.eks. navngives filerne i ét TIFF-dokument, bestående af to single page TIFF-filer, 1.tif, 2.tif. En lydsekvens, bestående af én MP3-fil, navngives 1.mp3.

4.F. Mappen *Schemas*

4.F.1 Mappen *Schemas* skal være opdelt i undermapperne standard og localShared.

4.F.2 Mappen *standard* skal indeholde skemaer for arkiveringsversionens indeksfiler, jf. bilag 8, samt W3C standard XML-skema, jf. http://www.w3.org/2001/XMLSchema.xsd

4.F.3 For skemaerne fileIndex.xsd, archiveIndex.xsd, contextDocumentationIndex.xsd, tableIndex.xsd, docIndex.xsd, researchIndex.xsd samt W3Cs standard XML-skema gælder, at der altid skal anvendes de skemaer, som Rigsarkivet stiller til rådighed. Skemaerne og deres navngivning må ikke ændres i arkiveringsversionen.

4.F.4 Mappen *localShared* skal indeholde eventuelle GML-skemaer, som ikke er placeret sammen med det pågældende GML-dokument, jf. 4.G.7.a.

Mappen Schemas skal indeholde to mapper, kaldt henholdsvis *standard* og *localShared*. Standardmappen skal indeholde de skemaer, som altid skal anvendes til validering af arkiveringsversionen.

Disse skemaer er udarbejdet af Rigsarkivet og kan downloades på hjemmesiden www.sa.dk.

Hvis arkiveringsversionen indeholder GML-filer (geodata) kan globale GML-skemaer brugt til validering af flere GML-filer i arkiveringsversionen og placeres i LocalShared mappen. Selve LocalShared mappen skal altid oprettes, også selvom den pågældende arkiveringsversion ikke indeholder GML-dokumenter.

4.G. Mappen *Documents*

4.G.1 Mappen Documents skal indeholde én eller flere dokumentsamlingsmapper, dog maksimalt 10.000.

4.G.2 Dokumentsamlingsmapperne navngives "docCollection[fortløbende nummer]", begyndende med 1. Navnet skal være unikt inden for *Documents*.

4.G.3 En dokumentsamlingsmappe må indeholde op til 10.000 dokumentmapper.

4.G.4 Hvert dokument skal i arkiveringsversionen tildeles et ID på op til 12 cifre. Dokumentets ID skal være unikt inden for Documents.

4.G.5 En dokumentmappe skal indeholde ét dokument, som består af én eller flere filer af samme format, og navngives med dokumentets ID. Foranstillede nuller

må ikke anvendes.

4.G.6 Et dokuments fil (eller filer) navngives fortløbende med et nummer, begyndende med 1 samt formatets ekstension. Foranstillede nuller må ikke anvendes.

4.G.7 For GML-filer lagres det relevante skema i samme mappe som GML-filen, og navngives med fortløbende nummer efterfulgt af .xsd, jf. dog 4.G.7.a. Foranstillede nuller må ikke anvendes.

4.G.7.a GML-skemaer kan alternativt lagres i den skema-mappe, som navngives localShared, jf. 4.F. GML-skemaer i mappen localShared navngives "localSchema[fortløbende nummer]", begyndende med 1.

4.G.8 Anvendelse af ekstensions

4.G.8.a Dokumenter i formatet TIFF skal have ekstension tif.

4.G.8.b Dokumenter i formatet MP3 skal have ekstension mp3.

4.G.8.c Dokumenter i formatet MPEG-2 og MPEG-4 skal have ekstension mpg.

4.G.8.d Dokumenter i formatet JPEG-2000 skal have ekstension jp2.

4.G.8.e Dokumenter i formatet GML skal have ekstension gml.

4.G.8.f Dokumenter i formatet WAVE skal have ekstension wav.

Bemærk, at alle ekstensions skal enten skrives med små bogstaver **eller** store bogstaver.

4.G.9 Muligheden for at fremsøge sagligt sammenhørende dokumenter skal overføres til arkiveringsversionen efter det modtagende arkivs nærmere anvisning.

Der tænkes her på typer af systemer, hvor sammenhængen mellem dokumenterne i et givet sagsforløb ikke fremgår af relationer mellem felterne i arkivsystemets tabeller (f.eks. en sagstabel og en akttabel), men er håndteret på anden måde gennem f.eks. dokumentkæder eller mappestrukturer. I sådanne tilfælde skal tilknytningen mellem dokumenter kunne genfindes i arkiveringsversionen, hvor den skal være konverteret til en relationel sammenhæng mellem tabeller suppleret med entydige navngivning og lagring af dokumenterne.

I forbindelse med afleveringsforberedelsen træffer den afleverende myndighed og det modtagende arkiv nærmere aftale om denne konvertering.

Bilag 5 – Dataindhold

Arkiveringsversionens dataindhold består af tabelindhold i standardiserede datatyper og af eventuelle digitale dokumenter, lyd, video og geodata i standardiserede formater.

5.A. Tabelindhold

5.A.1.a I overensstemmelse med den tabelstruktur, der i XML-instansen »tableIndex.xml« er defineret for hver tabel, skal hver tabel findes i en XML-instans navngivet »table[fortløbende nummer]. xml«.

5.A.1.b Den fortløbende nummerering begynder med 1. Foranstillede nuller må ikke anvendes.

5.A.2 Indholdet af de enkelte felter skal renses for eventuelle foran- og efterstillede blanktegn.

Indholdet af alle de tabeller, som skal indgå i arkiveringsversionen, skal lagres i selvstændige xml-filer og bruge et tilhørende xml-skema. Xml-filen og xml-skema nummereres fortløbende eksempelvis tabel1.xml med det tilhørende skema tabel1.xsd.

Se i øvrigt mappen ”Tables” i den grafiske oversigt over elementer og struktur i en arkiveringsversion i bilag 2.

Der må ikke indgå tabeller uden dataindhold i arkiveringsversionen, da de er uden historisk værdi og kan så tvivl om fejlkonvertering.

Eksempel på opmærkning af dataindholdet i en tabel samt det tilhørende xml-skema:

Uddrag af xml-opmærkningen af tabellen DOKTABEL(Table1):


```
  <?xml version="1.0" encoding="utf-8" ?>
- <table xmlns="http://www.sa.dk/xmlns/siard/1.0/schema0/table1.xsd"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.sa.dk/xmlns/siard/1.0/schema1/table1.xsd table1.xsd">
- <row>
  <c1>1</c1>
  <c2 xsi:nil="true" />
  <c3>Projektbeskrivelse</c3>
  <c4>1</c4>
    </row>
- <row>
  <c1>2</c1>
  <c2>1999-01-05</c2>
  <c3>Dagsorden for næste møde i Etisk Råd</c3>
  <c4>2</c4>
    </row>
- <row>
  <c1>3</c1>
  <c2>1999-04-07</c2>
  <c3>EU-kommisionens betænkning om kryptering</c3>
  <c4>3</c4>
    </row>
- <row>
```

```
  <c1>4</c1>
  <c2>1999-04-07</c2>
  <c3>Udkast til artikel</c3>
  <c4>4</c4>
     </row>
- <row>
  <c1>5</c1>
  <c2 xsi:nil="true" />
  <c3>Program for delegationsbesøget</c3>
  <c4>5</c4>
     </row>
- <row>
  <c1>6</c1>
  <c2>1999-05-27</c2>
  <c3>Brev til Etisk Råd om udnævnelse af medlem</c3>
  <c4>6</c4>
     </row>
```


xml-skema til validering af tabellen

```
?xml version="1.0" encoding="utf-8"?>
<xs:schema  xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns="http://www.sa.dk/xmlns/siard/1.0/schema0/table1.xsd" attributeFormDefault="unqualified"
elementFormDefault="qualified" targetNamespace="http://www.sa.dk/xmlns/siard/1.0/schema0/table1.xsd">
  <xs:element name="table">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" minOccurs="1" name="row" type="rowType">
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:complexType name="rowType">
    <xs:sequence>
      <xs:element minOccurs="1" name="c1"  type="xs:integer" nillable="false"/>
      <xs:element minOccurs="0" name="c2"  type="xs:date" nillable="true"/>
      <xs:element minOccurs="0" name="c3"  type="xs:string" nillable="true"/>
      <xs:element minOccurs="0" name="c4"  type="xs:integer" nillable="true"/>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```

Bemærk, at skemaet ikke behøver indgå I arkiveringsversionen, da det genereres i Rigsarkivets testprogram, ADA, i forbindelse med test ud fra de oplysninger, som er angivet i tableIndex.

5.B. Datatyper

5.B.1.a De standardiserede datatyper, som skal anvendes for tabelindhold, er angivet i figur 5.1. De er et uddrag af datatyper fra standarden SQL:1999 repræsenteret som datatyper i W3C XML Schema Language 1.0.

5.B.1.b Det er datatypen i W3C XML Schema Language 1.0, som skal anvendes.

Oversættelsen fra datatyper i SQL:1999 er angivet for at vise, hvorledes oversættelsen til datatyper i W3C XML Schema Language 1.0 skal foretages.

Figur 5.1 Tilladte datatyper for tabelindhold

Datatype	**SQL:1999** (angives i tableIndex.xml)	**XML-datatype** (angives i table[fortløbende nummer].xsd)	**Eksempel**
Tekst	CHARACTER, CHAR, CHARACTER VARYING, CHAR VARYING, VARCHAR, NATIONAL CHARACTER, NATIONAL CHAR, NCHAR NATIONAL CHARACTER VARYING, NATIONAL CHAR VARYING, NCHAR VARYING	string	Kristian Erslevs Gade
Heltal	INTEGER, INT, SMALLINT	integer	25 +45 -234 10000
Decimaltal	NUMERIC, DECIMAL, DEC	decimal	23.75 .10 -123.76 150000.25
	FLOAT	float	
	REAL, DOUBLE PRECISION	double	
Boolsk	BOOLEAN	boolean	1/0 true/false
Hexadecimal	CHARACTER, CHAR, CHARACTER VARYING, CHAR VARYING, VARCHAR, NATIONAL CHARACTER, NATIONAL CHAR, NCHAR, NATIONAL CHARACTER VARYING, NATIONAL CHAR VARYING, NCHAR VARYING	hexBinary	CC31E216BC2912DEB1E5CB196923B12E
Dato	DATE	date	YYYY-MM-DD 2010-05-21
Tidspunkt	TIME[WITH TIME ZONE]	time	hh:mm:ss 15:15:07
Tidsstempel	TIMESTAMP[WITH TIME ZONE]	dateTime	YYYY-MM-DDThh:mm:ss 2010-09-01T14:20:35
Tidsperiode	INTERVAL	duration	PnYnMnDTnHnMnS P3Y7M21DT3H2M45S

5.B.2 Datatypen *string* må kun indeholde ikke-opmærket tekst, som umiddelbart kan fortolkes.

5.B.3 Datatypen *boolean* kan jf. W3C kun antage værdierne 1; 0 eller *true; false.*

5.B.4 Datatyperne *date, time* og *dateTime* kan bruges med eller uden *Time Zone.*

Forklaring til Figur 5.1
Den første kolonne ”Datatype” i figur 5.1 angiver, hvad forskellige datatyper kaldes på dansk i dagligt sprogbrug, mens anden kolonne angiver mere specifikke betegnelser for datatyperne i SQL:1999. Tredje kolonne angiver betegnelser for XML-datatyper som de tilladte SQL:1999 datatyper svarer til.

Anvendelse og angivelse af XML-datatyperne
XML-datatypen skal anvendes i XML-instanserne for tabelindhold, f.eks. table1.xml. Det betyder at data med andre datatyper skal konverteres til disse XML-datatyper ved dannelse af arkiveringsversionen.
F.eks. skal datoværdier med den oprindelige datatype 03-12-2019 konverteres til XML datatypen 2019-12-03.

Selvom XML-datatypen skal anvendes for tabelindhold er det stadig specifikationen af SQL:1999 datatypen, som skal overholdes. Det betyder, at hvis XML-datatypen er rigere end SQL99 datatypen, er det stadig SQL:1999 datatypen, der skal anvendes. Dette er f.eks. tilfældet for datatypen TIME som i SQL:1999 kan have en præcision på op til 9 decimaler, mens der i XML-standarden for datatypen time ingen begrænsning er på antallet af decimaler. I dette tilfælde skal antallet af decimaler i XML dataformatet anvendt for tabelindhold begrænses til 9, så SQL:1999 standarden overholdes.

XML-datatyperne (string, integer, decimal osv.) skal desuden angives i de skemaer (f.eks. table1.xsd), der validerer tabeldatafilerne (f.eks. table1.xml). Disse skemaer genereres dog automatisk, når testprogrammet ADA anvendes.

Myndigheden/leverandøren skal ud fra egne data vurdere, hvilke XML-datatyper som skal anvendes for arkiveringsversionens tabelindhold. Dog må der ikke konverteres på tværs af de *generelle datatyper* angivet i kolonnen ”datatype” i figur 5.1. F.eks. må datoer ikke konverteres til tekst, hvis dataindholdet oprindeligt var en dato.

Bemærk, at:

- Alle felter uanset datatype skal renses for evt. foran- og efterstillede blanktegn, så det kun er det reelle feltindhold, som indgår i arkiveringsversionen.

Angivelse af SQL-datatyper i tableIndex.xml
Datatypen SQL:1999 skal angives i elementet ”type” i tableIndex.xml filen. Formålet med denne metadataregistrering er at muliggøre, at arkiveringsversionens XML-data kan genrejses automatisk som en relationel database på en SQL-server.

Hvis data stammer fra en SQL database og en kolonne oprindeligt havde datatypen VARCHAR før konvertering til XML-datatypen string i forbindelse med produktion af arkiveringsversionen, angives datatypen VARCHAR i elementet ”type” for denne kolonne i tableIndex.xml.

Hvis data stammer fra andre datakilder og derfor aldrig har eksisteret som SQL datatyper, angives den SQL:1999 datatype, som svarer bedst til det konverterede datas XML-datatype.

Myndigheden/leverandøren skal ud fra egne data vurdere, hvilke SQL:1999 datatyper, der bedst svarer til de konverterede datas XML-datatyper.

Bemærk, at:

- Angivelsen af datatype i tableIndex.xml skal skrives enten med store bogstaver ***eller*** små bogstaver for at kunne valideres med det tilhørende skema.

Eksempler på datarepræsentation i de tilladte datatyper

STRING
STRING bruges i felter med tekst som f.eks. vejnavne:

<c1>Kristian Erslevs Gade</c1>

Bemærk, at tekster kun må indeholde følgende styrekoder: tabulator, nylinje og vognretur. Tekster må ikke indeholde opmærkningssprog som HTML, RFT eller XML. Dette gælder dog ikke enkeltstående links til hjemmesider o. lign.

INTEGER
INTEGER anvendes til heltal som f.eks. husnumre:

<c2>101</c2>

DECIMAL
Et decimaltal, som f.eks. anvendes til at angive prisen på en ydelse:

<c3>999.50</c3>

Der kan angives op til 18 decimaler efter punktummet. Decimal kan også anvendes, hvis et tal ikke har decimaler efter punktum f.eks.:

<c4>1000</c4>

Bemærk, at XML-datatypen decimal ikke tillader komma som decimalmærke. Det er heller ikke tilladt at anvende videnskabelig notation.

BOOLEAN
Anvendes til at angive, om en værdi er enten sand eller falsk (true/false).

HEXBINARY
HexBinary anvendes typisk i felter med kontrolsummer efter MD5.

DATE
DATE anvendes til angivelse af dato i formatet “YYYY-MM-DD”, hvor

- YYYY er året
- MM er måneden
- DD er dagen

F.eks. vil angivelse af d. 21. maj 2010 som udbetalingsdato for en ydelse skulle angives som:

<c5>2010-05-21</c5>

Bemærk, alle elementer skal indgå i formatet.

TIME
TIME anvendes til angivelse af et tidspunkt i formatet "hh:mm:ss", hvor:

- hh angiver timen
- mm angiver minuttet

- ss angiver sekundet

F.eks. vil angivelse af klokken kvart over fire om eftermiddagen, som tidspunkt for en indlæggelse på et sygehus, skulle angives som

<c6>15:15:07</c6>

dateTime
dateTime anvendes til angivelse af en dato og et tidspunkt, et såkaldt tidsstempel, i formatet YYYY-MM-DDThh:mm:ss, hvor:

- YYYY angiver året
- MM angiver måneden
- DD angiver dagen
- T angiver starttidspunktet
- hh angiver time
- mm angiver minut
- ss angiver sekund

Formatet kan f.eks. anvendes til angivelse af starttidspunkt for en måling:

<c7>2010-09-01T14:20:35</c7>

Bemærk, at alle elementer skal indgå i formatet.

duration
Duration anvendes til at angive en tidsperiode (et tidsinterval) i formatet PnYnMnDTnHnMnS, hvor:

- P angiver perioden (obligatorisk)
- nY angiver antallet af år
- nM angiver antallet af måneder
- nD angiver antallet af dage
- T angiver start af en tidsafgrænsning/periode (obligatorisk, hvis timer, minutter og sekunder anvendes)
- nH angiver antallet af timer
- nM angiver antallet af minutter
- nS angiver antallet af sekunder

Duration kan f.eks. anvendes til angivelse af, at et patent har været gældende i 3 år, 7 måneder, 21 dage, 3 timer, 2 minutter og 45 sekunder.

<c9>P3Y7M21DT3H2M45S</c9>

5.C. Konvertering af tabelindhold til digitale dokumenter, lyd, video eller geodata

5.C.1 Tabelindhold skal overholde de angivne datatyper, jf. 5.B. Det følger heraf, at dataindhold i tabelform fra et it-system, som skal overføres til en arkiveringsversion og som ikke umiddelbart kan overholde dette krav, skal have sit dataindhold konverteret således:

5.C.1.a til digitale dokumenter, lyd, video eller geodata, idet indholdet konverteres til de formater, som fremgår af 5.E-5.G.

5.C.1.b til tabelindhold af datatypen string, jf. 5.B, idet andet indhold end det for datatypen tilladte slettes.

5.C.2 Det modtagende arkiv anviser, om et givet indhold skal håndteres iht. 5.C.1.a eller 5.C.1.b.

I nogle databaser lagres dokumenter, lyd, video og/eller geodata m.v. i binært format f.eks. som blobfelter i datatabellerne. Det giver problemer i forhold til arkiveringen, da de ikke kan lagres i de datatyper, som er angivet for dataindholdet i tabellerne.

De skal derfor enten udsøges og lagres i selvstændige filer i de formater, som ellers anvendes til dokumenter, lyd, video og geodata, eller *om muligt* konverteres til ren tekst, så datatypen "string" kan anvendes.
Det samme gælder tekstfelter, hvor der er anvendt opmærkningssprog som html, pdf, rft og xml i teksten. Her skal teksten enten konverteres til egentlige dokumenter eller konverteres til ren tekst, som kan lagres med datatypen "string".

Når der udarbejdes afleveringsbestemmelse for systemet, aftales det med modtagende arkiv, hvordan de konkrete forekomster skal håndteres.

Læs mere om håndtering af digitale dokumentarkiver i "Dokumentvejledningen", som findes på www.sa.dk.

5.D. Tekstformat

5.D.1 Unicode indkodning

5.D.1.a Data i arkiveringsversionens indeksfiler og tabelindhold skal være indkodet som well-formed UTF-8, som angivet i ISO/IEC 10646:2003 Annex D og som beskrevet i The Unicode Standard 5.1, kapitel 3.

5.D.1.b De indkodede karakterer skal være gyldige Unicode skalarværdier. Surrogater/RC-elementer og Unicode noncharacters må ikke anvendes.

5.D.1.c Ingen karakterer i Private Use Area må indkodes.

5.D.1.d Styrekodetegnene fra og med #x00 til og med #x1F er ikke tilladt med undtagelse af styrekodetegnene tabulator (TAB) #x09, nylinje (LF) #x0A og vognretur (CR) #x0D.

Alt tabelindhold, skemaer og indeksfiler skal have unicode (well-formed UTF8) som tegnsæt. Eneste undtagelse i anvendelsen af unicode i arkiveringsversionen er mappenavne (se bilag 2), hvor kun en delmængde af standarden må anvendes.

For beskrivelse af Uni-code standarden se:

- http://www.unicode.org/versions/Unicode5.1.0/
- http://www.unicode.org/versions/Unicode5.0.0/ch03.pdf

Der må ikke anvendes karakterer i arkiveringsversionen, som har været brugerdefinerede i den specifikke driftsversion af systemet.

Styrekodetegnene fra og med #x00 til #x1F betegnes som styrekodetegnsgruppen C1. Denne gruppe anvendes kun af ældre terminalbaserede systemer til styring af skærmen, f.eks. kommandoen "slet skærm" eller "bell". Disse tegn er ikke tilladte i arkiveringsversionen, dog med undtagelse af tabulator, linjeskift eller vognretur.

5.D.2 XML indkodning

5.D.2.a I henhold til XML-standarden gælder de regler for angivelse af de pågældende tegn, som fremgår af figur 5.2.

Figur 5.2 Angivelse af tegn i XML-standarden

Tegn	Entitet	Anvendelse
<	<	Obligatorisk
>	>	Anbefalet
&	&	Obligatorisk
’	'	Anbefalet
"	"	Anbefalet

5.D.2.b Tegnene fra og med #x7F til og med #x9F skal angives med deres decimale eller hexadecimale tegnreference.

5.D.2.c CDATA sektioner <![CDATA]]> må ikke anvendes.

5.E. Digitale dokumenter

5.E.1. Et digitalt dokument, jf. dog 5.F og 5.G, skal lagres i ét af følgende formater:

- **det grafiske bitmapformat TIFF, version 6.0 baseline.**
- **JPEG-2000 efter standarden ISO/IEC 15444-1:2004. Information technology - JPEG 2000 image coding system - Part 1: Core coding system.**

5.E.1.a Det er tilladt at benytte begge formater inden for samme arkiveringsversion.

5.E.2 Dokumenter i TIFF skal komprimeres efter følgende kompressionsregler:

5.E.2.a Sort/hvide dokumenter skal komprimeres med CCITT/TSS Grp3, Grp4, PackBit eller LZW.

5.E.2.b Dokumenter med gråtoner eller farver skal komprimeres med PackBit eller LZW.

Der kan anvendes to forskellige formater til arkivering af dokumenter, TIFF (version 6.0 baseline) og JPEG2000:

- TIFF 1 bit med kompression grp. 4 samt LZW er velegnet til sort/hvide dokumenter
- TIFF 4-24 bit med kompression LZW er velegnet til almindelige dokumenter, som indeholder tegninger og fotos i varierede størrelse og antal
- TIFF 24 bit er velegnet til tegninger, kort, fotos etc., hvor den eksakte bevaring af dokumentet er af betydning. Som eksempel kan nævnes satellitfoto af afgrøder, hvor tolkningen af farven på de enkelte pixels kan være af betydning
- JPEG2000 er velegnet til store tegninger, kort, fotos etc., såfremt farven på de enkelte pixels er af mindre betydning. Som eksempel kan nævnes indskannede matrikelkort eller affotograferede skibstegninger

Kvaliteten af den grafiske kopi skal svare til den kvalitet, som myndigheden selv anvendte, hvilket for digitalt fødte dokumenter svarer til en passende udskriftkvalitet (f.eks. 300 DPI). For indskannet materiale skal kvaliteten modsvare kvaliteten på det materiale, som danner grundlag for kopien.

Valg af kompressionstype og bitdybde (antal mulige farver) må gerne variere fra side til side inden for samme dokument.

Myndigheden skal sikre sig, at dokumenterne også efter konvertering til arkivformatet er læsbare f.eks. ved at kontrollere at kommaer fremstår korrekt i dokumentet, og ikke ligner punktummer. Hvis det er tilfældet, er opløsningen for lav, og dokumentet skal lagres med en højere opløsning.

Rigsarkivet anbefaler anvendelsen af grp. 4 og LZW-kompression eftersom dokumenter komprimeret med disse algoritmer fylder mindre end hvis der benyttes grp. 3 og PackBit.

Hvis der ved konverteringen af systemet til arkiveringsversionen opstår problemer med dokumenter, der ikke kan konverteres til arkivformaterne eller der er dokumenter, som helt mangler i systemet, skal det modtagende arkiv kontaktes, så der kan træffes aftale om håndtering af problemstillingen. Dette gælder kun for dokumenter, som er bestemt til bevaring jf. § 6 .

5.E.3 Dokumenter i TIFF RGB må udelukkende benytte følgende bitdybder: 1, 2, 4, 8, 24 og 32.

5.E.3.a Dokumenter i TIFF RGB må maksimalt benytte 3 farvekanaler med en maksimal bitdybde på 24 bit (8x8x8 bit) evt. suppleret med maksimalt en alfakanal (8 bit), således at den samlede bitdybde for en billedfil ikke kan overstige 32 bit.

5.E.4 Dokumenter i TIFF CMYK må udelukkende benytte følgende bitdybder: 1, 2, 4, 8, 32 og 40.

5.E.4.a Dokumenter i TIFF CMYK må maksimalt benytte 4 farvekanaler med en maksimal bitdybde på 32 bit (8x8x8x8 bit) evt. suppleret med maksimalt én alfakanal (8 bit), således at den samlede bitdybde for en billedfil ikke kan overstige 40 bit.

Hvad er bitdybde?
Bitdybden afspejler behovet for mængden af informationer (farver) som skal gemmes i et dokument.

Bitdybden angiver antallet af mulige farver i et digitalt dokument. TIFF RGB filer består af 3 farvekanaler med de 3 RGB farver rød, grøn og blå. TIFF CMYK filer består af 4 farvekanaler med de 4 CMYK farver cyan, magenta, yellow og key color (sort). Disse farvekanaler kan suppleres af en eller flere alfakanaler (ekstra kanal) med informationer om transparens, intensitet eller masker af forskellig type.

Følgende bitdybder er tilladte for TIFF-filer:

Tilladte bitdybder for TIFF-filer				
	Bitdybde	**Antal farver**	**Kompression**	**Typisk anvendelse**
FAX MONO	1 bit	2 farver, typisk sort og hvid	Grp3, Grp4	Dokumenter i sort/hvid, f.eks. fax, digitalt fødte rene tekstdokumenter uden farver.
PALETTE 1	1 bit	2 farver, typisk sort og hvid	PackBit eller LZW	Dokumenter i sort/hvid, f.eks. fax, digitalt fødte rene tekstdokumenter uden farver.
PALETTE 2	2 bit	4 farver	PackBit eller LZW	
PALETTE 4	4 bit	16 farver	PackBit eller LZW	Dokumenter med ganske få farver, max 16.
PALETTE 8	8 bit	256 farver	PackBit eller LZW	Kontordokumenter, regneark og andre dokumenter med billeder, grafik og betydningsbærende farver.
RGB 8,8,8	24 bit	ca. 16 mio farver	PackBit eller LZW	Fotografier og andre dokumenter med mange farvenuancer. Dokumenter bør scannes i 24 bit.
RGB 8,8,8,8	32 bit inkl. én alfakanal til transparens når det skønnes relevant	ca. 16 mio farver	PackBit eller LZW	
CMYK 8,8,8,8	32 bit		PackBit eller LZW	Fotografier og andre dokumenter med mange farvenuancer. Dokumenter bør scannes i 24 bit.
CMYK 8,8,8,8,8	40 bit inkl. én alfakanal til transparens når det skønnes relevant		PackBit eller LZW	

Hvorfor er store og visse bitdybder er problem?
En del TIFF viewere understøtter ikke visning af 16 bit og endnu færre kan håndtere 15 bit. Store bitdybder som 48 og 64 bit understøttes heller ikke i mange billedviewere. Dette vanskeliggør både validering af filerne under arkivering og visning af filerne til arkivets brugere.

Eksempler på RGB bitdybder med og uden alfakanal(er), som ikke er tilladte i en arkiveringsversion:

RGB 4,4,4 = 12 bit (extended gray 4096 nuancer af grå)
RGB 5,5,5 = 15 bit
RGB 5,5,6 = 16 bit

RGB	6,6,6	= 18 bit
RGB+A	12,12,12,12	= 48 bit (én alfakanal)
RGB	16,16,16	= 48 bit
RGB+A	16,16,16,16	= 64 bit (én alfakanal)

Hvad gør jeg hvis mine dokumenter har ikke tilladte bitdybder?
Hvis du har TIFF dokumenter med ikke tilladte bitdybder, findes der batchkonverteringsprogrammer der kan konvertere til den ønskede bitdybde.

Visse printerdrivere er i stand til at optælle antallet af farver i dokumentets sider og gemme i den mest optimale bitdybde når et digitalt født dokument gemmes/udskrives som TIFF.

Ved skanning af papirdokumenter til TIFF anbefales det at skanne med en bitdybde på 24 bit. Større scanningsprojekter med ensartet indhold anbefales at blive skannet til JPEG-2000, se mere herom i dokumentvejledningen.

5.E.5 TIFF dokumenters anvendelse af XResolution og YResolution (TIFF Tag 282 og 283), skal ske på en sådan måde, at forholdet mellem bredde og højde ved anvendelse af disse værdier (skalering) svarer til sidernes dimensioner i det oprindelige dokument.

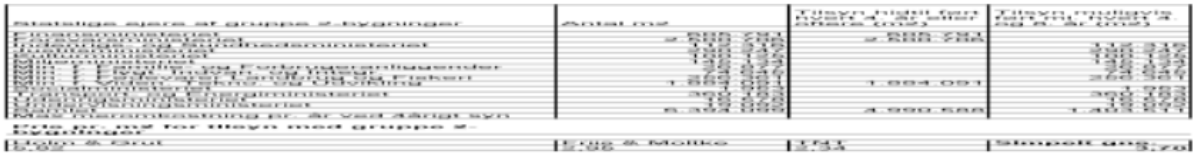
Hvad er XResolution og YResolution (DPI)?
DPI indikerer den størrelse (fysiske dimensioner) dokumentet havde i forbindelse med skanning og dermed en størrelse som vil være naturlig i forbindelse med reproduktion, f.eks. visning i en viewer eller udskrivning af dokumentet, når det skal bruges igen. DPI'en for et dokument anvendes således til skalering af dokumentet ved visning og print.

Angivelse af DPI (Dots Per Inch) håndteres i TIFF ved hjælp af begrebet *Resolution* hvor det er muligt at angive *XResolution* og *YResolution* sammen med en angivelse af *ResolutionUnit* som enten kan være *inch*, *cm* eller blot forholdet mellem x og y dimensionen. Værdierne er typisk identiske, f.eks. 300x300 pdi.

Hvad er skæv DPI?
Når en DPI er angivet skævt eller uens, f.eks. *600x96 dpi* skal dokumentet skalleres i dette forhold (600/96 = 6,25, dvs. skaleringsforholdet 1:6,25) når det vises i en viewer eller printes.

En "fotoviewer" (som Paint eller PhotoShop) der viser billeder og TIFF-dokumenter vil/skal <u>aldrig</u> tage hensyn til angivelse af DPI i modsætning til en "dokumentviewer" (som Windows billedfremviser eller IrfanView), som <u>altid</u> bør skalere dokumentet i det forhold som *XResolution* og *YResolution* angiver. Hvis et TIFF-dokument med skæv DPI vises i en fotoviewer bliver billedet derfor fladtrykt, som illustreret i figur 1.3.

Figur 1.3 TIFF-dokument i 600x96 dpi vist som billede/foto i en fotoviewer (WYSIWYG)



Hvornår er skæv DPI i et problem?
Jf. 5.E.5 skal *X*- og *YResolution* som udgangspunkt være identiske (f.eks. *300x300 dpi*) med mindre der er tale om GRP3 og GRP4.

Skæve DPI'er med ikke identisk *X*- og *YResolution* (f.eks. 600x96 dpi) er således tilladt så længe angivelsen sikrer at dokumentet vises korrekt, når det skaleres efter det angivne forhold (1:6,25) i en dokumentviewer, f.eks. ældre FAX dokumenter.

Hvis dokumentet vises forkert i en dokumentviewer, skal angivelsen af *X*- og *YResolution* rettes inden aflevering til arkiv. DPI kan ændres i nogle dokumentviewere.

Problemet vil ofte opstå i forbindelse med manuel ændring af DPI under konvertering af dokumenter eller redigering af billeder.

5.F. Lyd og video

5.F.1 Lydfiler skal lagres efter standarden MP3 DS/EN ISO/IEC 11172-3.

5.F.2 Det modtagende arkiv kan tillade, at lydfiler afleveres i formatet WAVE LPCM som specificeret i Multimedia Programming Interface and Data Specifications 1.0. IBM Corporation and Microsoft Corporation, August 1991. Dog begrænset til bitdybder, der er hele multipla af 8.

5.F.3 Videofiler skal lagres efter en af følgende standarder:

- **MPEG-2 DS/EN ISO/IEC 13818-2. Eventuel lyd indkodes som MP3, som specificeret i ISO/IEC 13818-3.**
- **MPEG-4 AVC DS/EN ISO/IEC 14496-10 (ITU-T H.264). Video indkodes som specificeret i ISO/IEC 14496-10. Eventuel lyd indkodes som AAC, som specificeret i ISO/IEC 14496-3. Video og lyd indpakkes i MPEG-4 formatet som defineret i ISO/IEC 14496-14.**

Lyd og video skal konverteres til de anviste formater på en sådan måde, at den oprindelige kvalitet i størst muligt omfang beholdes.

Der kan anvendes to forskellige lydformater ved aflevering til offentligt arkiv, MP3 og WAVE. Det vil normalt være MP3, som skal anvendes i arkiveringsversioner, men hvis myndigheden selv har anvendt WAVE til lagring af lyd i systemet, kan det modtagende arkiv tillade, at WAVE anvendes som arkiveringsformat. Det samme gælder for lydoptagelser, hvor en meget præcis lydgengivelse er vigtig. Et eksempel på dette er optagelser, hvor baggrundslyde har betydning for forståelsen, og konteksten for den pågældende optagelse.

Ved aflevering til Rigsarkivet fastsættes eventuel aflevering i WAVE-formatet i afleveringsbestemmelsen for systemet.

Hvis der ved konverteringen af systemet til arkiveringsversionen opstår problemer med lydfiler, der ikke kan konverteres til arkivformaterne, eller der er filer, som helt mangler i systemet skal det modtagende arkiv kontaktes og der skal træffes aftale om håndtering af problemstillingen.

5.G. Geodata

5.G.1 Data fra geografiske informationssystemer samt øvrige geodata lagres som GML-filer iht. specifikationerne, der er angivet af OGC (Open Geospatial Consortium) i ISO 19136 GML version 3.1.1.

5.G.1.a Geodatafiler journaliseret på sager i systemer med dokumenter kan konverteres til TIFF eller JPEG-2000, jf. 5.E., og behandles som et separat dokument iht. reglerne i bilag 4, 4.G. samt bilag 6, 6.C.4 - 6.C.6 efter aftale med det modtagende arkiv.

Geodatafiler journaliseret på sager i systemer med dokumenter er ofte særegne (udtræk af større) datasæt, der ved konvertering til GML hver især skal dokumenteres og valideres ved hjælp af et lokalt XML-skema. Såfremt det modtagende arkiv vurderer, at omkostningerne forbundet med dannelse af de lokale skemaer overstiger deres anvendelsespotentiale, kan de konverteres til GML 3.1.1, hvorefter den opmærkede tekst i GML-filen kan konverteres til TIFF eller JPEG-2000 og afleveres som sådan.

5.G.2 I undermappen *localShared* til mappen *Schemas*, placeres skemasamlingen fra Open Geospatial Consortium til GML version 3.1.1 i uændret form. Skemaerne kan hentes i en samlet zip-pakke på Rigsarkivets hjemmeside.

En arkiveringsversion skal være selvdokumenterende. Derfor medtages alle standardskemaer fra Open Geospatial Consortium, der er nødvendige til at validere GML 3.1.1-standarden i uændret form. Standardskemaerne skal bevare deres oprindelige navngivning og derfor ikke navngives iht. 4.G.7.a.

5.G.3 GML-filer, som er større end 1 GB, opdeles i mindre enheder efter det modtagende arkivs anvisning.

Rigsarkivet har af hensyn til behandling af geodata fastsat en vejledende øvre grænse på 1 GB pr. fil. Det er som udgangspunkt fastsat, at data udtrukket og/eller konverteret fra en databasetabel eller en geodatafil bevares *i samme omfang* i den resulterende GML-fil – uanset størrelse. Opdeling af data i mindre enheder med henblik på manuel håndtering af store datamængder opnås bedre ved udarbejdelse af forespørgsler, der gør det muligt at indlæse logiske inddelinger af data. Opdeling af data i mindre enheder, der er nødvendig for at gennemføre automatiske test, både kan og bør anvises.

5.G.4 Hver GML-fil, evt. efter opdeling, jf. 5.G.3, behandles som et separat dokument iht. reglerne i 4.G.

GML-filer skal lagres i en af arkiveringsversionens dokumentmapper *Documents*. Her skal reglerne for placering i dokumentsamlingsmapper iht. punkt 4.G overholdes. GML-filerne lægges i dokumentmapper, der navngives iht. punkterne 4.G.4 og 4.G.5, og selve filen navngives iht. punkt 4.G.6.

5.G.4.a Hver GML-fil opmærkes i *docIndex.xml* og elementet *gmlXsd* opmærkes med en reference til det lokale XML-skema, der angiver GML-filens syntaksregel, jf. bilag 4, punkt 4.C.6.a og 4.C.6.b.

Når en arkiveringsversion indeholder GML-filer, er det obligatorisk at opmærke elementet *gmlXsd* i indeksfilen *docIndex.xml* med en reference til det lokale skema, der dokumenterer og validerer filen. Værdien i elementet kan kun være enten ”1.xsd” eller navnet på et lokalt skema i *localShared* mappen navngivet ”localSchema” efterfulgt af et fortløbende nummer, samt afsluttet med ”.xsd”, f.eks. localSchema1.xsd.

5.G.4.b Kolonner i arkiveringsversionens øvrige tabeldata, der indeholder informationer om *Dokumentidentifikation* og *Lagringsform*, samt øvrige relevante særlige oplysninger, skal identificeres via elementet *functionalDescription* i *tableIndex.xml*, jf. bilag 6, punkt 6.C.4 - 6.C.6.

Når en arkiveringsversion indeholder GML-filer, er det obligatorisk at opmærke kolonner i tabeldata, der indeholder oplysninger om dokumentidentifikation af og lagringsform for GML-filerne.

Dokumentidentifikationen angives for den kolonne i arkiveringsversionen, som beskriver de digitale dokumenters entydige identifikation i arkiveringsversionens struktur. Dvs. den kolonne, der indeholder de ID'er på op til 12 cifre, som svarer til navnet på de dokumentmapper i arkiveringsversionen, hvor gml-filerne er placeret. Via de øvrige tabeldata, som f.eks. gml-filens navn, kan tabeldata og dokumentidentifikationen således anvendes til at fremsøge relevante gml-dokumenter i arkiveringsversionen.

Denne dokumentreference fra dokumentidentifikation i tabeldata til dokumentmapper med GML-filer er obligatorisk og skal ikke forveksles med de øvrige mulige relationer dokumenteret i lokale features i det lokale xml-skema til en gml-fil og kolonner i tabeldata beskrevet i 5.G.5.m.

5.G.5 Lokale XML-skemaer

Eksempel 1 viser et lokalt skema (1.xsd) med en simpel struktur, hvortil der er opmærket et navnerum for Rigsarkivet ("sa") samt nogle eksempeloplysninger. **(#)** undervejs i vejledningsteksten refererer noterne i venstre side af eksemplet ud for elementet, der korrekt udmønter de enkelte punkter.

5.G.5.a De nødvendige lokale XML-skemaer til dokumentation og validering af GML-filernes indhold skal indgå i arkiveringsversionen.

5.G.5.b Lokale XML-skemaer skal overholde syntaksen for skemaer i henhold til W3C XML Schema Definition Language (XSD) 1.1 Part 1.

5.G.5.c Lokale XML-skemaer skal indkodes i tegnsættet UTF-8, hvilket skal fremgå af XML-deklarationen. Anvendelsen af tegn er begrænset, jf. bilag 5 punkt 5.D.

Se **(1)** i eksempel 1.

5.G.5.d Et lokalt XML-skema placeres i samme mappe, som den GML-fil skemaet validerer, jf. reglerne i 4.G.7. Et lokalt XML-skema kan også fungere globalt for flere GML-filer i en arkiveringsversion, såfremt strukturen for arkiveringsversionens GML-filer er indeholdt i skemaets syntaksregler, og placeres i så tilfælde i undermappen *localShared* til mappen *Schemas*, jf. 4.G.7.a.

5.G.5.e I det lokale XML-skemas rodelement skal følgende navnerum opmærkes: gml, xlink, xsi og den afleverende myndigheds navnerum, jf. figur 5.3. Navnerummet »xxx« skal substitueres af den afleverende myndigheds *URI*.

Figur 5.3 Navnerum i XML-skemaets rodelement

```
<schema
xmlns="http://www.w3. org/2001/XMLSchema"
xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3. org/1999/xlink"
xmlns:xxx="http://www.xxx.dk/gml/xxx"
targetNamespace="http://www.xxx.dk/gml/xxx" elementFormDefault="qualified">
```

(2) Alle GML-filer skal referere et lokalt XML-skema, der opmærker anvendte navnerum og anviser den strukturelle opbygning af GML-filernes rene XML-del.

5.G.5.f Det lokale XML-skema skal indeholde et importelement til navnerummet for GML version 3.1.1.

(3) Det er obligatorisk at importere det specifikke navnerum for GML 3.1.1, der samtidig anviser den konkrete indgang til Open Geospatial Consortiums skemasæt (*gml.xsd*).

5.G.5.g Det lokale XML-skema skal indeholde en generel beskrivelse af GML-filens indhold dokumenteret ved brug af XML-dokumentationselementerne *annotation* og *documentation*. Omfanget og indholdet af dokumentationen aftales nærmere med det modtagende arkiv.

(5) Elementerne *annotation* og *documentation* indeholder på dette overordnede niveau en generel beskrivelse af, hvad GML-filen, som skemaet tilhører, indeholder. Beskrivelsen skal være så udførlig som muligt, så en fremtidig bruger har mulighed for at forstå GML-filens indhold og betydning. F.eks. hvad GML-filen viser på et kort og eventuelle tilknytninger til den resterende arkiveringsversion.

5.G.5.h I et lokalt XML-skema skal der altid være opmærket mindst en lokal GML-feature med tilhørende GML-geometri. Omfanget og indholdet af features aftales nærmere med det modtagende arkiv.

GML-filer bør indeholde alle relevante oplysninger, der er tilgængelige, om det geografiske objekt. **Eksempel 1** indeholder **(10, 12, 13)** tre lokale GML-features med **(14)** tilhørende GML-geometri. For hver lokal GML-feature angives oplysninger om navn og datatype, og om den er obligatorisk og/eller kan optræde flere gange, mens der for den tilhørende GML-geometri refereres til den klasse i GML-biblioteket, hvis beskrevne produktionsregler geometrien skal overholde. Der kan refereres *gml:Point* for punkter, *gml:LineString* for linjer og *gml:surfaceProperty* for arealer.

5.G.5.i Det lokale XML-skemas GML-features og geometrier opmærkes som en sekvens af features med elementet *sequence* og omsluttes i GML-filen af et XML-samleelement, der navngives »GEOMETRI«.

(9) Elementet *sequence* er obligatorisk og angiver, at det efterfølgende indhold har en fast rækkefølge, der gentager sig for hvert samleelement *GEOMETRI* indlejret *gml:featureMember* i selve GML-filen.

5.G.5.j Samleelementet *GEOMETRI* defineres i det lokale XML-skema og nedarver fra klassen *gml:_Feature* opmærket ved brug af XML-attributten *substitutionGroup* samt opmærkes med elementerne *complexType* og *complexContent* og placeres i den afleverende myndigheds navnerum.

(4) Elementet *GEOMETRI* er obligatorisk og skal via attributten *type* placeres i den afleverende myndigheds navnerum som opmærket i rodelementet. **(6)** Endvidere skal det angives, at elementet *GEOMETRI* er en kompleks type, og at **(7)** underelementerne består af et komplekst indhold.

5.G.5.k Det lokale XML-skema skal nedarve egenskaberne fra GML-klassen *gml:AbstractFeatureType*.

(8) Det skal angives, at det efterfølgende indhold nedarver egenskaberne fra klassen *gml:AbstractFeatureType* og skal overholde de for klassen beskrevne produktionsregler.

5.G.5.l Myndighedens lokale GML-features samt geometrier skal beskrives og dokumenteres ved brug af XML-dokumentationselementer *annotation* og *documentation*. Omfanget og indholdet af dokumentationen aftales nærmere med det modtagende arkiv.

(11) For hver lokal GML-feature skal der indsættes en beskrivelse. Beskrivelsen skal omkranses af elementerne *annotation* og *documentation*. Beskrivelserne skal være så udførlige som muligt, således at en fremtidig bruger har mulighed for at forstå indhold og betydning af de lokale GML-features.

5.G.5.m Der skal til den enkelte *gml:featureMember* i GML-filen etableres en entydig sammenkædning til arkiveringsversionens øvrige tabeldata, hvis en sådan eksisterer. Sammenkædningen mellem tabeldata og GML-instansen skal dokumenteres i det lokale XML-skema og i *tableIndex.xml* for de lokale features og kolonner, som udgør referencen. Mulige løsningsmodeller for etablering af referencer aftales nærmere med det modtagende arkiv.

(11) Det er obligatorisk, at der findes en nøgle, der entydigt identificerer det enkelte *gml:featureMember* og at denne beskrives i det lokale XML-skema i *annotation* og *documentation* elementer for den eller de GML feature(s), der udgør nøglen. Eksisterende nøgler i driftssystemet kan med fordel anvendes.

Den entydige nøgle kan referere tabeldata i arkiveringsversionen. Hvis en sådan sammenkædning til arkiveringsversionens øvrige tabeldata findes, f.eks. mellem et geografisk objekt i gml-filen og en sag i tabeldata, skal den dokumenteres i det lokale XML-skema i beskrivelsen af den lokale feature der udgør nøglen. Endvidere skal denne kobling beskrives i feltinformationen i tableIndex.xml, for det felt, som GML-filens unikke nøgle refererer til.

Eksempel 1: Lokalt XML-skema (1.xsd)

```
(1)  <?xml version="1.0" encoding="UTF-8"?>

(2)  <schema
       xmlns="http://www.w3.org/2001/XMLSchema"
       xmlns:gml="http://www.opengis.net/gml"
       xmlns:xlink="http://www.w3.org/1999/xlink"
       xmlns:xxx="http://www.sa.dk/gml/sa"
       targetNamespace="http://www.sa.dk/gml/sa" elementFormDefault="qualified">

(3)  <import namespace="http://www.opengis.net/gml"
       schemaLocation="http://schemas.opengis.net/gml/3.1.1/base/gml.xsd"/>

(4)  <element name="GEOMETRI" type="sa:GEOMETRIType" substitutionGroup="gml:_Feature"/>

(5)  <annotation>
      <documentation>Geodata i GML-filen er linjer som angiver placering af vejstrækninger i Danmark.
      GML-filens geodata kan kobles til sager over GML-filens unikke nøgle SagsNr som matcher
      sagsnummeret i feltet Sagsnummer i tabellen Sag i den øvrige del af
      arkiveringsversionen.</documentation>
     </annotation>

(6)  <complexType name="GEOMETRIType">
(7)  <complexContent>
(8)  <extension base="gml:AbstractFeatureType">
(9)  <sequence>

(10) <element name="SagsNr" type="string" minOccurs="1" maxOccurs="1">
(11)   <annotation>
        <documentation>Unikt sagsnummer på den sag, som geografien
        (linjen/vejstrækningen/gml:LineString) er tilknyttet. SagsNr kan kobles til feltet ”Sagsnummer” i
        tabellen ”Sag”. SagsNr bruges således til at knytte geografien til informationer i den øvrige del af
        arkiveringsversionen.</documentation>
       </annotation>
      </element>

(12) <element name="VejNavn" type="string" minOccurs="1" maxOccurs="1">
       <annotation>
        <documentation>Vejens navn.</documentation>
       </annotation>
      </element>

(13) <element name="KommuneNummer" type="integer" minOccurs="1" maxOccurs="1">
       <annotation>
        <documentation>Kommunenummer for den kommune som vejen er beliggende i.</documentation>
       </annotation>
      </element>

(14) <element ref="gml:LineString" minOccurs="1">
       <annotation>
        <documentation>Linje som angiver geografisk placering af en vejstrækning.</documentation>
       </annotation>
      </element>

      </sequence>
      </extension>
      </complexContent>
      </complexType>
      </schema>
```

5.G.6 GML filer

Eksempel 2 angiver en GML-fil tilhørende det lokale skema i **eksempel 1**, der anvender det i skemaet opmærkede navnerum og indeholder eksempeloplysninger om koordinat- og referencesystem samt enkelte instanser af den i skemaet opmærkede sekvens. **(#)** undervejs i vejledningsteksten refererer noterne i venstre side af eksemplet ud for elementet, der korrekt udmønter de enkelte punkter.

5.G.6.a Alle GML-filer i en arkiveringsversion skal overholde de gældende syntaksregler for henholdsvis XML version 1.0 og GML version 3.1.1.

5.G.6.b GML-filer skal indkodes i tegnsættet UTF-8, hvilket skal fremgå af XML-deklarationen. Anvendelsen af tegn er begrænset, jf. bilag 5 punkt 5.D.

(1) XML-deklaration er obligatorisk og angiver, at det er et XML-dokument. XML-deklarationen skal altid indlede GML-filen uden foranstillede tegn, og man skal særligt sikre sig, at GML-instansen rent faktisk bliver indkodet i tegnsættet UTF-8.

5.G.6.c Geometrier og features fra samme kortmateriale/tabel skal så vidt muligt holdes samlet inden for den samme GML-fil. Afvigelser herfra aftales nærmere med det modtagende arkiv.

Se vejledningsteksten til punkt 5.G.3.

5.G.6.d Som rodelement skal anvendes et *gml:FeatureCollection* element, jf. figur 5.4.

Figur 5.4 Navnerum i GML-filens rodelement

```
<gml:FeatureCollection
xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3. org/1999/xlink"
xmlns:xsi="http://www.w3. org/2001/XMLSchema-instance"
xmlns:xxx="http://www.xxx.dk/gml/xxx"
xsi:schemaLocation="http://www.xxx.dk/gml/xxx . /1.xsd">
```

5.G.6.e I rodelementet skal følgende navnerum opmærkes: gml, xlink, xsi og den afleverende myndigheds navnerum, jf. figur 5.4. Navnerummet »xxx« skal substitueres af den afleverende myndigheds *URI.*

5.G.6.f I rodelementet skal angives en lokal skemareference *xsi:schemaLocation* med angivelse af en relativ sti for placeringen af det lokale skema, jf. figur 5.4 og 5.G.5.d.

(2) Alle GML-filer skal anvende det obligatoriske rodelement gml:FeatureCollection, der opmærker de anvendte navnerum og refererer med attributten *xsi:schemaLocation* den konkrete placering af det lokale XML-skema, der dokumenterer og validerer GML-filen. I eksemplet er skemaet placeret i samme mappe som GML-filen. Hvis det lokale XML-skema derimod placeres i mappen *localShared* angives placeringen som "../../../Schemas/localShared/localschema1.xsd", hvis skemafilen navngives localschema1.xsd.

5.G.6.g Alle GML-filer skal opmærkes med et koordinat- og referencesystem. »EPSG:25832« alias »ETRF89 / ETRS89 UTM zone 32N« skal anvendes som opmærkning, hvis de afleverede geodata er placeret i Jylland, på Fyn eller Sjælland samt farvandet heromkring, dvs. inden for punkterne angivet i *lowerCorner* og *upperCorner*, jf. punkt 5.G.6.i. Er dette ikke tilfældet, kan anvendes et andet koordinat- og referencesystem efter aftale med det modtagende arkiv.

5.G.6.h Koordinat- og referencesystemet angives ved brug af GML-elementet *gml:boundedBy* med underelementerne *Envelope* og *lower-* og *upperCorner*. Denne opmærkning er gældende globalt for hele GML-filen og skal ikke angives lokalt i de enkelte *gml:featureMember* elementer.

5.G.6.i Geografierne i GML-filen skal placeres inden for punkterne *lowerCorner* (212481.60 6019669.40) og *upperCorner* (961440.75 6510422.51). En anden eller mere præcis angivelse af disse yderpunkter kan anvendes efter aftale med det modtagende arkiv.

5.G.6.j Geografiens dimension skal opmærkes ved brug af GML-attributten *srsDimension* i GML-elementet "*Envelope*" sammen med GML-attributten *srsName*, der angiver kortmaterialets koordinat- og referencesystem.

(3) Elementet *gml:boundedBy* og dets underelementer er obligatoriske og anvendes til at angive, hvilket område, de opmærkede geografier befinder sig inden for, mens *gml:Envelope* ved hjælp af attributterne *srsName* og *srsDimension* angiver hhv. koordinat- og referencesystem (datum) og antal dimensioner. Et datum er et sæt af referencepunkter på jordens overflade, der anvendes til en korrekt placering af GML-filens geografiske objekter, mens antal dimensioner angiver, om der i forventes to eller flere – f.eks. en højdeangivelse – dimensioner pr. koordinatsæt.

5.G.6.k En GML-fil skal indeholde mindst et *gml:featureMember* element. I *gml:featureMember* placeres samleelementet *GEOMETRI,* der indeholder GML-filens lokale features og geometrier. Elementet *GEOMETRI* skal være inkluderet i den afleverende myndigheds navnerum.

(4) Elementet *gml:featureMember* er det overordnede element til at samle lokale GML-features og repræsenterer en instans af den i skemaet opmærkede sekvens. Inden for rodelementet kan der opmærkes et vilkårligt antal, dog altid mindst et, gml:featureMember.

(5) Elementet GEOMETRI er obligatorisk samleelement for alle lokale GML-features, der placeres i den afleverende myndigheds navnerum.

(6, 7, 8) Elementerne opmærker de i det lokale XML-skema angivne lokale GML-features, hhv. SagsNr, VejNavn og KommuneNummer og placeres i den afleverende myndigheds navnerum.

(9) Elementet indeholder den i det lokale XML-skema angivne GML-geometri. I eksemplet er det gml:LineString, der refererer til en klasse i GML-biblioteket, hvis produktionsregler geometrien skal overholde. Der skal ved opmærkning af linjer anvendes underelementet *gml:posList*, der angiver, at geometrien består af en liste af mindst to koordinatsæt.

Hvis geografien er et areal refereres klassen *gml:surfaceProperty*, og der skal ved opmærkning anvendes underelementerne *gml:Surface*, *gml:patches*, *gml:PolygonPatch*, *gml:exterior*, *gml:LinearRing* og

gml:posList, der blandt andet angiver, at geometrien består af en liste af minimum fire koordinatsæt, hvor første og sidste koordinatsæt er sammenfaldende. For at opmærke arealets eventuelle indre grænse(r), skal elementet *gml:interior* anvendes efter afslutning af *gml:exterior*. Der kan være én *gml:exterior*, men mange *gml:interior*.

Eksempel på angivelse af areal med indre og ydre grænser

```
<gml:surfaceProperty>
<gml:Surface>
  <gml:patches>
   <gml:PolygonPatch>
    <gml:exterior>
     <gml:LinearRing>
      <gml:posList>ydre grænse for areal</gml:posList>
     </gml:LinearRing>
    </gml:exterior>
    <gml:interior>
     <gml:LinearRing>
      <gml:posList>indre grænse for areal</gml:posList>
     </gml:LinearRing>
   </gml:interior>
   </gml:PolygonPatch>
  </gml:patches>
</gml:Surface>
</gml:surfaceProperty>
```

Eksempel 2: GML-fil (1.gml) tilhørende det lokale skema i eksempel 1

```
(1)  <?xml version="1.0" encoding="UTF-8"?>

(2)  <gml:FeatureCollection
       xmlns:gml="http://www.opengis.net/gml"
       xmlns:xlink="http://www.w3.org/1999/xlink"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xmlns:sa="http://www.sa.dk/gml/sa"
       xsi:schemaLocation="http://www.sa.dk/gml/sa ./1.xsd">

(3)  <gml:boundedBy>
       <gml:Envelope srsName="EPSG:25832" srsDimension="2">
         <gml:lowerCorner>212481.60 6019669.40</gml:lowerCorner>
         <gml:upperCorner>961440.75 6510422.51</gml:upperCorner>
       </gml:Envelope>
     </gml:boundedBy>

(4)  <gml:featureMember>
(5)    <sa:GEOMETRI>
(6)      <sa:SagsNr>2001-01000</sa:SagsNr>
(7)      <sa:VejNavn>Kristian Erslevs Gade</sa:VejNavn>
(8)      <sa:KommuneNummer>101</sa:KommuneNummer>
(9)      <gml:LineString>
           <gml:posList>724297.34 6174822.74 724372.67 6174738.59</gml:posList>
         </gml:LineString>
       </sa:GEOMETRI>
     </gml:featureMember>

     <gml:featureMember>
       <sa:GEOMETRI>
         <sa:SagsNr>2001-07501</sa:SagsNr>
         <sa:VejNavn>Kalvebod Brygge</sa:VejNavn>
         <sa:KommuneNummer>101</sa:KommuneNummer>
         <gml:LineString>
           <gml:posList>724944.42 6175249.02 724881.34 6175219.90 723779.41
6174147.20</gml:posList>
         </gml:LineString>
       </sa:GEOMETRI>
     </gml:featureMember>

     <!-- angivelse af eventuelle yderligere instanser af den i skemaet opmærkede sekvens -->

     </gml:FeatureCollection>
```

5.H. Komprimering

5.H.1 En arkiveringsversion må ikke være komprimeret ud over den kompression, som er anvist for eller følger af de dokumentformater, der kræves til brug for arkiveringsversionen, jf. 5.E og 5.F.

En arkiveringsversion må ikke komprimeres ud over den kompression, som er en del af dokumentformaterne. Det betyder f.eks., at arkiveringsversionen ikke må zip-komprimeres.

5.I. Optimering

5.I.1 Det modtagende arkiv kan i konkrete tilfælde træffe afgørelse om nedbringelse af dokumenters pladsforbrug, f.eks. ved at anvise anvendelse af optimal bitdybde.

Som eksempel kan nævnes dokumenter gemt som TIFF 24 LZW, hvor alle eller hovedparten af siderne består af ren tekst. Sådanne dokumenter kan optimeres ved at benytte grp. 4 for de sider, som er ren tekst, og LZW-kompression af varierende bitdybde for de sider, som indeholder farve.

Myndigheder, som har systemer med store dokumentmængder, skal være særligt opmærksomme på denne problematik. Det gælder f.eks. ved aflevering af ESDH-koncernløsninger. Myndighederne skal således i forbindelse med produktion af arkiveringsversionen sikre sig, at de konverterede dokumenter ikke kommer til at fylde unødigt meget i arkiveringsversionen.

5.J. Ingen forringelse

5.J.1 I forbindelse med generering af arkiveringsversionen må der ikke ske en kvalitetsmæssig forringelse af dokumenterne, herunder lyd og video, ud over hvad der måtte være en konsekvens af det krævede format til brug for arkiveringsversionen eller de anvisninger, som i øvrigt gives af det modtagende arkiv.

Det er vigtigt, at dokumenter, lyd og video leveres i den bedst mulige kvalitet, så der må ikke ske forringelser udover hvad der måtte være konsekvensen af konverteringen til arkivformatet.

Bilag 6 – Information om arkiveringsversionen

Informationen om arkiveringsversionen består af følgende elementer:

- **Arkivbeskrivelsesfil, jf. 6.A.**
- **Kontekstdokumentation, jf. 6.B.**
- **Data om arkiveringsversionens tabeller (tabelindeks), jf. 6.C.**
- **Evt. SQL-forespørgsler, jf. 6.D.**

Det er en vigtig forudsætning for den fremtidige anvendelse af data, at arkiveringsversionen indeholder information til dokumentation af data, der er overført fra it-systemet eller forskningsdatasættet.

For det første skal arkiveringsversionen indeholde information om datas oprindelse, brug og administrative eller forskningsmæssige sammenhæng samt om it-systemet. Denne information skal findes i arkivbeskrivelsesfilen og kontekstdokumentationen.

Formålet med **arkivbeskrivelsesfilen** er at give brugeren et struktureret og nemt tilgængeligt overblik over datas oprindelse, brug og administrative/forskningsmæssige sammenhæng. Arkivbeskrivelsesfilen udarbejdes i forbindelse med produktion af arkiveringsversionen. Den er ikke omfattet af arkivlovens tilgængelighedsfrister og kan straks offentliggøres som søgemiddel til arkivernes samlinger.

Formålet med **kontekstdokumentationen** er at give brugeren en dybtgående viden om data ved hjælp af dokumenter, der er udarbejdet i forbindelse med anskaffelse og brug af it-systemet eller undervejs i forskningsprocessen. Hvis myndigheden ikke har en tilstrækkelig dokumentation, skal den udarbejdes i forbindelse med produktion af arkiveringsversionen. I henhold til arkivlovgivningen er kontekstdokumentationen som udgangspunkt belagt med en tilgængelighedsfrist på 20 år. Det betyder, at såfremt en arkivbruger ønsker at gøre brug af materialet, før der er gået 20 år, skal den afleverende myndighed give sit samtykke, før materialet kan udleveres, jf. § 33 i arkivloven.

For det andet skal arkiveringsversionen indeholde information om datas struktur og betydning. Denne information skal findes i tabelindeks. **Tabelindeks** er udformet med henblik på, at det skal være muligt automatisk at trække hovedparten af oplysningerne ud af it-systemets database. Dette gælder også sammenhængen mellem tabeller i form af views. Det er ofte nødvendigt at udarbejde SQL-forespørgsler til dokumentation af visse sammenhænge i arkiveringsversionen. Forespørgslerne skal kunne køres direkte på arkiveringsversionens tabeller.

6.A. Arkivbeskrivelsesfil

6.A.1 Enhver arkiveringsversion skal indeholde en arkivbeskrivelsesfil med angivelse af oplysninger iht. figur 6.1.

6.A.2 Arkivbeskrivelsesfilen navngives archiveIndex.xml og skal overholde det tilhørende skema, jf. bilag 8.

6.A.3 Indholdet af arkivbeskrivelsesfilen fastlægges efter drøftelse mellem den afleverende myndighed og det modtagende arkiv.

Figur 6.1 Oplysninger i archiveIndex.xml, administrative data

Vejledning i denne figur er tilpasset administrative data fra IT-systemer og dokumentsamlinger, der afleveres efter bilag 1-8.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveInforma-tionPackageID	Arkiverings-versionsID	Entydigt ID som tildeles arkiveringsversioner af Rigsarkivet, og som består af arkivkode og løbenummer.	Arkiverings-versionsID, som defineret i denne bekendtgørelse	1	Ja	Oplyses af det modtagende arkiv, f.eks. AVID.SA.12345. Består af præfixet AVID (punktum), en kode på 2-4 bogstaver for det modtagende arkiv (punktum) og et løbenummer.
archiveInforma-tionPackageID-Previous	Tidligere aflevering	Entydigt arkiveringsversionsID på tidligere aflevering fra samme it-system	Arkiverings-versionsID, som defineret i denne eller tidligere bekendtgørelse	0-1	Nej	Oplyses af det modtagende arkiv, f.eks. AVID.SA.12322.
archivePeriodStart	Arkiverings-version startdato	Startdato for de afleverede data	År, år-måned, eller år-måned-dag	1	Ja	Datering skal være så præcis som mulig, f.eks. 2006-01-01. Datering kan være et årstal, f.eks. 1867, hvis der er foretaget indtastning af gamle data eller indskanning af gamle dokumenter.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archivePeriodEnd	**Arkiverings-version slutdato**	**Slutdato for de afleverede data**	**År, år-måned, eller år-måned-dag**	**1**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2010-12-31.
documentPeriodeStart	**Startdato for dokumenter**	**Startdato for digitale dokumenter i arkiveringsversionen**	**År, år-måned eller år-måned-dag**	**0-1**	**Ja, hvis arkiverings-versionen indeholder digitale dokumenter**	Nogle systemer har ikke dokumenter for hele den periode de har været i drift. Det kan derfor være nødvendigt at angive hvilken periode dokumenterne dækker, da den kan være forskellig fra perioden for øvrige data.
documentPeriodeEnd	**Slutdato for dokumenter**	**Slutdato for digitale dokumenter i arkiveringsversionen**	**År, år-måned eller år-måned-dag**	**0-1**	**Ja, hvis arkiverings-versionen indeholder digitale dokumenter**	Nogle systemer har ikke dokumenter for hele den periode de har været i drift. Det kan derfor være nødvendigt at angive hvilken periode dokumenterne dækker, da den kan være forskellig fra perioden for øvrige data.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveInforma-tionPacketType	**Slutaflevering**	**Angivelse af, om arkiveringsversionen er en slutaflevering**	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis it-systemet på tidspunktet for produktion af arkiveringsversionen er lukket for opdateringer. Har ligeledes værdien *true*, hvis Rigsarkivet på tidspunktet for produktion af arkiveringsversionen har truffet bestemmelse om, at der ikke fremover skal afleveres arkiveringsversioner af data. Har værdien *false* i øvrige tilfælde.
creatorName	**Arkivskaber**	**Den eller de myndigheder, institutioner, organisationer eller personer, der har skabt data. Kan også være organisatoriske enheder inden for en myndighed eller andre bidragydere**	**Fri tekst**	**1-m**	**Ja**	Det aftales med det modtagende arkiv, hvilke arkivskabere der skal anføres.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
creationPeriodStart	**Arkivskaber startdato**	**For hver arkivskaber angives, hvornår arkivskaber er begyndt at tilføje data**	**År, år-måned, eller år-måned-dag**	**1-m**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2008-06-01.
creationPeriodEnd	**Arkivskaber slutdato**	**For hver arkivskaber angives, hvornår arkivskaber er ophørt med at tilføje data**	**År, år-måned, eller år-måned-dag**	**1-m**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2010-12-31.
archiveType	**Arkiverings-versions-periodetype**	**Angivelse af, om der er tale om en afsluttet arkivperiode (årgang)**	**Boolsk værdi**	**1**	**Ja**	Oplyses af det modtagende arkiv. Har værdien *true*, hvis der er tale om en afsluttet arkivperiode/årgang. Det kan f.eks. være et ESDH-system, hvor der er foretaget periodeskift, eller en arkiveringsversion, der indeholder statistiske data for ét år. Data og dokumenter fra en afsluttet periode skal ikke afleveres igen. Har værdien *false*, hvis der er tale om et øjebliksbillede. Det kan f.eks. være CPR-registret, hvor der løbende akkumuleres data. De afleverede data vil også indgå i næste arkiveringsversion.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveTypsClosedFiles	**Arkiveringsversion med lukkede sager**	**Angivelse af, om der for it-systemer med dokumenter er tale om et øjebliksbillede af metadata, men kun dokumenter fra lukkede sager**	**Boolsk værdi**	**1**	**Ja, hvis arkiverings-versionen indeholder digitale dokumenter**	Oplysning til kommende brugere om at arkiveringsversionen kun indeholde lukkede sager og der således kan være sager fra samme periode, som ikke er afleveret, da de har været åbne på afleveringstidspunktet.
systemName	**Systemnavn**	**Det officielle navn på it-system eller datakilde, hvor alle forkortelser er opløst**	**Fritekst**	**1**	**Ja**	Et officielt navn kan f.eks. være Bygnings- og Boligregistret eller Det Centrale Personregister. Det aftales med det modtagende arkiv, hvilket navn der skal anvendes, hvis arkiveringsversionen indeholder data fra flere it-systemer, eller hvis data fra ét it-system deles på flere arkiveringsversioner.
alternativeName	**Alternativt systemnavn**	**Alternative navne på it-system eller datakilde**	**Fritekst**	**0-m**	**Nej**	Oplysningen er ikke obligatorisk, men kan f.eks. anvendes, hvis it-systemet i daglig tale benævnes med et kaldenavn eller en forkortelse. Det kan f.eks. være BBR eller CPR.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
systemPurpose	**Systemformål**	**Beskrivelse af formålet med at oprette og anvende it-systemet eller indsamle data**	**Fritekst**	**1**	**Ja**	Teksten skal besvare spørgsmålet om, hvorfor der er brugt ressourcer på at oprette og anvende it-systemet f.eks. er formålet med Landspatientregistret bl.a. at: ”danne grundlag for Sundhedsstyrelsens løbende sygehusstatistik og forsyne de myndigheder, der er ansvarlig for sygehusplanlægningen med det nødvendige beslutningsgrundlag samt at indgå som grundmateriale i forbindelse med overvågning af hyppigheden af forskellige sygdomme og behandlinger". Dokumenter med detaljerede oplysninger indgår i kontekstdokumentation.

systemContent	**Systemindhold**	**Beskrivelse af den centrale population og centrale variable i data**	**Fritekst**	**1**	**Ja**	Teksten skal besvare spørgsmålet om, hvem eller hvad der primært registreres hvilke data om. F.eks. et tilskudsadministrativt system, hvor den centrale population er ansøgere til og modtagere af tilskud til miljøforbedringer. De centrale variable er modtageridentifikation, projektbeskrivelse, afrapportering samt beløb. Det vil være en hjælp til genanvendelse af data i arkiveringsversionen, hvis der henvises til de tabeller, hvor de centrale variable findes. Hvis der i et stort system er flere væsentlige populationer, kan de nævnes. F.eks. er den centrale population i CPR alle personer, der siden 1968 bor eller har boet i Danmark/været tilmeldt en dansk kommune. Men herudover har CPR som en central population det officielle vejnavneregister, hvor centrale variable er vejnavn, beliggenhed og tilknytning til forskellige administrative inddelinger.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
						Dokumenter med detaljerede oplysninger indgår i kontekstdokumentation.
regionNum	**Regionsnumre**	**Angivelse af, om der i systematisk form er registreret regionsnumre i data**	**Boolsk værdi**	**1**	**Ja**	For hver af de angivne identifikatorer skal det oplyses, om den er registreret i it-systemet. Har værdien *true*, hvis den pågældende identifikator er registreret systematisk. Har værdien *false*, hvis den pågældende identifikator ikke er registreret systematisk. Hvis der i arkiveringsversionen findes andre fagspecifikke identifikatorer, der anvendes i flere it-systemer, f.eks. et geografisk referencesystem eller Dansk branchekode, kan det angives i forbindelse med oplysninger om central population og centrale variable (systemindhold/systemContent).
komNum	**Kommune-numre**	**Angivelse af, om der i systematisk form er registreret kommunenumre i data**	**Boolsk værdi**	**1**	**Ja**	
cprNum	**CPR-numre**	**Angivelse af, om der i systematisk form er registreret CPR-numre i data**	**Boolsk værdi**	**1**	**Ja**	
cvrNum	**CVR-numre**	**Angivelse af, om der i systematisk form er registreret CVR-numre i data**	**Boolsk værdi**	**1**	**Ja**	
matrikNum	**Matrikelnumre**	**Angivelse af, om der i systematisk form er registreret matrikelnumre i data**	**Boolsk værdi**	**1**	**Ja**	

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
bbrNum	**BBR-numre**	**Angivelse af, om der i systematisk form er registreret BBR-numre i data**	**Boolsk værdi**	**1**	**Ja**	
whoSygKod	**WHOs sygdomskoder**	**Angivelse af, om der i systematisk form er registreret WHO-sygdomskoder i data**	**Boolsk værdi**	**1**	**Ja**	
sourceName	**Datakilde**	**Det officielle navn på andre datakilder, som har leveret data til datasættet (opslag, overførsel, samkøring osv.)**	**Fritekst**	**0-m**	**Nej**	Det modtagende arkiv træffer aftale med den afleverende myndighed om, hvorvidt denne oplysning skal indgå. Teksten skal besvare spørgsmålet om, fra hvilke andre it-systemer data stammer. I kontekstdokumentationen kan indgå dokumenter med detaljerede oplysninger, f.eks. om hvilke data (kolonner) der er overført fra andre systemer, og hvordan overførsel er sket.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
userName	**Databrugere**	**Andre it-systemer som har brugt data fra it-systemet (opslag, overførsel, samkøring osv.)**	**Fritekst**	**0-m**	**Nej**	Det modtagende arkiv træffer aftale med den afleverende myndighed om, hvorvidt denne oplysning skal indgå. Teksten skal besvare spørgsmålet om, hvorvidt data anvendes i andre it-systemer. I kontekstdokumentationen kan indgå dokumenter med detaljerede oplysninger, f.eks. om hvilke data (kolonner) der er anvendt i andre systemer, og hvordan data er anvendt.
predecessorName	**Forgænger-systemer**	**Systemer, der tidligere har varetaget samme funktion**	**Fritekst**	**0-m**	**Nej**	Det modtagende arkiv træffer aftale med den afleverende myndighed om, hvorvidt denne oplysning skal indgå. Et forgængersystem til et ESDH-system kan f.eks. være en elektronisk journal samt et papirarkiv. Forgængersystemet til CPR-registret er folkeregisterkortene.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
formVersion	**FORM-version**	**Angivelse af, hvilken version af FORM (den fællesoffentlige forretningsreference model) nedenstående kategorisering er hentet fra**	**Gyldige FORM-versionsnumre**	**0-1**	**Ja, for it-system-er fra offentlige myndigheder**	F.eks. version 2.0. På Rigsarkivets hjemmeside www.sa.dk er der link til forretningsreferencemodellen FORM.
formClass	**FORM-klassifikation**	**Kategorisering af arkiveringsversionen iht. FORM**	**Enhver gyldig FORM-reference iht. den angivne version af FORM**	**0-m**	**Ja, for it-system-er fra offentlige myndigheder**	Et eller flere FORM-numre fra FORM's opgavekatalog. F.eks. kan der for en fiktiv arkiveringsversion med data vedrørende FødevareErhvervs ressortområde angives de 5 FORM-numre: • 34.25.15.10 • 34.25.05.25 • 34.25.15.12 • 34.30.05.05 • 34.35.05.05

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
formClassText	**FORM-klassifikation klartekst**	**FORM-klassifikation i klartekst**	**Fri tekst**	**0-m**	**Ja, for it-system-er fra offentlige myndigheder**	Det eller de angivne FORM-numre i klartekst jf. den anvendte version af FORM opgavekataloget. F.eks. vil der til de ovenfor nævnte FORM-numre skulle angives nedenstående 5 tekster: • Foderstoffer • Landdistrikter • Genetisk modificerede organismer (GMO) i planteavl og foderproduktion • Fødevaresikkerhed • Engros- og detailhandel
containsDigital-Documents	**Digitale dokumenter indeholdt**	**Angivelse af, om der er digitale dokumenter i arkiveringsversionen, udover kontekstdokumentation**	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis det er en arkiveringsversion med digitale dokumenter. Har værdien *false*, hvis det er en arkiveringsversion uden digitale dokumenter (dog vil der være dokumenter med kontekstdokumentation).
containsGeodata	**GML-filer indeholdt**	**Angivelse af, om der er GML-filer i arkiveringsversionen**	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis det er en arkiveringsversion med struktureret GML.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
containsResearchData	Forskningsdata indeholdt	Angivelse af, om arkiveringsversionen indeholder forskningsdata	Boolsk værdi	1	Ja	Har værdien *"true",* hvis der er tale om forskningsdata i en arkiveringsversion Har også værdien "true", hvis forskningsdata er journaliseret på en sag i det afleverede ESDH-system
researchSIP	Forskningsdata afleveret iht. bilag 9	Angivelse af, om data er afleveret iht. bilag 9	Boolsk værdi	1	Ja	Har værdien *"true",* hvis der er tale om visse typer af forskningsdata som afleveres som en afleveringspakke i iht. bilag 9. Har værdien *"false",* hvis der er tale om data som afleveres som en arkiveringsversion iht. bilag 1-8.
documentsDisposal	Kassation	Angivelse af, om der er foretaget systematisk kassation af dokumenter inden eller i forbindelse med produktion af arkiveringsversionen	Boolsk værdi	1	Ja	Har værdien *"true",* hvis der er foretaget kassation i henhold til fastsatte bestemmelser om kassation f.eks. en bevarings- og kassations bestemmelse for et statslige system eller via bk-koder i KLE for kommunale systemer. Omfanget af kassationen beskrives i kontekstdokumentationen

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
searchRelatedOther Records	**Søgemiddel til andre sager eller dokumenter**	**Angivelse af, om arkiveringsversionen er et nødvendigt søgemiddel til papirsager/dokument er eller sager/dokumenter i et andet it-system**	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis data i arkiveringsversionen er nødvendige for at kunne fremfinde sager/dokumenter på papir eller i et andet it-system. Det mest udbredte eksempel er et elektronisk journalsystem, der er nødvendig for at kunne fremfinde sager på papir. Har værdien *false* i øvrige tilfælde.
relatedRecordsName	**Adresserede arkivalier**	**Henvisning til de arkivalier, som arkiveringsversionen er søgemiddel til**	**Fritekst**	**0-m**	**Ja, hvis der er svaret ja på fore-gående spørgs-mål**	Feltet udfyldes kun, hvis Søgemiddel til andre sager eller dokumenter / searchRelatedOtherRecord s har værdien *true*. Et eksempel kan være ”Søgemiddel til tilskudsadministrative papirsager vedr. miljøforbedring.” Bemærk, at hvis arkiveringsversionen er søgemiddel til flere arkivaliegrupper eller it-systemer, skal disse angives enkeltvis (ikke som en lang tekst).

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
systemFileConcept	**Eksistens af sagsbegreb i it-systemet**	**Angivelse af, at it-systemet har et sagsbegreb, forstået som en registreret sammenhæng mellem sagligt sammenhørende dokumenter**	**Boolsk værdi. Kan kun være *true*, hvis der er digitale dokumenter i it-systemet**	**1**	**Ja**	Har værdien *true*, når der er registreret en saglig sammenhæng mellem digitale dokumenter. Har værdien *false* i alle andre tilfælde, også når der registreret en saglig sammenhæng mellem papirdokumenter, således at værdien vil være *false* for et elektronisk journalsystem.
multipleData-Collection	**SOA arkitektur**	**Angivelse af, om it-systemet er sammensat af data og eventuelt dokumenter fra flere forskellige it-systemer i en serviceorienteret arkitektur**	**Boolsk værdi**	**1**	**Ja**	Har f.eks. værdien *true* ved et 'virtuelt' system, der på brugerniveau fremstår som et samlet system til løsning af en eller flere forvaltningsmæssige opgaver, men hvor data og evt. dokumenter kommer fra og lagres i flere forskellige systemer. Hvis værdien er *true*, skal der være en tekst om sammensætning i Datakilde / sourceName og Databruger / userName. Dokumenter med detaljerede oplysninger skal indgå i kontekstdokumentation.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
personalData-RestrictedInfo	**Persondata**	**Angivelse af, om der i arkiveringsversionen findes følsomme personoplysninger iht. databeskyttelsesforordningen**	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis der i arkiveringsversionen findes følsomme personoplysninger (oplysninger om racemæssig eller etnisk baggrund, politisk, religiøs eller filosofisk overbevisning, fagforeningsmæssige tilhørsforhold, oplysninger om helbredsmæssige og seksuelle forhold samt oplysninger om strafbare forhold og væsentlige sociale problemer).
otherAccessTypeRestrictions	**Fristforlængende oplysninger**	**Angivelse af, om der i arkiveringsversionen findes oplysninger, der kan betinge længere tilgængelighedsfrist i øvrigt**	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis tilgængelighedsfristen er forlænget. Hvis den afleverende myndighed ønsker forlængelse af tilgængelighedsfristen, skal det i forbindelse med aflevering drøftes med det modtagende arkiv, jf. lovbekendtgørelse nr. 1035 af 21. august 2007 (Arkivloven), § 27.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveApproval	**Godkendelses-arkiv**	**Angivelse af, hvilket offentligt arkiv, der godkender arkiveringsversionen**	**Identifikation af det pågældende arkiv (2-4 tegn)**	**1**	**Ja**	Oplyses af det modtagende arkiv.
archiveRestrictions	**Tilgængelig-hedsbegræns-ninger**	**Angivelse af nærmere bestemmelser for adgang til materialet. Feltet kan anvendes efter det modtagende arkivs nærmere retningslinjer**	**Fritekst**	**0-1**	**Nej**	Det modtagende arkiv træffer aftale med den afleverende myndighed om, hvorvidt og med hvilke oplysninger dette felt skal udfyldes.

Figur 6.1 Oplysninger i archiveIndex.xml, forskningsdata

Vejledning i denne figur er tilpasset forskningsdata, der afleveres efter bilag 1-8.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveInforma-tionPackageID	**Arkiverings-versionsID**	Entydigt ID som tildeles arkiveringsversioner af Rigsarkivet, og som består af arkivkode og løbenummer.	**Arkiverings-versionsID, som defineret i denne bekendtgørelse**	**1**	**Ja**	Oplyses af det modtagende arkiv, f.eks. AVID.SA.12345 Består af præfixet AVID (punktum), en kode på 2-4 bogstaver for det modtagende arkiv (punktum) og et løbenummer
archiveInforma-tionPackageID-Previous	**Tidligere aflevering**	Entydigt arkiveringsversionsID på tidligere aflevering fra samme forskningsprojekt.	**Arkiverings-versionsID, som defineret i denne eller tidligere bekendtgørelse**	**0-1**	**Nej**	Oplyses af det modtagende arkiv, f.eks. AVID.SA.12322 ArkiveringsversionsID på tidligere afleveringer af forskningsprojektet med kontinuerlig dataindsamling (f.eks. kohorter og dataindsamling i forbindelse med overvågning af fænomener).
archivePeriodStart	**Arkiverings-version startdato**	Startdato for indsamlingen af de afleverede data.	**År, år-måned, eller år-måned-dag**	**1**	**Ja**	Angiver hvornår dataindsamlingen er påbegyndt. Dateringen skal angives så præcist som muligt, f.eks. 2006-01-01.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archivePeriodEnd	**Arkiverings-version slutdato**	Slutdato for indsamling af de afleverede data.	**År, år-måned, eller år-måned-dag**	**1**	**Ja**	Angiver hvornår dataindsamlingen er afsluttet. Dateringen skal angives så præcist som muligt, f.eks. 2010-12-31.
documentPeriodeStart	**Startdato for dokumenter**	Startdato for digitale dokumenter i arkiveringsversionen	**År, år-måned eller år-måned-dag**	**0-1**	**Ja, hvis arkiverings-versionen indeholder digitale dokumenter**	Angiver startdato for det først daterede digitale dokument i arkiveringsversionen, f.eks. datoen for det første interview.
documentPeriodeEnd	**Slutdato for dokumenter**	Slutdato for digitale dokumenter i arkiveringsversionen	**År, år-måned eller år-måned-dag**	**0-1**	**Ja, hvis arkiverings-versionen indeholder digitale dokumenter**	Angiver slutdato for det sidst daterede digitale dokument i arkiveringsversionen, f.eks. datoen for det sidste interview.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveInforma-tionPacketType	**Slutaflevering**	Angivelse af, om arkiveringsversionen er sidste aflevering af data, herunder også sidste dataaflevering fra et forskningsprojekt med kontinuerlig dataindsamling over en længere periode (evt. kohorte).	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis datasættet på tidspunktet for produktion af arkiveringsversionen er lukket for ny dataindsamling/ opdateringer Har ligeledes værdien *true*, hvis Rigsarkivet på tidspunktet for produktion af arkiveringsversionen har truffet bestemmelse om, at der ikke fremover skal afleveres data fra forskningsprojektet. Har værdien *false* i øvrige tilfælde.
creatorName	**Arkivskaber**	Både forskningsinstitution og primærforsker, som har skabt datasættet.	**Fri tekst**	**1-m**	**Ja**	Det aftales med det modtagende arkiv, hvilke dataskabere der skal anføres.
creationPeriodStart	**Arkivskaber startdato**	Det angives hvornår dataindsamlingen er påbegyndt for hver af de dataskabere/institutioner, der har bidraget med indsamling af data til projektet.	**År, år-måned, eller år-måned-dag**	**1-m**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2008-06-01.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
creationPeriodEnd	**Arkivskaber slutdato**	Det angives hvornår dataindsamlingen er afsluttet for hver af de dataskabere/institutioner, der har bidraget med indsamling af data til projektet.	**År, år-måned, eller år-måned-dag**	**1-m**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2010-12-31.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveType	**Arkiverings-versions-periodetype**	Angivelse af om der er tale om en afsluttet aflevering	**Boolsk værdi**	**1**	**Ja**	Oplyses af det modtagende arkiv. Har værdien *true*, hvis der er tale om en afsluttet aflevering *(= afsluttet arkivperiode*). Det kan f.eks. være: en tværsnitsundersøgelse af befolkningens holdninger til diverse emner eller en afsluttet naturvidenskabelig observationsundersøgelse. Har værdien *false (=Øjebliksbillede)*, hvis der er tale om en kontinuerlig aflevering, hvor dataindsamling er kontinuerlig over en længere periode og hvor der skal ske kontinuerlig aflevering af data i form af et øjebliksbillede, hvor de afleverede data også vil indgå i næste afleveringspakke. Det kan f.eks. være en kohorteundersøgelse eller forløbsundersøgelser eller dataindsamling ifm. overvågning af fænomener, hvor der løbende akkumuleres data.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveTypsClosedFiles	**Arkiveringsversion med lukkede sager**	Angivelse af, om der for it-systemer med dokumenter er tale om et øjebliksbillede af metadata, men kun dokumenter fra lukkede sager	**Boolsk værdi**	**1**	**Ja, hvis arkiverings-versionen indeholder digitale dokumenter**	Angives som "*false*" ved aflevering af forskningsdata
systemName	**Systemnavn**	Den officielle danske titel på forskningsprojektet, hvor alle forkortelser er opløst.	**Fritekst**	**1**	**Ja**	Forskningsprojektets titel på dansk.
alternativeName	**Alternativt systemnavn**	Den officielle engelske titel på forskningsprojektet, hvor alle forkortelser er opløst	**Fritekst**	**0-m**	**Nej**	Forskningsprojektets titel på engelsk skal angives. Andre kaldenavne eller forkortelser for forskningsprojektet kan også angives.
systemPurpose	**Systemformål**	Beskrivelse af forskningsprojektets formål	**Fritekst**	**1**	**Ja**	Forskningsprojektets formål. F.eks. forskningsprojektets abstract. Beskrivelsen skal redegøre for projektets centrale problemstillinger. Dokumenter med detaljerede oplysninger indgår i kontekstdokumentation.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
systemContent	**Systemindhold**	Beskrivelse af den centrale population og centrale variable i data	**Fritekst**	**1**	**Ja**	Beskrivelsen skal besvare spørgsmålet om, hvem eller hvad (= den centrale population) der primært registreres hvilke data om (= de centrale variable). F.eks. et forskningsprojekt hvor befolkningen i Danmark (= population) undersøges for holdning til Brexit (her vil være en vifte af centrale variable) eller et forskningsprojekt om størrelsen på ål i Lillebælt, hvor ål = population og centrale variable = længde og vægt.
regionNum	**Regionsnumre**	Angivelse af, om der i systematisk form er registreret regionsnumre i data	**Boolsk værdi**	**1**	**Ja**	For hver af de angivne identifikatorer skal det oplyses, om den er registreret i datasættet. Har værdien *true*, hvis den pågældende identifikator er registreret systematisk. Har værdien *false*, hvis den pågældende identifikator ikke er registreret systematisk. Hvis der i datasættet findes andre fagspecifikke identifikatorer, der anvendes i flere datasæt eller registre, f.eks. et geografisk
komNum	**Kommune-numre**	Angivelse af, om der i systematisk form er registreret kommunenumre i data	**Boolsk værdi**	**1**	**Ja**	
cprNum	**CPR-numre**	Angivelse af, om der i systematisk form er registreret CPR-numre i data	**Boolsk værdi**	**1**	**Ja**	
cvrNum	**CVR-numre**	Angivelse af, om der i systematisk form er registreret CVR-numre i data	**Boolsk værdi**	**1**	**Ja**	

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
matrikNum	**Matrikelnumre**	Angivelse af, om der i systematisk form er registreret matrikelnumre i data	**Boolsk værdi**	**1**	**Ja**	referencesystem, Dansk branchekode, eller beskæftigelseskoder for baggrundsvariable, diagnosekoder eller medicinkoder kan det angives i forbindelse med oplysninger om central population og centrale variable (i *systemContent*).
bbrNum	**BBR-numre**	Angivelse af, om der i systematisk form er registreret BBR-numre i data	**Boolsk værdi**	**1**	**Ja**	
whoSygKod	**WHOs sygdomskoder**	Angivelse af, om der i systematisk form er registreret WHO-sygdomskoder i data	**Boolsk værdi**	**1**	**Ja**	
sourceName	**Datakilde**	Eventuelle registre der er udtrukket data fra/ andre datasæt der er udtrukket data fra (opslag, overførsel, samkøring osv.)	**Fritekst**	**0-m**	**Nej**	Teksten skal besvare spørgsmålet om der i forskningsprojektet er anvendt data udtrukket fra registre eller andre datasæt og fra hvilke. I kontekstdokumentationen kan indgå dokumenter med detaljerede oplysninger, f.eks. om hvilke data (kolonner) der er overført fra andre registre eller datasæt og hvordan overførsel er sket.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
userName	**Databrugere**	Eventuelle andre forskningsprojekter, der anvender dette projekts data	**Fritekst**	**0-m**	**Nej**	Teksten skal besvare spørgsmålet om, hvorvidt forskningsprojektets data anvendes i andre forskningsprojekter samt i hvilke. I kontekstdokumentationen kan indgå dokumenter med detaljerede oplysninger, f.eks. om hvilke data (kolonner) der er anvendt i andre forskningsprojekter, og hvordan data er anvendt.
predecessorName	**Forgænger-systemer**	Eksisterende forskningsprojekt som dette forskningsprojekt følger op på.	**Fritekst**	**0-m**	**Nej**	Hvis der er tale om en opfølgende forskningsundersøgelse til et tidligere udført forskningsprojekt, skal titlen på det tidligere udførte forskningsprojekt angives.
formVersion	**FORM-version**	Angivelse af, hvilken version af FORM (den fællesoffentlige forretningsreferencemodel) nedenstående kategorisering er hentet fra	**Gyldige FORM-versionsnumre**	**0-1**	**Ja, for it-system-er fra offentlige myndigheder**	Udfyldes ikke for forskningsdata

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
formClass	**FORM-klassifikation**	Kategorisering af arkiveringsversionen iht. FORM	**Enhver gyldig FORM-reference iht. den angivne version af FORM**	**0-m**	**Ja, for it-system-er fra offentlige myndigheder**	Udfyldes ikke for forskningsdata
formClassText	**FORM-klassifikation klartekst**	FORM-klassifikation i klartekst	**Fri tekst**	**0-m**	**Ja, for it-system-er fra offentlige myndigheder**	Udfyldes ikke for forskningsdata
containsDigital-Documents	**Digitale dokumenter indeholdt**	Angivelse af, om der er digitale dokumenter i arkiveringsversionen, udover kontekstdokumentation	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis det er en arkiveringsversion med digitale dokumenter. F.eks. ved aflevering af et kvalitativt forskningsprojekt, hvor interviews findes i form af lydfiler eller dokumenter. Har værdien *false*, hvis det er en arkiveringsversion uden digitale dokumenter (dog vil der være dokumenter med kontekstdokumentation).
containsGeodata	**GML-filer indeholdt**	Angivelse af, om der er GML-filer i arkiveringsversionen	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis det er en arkiveringsversion med struktureret GML. F.eks. ved aflevering af et forskningsprojekt, der indeholder georefererede data.

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
containsResearchData	**Forskningsdata indeholdt**	Angivelse af, om arkiveringsversionen indeholder forskningsdata	**Boolsk værdi**	**1**	**Ja**	Har værdien *"true",* hvis der er tale om forskningsdata i en arkiveringsversion
researchSIP	**Forskningsdata afleveret iht. bilag 9**	Angivelse af, om data er afleveret iht. bilag 9	**Boolsk værdi**	**1**	**Ja**	Har værdien *"true",* hvis der er tale om visse typer af forskningsdata som afleveres som en afleveringspakke iht. bilag 9. Har værdien *"false",* hvis der er tale om forskningsdata som afleveres som en arkiveringsversion iht. bilag 1-8.
documentsDisposal	**Kassation**	Angivelse af, om der er foretaget systematisk kassation af dokumenter inden eller i forbindelse med produktion af arkiveringsversionen	**Boolsk værdi**	**1**	**Ja**	Har værdien "*false"* i for forskningsdata
searchRelatedOtherRecords	**Søgemiddel til andre sager eller dokumenter**	Angivelse af, om arkiveringsversionen er et nødvendigt søgemiddel til papirsager/dokumenter eller sager/dokumenter i et andet it-system	**Boolsk værdi**	**1**	**Ja**	Har værdien "*false"* i for forskningsdata
relatedRecordsName	**Adresserede arkivalier**	Henvisning til de arkivalier, som arkiveringsversionen er søgemiddel til	**Fritekst**	**0-m**	**Ja, hvis der er svaret ja på fore-gående spørgs-mål**	Udfyldes ikke for forskningsdata

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
systemFileConcept	**Eksistens af sagsbegreb i it-systemet**	Angivelse af, at it-systemet har et sagsbegreb, forstået som en registreret sammenhæng mellem sagligt sammenhørende dokumenter	**Boolsk værdi. Kan kun være *true*, hvis der er digitale dokumenter i it-systemet**	**1**	**Ja**	Har værdien ”*false”* i for forskningsdata
multipleData-Collection	**SOA arkitektur**	Angivelse af, om it-systemet er sammensat af data og eventuelt dokumenter fra flere forskellige it-systemer i en serviceorienteret arkitektur	**Boolsk værdi**	**1**	**Ja**	Har værdien ”*false”* i for forskningsdata
personalData-RestrictedInfo	**Persondata**	Angivelse af, om der i data findes personfølsomme oplysninger i henholdt til databeskyttelsesforordningen	**Boolsk værdi**	**1**	**Ja**	Personhenførbare oplysninger. Har værdien ”*true”*, hvis der i data findes følsomme personoplysninger (oplysninger om racemæssig eller etnisk baggrund, politisk, religiøs eller filosofisk overbevisning, fagforeningsmæssige tilhørsforhold, oplysninger om helbredsmæssige og seksuelle forhold samt oplysninger om strafbare forhold og væsentlige sociale problemer).

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
otherAccessTypeRestrictions	**Fristforlængende oplysninger**	Angivelse af, om der i data og den tilhørende dokumentation findes oplysninger, der kan betinge længere tilgængelighedsfrist i øvrigt	**Boolsk værdi**	**1**	**Ja**	Særlige tilgængelighedsbestemmelser. Har værdien ”*true”*, hvis tilgængelighedsfristen er forlænget. Hvis den afleverende myndighed ønsker forlængelse af tilgængelighedsfristen, skal det i forbindelse med aflevering drøftes med det modtagende arkiv, jf. lovbekendtgørelse nr. 1035 af 21. august 2007 (Arkivloven), § 27.
archiveApproval	**Godkendelses-arkiv**	Angivelse af, hvilket offentligt arkiv, der godkender arkiveringsversionen	**Identifikation af det pågældende arkiv (2-4 tegn)**	**1**	**Ja**	Arkiv der godkender arkiveringsversionen. Oplyses af det modtagende arkiv. Hvis det modtagende arkiv f.eks. er Rigsarkivet skrives forkortelsen SA (= Statens Arkiver).

Elementnavn	Betegnelse	Beskrivelse	Udfaldsrum	Fore-komst	Obligatorisk	Vejledning
archiveRestrictions	**Tilgængelig-hedsbegræns-ninger**	Angivelse af nærmere bestemmelser for adgang til materialet. Feltet kan anvendes efter det modtagende arkivs nærmere retningslinjer	**Fritekst**	**0-1**	**Nej**	Beskrivelse af betingelser for adgang til data. Det modtagende arkiv træffer aftale med den afleverende myndighed om, hvorvidt og med hvilke oplysninger dette felt skal udfyldes.

6.B. Kontekstdokumentation

6.B.1 Enhver arkiveringsversion skal indeholde dokumenter, som dokumenterer it-systemets administrative funktion samt struktur og funktionalitet eller forskningsdatas ophav og opbygning.

6.B.2 Det modtagende arkiv fastlægger efter drøftelse med den afleverende myndighed, hvilke dokumenter, der skal afleveres, herunder hvilke punkter i figur 6.2, som ikke er relevante at dokumentere i den konkrete aflevering.

6.B.3.a Dokumenterne indplaceres i en eller flere af de kategorier, som fremgår af figur 6.2.

6.B.3.b Oplysninger om kategoriseringen registreres i indeksfilen contextDocumentationIndex, jf. 4.C.4.a.

6.B.4 Dokumenterne skal lagres i et af de i arkiveringsversionen tilladte dokumentformater, jf. 5.E-5.F.

Kontekstdokumentationen består af dokumenter og evt. lyd og video. Dokumenterne bør være skabt i forbindelse med anskaffelse og brug af systemet eller opsamlet i forbindelse med et forskningsprojekt, men kan også udarbejdes i forbindelse med produktion af arkiveringsversionen, hvis de ikke findes i forvejen. Der træffes aftale mellem myndigheden og det modtagende arkiv om, hvilken dokumentation der skal indgå.

Navngivning og placering i arkiveringsversionens struktur af dokumenterne med kontekstdokumentation fremgår ikke af dette bilag, men kan læses i bilag 4, punkt 4.E.

Der skal udarbejdes en fortegnelse over dokumenterne i kontekstdokumentation. I fortegnelsen er det obligatorisk at anføre dokumentidentifikation, titel og en eller flere kategorier (jf. nedenfor). Herudover kan der anføres oplysninger om dokumentbeskrivelse, én dato, et eller flere forfatternavne samt en eller flere forfatterinstitutioner. Den nærmere udformning af fortegnelsen fremgår ikke af dette bilag, men kan læses i bilag 4, punkt 4.C.4.

Dokumentkategorier i kontekstdokumentationsfilen

For hvert enkelt dokument skal anføres mindst en af nedenstående dokumentkategorier, og der er ingen begrænsninger for, hvor mange dokumentkategorier der kan anføres for det enkelte dokument. F.eks. kan instruksen for et ESDH-system indeholde oplysninger om en række af kategorierne under 1 (administrativ brug af it-systemet).

Der træffes nærmere aftale om placering i kategorier med det modtagende arkiv.

Kategoriseringen af dokumenter i kontekstdokumentationsfilen skal hjælpe fremtidige brugere til nemt at finde relevant dokumentation til netop deres behov. Det vil være forskelligt hvor mange kategorier, der er anvendt for forskellige typer af data.

1. For administrative data vil der primært være dokumentation i kategori 1-6
2. For forskningsdata vil der primært være dokumentation i kategori 3-7

Alle kategorier vil således ikke være relevante for alle typer af data.

Figur 6.2 Dokumentkategorier i contextDocumentationIndex.xml

1. Dokumentation vedrørende administrativ brug af it-systemet			
	Elementnavn	**Beskrivelse**	**Vejledning**
1.a	**systemPurpose**	**It-systemets formål**	Dokumenter, der belyser, hvorfor der er brugt ressourcer på at oprette og drive it-systemet Det kan f.eks. være kravspecifikation, instruks, beskrivelse på hjemmeside, lov- og regelgrundlag eller beslutningsgrundlag for bevilling
1.b	**systemRegulations**	**It-systemets lov- og regelgrundlag**	For det første love, bekendtgørelser m.v, hvis regler administreres ved hjælp af it-systemet. Det kan f.eks. være selskabsskatteloven (dokumentation til Selskabsskatteregistret) eller EU-forordninger (dokumentation til et tilskudsadministrativt system hos Fødevareerhverv) For det andet love, bekendtgørelse mv., der fastsætter oprettelse og drift af it-systemet. Det kan f.eks. være CPR-loven eller Tinglysningsloven. Det aftales med det modtagende arkiv, om selve dokumenterne skal indgå i arkiveringsversionen, eller om der skal udarbejdes en detaljeret liste.
1.c	**systemContent**	**It-systemets indhold, population og særlige begreber**	For det første dokumenter, der giver flere oplysninger end i arkivbeskrivelsesfilen om, hvem/hvad der registreres hvilke data om. Det kan f.eks. være (del af) systembeskrivelse, Danmarks Statistiks varedeklarationer eller

			bekendtgørelse om ejeres pligt til at afgive oplysninger til BBR. For det andet dokumenter, der indeholder definitioner af fagspecifikke begreber og ligeledes begreber, der kun anvendes hos den pågældende myndighed / i det pågældende it-system. Det kan f.eks. være definitionen af CVR-numre.
1.d	**systemAdministrativeFunctions**	**It-systemets administrative funktioner**	Dokumenter, der belyser, hvilke opgaver hos myndigheden it-systemet har løst eller understøttet løsningen af. Det kan evt. også være opgaver, der er afledt af de primære administrative funktioner, f.eks. afrapportering til overordnet myndighed. Eksempler på dokumenter kan være instruks, brugervejledning, (del af) kravspecifikation eller retningslinjer for udtræk til statistisk afrapportering.
1.e	**systemPresentationStructure**	**It-systemets præsentationsstruktur**	Dokumenter, der belyser it-systemets grænseflade til slutbrugeren. Det kan f.eks. være skærmbilleder eller udskrifter. Det kan lette genanvendelsen af data, hvis skærmbilleder og udskrifter indeholder oplysning om navne på de tabeller og kolonner, hvorfra data på skærmbillederne og udskrifterne trækkes. En anden mulighed er at referere til views eller SQL-forespørgsler i tabelindeks, hvis sådanne findes.
1.f	**systemDataProvision**	**Tilvejebringelse af data**	Dokumenter, der belyser, hvor data stammer fra, og hvordan data er blevet registreret i it-systemet.

			Det kan f.eks. være instrukser med registreringsvejledning, papir- eller web-baserede blanketter med vejledning til borgere, virksomheder eller myndigheder, specifikation af opslag i eller overførsel af data fra andre it-systemer
1.g	**systemDataTransfer**	**Videregivelse af data**	Dokumenter, der belyser, hvilke myndigheder eller borgere og virksomheder data er videregivet til. Herunder oplysninger om, hvilke data der er videregivet og på hvilken måde. Det kan f.eks. være instruks, bekendtgørelse eller lignende, der fastsætter pligt til overførsel, specifikation af overførsel af data til andre it-systemer eller salgsoplysninger på hjemmeside.
1.h	**systemPreviousSubsequentFunctions**	**Data og funktioner fælles med forgænger- og efterfølgersystemer**	Det kan f.eks. være kravspecifikation eller systembeskrivelse.
1.i	**systemAgencyQualityControl**	**Myndighedens egen kvalitetskontrol**	Dokumenter, der belyser regler for automatisk eller manuel kontrol, evt. ved borgeres og virksomheders indtastning i it-systemet. Det kan f.eks. være brugervejledning eller kravspecifikation.
1.j	**systemPublication**	**Publikation af og om data**	For det første publikationer, hvor it-systemets data er anvendt som grundlag. For det andet publikationer, hvor it-systemets data og deres validitet vurderes. Det kan også være interne rapporter. Det aftales med det modtagende arkiv, om publikationer skal indgå, eller om der skal laves en liste med ISBN-nr. (for publicerede) og

			journalnummer (for upublicerede rapporter).
1.k	**systemInformationOther**	**Andet**	
1.l	**systemTaxonomy**	**Registreringssystematik**	Det kan f.eks. være en journalplan, emneordsliste, indholdsfortegnelse, mappehierarki eller anden oversigt, der dokumenterer hvordan sager, dokumenter, fotos og lignende er grupperet eller registreret.
1.m	**systemInstruction**	**Instruks for anvendelse af systemet**	Det kan f.eks. være brugervejledninger til systemet, instrukser, arbejdsgangsbeskrivelser, procedurer og lignende
2. Dokumentation vedrørende it-systemets tekniske udformning, drift og udvikling			
	Elementnavn	**Beskrivelse**	**Vejledning**
2.a	**operationalSystemInformation**	**Driftsversionens opbygning**	Det kan f.eks. være ER-diagram, datadicitionary, kravspecifikation, systemdokumentation eller ændringsdokumenter til systemdokumentationen.
2.b	**operationalSystemConvertedInformation**	**Konvertering hos myndigheden**	Kategorien skal anvendes, hvis data er konverteret fra et system til et andet. I givet fald kan f.eks. konverteringsbeskrivelsen lægges ind.
2.c	**operationalSystemSOA**	**Dokumentation af sammensætning af data og eventuelle dokumenter fra flere forskellige it-systemer i en serviceorienteret arkitektur**	Dokumenter med detaljerede oplysninger, hvis værdien i SOA-arkitektur/ multipleDataCollection i arkivbeskrivelsesfilen er angivet som *true*.
2.d	**operationalSystemInformationOther**	**Andet**	
3. Dokumentation vedrørende arkivskabers aflevering af data			
	Elementnavn	**Beskrivelse**	**Vejledning**

3.a	**archivalProvisions**	**Arkivets bestemmelser, herunder afleveringsbestemmelse**	Dokumenterne vil være udarbejdet af det modtagende arkiv, der enten sender dem til den afleverende myndighed eller selv lægger dem ind i arkiveringsversionen. Det kan f.eks. være bevarings- og kassationsbestemmelser, notat om tabeller, der ikke medtages i arkiveringsversionen, evt. dispensationer eller notat om særlige tilgængelighedsforhold.
3.b	**archivalTransformationInformation**	**Dokumentation af konvertering fra driftsversion til arkiveringsversion**	Dokumenterne kan udarbejdes i forbindelse med produktion af arkiveringsversion. Det kan være dokumenter, der beskriver ændringer af it-systemets tabelstruktur ved overførsel til arkiveringsversion, eller dokumenter vedr. problemer med konvertering af dokumenter til TIFF.
3.c	**archivalInformationOther**	**Andet**	
4. Dokumentation vedrørende arkivets modtagelse af data (udfyldes af modtagende arkiv)			
	Elementnavn	**Beskrivelse**	**Vejledning**
4.a	**archivistNotes**	**Arkivarnoter**	Disse kategorier skal ikke anvendes af den afleverende myndighed.
4.b	**archivalTestNotes**	**Testnoter**	
4.c	**archivalInformationOther**	**Andet**	
5. Dokumentation vedrørende arkivets bevaring af arkiveringsversionen (udfyldes af det modtagende arkiv)			
	Elementnavn	**Beskrivelse**	**Vejledning**
5.a	**archivalMigrationInformation**	**Konvertering hos arkivet**	Denne kategori skal ikke anvendes af den afleverende myndighed.
5.b	**archivalInformationOther**	**Andet**	
6. Anden dokumentation			

	Elementnavn	Beskrivelse	Vejledning
6.a	**informationOther**	**Andet**	
7. Dokumentation af forskningsdata			
	Elementnavn	**Beskrivelse**	**Vejledning**
7.a	**researchProjectDescription**	**Projektbeskrivelse gældende for de afleverede data**	Projektbeskrivelser af forskningsprojektet, fondsansøgning o. lign.
7.b	**researchQuestionnaire**	**Spørgeskema, interviewguide og/eller registreringsskema anvendt til at indsamle og analysere de afleverede data**	Spørgeskemaer, interviewguides mv. som kan bidrage til forståelse af nærværende data
7.c	**researchProtocol**	**Protokoller og metoderapporter**	Metodebeskrivelser, forskningsprotokoller, beskrivelser af måleudstyr osv
7.d	**researchPublication**	**Publikationer, som er udgivet på basis af de afleverede data**	Liste over publikationer med forskningsresultater baseret på nærværende data
7.e	**researchInformationOther**	**Andet**	

6.C. Data om arkiveringsversionens tabeller

6.C.1 En arkiveringsversion skal indeholde dokumentation af arkiveringsversionens tabeller og relationer (tabelindeks). Tabelindekset skal indeholde de oplysninger, som fremgår af figur 6.3 nedenfor.

Formålet med tabelindeks er at dokumentere:

- Dataindhold i tabeller og kolonner i arkiveringsversionen
- Relationer mellem tabeller ved angivelse af fremmednøgler
- Sammenhænge vedrørende brugen af systemet ved oplysninger om enten views (fra it-systemets database) eller SQL-forespørgsler (refererer til arkiveringsversionen)
- Kolonner med særlige informationer i arkiveringsversioner, der indeholder digitale dokumenter

Navngivning og nærmere retningslinjer for udformning af tabelindeks fremgår ikke af dette bilag, men kan læses i bilag 4, punkt 4.C.5.

Da det forventes, at oplysningerne i tabelindeks i mange tilfælde vil kunne uddrages automatisk fra it-systemets database, er det i nedenstående figur ved en række oplysninger angivet, at oplysningen kun er obligatorisk, hvis den findes i databasen.

Figur 6.3 Oplysninger i tableIndex.xml

1. Oplysninger om databasen				
	Elementnavn	**Beskrivelse**	**Forekomst pr. database**	**Obligatorisk**
1.a	**version**	**Format version, altid "1.0"**	**1**	**Ja**
1.b	**dbName**	**Kort navn på databasen**	**0-1**	**Ja, hvis oplysningen findes i databasen**
1.c	**databaseProduct**	**Navn på og version af det databaseprodukt, hvorfra arkivering af data er sket**	**0-1**	**Ja, hvis oplysningen findes i databasen**
2. Oplysninger om tabeller og views				
	Elementnavn	**Beskrivelse**	**Forekomst pr. arkiverings-version**	**Obligatorisk**
2.a	**tables**	**Liste over tabeller i databasen**	**1**	**Ja**
2.b	**views**	**Liste over views i databasen**	**0-1**	**Ja, hvis der indgår views i arkiverings-versionen**
3. Oplysninger om de enkelte tabeller				
	Elementnavn	**Beskrivelse**	**Forekomst pr. tabel**	**Obligatorisk**
3.a	**name**	**Tabelnavn. Må ikke begynde med et tal.**	**1**	**Ja**
3.b	**folder**	**Navn på mappen som indeholder tabel og tilhørende skema**	**1**	**Ja**
3.c	**description**	**Beskrivelse af tabellens indhold og betydning**	**1**	**Ja**
3.d	**columns**	**Liste over kolonner i tabellen**	**1-m**	**Ja**
3.e	**primaryKey**	**Tabellens primærnøgle**	**1**	**Ja**

3.f	**foreignKey**	**Fremmednøgle i tabel**	**0-m**	**Ja, hvis der findes fremmednøgler**
3.g	**rows**	**Angivelse af antal rækker i tabellen**	**1**	**Ja**

Tabelnavn (name) er det navn, der anvendes i it-systemets database. Hvis der i forbindelse med produktion af arkiveringsversionen dannes nye tabeller (f.eks. for at overholde kravene til relationel struktur, eller fordi it-systemets data ikke er lagret i en relationel database), er kravet til navngivning, at tabelnavnet skal begynde med et bogstav og kun må indeholde bogstaver, tal og _.

Navngivning af folder fremgår ikke af dette bilag, men af bilag 4, punkt 4.D.2. Det fremgår, at folder skal navngives table[fortløbende nummer]. Den fortløbende nummerering begynder med 1, og der må ikke anvendes foranstillede nuller.

Description skal udfyldes så udførligt som muligt, således at en fremtidig bruger har mulighed for at forstå tabellens indhold og betydning, f.eks. om tabellen indeholder den centrale population eller centrale variable, eller om tabellen er en kodetabel i it-systemets database eller en kodetabel skabt til arkiveringsversionen.

4. Oplysninger om kolonner				
	Elementnavn	**Beskrivelse**	**Forekomst pr. kolonne**	**Obligatorisk**
4.a	**name**	**Kolonnenavn**	**1**	**Ja**
4.b	**columnID**	**Entydig identifikation af kolonne. Begynder med bogstavet 'c' efterfulgt af nummeret på kolonnens plads i tabellen.**	**1**	**Ja**
4.c	**type**	**SQL:1999 datatype**	**1**	**Ja**
4.d	**typeOriginal**	**Original datatype**	**0-1**	**Ja, hvis oplysningen findes i databasen**
4.e	**defaultValue**	**Defaultværdi for felter i kolonnen**	**0-1**	**Ja, hvis oplysningen findes i databasen**
4.f	**nullable**	**Angivelse af, om felter i kolonnen kan være "NULL". Boolsk værdi.**	**1**	**Ja**
4.g	**description**	**Beskrivelse af kolonnens indhold**	**1**	**Ja**
4.h	**functional-Description**	**Angivelse af kolonnens funktion iht. figur 6.4, figur 6.5 og figur 6.6 nedenfor.**	**0-m**	**Ja, hvis kolonnen har en af de angivne funktioner**

Kolonnenavn (name) er det navn, der anvendes i it-systemets database. Hvis der i forbindelse med produktion af arkiveringsversionen dannes nye kolonner (f.eks. for at overholde kravet om at alle nøglefelter skal have en entydig identifikator), er kravet til navngivning, at kolonnenavnet skal begynde med et bogstav og kun må indeholde bogstaver, tal og _.

Kolonnen skal tildeles et nummer (columnID), der begynder med bogstavet 'c' efterfulgt af nummeret på kolonnen i tabellen. Den rækkefølge for kolonner, der angives med nummereringen, skal svare til den rækkefølge for kolonner, der defineres i xml-skemaet for den pågældende tabel. Navngivning og nærmere retningslinjer for udformningen af xml-skemaet fremgår ikke af dette bilag, men kan læses i bilag 4, punkt 4.D.

Datatypen (type) skal angives som en SQL-datatype. SQL-datatyperne kan ses i bilag 5, figur 5.1. Der skal være overensstemmelse mellem den angivne SQL-datatype og den XML-datatype, der angives i xml-skemaet for den pågældende tabel.

Description skal udfyldes så udførligt som muligt, således at en fremtidig bruger har mulighed for at forstå, hvilken oplysning der er registreret i kolonnen.

5. Oplysninger om primærnøgler				
	Elementnavn	**Beskrivelse**	**Forekomst pr. nøgle i tabel**	**Obligatorisk**
5.a	**name**	**Navn på primærnøgle. Navnet skal være unikt inden for arkiveringsversionen og være i overensstemmelse med reglerne for *constraint names* i standarden SQL1999**	**1**	**Ja**
5.b	**column**	**Oprindeligt navn på kolonner i primærnøglen**	**1-m**	**Ja**
6. Oplysninger om fremmednøgler				
	Elementnavn	**Beskrivelse**	**Forekomst pr. nøgle**	**Obligatorisk**
6.a	**name**	**Navn på fremmednøgle. Navnet skal være unikt inden for arkiveringsversionen og leve op til reglerne for *constraint names* i standarden SQL:1999**	**1**	**Ja**
6.b	**referencedTable**	**Den tabel, som fremmednøglen refererer til.**	**1**	**Ja**

6.c	**reference**		**Oprindeligt navn på de kolonner, der indgår i fremmednøglen**	**1-m**	**Ja**
6.d		**column**	**Navn på kolonnen, der refereres fra**	**1-m**	**Ja**
6.e		**referenced**	**Navn på kolonnen, der refereres til**	**1-m**	**Ja**

For kolonner, der indgår i primær- og fremmednøgler, skal der angives oprindeligt navn fra it-systemets database. Hvis det ikke er muligt, fordi der i forbindelse med produktion af arkiveringsversionen er dannet kolonner, der ikke findes i it-systemets database, må det aftales med det modtagende arkiv, hvordan dette håndteres. Enten kan det navn på kolonnen, der er defineret i tabelindeks (name) benyttes, eller definitionen af fremmednøglen kan udelades.

7. Oplysninger om views og forespørgsler				
	Elementnavn	**Beskrivelse**	**Forekomst pr. forespørgsel eller view**	**Obligatorisk**
7.a	**name**	**Navn på view**	**1**	**Ja**
7.b	**queryOriginal**	**Original SQL-forespørgsel, som definerer view eller SQL-forespørgsel, som er dannet specifikt til arkiveringsversionen**	**1**	**Ja**
7.c	**description**	**Indhold og betydning af view eller SQL-forespørgsel**	**0-1**	**Ja, hvis de kan trækkes fra databasen eller hvis der er tale om forespørgsler, der er særligt defineret til arkiveringsversionen**

6.C.2 Systemviews medtages ikke.

6.C.3 Det modtagende arkiv kan anvise, at de væsentligste views skal have en beskrivelse, jf. figur 6.3, 7.c.

Hvis der i databasen er defineret views, der automatisk kan trækkes ud, skal de indgå i arkiveringsversionen. Der er ingen krav til, hvilken SQL-standard der er anvendt.

Systemviews skal ikke medtages. Med systemviews menes views, der ikke er rettet mod slutbrugernes anvendelse af it-systemet, f.eks. views til styring af brugerrettigheder eller views til opsætning af systemet. Hvis det kræver et større manuelt arbejde at udskille systemviews, medtages de i arkiveringsversionen.

Det modtagende arkiv kan anvise, at de væsentligste views skal have en beskrivelse. Med væsentligste menes de views, der er centrale for slutbrugernes anvendelse af it-systemet.

Hvis det modtagende arkiv vurderer, at de væsentligste views skal have en beskrivelse, kan den afleverende myndighed blive bedt om at udpege views, der opfylder bestemte kriterier. Kriterierne kan f.eks. være:

- Views, der angiver, hvilke data der skal vises på skærmbilleder. Udgangspunktet kan f.eks. være de skærmbilleder, der indgår i kontekstdokumentationen under kategorien Præsentationsstruktur.
- Views, der angiver, hvilke data der skal udskrives i rapporter. Udgangspunktet kan ligeledes være rapporter i kontekstdokumentationen.
- Views, der dokumenterer, i hvilke tabeller og kolonner data om den eller de centrale populationer findes.

6.C.4 Der skal opmærkes særlige oplysninger for it-systemer med registrering af oplysninger om dokumenter.

6.C.5 For arkiveringsversioner af it-systemer, som nævnt i 6.C.4, skal kolonner, der indeholder særlige informationer, jf. figur 6.4 - figur 6.6, identificeres via elementet functionalDescription, jf. figur 6.3, 4.h.

6.C.6 De særlige oplysninger, der er angivet i figur 6.6, skal opmærkes i det omfang, de er registreret i it-systemet. Hvis oplysningerne i figur 6.6 ikke findes registreret i it-systemet, skal der i forbindelse med afleveringen aftales eventuelle alternative opmærkninger, der sikrer identifikation af sagligt sammenhørende dokumenter, jf. gældende bestemmelser herom.

Bemærk, reglerne for opmærkning af særlige oplysninger gælder ikke for forskningsdata

Figur 6.4 functionalDesciptions i IT-systemer, der deles af flere myndigheder

8. Opmærkning af kolonner med faste oplysninger i arkiveringsversioner af it-systemer med registrering af dokumenter, der deles af flere myndigheder	
Oplysning	**Beskrivelse**
Myndighedsidentifikation	**Den eller de kolonner i arkiveringsversionen, som indeholder oplysninger om hvilken myndighed, der har registreret sagen eller dokumentet**

I en arkiveringsversion af et it-system med dokumenter (digitale eller papirbaserede), hvori flere myndigheder registrerer sager og dokumenter, skal der ske opmærkning med functionalDescription for en eller flere kolonner, der indeholder myndighedsidentifikationen.

Figur 6.5 functionalDesciptions i IT-systemer med dokumenter

9. Opmærkning af kolonner med faste oplysninger i arkiveringsversioner af it-systemer med dokumenter		
Oplysning	**Beskrivelse**	**Udfaldsrum i tabeldata**
Dokumentidentifikation	**Bruges til at angive den eller de kolonner i arkiveringsversionen, som beskriver de digitale dokumenters entydige identifikation i arkiveringsversionens struktur**	**ID på op til 12 cifre, som findes blandt de navngivne dokumentmapper og er unikt inden for mappen *Documents*, jf. 4.G.4**
Lagringsform	**Bruges til at angive den eller de kolonner i arkiveringsversionen, som beskriver, om dokumentet er helt eller delvist digitalt, på papir eller ikke har tilknyttet en fil**	**Helt eller delvis digitalt = 1, papir = 2, ikke relevant = 3**
Afleveret	**Bruges ved aflevering af øjebliksbilleder m.v. til at angive den eller de kolonner i arkiveringsversionen, som beskriver, om dokumentet allerede er afleveret i en tidligere arkiveringsversion**	**Tidligere afleveret = 1, ikke tidligere afleveret = 2**

I en arkiveringsversion af et it-system med digitale dokumenter skal der ske opmærkning med functionalDescription for kolonner, der indeholder oplysning om dokumentidentifikation, lagringsform og afleveret.

Med dokumentidentifikation menes den kolonne, der indeholder ID, der entydigt udpeger det enkelte dokument. Det entydige dokumentID er ligeledes registreret i docIndex.xml og bruges til navngivning af den mappe, hvori det digitale dokument er placeret. På denne måde skabes der i en arkiveringsversion af et it-system med digitale dokumenter en forbindelse mellem registreringerne om dokumentet og selve dokumentet.

Med lagringsformen papir skal forstås alt materiale, der ikke er lagret digitalt i arkiveringsversionen. Med ”ikke relevant” menes, at der ikke er tilknyttet en fil til registreringen – og aldrig har været det. Det kan f.eks. være en registrering, som har fungeret som et telefonnotat.

Figur 6.6 functionalDesciptions i IT-systemer med registrering af dokumenter

10. Opmærkning af kolonner med oplysninger i arkiveringsversioner af it-systemer med registrering af dokumenter	
Oplysning	**Beskrivelse**
Sagsidentifikation	**Den eller de kolonner i arkiveringsversionen, som beskriver sagernes entydige identifikation**

Sagstitel	**Den eller de kolonner i arkiveringsversionen, som indeholder sagernes titler**
Dokumenttitel	**Den eller de kolonner i arkiveringsversionen, som indeholder dokumenternes titler/beskrivelser**
Dokumentdato	**Den eller de kolonner i arkiveringsversionen, som indeholder oplysninger om dokumenternes afsendelses- og modtagelsesdatoer**
Afsender/modtager	**Den eller de kolonner i arkiveringsversionen, som indeholder oplysninger om et dokuments afsender eller modtager**
Digital signatur	**Den eller de kolonner i arkiveringsversionen, som indeholder oplysninger, der er uddraget fra en digital signatur**
FORM	**Den eller de kolonner i arkiveringsversionen, som indeholder reference til FORM (Den fællesoffentlige forretningsreferencemodel)**
Kassation	**Den eller de kolonner i arkiveringsversionen, som indeholder oplysninger om bevaring og kassation**

I en arkiveringsversion af et it-system med dokumenter (digitale eller papirbaserede) skal der ske opmærkning med functionalDescription for kolonner, der indeholder oplysninger om sagsidentifikation, sagstitel, dokumenttitel, dokumentdato, afsender/modtager, digital signatur, FORM-reference og kassationskoder, **hvis** disse oplysninger findes. Hvis der i et it-system med dokumenter ikke er kolonner med oplysninger om sagsidentifikation og sagstitel, aftales det med det modtagende arkiv, om der skal foretages alternativ opmærkning, således at sagligt sammenhørende dokumenter kan identificeres.

Der kan ske opmærkning af flere kolonner med den samme functionalDescription, således at f.eks. både dokumentdato og dokumentets modtagelsesdato kan opmærkes.

6.D. SQL-forespørgsler

6.D.1 Det modtagende arkiv kan fastsætte, at der til en arkiveringsversion skal defineres et antal SQL-forespørgsler til dokumentation af bestemte sammenhænge i arkiveringsversionen.

6.D.2 SQL-forespørgslerne udformes efter standarden SQL:1999 (core).

6.D.3 Forespørgslerne placeres i "Oplysninger om views og forespørgsler" i tabelindekset jf. figur 6.3, 7, og navngives efter myndighedens eget valg, dog således at navnet på de pågældende forespørgsler begynder med "AV".

Som supplement til eller i stedet for views fra databaser, kan det modtagende arkiv fastsætte, at der skal defineres SQL-forespørgsler til arkiveringsversionen. SQL-forespørgslerne navngives, defineres og forklares i lighed med views i tabelindeks (se figur 6.3, punkt 7 i dette bilag). SQL-forespørgslens navn (name) skal begynde med AV, må kun indeholde bogstaver, tal og _ samt højst have en længde på 128 tegn.

Formålet med forespørgslerne er at vise sammenhænge i data, som det kan være svært at vise på anden vis. Forespørgslerne skal i modsætning til views kunne køres direkte på arkiveringsversionen.

Eksempler på arkiveringsversioner, hvor der kan være behov for at definere SQL-forespørgsler:

- Der er ikke views i databasen.
- Der er sket ændringer i databasestrukturen, f.eks. for at kunne overholde kravet om minimum 1. normalform, eller fordi der er kasseret tabeller.
- De eksisterende views dokumenterer ikke, i hvilke tabeller og kolonner den eller de centrale populationer og tilhørende centrale variable findes, og evt. hvorledes tabellerne er relateret (jf. dette bilag, punkt A, figur 1 samt punkt B, figur 2, I.3).
- De eksisterende views dokumenterer ikke, fra hvilke tabeller og kolonner data i variable felter i dokumentationsstrukturen hentes, f.eks. ved dannelsen af skærmbilleder og udskrifter (jf. dette bilag, punkt B, figur 2, I.5).

Bilag 7 – Afleveringsmedie

7.A.1 Arkiveringsversioner kan afleveres på CD-R, DVD-R, BD-R eller USB-medie.

7.A.2 Antallet af CD-R, DVD-R eller BD-R i en aflevering må ikke overstige 10, med mindre andet aftales mellem afleverende myndighed og modtagende arkiv.

7.A.3 Ved aflevering på USB-medie af arkiveringsversioner over 4 GB skal der anvendes USB 3.0 (USB 3.1 Gen1) eller hurtigere USB-standard.

7.A.4 Den afleverende myndighed og det modtagende arkiv kan træffe aftale om aflevering på andre medier eller andre metoder til transport af data.

7.A.5 Medier med følsomme personoplysninger krypteres efter anvisning fra det modtagende arkiv.

Det modtagende arkiv fastsætter hvordan data skal overdrages til arkivet ud fra ovenstående retningslinjer.

Bilag 8 – Skemaer

Færdige skemaer til brug for skabelsen af en arkiveringsversion kan hentes fra Rigsarkivets hjemmeside.

Skemaer til validering af tabelindhold m.v. kan hentes på www.sa.dk

Bilag 9 – Afleveringspakke for visse typer af forskningsdata

Læsevejledning

Bemærk at denne vejledning til bilag 9 primært er et opslagsværk. Her kan du finde svar på hvordan kravene skal forstås og tolkes. Dette kan være anvendeligt, hvis du under produktion og test af en afleveringspakke finder specifikke fejl, der skal rettes.

Der henvises til Rigsarkivets øvrige vejledninger, f.eks. *"Quickguide - til produktion og test af en afleveringspakke med ASTA"* og *"Brugervejledning til ASTA"*. Quickguiden giver dig en lettere introduktion til kravene i bilag 9 og ASTA brugervejledningen anviser hvordan du skaber og tester en afleveringspakke af udtræk fra statistikfiler i formaterne SAS, Stata og SPSS med Rigsarkivets værktøj ASTA. Rigsarkivets har også udarbejdet en vejledning om, hvordan du kan skabe en afleveringspakke med data fra regneark og csv-filer.

Se også eksempelafleveringspakken FD.18005 med forskningsdata fra alle tre typer statistikfiler på Rigsarkivets hjemmeside www.sa.dk.

Formålet med bilag 9 – at sikre bevaring og genanvendelse af forskningsdata skabt i statistikformat

Statistikdata er den type forskningsdata som oftest afleveres til arkiv og Rigsarkivet ønsker med dette bilag 9, at styrke indsatsen for bevaring af forskningsdata skabt i statistikformat. Reglerne skal sikre, at de bevaringsværdige forskningsdata afleveres til arkivet således, at der i forhold til format, struktur og dokumentation sker en sikker bevaring, samtidig med, at mulighederne for genanvendelse understøttes.

Reglerne i bilag 9 sikrer, at statistikdata afleveres i en systemuafhængig standard, som indebærer, at data altid afleveres i en fast mappestruktur, i særlige formater for data og metadata og med specifikke datatyper. Med aflevering efter standarden i bilag 9 kan arkivet konvertere de afleverede data til et systemuafhængigt bevaringsformat, som betyder, at Rigsarkivet altid vil kunne konvertere de arkiverede data til nye formater, som kan læses, også når der ikke længere findes statistikprogrammer som f.eks. SAS og STATA. Det betyder også, at forskere, som senere ønsker at få statistikdata udleveret til brug i nye videnskabelige undersøgelser, vil kunne få data udleveret i de nye statistikformater, der findes på markedet.

Reglerne i bilag 9 sikrer også, at der stilles specifikke krav til dokumentationen af de arkiverede data, således at forskere og andre arkivbrugere i fremtiden vil kunne forstå og genbruge data. Det gøres ved krav til metadata og kontekstdokumentation, der udgøres af følgende elementer:

- ***Kontekstdokumenter***, der på et overordnet plan beskriver den kontekst data er skabt i. Det kan bl.a. være en projektbeskrivelse, en metoderapport, et eksempel på et spørgeskema etc.
- ***Kontekstdokumentationsfil***, der giver oplysninger om de enkelte kontekstdokumenter f.eks. dokumenttitel, dato, dokumentkategori og forfatter.
- ***Arkivbeskrivelsesfil***, der f.eks. beskriver hvilke data der er afleveret, hvilke adgangskriterier der er for udlevering af data, hvilket formål data er indsamlet til og hvilken periode dataindsamlingen strækker sig over. Oplysningerne i arkivbeskrivelsesfilen vil blive anvendt af forskere og andre arkivbrugere til at fremsøge data i Rigsarkivets søgekatalog. Da arkivbeskrivelsesfilen ikke er omfattet af arkivlovens tilgængelighedsfrister kan den straks offentliggøres som søgemiddel til Rigsarkivets samling af forskningsdata.
- ***Metadatafil*** med informationer om statistikfilens indhold, herunder f.eks. variabelnavne, variabelbeskrivelser, value labels, koder for manglende værdier og dataformater.

Bilag 9 anviser hvordan disse krav sikres ved at statistikdata afleveres i form af en særlig "afleveringspakke".

Hvad er en afleveringspakke?

Forskningsdata skabt eller behandlet i statistikprogrammer eller tilsvarende, skal afleveres som en afleveringspakke, der overholder kravene i bilag 9.

En afleveringspakke er et afleveringsformat, som indeholder de informationer, data og metadata, der er nødvendige for at kunne konvertere til bevaringsformat.

En afleveringspakke skal først og fremmest indeholde data- og metadataudtræk fra selve statistikfilen. Derudover skal den også indeholde en arkivbeskrivelsesfil og en kontekstdokumentationsfil samt de tilhørende kontekstdokumenter. Indhold og struktur af en afleveringspakke fremgår af figur 9.1.

Andre typer af forskningsdata skal afleveres til Rigsarkivet i de formater og strukturer og med den dokumentation, som fremgår af bekendtgørelsens bilag 1-8. Dette gælder f.eks. forskningsdata i databaser eller kvalitative undersøgelser, f.eks. lydfiler eller transskriptioner af interviews eller andre dokumentsamlinger. Forskningsdata i form af mindre dokumentsamlinger kan efter aftale med det modtagende arkiv også indgå som en del af afleveringen af institutionens ESDH-system (system med dokumenter), jf. bilag 3.

Programmer der kan bruges til at skabe afleveringspakken

Rigsarkivet stiller en række programmer til rådighed, som kan anvendes i processen med at skabe afleveringspakken:

- Programmet ***Skab archiveIndex.exe***, der anvendes til at lave arkivbeskrivelsesfilen archiveIndex.xml, som skal indgå i afleveringspakken.
- Programmet ***Skab contextDocumentationIndex.exe***, der anvendes til at lave kontekstdokumentationsfilen contextDocumentationIndex.xml, som skal indgå i afleveringspakken.
- Programmet ***Asta (Aflevering af Statistikfiler Til Arkiv)***, der kan anvendes til først at skabe afleveringspakken med udtræk fra statistikformaterne SAS, SPSS, STATA og efterfølgende teste om afleveringspakken overholder krav specificeret i bilag 9.

Alle programmer kan tilgås fra Rigsarkivets hjemmeside www.sa.dk.

Hvis du har et regneark kan afleveringspakken eventuelt laves i hånden. Metadatafilen skal da udfyldes manuelt i en teksteditor. CSV-datafilen vil kunne udtrækkes fra regnearket. Se Rigsarkivets vejledning om udtræk fra regneark til en afleveringspakke på Rigsarkivets hjemmeside www.sa.dk.

Hvis myndigheden ikke selv kan producere en afleveringspakke kan en leverandør købes til at producere den.

Hvem har ansvaret for at lave afleveringspakken?

Det er myndighedens opgave at sørge for, at der skabes og afleveres en afleveringspakke til det modtagende arkiv af de bevaringsværdige data.

Arbejdet med at skabe afleveringspakken kan eventuelt udføres af en datamanager i tæt samarbejde med forskeren eller en anden person, som har indgående kendskab til data. En datamanager vil oftest have de relevante tekniske kundskaber og desuden have rettigheder til at behandle personfølsomme data.

Hvad sker der med data efter afleveringspakken er afleveret?

En afleveringspakke med statistikdata er ikke umiddelbart klar til den endelige arkivering. Afleveringspakken er kun et mellemstadie på vejen mod at konvertere de originale statistikdata til Rigsarkivets bevaringsformat som kaldes en arkiveringsversion og er defineret i bilag 1-8. Bevaringsformatet sikrer, at data kan

langtidsbevares. Når det modtagende arkiv (f.eks Rigsarkivet) har modtaget afleveringspakken er det efterfølgende arkivets ansvar at udføre den sidste konvertering til det endelige bevaringsformat.

Udlevering af forskningsdata i statistikformat

Når forskere og andre arkivbrugere ønsker data udleveret igen, vil det modtagende arkiv sikre, at data fra statistikprogrammer afleveret efter bilag 9, kan udleveres til brugerne i statistikformat igen. For generelle regler vedr. tilgængelighedsfrister og adgang til ikke umiddelbart tilgængelige arkivalier se Arkivloven kapitel 6 og kapitel 7.

Hvis data med personfølsomme oplysninger afleveres til det modtagende arkiv, vil de være beskyttet af arkivlovens bestemmelser. Det betyder, at data først er umiddelbart tilgængelige efter 75 år. En bruger kan dog søge om adgang til dem inden da. De første 20 år efter data er skabt, vil myndigheden blive spurgt om der kan gives adgang – og hvis ja, så skal Datatilsynet også høres inden brugeren får adgang til oplysninger om rent private forhold. Efter 20 år er det arkivet som kan give adgang til data – og igen skal Datatilsynet høres, inden data stilles til rådighed. Der kan som hovedregel fastsættes vilkår for brug af data, som der gives adgang til jf. Arkivlovens § 23, stk. 1.

Vejledning til punkterne i bekendtgørelsens bilag 9

9.A. Aflevering af forskningsdata

9.A.1 Reglerne i dette bilag gælder kun for data, som er skabt i forbindelse med forskning med anvendelse af videnskabelig metode, og som er skabt eller bearbejdet i statistikprogrammer eller tilsvarende.

Reglerne i bilag 9 gælder kun data, som er indsamlet i forbindelse med en forskningsproces og kun hvis disse forskningsdata er skabt eller bearbejdet i statistikprogrammer og lignende værktøjer, der kan fortage statistiske analyser.

Forskningsdata, som er skabt eller bearbejdet i statistikprogrammer, kan f.eks. være spørgeskemaundersøgelser eller målinger/registreringer, bl.a. biologiske, fysiologiske og neurologiske målinger, i formaterne SAS, STATA, SPSS og R.

Data anvendt til statistisk analyse kan også være skabt i andre tabulær formater end statistikformat, f.eks. regneark eller CSV-filer. Disse skal også afleveres efter bilag 9.

Findes forskningsdata anvendt til statistisk analyse derimod i en relationel database, skal disse data afleveres i henhold til reglerne i bilag 1-8.

Det modtagende arkiv afgør om dine forskningsdata skal afleveres efter reglerne i bilag 9. Det fremgår af den afleveringsbestemmelse du modtager fra arkivet, hvis dine forskningsdata skal afleveres efter bilag 9.

9.A.2 Afleveringspakken konverteres til en arkiveringsversion, jf. reglerne i bilag 1-8, af det modtagende arkiv.

En arkiveringsversion med statistikdata afleveret efter bilag 9 skal overholde gældende regler i bilag 1-8, herunder specifikt indeholde indeksfilen researchIndex.xml, jf. punkt 4.C.1.c og 4.C.7.a.

9.B. Afleveringspakkens mappestruktur

9.B.1 I roden af filsystemet på afleveringsmediet, jf. bilag 7, skal der være placeret en mappe navngivet med afleveringspakkens navn. Afleveringspakkens navn består af præfikset »FD.« samt et unikt løbenummer for afleveringspakken.

9.B.2 Løbenummeret for afleveringspakken udleveres af Rigsarkivet.

En afleveringspakke har en fast mappestruktur. I roden af mappestrukturen placeres en mappe, som navngives med afleveringspakkens navn.

Afleveringspakkens navn består af præfikset »FD.« og et unikt løbenummer på minimum 5 cifre:
F.eks. FD.18999

Løbenummeret udleveres af det modtagende arkiv og fremgår af afleveringsbestemmelsen, som du modtager fra arkivet.

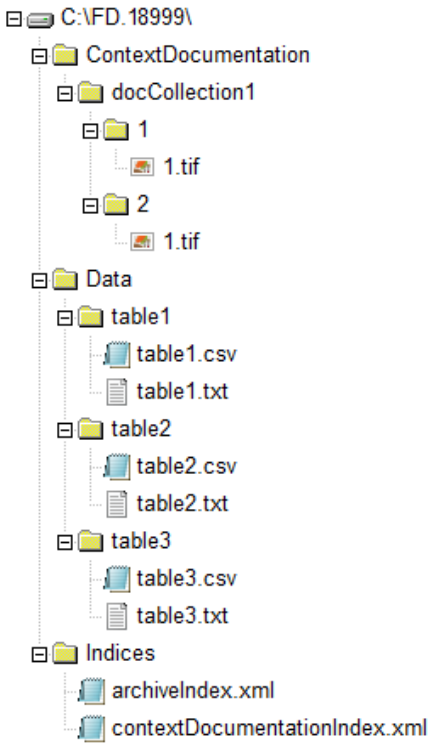
Løbenummeret i afleveringspakkens navn svarer til løbenummeret i det arkiveringsversions-ID, afleveringen får tildelt når afleveringspakken med statistikdata konverteres til en arkiveringsversion af det modtagende arkiv. Dvs. FD.18999bliver til AVID.SA.18999.1.

9.B.3 Afleveringspakkens indhold fordeles i mapper, som angivet i figur 9.2.

9.B.4 Mapperne skal navngives som angivet i figur 9.2.

Der skal være tre undermapper under rodmappen, som navngives henholdsvis *ContextDocumentation*, *Data* og *Indices*.

Figur 9.1 Grafisk oversigt over elementer og struktur i en afleveringspakke



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

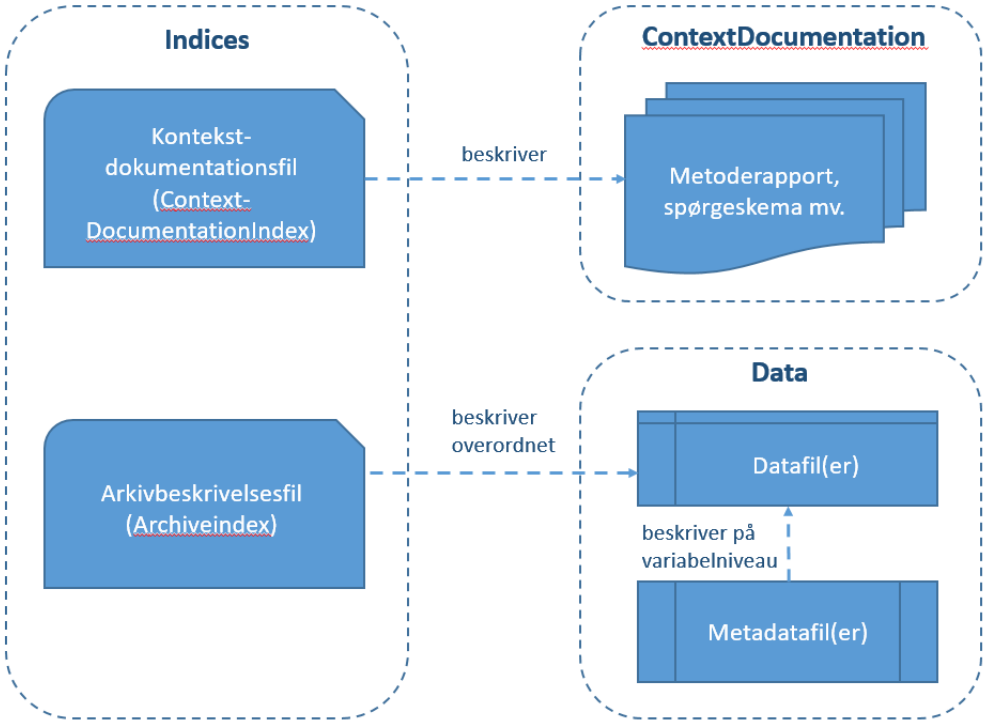
Mappen **Data** indeholder både en datafil og en metadatafil, der begge overholder kravene i bilag 9. Data udtrukket fra de originale statistikfiler skal afleveres i en semikolonsepareret csv-fil (**table1.csv**). Metadata udtrukket fra statistikfilen, f.eks. variabelbeskrivelser (variable labels), svarkategorier (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.

Figur 9.2 Afleveringspakkens mapper

Navn på mappe	Indhold
ContextDocumentation	Kontekstdokumentation, jf. 9.D
Data	Datafil og metadatafil, jf. 9.E
Indices	Indeksfiler, jf. 9.C

Afleveringspakkens tre mapper og sammenhængen mellem dem



9.C. Mappen Indices

9.C.1 Mappen *Indices* skal indeholde følgende indeksfiler med oplysninger om afleveringspakken og dens indhold:

- **archiveIndex.xml**
- **contextDocumentationIndex.xml**

Disse to indeksfiler skabes af myndigheden og godkendes af det modtagende arkiv før afleveringspakken produceres. Rigsarkivet stiller indtastningsprogrammer til indeksfilerne til rådighed på www.sa.dk.

9.C.2 Indeksfilerne skal overholde deres tilhørende skema, jf. bilag 8.

Når indeksfilerne udfyldes ved brug af de indtastningsprogrammer, som Rigsarkivet stiller til rådighed på www.sa.dk, valideres indtastningerne automatisk op imod de gældende skemaer for indeksfilerne, både når de udfyldes og gemmes.

9.C.3 Arkivbeskrivelsesfilen archiveIndex.xml skal overholde reglerne i bilag 6 punkt 6.A.

Arkivbeskrivelsesfilen *archiveIndex.xml* er en metadatafil, der indeholder generelle oplysninger om de data, der afleveres. F.eks. datasættets navn, start og slutdato for data, navn på afleverende myndighed, oplysninger om tilgængelighedsfrister osv.

Reglerne vedrørende arkivbeskrivelsesfilen er fremsat i bilag 6, punkt 6.A.1-3 og er beskrevet herunder:

1. Enhver afleveringspakke skal indeholde en arkivbeskrivelsesfil med angivelse af oplysninger iht. Figur 6.1, som vises herunder. Den viste figur er tilpasset aflevering af forskningsdata efter bilag 9. Af figuren fremgår også hvilke oplysninger der er obligatoriske at udfylde samt øvrig vejledning til udfyldelse af oplysningerne.

2. Arkivbeskrivelsesfilen navngives archiveIndex.xml og skal overholde det tilhørende skema, jf. bilag 8.

3. Indholdet af arkivbeskrivelsesfilen fastlægges efter drøftelse med den afleverende myndighed og den modtagende arkiv.

Figur 6.1 Oplysninger i archiveIndex.xml

Vejledning i denne figur er tilpasset forskningsdata der afleveres efter bilag 9.

Elementnavn	Betegnelse i figur 6.1.a	Beskrivelse	Udfaldsrum	Fore-komst	Obliga-torisk	Vejledning
archiveInforma-tionPackageID	Afleveringspak-keløbenummer	Entydigt ID som tildeles afleveringspakken af Rigsarkivet, som består af arkivkode og løbenummer.	**Afleveringspakkeløbenummer, som defineret i denne bekendtgørelse**	**1**	**Ja**	Afleveringspakkeløbenummer oplyses af det modtagende arkiv, f.eks. 12345 Arkivkoden er f.eks. SA (= Statens Arkiver), hvis det modtagende arkiv f.eks. er Rigsarkivet.
archiveInforma-tionPackageID-Previous	Tidligere afleveringer (f.eks. kohorte)	Entydigt ArkiveringsversionsID på tidligere aflevering fra samme forskningsprojekt.	**ArkiveringsversionsID, som defineret i denne eller tidligere bekendtgørelse.**	**0-m**	**Nej**	Oplyses af det modtagende arkiv. ArkiveringsversionsID (f.eks. AVID.SA.12345) på tidligere afleveringer af forskningsprojektet med kontinuerlig dataindsamling (f.eks. kohorter og dataindsamling ifm overvågning af fænomener).
archivePeriod-Start	Dataindsamling startdato	Startdato for indsamlingen af de afleverede data.	**År, år-måned, eller år-måned-dag.**	**1**	**Ja**	Dateringen skal angives så præcist som muligt, f.eks. 2006-01-01.
archivePeriodEnd	Dataindsamling slutdato	Slutdato for indsamling af de afleverede data.	**År, år-måned, eller år-**	**1**	**Ja**	Dateringen skal angives så præcist som muligt, f.eks. 2010-12-31.

archiveInformationPacketType	Slutaflevering	Angivelse af, om afleveringspakken er sidste aflevering af data, herunder også sidste dataaflevering fra et forskningsprojekt med kontinuerlig dataindsamling over en længere periode (evt. kohorte).	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis datasættet på tidspunktet for produktion af afleveringspakken er lukket for ny dataindsamling/opdateringer. Dette gælder f.eks. den afsluttende dataindsamling fra en kohorteundersøgelse. Har ligeledes værdien *true*, hvis Rigsarkivet på tidspunktet for produktion af afleveringspakken har truffet bestemmelse om, at der ikke fremover skal afleveres data fra forskningsprojektet. Har værdien *false* i øvrige tilfælde.
creatorName	Forskningsinstitutionen og Primærforsker	Både forskningsinstitutionen og primærforsker, som har skabt datasættet.	**Fri tekst**	**1-m**	**Ja**	Det aftales med det modtagende arkiv, hvilke dataskabere der skal anføres.
creationPeriodStart	Dataindsamlings startdato for hver dataskaber	Det angives hvornår data indsamlingen er påbegyndt for hver af de dataskabere/institutioner, der har bidraget med indsamling af data til projektet.	**År, år-måned, eller år-måned-dag**	**1-m**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2008-06-01.
creationPeriodEnd	Dataindsamlings slutdato for hver dataskaber	Det angives hvornår data indsamlingen er afsluttet for hver af de dataskabere/institutioner, der har bidraget med indsamling af data til projektet.	**År, år-måned, eller år-måned-dag**	**1-m**	**Ja**	Datering skal være så præcis som mulig, f.eks. 2010-12-31.
archiveType	Arkivtype	Angivelse af om der er tale om en afsluttet aflevering eller en kontinuerlig aflevering.	**Boolsk værdi**	**1**	**Ja**	Oplyses af det modtagende arkiv. Har værdien *true*, hvis der er tale om en afsluttet aflevering *(= afsluttet arkivperiode*).

						Det kan f.eks. være: en tværsnitsundersøgelse af befolkningens holdninger til diverse emner eller en afsluttet naturvidenskabelig observationsundersøgelse. Har værdien *false (=Øjebliksbillede)*, hvis der er tale om en kontinuerlig aflevering, hvor dataindsamling er kontinuerlig over en længere periode og hvor der skal ske kontinuerlig aflevering af data i form af et øjebliksbillede, hvor de afleverede data også vil indgå i næste afleveringspakke. Det kan f.eks. være en kohorteundersøgelse eller forløbsundersøgelser eller dataindsamling ifm. overvågning af fænomener, hvor der løbende akkumuleres data.
systemName	Forskningspro-jektets titel	Den officielle danske titel på forskningsprojek-tet, hvor alle forkortelser er opløst.	**Fritekst**	**1**	**Ja**	Forskningsprojektets titel på dansk.
alternativeName	Forskningspro-jektets engelske titel	Den officielle engelske titel på forskningsprojek-tet, hvor alle forkortelser er opløst	**Fritekst**	**0-m**	**Nej**	Forskningsprojektets titel på engelsk skal angives. Andre kaldenavne eller forkortelser for forskningsprojektet kan også angives.
systemPurpose	Forskningspro-jektets formål	Beskrivelse af forskningsprojektets formål (evt. = forskningsprojektets abstract)	**Fritekst**	**1**	**Ja**	Forskningsprojektets formål. F.eks. forskningsprojektets abstract. Beskrivelsen skal redegøre for projektets centrale problemstillinger.

						Dokumenter med detaljerede oplysninger indgår i kontekstdokumentation.
systemContent	Beskrivelse af central population og centrale variable	Beskrivelse af den centrale population og centrale variable i data	**Fritekst**	**1**	**Ja**	Beskrivelsen skal besvare spørgsmålet om, hvem eller hvad (= den centrale population) der primært registreres hvilke data om (= de centrale variable). F.eks. et forskningsprojekt befolkningen i Danmark (= population) undersøges for holdning til Brexit (der vil være en vifte af centrale variable) eller et forskningsprojekt om størrelsen på ål i Lillebælt, hvor ål = population og centrale variable = længde og vægt.
regionNum	Regionsnumre	Angivelse af, om der i systematisk form er registreret regionsnumre i datasættet	**Boolsk værdi**	**1**	**Ja**	For hver af de angivne identifikatorer skal det oplyses, om den er registreret i datasættet. Har værdien *true*, hvis den pågældende identifikator er registreret systematisk. Har værdien *false*, hvis den pågældende identifikator ikke er registreret systematisk. Hvis der i datasættet findes andre fagspecifikke identifikatorer, der anvendes i flere datasæt eller registre, f.eks. et geografisk referencesystem, Dansk branchekode, eller beskæftigelseskoder for baggrundsvariable, diagnosekoder eller medicinkoder kan det angives i forbindelse med oplysninger om central population og centrale variable (Beskrivelse af central population og centrale variable i *systemContent*).
komNum	Kommune-numre	Angivelse af, om der i systematisk form er registreret kommunenumre i datasættet	**Boolsk værdi**	**1**	**Ja**	
cprNum	CPR-numre		**Boolsk værdi**	**1**	**Ja**	

		Angivelse af, om der i systematisk form er registreret CPR-numre i datasættet				
cvrNum	CVR-numre	Angivelse af, om der i systematisk form er registreret CVR-numre i datasættet	**Boolsk værdi**	**1**	**Ja**	
matrikNum	Matrikelnumre	Angivelse af, om der i systematisk form er registreret matrikelnumre i datasættet	**Boolsk værdi**	**1**	**Ja**	
bbrNum	BBR-numre	Angivelse af, om der i systematisk form er registreret BBR-numre i datasættet	**Boolsk værdi**	**1**	**Ja**	
whoSygKod	WHOs sygdomskoder	Angivelse af, om der i systematisk form er registreret WHO-sygdomskoder i datasættet	**Boolsk værdi**	**1**	**Ja**	
sourceName	Evt. registre/andre datasæt der er udtrukket data fra	Eventuelle registre der er udtrukket data fra/ andre datasæt der er udtrukket data fra (opslag, overførsel, samkøring osv.)	**Fritekst**	**0-m**	**Nej**	Beskrivelsen skal besvare spørgsmålet om der er anvendt data udtrukket fra registre eller andre datasæt og fra hvilke. I kontekstdokumentationen kan indgå dokumenter med detaljerede oplysninger, f.eks. om hvilke data (kolonner) der er overført fra andre registre eller datasæt og hvordan overførsel er sket.

userName	Evt. andre projekter, der anvender dette projekts data	Eventuelle andre forskningsprojekter, der anvender dette projekts data	**Fritekst**	**0-m**	**Nej**	Beskrivelsen skal besvare spørgsmålet om, hvorvidt data anvendes i andre forskningsprojekter samt i hvilke. I kontekstdokumentationen kan indgå dokumenter med detaljerede oplysninger, f.eks. om hvilke data (kolonner) der er anvendt i andre forskningsprojekter, og hvordan data er anvendt.
predecessorName	Opfølgning på eksisterende forskningspro-jekt	Eksisterende forskningsprojekt som dette forskningsprojekt følger op på.	**Fritekst**	**0-m**	**Nej**	Hvis der er tale om en opfølgende forskningsundersøgelse til et tidligere udført forskningsprojekt, skal titlen på det tidligere udførte forskningsprojekt angives.
archiveApproval	Arkiv der godkender afleveringspak-ken (f.eks. SA)	Angivelse af, hvilket offentligt arkiv, der godkender afleveringspakken	**Identifikation af det pågældende arkiv (2-4 tegn)**	**1**	**Ja**	Oplyses af det modtagende arkiv. Hvis det modtagende arkiv er Rigsarkivet skrives forkortelsen SA (= Statens Arkiver).
personalData-RestrictedInfo	Personhenfør-bare oplysninger	Angivelse af, om der i data findes personfølsomme oplysninger i henholdt il databeskyttelsesforor dningen	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis der i data findes følsomme personoplysninger (oplysninger om racemæssig eller etnisk baggrund, politisk, religiøs eller filosofisk overbevisning, fagforeningsmæssige tilhørsforhold, oplysninger om helbredsmæssige og seksuelle forhold samt oplysninger om strafbare forhold og væsentlige sociale problemer).
otherAccessType Restrictions	Særlige tilgængeligheds-bestemmelser	Angivelse af, om der i data og den tilhørende dokumentation findes oplysninger, der kan betinge længere tilgængelighedsfrist i øvrigt	**Boolsk værdi**	**1**	**Ja**	Har værdien *true*, hvis tilgængelighedsfristen er forlænget. Hvis den afleverende myndighed ønsker forlængelse af tilgængelighedsfristen, skal det i forbindelse med aflevering drøftes med det modtagende arkiv, jf. lovbekendtgørelse nr. 1035 af 21. august 2007 (Arkivloven), § 27.

archiveRestrictions	Beskrivelse af betingelser for adgang til data	Angivelse af nærmere bestemmelser for adgang til materialet. Feltet kan anvendes efter det modtagende arkivs nærmere retningslinjer.	**Fritekst**	**0-1**	**Nej**	Det modtagende arkiv træffer aftale med den afleverende myndighed om, hvorvidt og med hvilke oplysninger dette felt skal udfyldes.

Figur 6.1.a Skærmbilleder af Rigsarkivets indtastningsprogram til udfyldning af arkivbeskrivelsesfilen jf. bilag 9

Skab ArchiveIndex ver. 2.2.7275

Filer

Archive Index | Html

Afleveringsregler

- Arkiveringsversion jvf. bilag 1-8 i bekendtgørelse
- Forskningsdata afleveret i henhold til bilag 9 i bekendtgørelse

Afleveringspakke

Afleveringspakkeløbenummer — Arkivkode (2-4 bogstaver) — Løbenummer

Dataindsamling startdato — AAAA MM DD

Dataindsamling slutdato — AAAA MM DD

Slutaflevering

Forskningsinstitutionen og Primærforsker

Dataindsamlings startdato for hver dataskaber — AAAA MM DD

Dataindsamlings slutdato for hver dataskaber — AAAA MM DD

Dataskabere — Tilføj | Opdater | Fjern

Arkivtype — Afsluttet aflevering — Kontinuerlig aflevering (øjebliksbillede)

Tidligere afleveringer (fx kohorte) — Tilføj | Opdater | Fjern

Forskningsprojektet

Forskningsprojektets titel

Forskningsprojektets engelske titel

Tilføj | Opdater | Fjern

Eventuelle alternative navne

Forskningsprojektets formål

Beskrivelse af central population og centrale variable

Afleveringspakken indeholder

Regionsnumre — Matrikelnumre

Kommune-numre — BBR-numre

CPR-numre — WHOs sygdomskoder

CVR-numre

Afleveringspakkens datakilder

Evt. registre/andre datasæt der er udtrukket data fra — Tilføj | Opdater | Fjern

Evt. andre projekter, der anvender dette projekts data

Evt. andre projekter, der anvender dette projekts data — Tilføj | Opdater | Fjern

Opfølgning på eksisterende forskningsprojekt

Opfølgning på eksisterende forskningsprojekt — Tilføj | Opdater | Fjern

Adgangsoplysninger

Arkiv der godkender afleveringspakken (fx SA)

Personhenførbare oplysninger

Særlige tilgængelighedsbestemmelser

Beskrivelse af betingelser for adgang til data

Gem

Nyt archiveIndex

9.C.4 Kontekstdokumentationsfilen contextDocumentationIndex.xml skal overholde reglerne i bilag 4 punkt 4.C.4.a samt 4.C.4.b.

Kontekstdokumentationsfilen *contextDocumentationIndex.xml* er en metadatafil, der indeholder oplysninger om kontekstdokumenterne, der vedlægges afleveringspakken jf. 9.D., f.eks. dokumenttitel, dokumentdato, forfatter og dokumentkategori. Titel og dokumentkategori er obligatorisk at udfylde.
Reglerne vedrørende kontekstdokumentationsfilen er fremsat i bilag 4, punkt 4.C.4.a, 4.C.4.b samt 6.B. og er beskrevet herunder:

1. Enhver afleveringspakke skal indeholde kontekstdokumenter, som på et overordnet plan beskriver den kontekst data er skabt i. Det kan bl.a. bestå af en projektbeskrivelse, en metoderapport, et eksempel på et spørgeskema etc., jf. 6.B.1.

2. Det modtagende arkiv fastlægger efter drøftelse med den afleverende myndighed, hvilke kontekstdokumenter der skal afleveres, herunder hvilke punkter, specificeret i bekendtgørelsens figur 6.2, som ikke er relevante at dokumentere i den konkrete aflevering, jf. 6.B.2.

   Hvis der ikke er en tilstrækkelig dokumentation, skal den udarbejdes i forbindelse med produktion af afleveringspakken. I henhold til arkivlovgivningen er kontekstdokumentationen som udgangspunkt belagt med en tilgængelighedsfrist på 20 år. Det betyder, at såfremt en arkivbruger ønsker at gøre brug af materialet, før der er gået 20 år, skal den afleverende myndighed give sit samtykke, før materialet kan udleveres, jf. § 33 i arkivloven.

3. Kontekstdokumentationsfilen *contextDocumentationIndex.xml* skal indeholde et indeks over de kontekstdokumenter, som findes i afleveringspakkens kontekstdokumentation, jf. 4.C.4.a.

4. For hvert kontekstdokument skal der i kontekstdokumentationsfilen angives specifikke oplysninger[1], jf. 4.C.4.b og 6.B.3.b og vælges én eller flere dokumentkategorier[2], jf. 6.B.3.a og figur 6.2.

   Oplysninger i kontekstdokumentationsfilen:

   Aktuelt dokument
   - 'Dokumenttitel' (Obligatorisk): f.eks. *Projektbeskrivelse*
   - 'Dokumentbeskrivelse': f.eks. *Projektbeskrivelse af forskningsprojektet*
   - 'Dato': f.eks. *2019-01-04*

   Forfattere
   - 'Forfatternavn': f.eks. *Professor Erik Eriksen* (udlades ved Rigsarkivets Afleveringsbestemmelse)
   - 'Forfatterinstitution': f.eks. *Københavns Universitet* (Ved dokumentet 'Rigsarkivet Afleveringsbestemmelse' skrives her 'Rigsarkivet').

   Dokumentkategorier
   - Der skal være valgt mindst én kategori for hvert dokument.

   Det er primært dokumentkategorierne 7.a-7.e i figur 6.2, der er relevante for forskningsdata. Dog kan andre dokumentkategorier være relevante, f.eks. skal Rigsarkivets Afleveringsbestemmelse' tilknyttes dokumentkategorien *3.a 'Arkivets bestemmelse, herunder afleveringsbestemmelse'*.

[1] Oplysningerne fremgår også af figur 4.3 i bilag 4.
[2] Dokumentkategorierne fremgår også af figur 6.2 i bilag 6.

Figur 6.2 Dokumentkategorier i contextDocumentationIndex.xml
Primært dokumentkategorierne 7.a-7.e er relevante for forskningsdata

1. Dokumentation vedrørende administrativ brug af it-systemet		
	Elementnavn	**Beskrivelse**
1.a	systemPurpose	It-systemets formål
1.b	systemRegulations	It-systemets lov- og regelgrundlag
1.c	systemContent	It-systemets indhold, population og særlige begreber
1.d	systemAdministrativeFunctions	It-systemets administrative funktioner
1.e	systemPresentationStructure	It-systemets præsentationsstruktur
1.f	systemDataProvision	Tilvejebringelse af data
1.g	systemDataTransfer	Videregivelse af data
1.h	systemPreviousSubsequentFunctions	Data og funktioner fælles med forgænger- og efterfølgersystemer
1.i	systemAgencyQualityControl	Myndighedens egen kvalitetskontrol
1.j	systemPublication	Publikation af og om data
1.k	systemInformationOther	Andet
1.l	systemTaxonomy	Registreringssystematik
1.m	systemInstruction	Instruks for anvendelse af systemet
2. Dokumentation vedrørende it-systemets tekniske udformning, drift og udvikling		
	Elementnavn	**Beskrivelse**
2.a	operationalSystemInformation	Driftsversionens opbygning
2.b	operationalSystemConvertedInformation	Konvertering hos myndigheden
2.c	operationalSystemSOA	Dokumentation af sammensætning af data og eventuelle dokumenter fra flere forskellige it-systemer i en serviceorienteret arkitektur
2.d	operationalSystemInformationOther	Andet
3. Dokumentation vedrørende arkivskabers aflevering af data		
	Elementnavn	**Beskrivelse**
3.a	archivalProvisions	Arkivets bestemmelser herunder afleveringsbestemmelse
3.b	archivalTransformationInformation	Dokumentation af konvertering fra driftsversion til arkiveringsversion
3.c	archivalInformationOther	Andet
4. Dokumentation vedrørende arkivets modtagelse af data (udfyldes af modtagende arkiv)		
	Elementnavn	**Beskrivelse**
4.a	archivistNotes	Arkivarnoter
4.b	archivalTestNotes	Testnoter
4.c	archivalInformationOther	Andet
5. Dokumentation vedrørende arkivets bevaring af arkiveringsversionen (udfyldes af det modtagende arkiv)		
	Elementnavn	**Beskrivelse**
5.a	archivalMigrationInformation	Konvertering hos arkivet
5.b	archivalInformationOther	Andet

6. Anden dokumentation		
	Elementnavn	**Beskrivelse**
6.a	informationOther	Andet
7. Dokumentation af forskningsdata		
	Elementnavn	**Beskrivelse**
7.a	researchProjectDescription	Projektbeskrivelse gældende for de afleverede data
7.b	researchQuestionnaire	Spørgeskema, interviewguide og/eller registreringsskema anvendt til at indsamle og analysere de afleverede data
7.c	researchProtocol	Protokoller og metoderapporter
7.d	researchPublication	Publikationer, som er udgivet på basis af de afleverede data
7.e	researchInformationOther	Andet

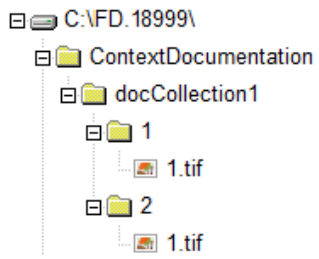
9.D. Mappen *ContextDocumentation*

9.D.1 Mappen *ContextDocumentation* skal indeholde kontekstdokumentation, jf. 4.E.

Kontekstdokumentationsfilen er beskrevet under 9.C.4. Her beskrives udelukkende regler for placering og navngivning af kontekstdokumenterne, jf. 4.E.

Selve kontekstdokumenterne skal placeres i undermappen *ContextDocumentation* i afleveringsmappen.

Illustration af mappen *ContextDocumentation* og dens indhold i afleveringspakken



Reglerne vedrørende placering af kontekstdokumenterne i afleveringspakken er følgende[3]:

Regler til Dokumentsamlingsmappen
Mappen *ContextDocumentation* skal indeholde en undermappe kaldet en dokumentsamlingsmappe. Dokumentsamlingsmappen navngives "docCollection[fortløbende nummer]", begyndende med 1.

En dokumentsamlingsmappe med kontekstdokumentation må indeholde op til 10.000 dokumentmapper, jf. 4.E.2. Der vil derfor sjældent være behov for flere dokumentsamlingsmapper, men opstår behovet skal næste dokumentsamlingsmappe navngives med det næste fortløbende nummer (*docCollection2*), således at navnet på hver dokumentsamlingsmappe er unikt inden for *ContextDocumentation* mappen, jf. 4.E.3.

Regler til dokumentmapperne
I dokumentsamlingsmappen *docCollection1* placeres en eller flere undermapper, kaldet dokumentmapper, hvor der placeres én mappe til hvert kontekstdokument, 4.E.2.

Hvert kontekstdokument er i kontekstdokumentationsfilen tildelt et dokumentID. Når kontekstdokumenterne placeres i afleveringspakkens *docCollection1* mappe, skal de placeres i overensstemmelse med

[3] Reglerne for placeringen af kontekstdokumenterne i afleveringspakken er desuden angivet i bilag 4, punkt 4.E.

rækkefølgen angivet i kontekstdokumentationsfilen, f.eks. hvis kontekstdokumentet *"Afleveringsbestemmelse"* har dokumentID *1* skal det placeres i dokumentmappen navngivet "1", jf. 4.E.4 og 4.E.5.

Regler for kontekstdokumenterne
En dokumentmappe skal indeholde ét dokument, som består af én eller flere filer af samme format, jf. 4.E.5.

Et dokuments fil (eller filer) navngives fortløbende med et nummer, begyndende med 1 samt formatets ekstension, jf. 4.G.6., f.eks. 1.tif. Hvis dokumentet består af flere filer navgives de *1.tif*, *2.tif*, *3.tif* osv.

Dokumenterne skal afleveres i ét af de i arkiveringsversionen tilladte dokumentformater, jf. 6.B.4 samt 5.E.-5.F. Følgende dokumentformater er tilladt:

- .tif (TIFF) og .jp2 (JPEG-2000) til dokumenter og billeder
- .mp3 (MP3) og .wav (WAVE) til lydfiler
- .mpg (MPEG-2 og MPEG-4) til video

Det er vigtigt, at dokumenterne konverteres efter reglerne efter bekendtgørelsens bilag 5.E. og 5.F., som er beskrevet herunder. Se også Rigsarkivets vejledning i konvertering af dokumenter til TIFF på Rigsarkivets hjemmeside www.sa.dk.

5.E. Digitale dokumenter

5.E.1 Et digitalt dokument, jf. dog 5.F. og 5.G., skal lagres i et af følgende formater:

- **det grafiske bitmapformat TIFF, version 6.0 baseline.**
- **JPEG-2000 efter standarden ISO/IEC 15444-1:2004. Information technology - JPEG 2000 image coding system - Part 1: Core coding system.**

5.E.1.a Det er tilladt at benytte begge formater inden for samme arkiveringsversion.

5.E.2 Dokumenter i TIFF skal komprimeres efter følgende kompressionsregler:

5.E.2.a Sort/hvide dokumenter skal komprimeres med CCITT/TSS Grp3, Grp4, PackBit eller LZW.

5.E.2.b Dokumenter med gråtoner eller farver skal komprimeres med PackBit eller LZW.

5.E.3 Dokumenter i TIFF RGB må udelukkende benytte følgende bitdybder: 1, 2, 4, 8, 24 og 32.

5.E.3.a Dokumenter i TIFF RGB må maksimalt benytte 3 farvekanaler med en maksimal bitdybde på 24 bit (8x8x8 bit) evt. suppleret med maksimalt en alfakanal (8 bit), således at den samlede bitdybde for en billedfil ikke kan overstige 32 bit.

5.E.4 Dokumenter i TIFF CMYK må udelukkende benytte følgende bitdybder: 1, 2, 4, 8, 32 og 40.

5.E.4.a Dokumenter i TIFF CMYK må maksimalt benytte 4 farvekanaler med en maksimal bitdybde på 32 bit (8x8x8x8 bit) evt. suppleret med maksimalt én alfakanal (8 bit), således at den samlede bitdybde for en billedfil ikke kan overstige 40 bit.

5.E.5 TIFF dokumenters anvendelse af XResolution og YResolution (TIFF Tag 282 og 283), skal ske på en sådan måde, at forholdet mellem bredde og højde ved anvendelse af disse værdier (skalering) svarer til sidernes dimensioner i det oprindelige dokument.

Der kan anvendes to forskellige formater til arkivering af dokumenter, TIFF (version 6.0 baseline) og JPEG2000:

- TIFF 1 bit med kompression Grp4 samt LZW er velegnet til sort/hvide dokumenter
- TIFF 4-24 bit med kompression LZW er velegnet til almindelige dokumenter, som indeholder tegninger og fotos i varierede størrelse og antal
- TIFF 24 bit er velegnet til tegninger, kort, fotos etc., hvor den eksakte bevaring af dokumentet er af betydning. Som eksempel kan nævnes satellitfoto af afgrøder, hvor tolkningen af farven på de enkelte pixels kan være af betydning
- JPEG-2000 er overordentlig velegnet til store tegninger, kort, fotos etc., såfremt farven på de enkelte pixels er af mindre betydning. Som eksempel kan nævnes indscannede matrikelkort eller affotograferede skibstegninger

Kvaliteten af den grafiske kopi skal svare til den kvalitet, som myndigheden selv anvendte, hvilket for digitalt fødte dokumenter svarer til en passende udskriftkvalitet (f.eks. 300 DPI). For indscannet materiale skal kvaliteten modsvare kvaliteten på det materiale, som danner grundlag for kopien.

Valg af kompressionstype og bitdybde (antal mulige farver) må gerne variere fra side til side inden for samme dokument.

Myndigheden skal sikre sig, at dokumenterne også efter konvertering til arkivformatet er læsbare f.eks. ved at kontrollere at kommaer fremstår korrekt i dokumentet, og ikke ligner punktummer. Hvis det er tilfældet, er opløsningen for lav, og dokumentet skal lagres med en højere opløsning.

Rigsarkivet anbefaler anvendelsen af Grp4 og LZW-kompression eftersom dokumenter komprimeret med disse algoritmer fylder mindre end hvis der benyttes Grp3 og PackBit.

5.F. Lyd og video

5.F.1 Lydfiler skal lagres efter standarden MP3 DS/EN ISO/IEC 11172-3.

5.F.2 Det modtagende arkiv kan tillade, at lydfiler afleveres i formatet WAVE LPCM som specificeret i *Multimedia Programming Interface and Data Specifications 1.0. IBM Corporation and Microsoft Corporation, August 1991*. Dog begrænset til bitdybder, der er hele multipla af 8.

5.F.3 Videofiler skal lagres efter en af følgende standarder:

- **MPEG-2 DS/EN ISO/IEC 13818-2. Eventuel lyd indkodes som MP3, som specificeret i ISO/IEC 13818-3.**
- **MPEG-4 AVC DS/EN ISO/IEC 14496-10 (ITU-T H.264). Video indkodes som specificeret i ISO/IEC 14496-10. Eventuel lyd indkodes som AAC, som specificeret i ISO/IEC 14496-3. Video og lyd indpakkes i MPEG-4 formatet som defineret i ISO/IEC 14496-14.**

Kontekstdokumenter i form af lyd og video skal konverteres til de anviste formater på en sådan måde, at den oprindelige kvalitet i størst muligt omfang beholdes.

Der kan anvendes to forskellige lydformater ved aflevering til offentligt arkiv, MP3 og WAVE. Det vil normalt være MP3, som skal anvendes, men hvis myndigheden selv har anvendt WAVE kan det modtagende arkiv tillade, at WAVE anvendes som arkiveringsformat. Det samme gælder for lydoptagelser, hvor en meget præcis lydgengivelse er vigtig. Et eksempel på dette er optagelser, hvor baggrundslyde har betydning for forståelsen, og konteksten for den pågældende optagelse.

9.E. Mappen Data

9.E.1 En afleveringspakke skal indeholde et eller flere datasæt. Hvert datasæt skal bestå af en datafil og en tilhørende metadatafil.

Hvis aflevering af forskningsdata kun består af én statistikfil svarer det til ét datasæt. Hvis et datasæt består af flere statistikfiler, vil afleveringspakken bestå af flere datasæt med et datasæt for hver statistikfil.

Datafilen indeholder data fra statistikfilen, regnearket, csv-filen eller tilsvarende.

Metadatafilen indeholder metadata fra selve statistikfilen eller regnearket, f.eks. variabelnavne, variabelbeskrivelser og value labels. Derudover indeholder den informationerne SYSTEMNAVN, DATAFILNAVN og DATAFILBESKRIVELSE samt eventuelt NØGLEVARIABEL og REFERENCE som indtastes af arkivskaberen, jf. figur 9.4.

Hvis der findes vigtige informationer i statistikfilen, som ikke kan udtrækkes til metadatafilen i afleveringspakken, f.eks. noter om hvilken kilde en variabel stammer fra i notefelter til variable, kan disse informationer placeres i et kontekstdokument i afleveringspakken.

9.E.2 Datafil og metadatafil placeres i mappen Data i en undermappe, der navngives efter bilag 4 punkt 4.D.2.a samt 4.D.2.b.

9.E.2.a Datafil navngives med undermappens navn efterfulgt af ekstensionen ».csv«.

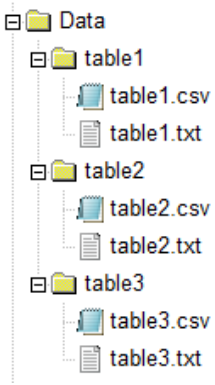
9.E.2.b Metadatafil navngives med undermappens navn efterfulgt af ekstensionen ».txt«.

Hver datafil og dennes tilhørende metadatafil placeres i en undermappe i mappen *Data*.

Undermappen navngives tabel [fortløbende nummer], f.eks. table1. Den fortløbende nummerering begynder med 1. Foranstillede nuller må ikke anvendes, jf. 4.D.2.a samt 4.D.2.b.

Datafilen navngives med undermappens navn, efterfulgt af ekstensionen ».csv«, f.eks.: tabel 1.csv.
Metadatafil navngives med undermappens navn, efterfulgt af ekstensionen ».txt«, f.eks.: tabel 1.txt.

Illustration af mappen *Data* og dens indhold i afleveringspakken



9.F. Tekstformat

9.F.1 Tegnsættet for henholdsvis datafil og metadatafil skal være indkodet som UTF-8, som angivet i ISO/IEC 10646:2003 Annex D og som beskrevet i *The Unicode Standard 5.1*, kapitel 3, og skal i øvrigt overholde bestemmelserne i bilag 5 punkt 5.D.1.b – 5.D.1.d.

Tegnsættet for både data- og metadatafilen skal være indkodet som UTF-8. I Rigsarkivets særskilte vejledning om UTF-8 kan du læse mere om hvad UTF-8 indkodning er, hvordan du tjekker om tegnsættet er indkodet som UTF-8 samt hvordan du indkoder tegnsæt som UTF-8 i statistikprogrammerne SAS, Stata og SPSS eller i en teksteditor som NotePad++.

Rigsarkivets udtræksværktøj ASTA, der kan anvendes til at skabe og teste afleveringspakken, tester ikke automatisk for om tegnsættet i data- og metadatafilen er indkodet som UTF-8. Hvis dette ikke er tilfældet skal det opdages i den manuelle visuelle test af afleveringspakken, f.eks. ved at teste om nationale karakterer som æ, ø og å vises korrekt.

9.G. Datafil

9.G.1 Datafil afleveres som en semikolonsepareret tekstfil i henhold til RCF 4180 Common Format and MIME Type for CSV Files, der overholder syntaksen beskrevet som EBNF i figur 9.12.

Datafilen skal afleveres som en semikolonsepareret tekstfil, der overholder syntaksen beskrevet som EBNF i figur 9.12. EBNF'en overholder men begrænser også csv-standarden RCF4180.

Eksempel på datafil (table1.csv) der overholder EBNF i figur 9.12

table1.csv - Notesblok

Filer Rediger Formater Vis Hjælp

```
barn_id;koen;vaegt_idag;slut_dato;hobby;laegebesoeg;maaling;aarsag
11;2;78.2;2017/01/08;4;30-MAR-2002 15:31:22;12:02:57;blødning i urinen
204;1;52.3;2018/05/11;10;03-SEP-2008 12:00:01;4:31:36;smerte når jeg går på toilettet
10095;1;83.1;2013/02/02;9;01-MAR-1999 05:10:45;1:42:20;svaghed
10098;1;56.0;2017/02/02;4;15-NOV-2009 11:15:00;15:46:35;svært at trække vejret
10115;1;74.5;2013/08/28;5;02-JAN-1996 10:44:20;11:15:55;træt hele tiden
10116;1;80.8;2014/12/20;4;31-OCT-2001 23:51:30;18:23:05;"ingen symptomer; men vil gerne tjekkes"
10118;1;102.4;2013/02/18;5;10-JUN-1965 01:26:10;10:01:06;træthed
10119;2;60.2;2015/07/16;2;06-NOV-2008 20:07:30;18:42:26;besvimer
10120;1; ;2018/01/11;1;18-APR-1992 18:00:01;22:34:57;konstant mangel på vitaminer
10121;1;80.0;2017/02/10;3; ;20:06:18;vil gerne tjekkes
10125;1;95.2;2018/07/08;10;03-SEP-2001 02:10:01;6:44:57;smerte
1013;2;64.5;2017/10/09;3;01-FEB-1995 08:20:45;8:55:41;blødning i urinen
10139;1;87.7;2013/08/28;3;30-AUG-2012 18:10:22;8:45:18;"får ""kvalme"" når jeg spiser"
10140;2;50.1;2014/12/20;3;03-NOV-2018 11:22:01;15:48:36;kan ikke trække vejret
10142;2;56.8;2017/02/07;4;01-JAN-1991 07:15:56;19:42:31;træthed
12906;9;90.8;2014/11/08;9;30-MAY-2001 15:31:22;15:57:57;oppustet mave
16434;1;96.7;2015/02/28;9;03-DEC-2010 10:00:01;6:26:36;konstant blødning - menstruation
4520;1;59.1;2017/11/13;5;30-OCT-2009 17:31:20;17:57:55;smerte i maven
4574;1; ;2013/09/23;4;03-APR-1998 17:10:01;6:01:35;fejler ikke noget
9382;2;49.1;2018/05/11;3;01-JUL-1996 08:10:45;12:37:20;lavt immunforsvar
```

9.G.1.a Første linje i datafilen skal altid angive alle variabelnavne, angivet i samme rækkefølge som i metadatafilen.

Første linje i datafilen skal angive variabelnavne på alle variable i den oprindelige statistikfil eller regneark, som dataudtrækket er lavet fra. Rækkefølgen af variabelnavne i datafilen skal stemme overens med rækkefølgen af variabelnavne angivet i metadatafilen under etiketten VARIABEL.

Variabelnavne skal i datafilen adskilles af semikolon, men der anvendes ikke semikolon efter sidste variabelnavn.

Variabelnavne skal overholde SQL99 standarden (ISO/IEC 9075:1999 - Database Language SQL) for angivelse af tabel- og kolonnenavne. Dvs. at variabelnavn må aldrig begynde med et tal, men må godt være en blanding af bogstaver, tal samt underscore på max 128 tegn. Hvis variabelnavnet er et reserveret ord i SQL:1999, skal det enten omdøbes eller omkranses af dobbeltapostrof.

Reserverede ord i SQL 99 standarden

<reserved word> ::=

```
ABSOLUTE | ACTION | ADD | ADMIN | AFTER | AGGREGATE | ALIAS | ALL | ALLOCATE |
ALTER | AND | ANY | ARE | ARRAY | AS | ASC | ASSERTION | AT | AUTHORIZATION |
BEFORE | BEGIN | BINARY | BIT | BLOB | BOOLEAN | BOTH | BREADTH | BY
| CALL | CASCADE | CASCADED | CASE | CAST | CATALOG | CHAR | CHARACTER | CHECK |
CLASS | CLOB | CLOSE | COLLATE | COLLATION | COLUMN | COMMIT | COMPLETION |
CONNECT | CONNECTION | CONSTRAINT | CONSTRAINTS || CONSTRUCTOR | CONTINUE |
CORRESPONDING | CREATE | CROSS | CUBE | CURRENT | CURRENT_DATE | CURRENT_PATH |
CURRENT_ROLE | CURRENT_TIME | CURRENT_TIMESTAMP | CURRENT_USER | CURSOR | CYCLE
```

| DATA | DATE | DAY | DEALLOCATE | DEC | DECIMAL | DECLARE | DEFAULT | DEFERRABLE | DEFERRED | DELETE | DEPTH | DEREF | DESC | DESCRIBE | DESCRIPTOR | DESTROY | DESTRUCTOR | DETERMINISTIC | DICTIONARY | DIAGNOSTICS | DISCONNECT | DISTINCT | DOMAIN | DOUBLE | DROP | DYNAMIC | EACH | ELSE | END | END-EXEC | EQUALS | ESCAPE | EVERY | EXCEPT | EXCEPTION | EXEC | EXECUTE | EXTERNAL | FALSE | FETCH | FIRST | FLOAT | FOR | FOREIGN | FOUND | FROM | FREE | FULL | FUNCTION | GENERAL | GET | GLOBAL | GO | GOTO | GRANT | GROUP | GROUPING | HAVING | HOST | HOUR | IDENTITY | IGNORE | IMMEDIATE | IN | INDICATOR | INITIALIZE | INITIALLY | INNER | INOUT | INPUT | INSERT | INT | INTEGER | INTERSECT | INTERVAL | INTO | IS | ISOLATION | ITERATE | JOIN | KEY | LANGUAGE | LARGE | LAST | LATERAL | LEADING | LEFT | LESS | LEVEL | LIKE | LIMIT | LOCAL | LOCALTIME | LOCALTIMESTAMP | LOCATOR | MAP | MATCH | MINUTE | MODIFIES | MODIFY | MODULE | MONTH | NAMES | NATIONAL | NATURAL | NCHAR | NCLOB | NEW | NEXT | NO | NONE | NOT | NULL | NUMERIC | OBJECT | OF | OFF | OLD | ON | ONLY | OPEN | OPERATION | OPTION | OR | ORDER | ORDINALITY | OUT | OUTER | OUTPUT | PAD | PARAMETER | PARAMETERS | PARTIAL | PATH | POSTFIX | PRECISION | PREFIX | PREORDER | PREPARE | PRESERVE | PRIMARY | PRIOR | PRIVILEGES | PROCEDURE | PUBLIC | READ | READS | REAL | RECURSIVE | REF | REFERENCES | REFERENCING | RELATIVE | RESTRICT | RESULT | RETURN | RETURNS | REVOKE | RIGHT | ROLE | ROLLBACK | ROLLUP | ROUTINE | ROW | ROWS | SAVEPOINT | SCHEMA | SCROLL | SCOPE | SEARCH | SECOND | SECTION | SELECT | SEQUENCE | SESSION | SESSION_USER | SET | SETS | SIZE | SMALLINT | SOME| SPACE | SPECIFIC | SPECIFICTYPE | SQL | SQLEXCEPTION | SQLSTATE | SQLWARNING | START | STATE | STATEMENT | STATIC | STRUCTURE | SYSTEM_USER | TABLE | TEMPORARY | TERMINATE | THAN | THEN | TIME | TIMESTAMP | TIMEZONE_HOUR | TIMEZONE_MINUTE | TO | TRAILING | TRANSACTION | TRANSLATION | TREAT | TRIGGER | TRUE | UNDER | UNION | UNIQUE | UNKNOWN | UNNEST | UPDATE | USAGE | USER | USING | VALUE | VALUES | VARCHAR | VARIABLE | VARYING | VIEW | WHEN | WHENEVER | WHERE | WITH | WITHOUT | WORK | WRITE | YEAR | ZONE

9.G.1.b Hvis semikolon indgår i en værdi for en variabel, skal hele værdien omsluttes 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 omsluttes med dobbelt apostroffor.

Hvis semikolon indgår i en værdi for en variabel i et tekstfelt, skal hele teksten omsluttes med dobbelt apostrof (dobbelte anførselstegn også kaldet gåseøjne).

Eksempel på udtræk af en variabelværdi der indeholder semikolon

Variabelværdien *Antal hjorte; 6* skal i datafilen angives som *”Antal hjorte; 6”*

Hvis dobbeltapostrof indgår i en værdi for en variabel i et tekstfelt, skal dobbeltapostroffen foranstilles med en dobbelt apostrof samtidig med at hele teksten omsluttes med dobbelt apostrof.

Dobbeltapostroffen der anvendes skal have unicode-værdien U+0022. Hvis der anvendes andre former for dobbeltapostrof (f.eks. højrestillet dobbelt apostrof med unicode-værdien U+201D og UTF-8 hex værdien E2 80 9D eller venstrestillet dobbeltapostrof med unicode-værdien U+201C og UTF-8 hex-værdien E2 80 9C) bliver disse ekstra dobbelt apostroffer ikke fjernet igen, når arkivet konverterer data i afleveringspakken videre til bevaringsformat.

Eksempler på udtræk af variabelværdier der indeholder dobbelt apostrof

Variabelværdien *Antal ”dyr” som kan tale* skal i datafilen angives som *”Antal ””dyr”” som kan tale”*
Variabelværdien *”Hurra”* skal i datafilen angives som *””””Hurra””””*

9.G.1.c Som linjeseparator skal anvendes en af følgende metoder for linjeskift: »CR+LF« (U+000D) samt (U+000A) eller »CR« (U+000D) eller »LF« (U+000A). Indhold i variable i datafilen må ikke indeholde linjeskift.

Rækker i datafilen skal adskilles med linjeskift. Når der angives et linjeskift i en tekstfil angives en usynlig linjeseparator med en af følgende unicode værdier »CR+LF« (U+000D) samt (U+000A) eller »CR« (U+000D) eller »LF« (U+000A). Hvilke metoder der anvendes for linjeskift i en tekstfil afhænger af det styresystem der arbejdes på.

I hver linje i datafilen skal angives alle variabelværdier, svarende til antallet af variabelnavne angivet i første linje, før der angives et linjeskift. Da variabelværdierne adskilles med semikolon, betyder det at der i alle linjer i datafilen skal være præcis lige mange semikolon anvendt før et linjeskift.

Variabelværdier i datafilen må ikke indeholde linjeskift, da dette vil medføre at linjen indeholder færre semikolon end tilladt.

Eksempel på datafil med korrekte linjeskift

```
ID;Alder;Vægt;Indkomst;Stilling
00001;45;70,4;45000;IT-medarbejder
00002;63;85,3;130000;Leder
00003;21;76,2;23000;Postbud
```

9.G.2 En manglende værdi kan være en af følgende tre typer: Manglende værdi (tom) (jf. 9.G.2.a), specialkode for manglende værdi (jf. 9.G.2.d) eller brugerdefineret kode for manglende værdi (jf. 9.I.6).

9.G.2.a Manglende værdier (tom) i datafilen skal enten repræsenteres som ingen værdi »«, eller et mellemrum » « (U+0020).

9.G.2.b I en datafil må der konsekvent kun anvendes enten specialkoder eller brugerdefinerede koder for manglende værdier.

9.G.2.c Specialkoder for manglende værdier må kun anvendes for kategoriske og numeriske variable.

9.G.2.d Specialkoder for manglende værdier må kun anvendes for heltal og decimaltal, og skal angives enten som en værdi fra A-Z eller .a-.z.

Hvad er koder for manglende værdier?
Koder for manglende værdier anvendes i statistikfiler og kaldes også missing values.

Typisk skelnes mellem følgende tre typer af koder for manglende værdier i spørgeskemaundersøgelser:

- ***Irrelevant***: Spørgsmålet er ikke besvaret, fordi respondenten ikke skal svare grundet et filter i spørgeskemaet. Et filter i et spørgeskema anvendes, når en respondent ikke skal besvare det efterfølgende spørgsmål, hvis der svares noget specifikt på det foregående spørgsmål. F.eks. hvis respondenten svarer ”Mand” på spørgsmålet om hans køn, skal han efterfølgende ikke besvare spørgsmålet ”Hvor mange børn har du født”.
- ***Uoplyst***: Spørgsmålet er ikke besvaret, fordi respondenten ikke ønskede at svare.
- ***Deltager ikke*** Spørgsmålet er ikke besvaret, fordi respondenten ikke er en del af undersøgelsen (out of sample) selvom respondenten fremgår af statistikfilen.

Præcise informationer om hvorfor værdier mangler er vigtige for at kunne opnå pålidelige analyseresultater, når data genanvendes. Det er vigtigt at vide hvor godt besvaret et spørgeskema er. Hvis værdier mangler, fordi respondenterne ikke ønskede at besvare spørgsmålene, er data mangelfulde og ikke nær så pålidelige, som hvis værdierne blot mangler grundet spørgsmålsfiltre, fordi respondenterne ikke skulle svare.

Når koder for manglende værdier anvendes på ikke kategoriske variable, men reelle værdier (numeriske variable) som f.eks. alder, antal eller pris, er det vigtigt at koden angives som en missing value i statistikfilen og ikke som en reel værdi. Hvis koden 999 i en variabel med alder ikke er opmærket som en kode for en manglende værdi, vil værdien 999 indgå i aldersgennemsnittet, når der i forbindelse med en statistisk analyse laves et gennemsnit at variablens indhold.

Hvordan angives manglende værdier i datafilen?
En manglende værdi kan angives på forskellige måder i forskellige statistikprogrammer.
I afleveringspakkens datafil kan manglende værdier dog kun angives på følgende 3 måder:

1. **Manglende værdi (tom)**
   Hvis feltet blot er tomt, er værdien fraværende og vi ved ikke hvorfor. Den tomme værdi angives enten som en helt tom værdi »« eller som et mellemrum » «. Det er således ikke tilladt at angive sysmis (punktum) ».« eller værdierne ”NULL”, ”n.a.” osv.

2. **Specialkode for manglende værdi**
   En specialkode for en manglende værdi er en kode der angiver hvorfor en værdi i en variable mangler.

   Brugeren af statistikprogrammet kan selv definere betydningen af specialkoderne, men kan kun anvende specialkoder med værdier der ligger inden for de tilladte udfaldsrum. Udfaldsrummet for specialkoder er A-Z i statistikprogrammet SAS og .a-.z i statistikprogrammet Stata.

   Eksempler på specialkoder for manglende værdier:

   .i = irrelevant
   .u = uoplyst
   .f = fejl

   Det særlige ved specialkoder er, at de i statistikprogrammerne SAS og Stata angives i en *special numeric* datatype, dvs. en datatype som er defineret som numerisk, men også kan indeholde tekst i form af disse specialkoder.

   Specialkoder må kun anvendes for kategoriske eller numeriske variable af datatypen heltal eller decimaltal. Dvs. at de ikke må anvendes i variable med datatyperne tekst, tidspunkt, dato og tidsstempel.

3. **Brugerdefineret kode for manglende værdi**
   En brugerdefineret kode for en manglende værdi er en kode der angiver hvorfor værdien er manglende. Modsat specialkoder, som skal vælges fra værdier i faste udfaldsrum, kan de brugerdefinerede koder frit defineres af brugeren med bogstaver, tal, specialtegn osv.

   Brugerdefinerede koder anvendes typisk i statistikprogrammet SPSS.

   Eksempler på brugerdefinerede koder for manglende værdier:

   9 = uoplyst
   10 = irrelevant
   oos = out of sample
   fejl = fejl i besvarelse

   Læs mere om krav til brugerdefinerede koder for manglende værdier i punkt 9.I.7.

Hvad er forskellen på kategoriske og numeriske variable?

Kategorisk variabel
En kategorisk variabel indeholder værdier, som angiver koder for kategorier.
Til en kategorisk variabel er knyttet en række kodeforklaringer (value labels eller format). Dvs. værdien i variablen er en kode, hvis betydning fremgår af en kodeforklaring.
I spørgeskemaundersøgelser svarer disse koder og kodeforklaringer ofte til svarkategorier til et spørgsmål.

Eksempel på en kategorisk variabel

Variabelnavn: Ægteskab
Variabelbeskrivelse: Er du gift?

Kodeliste (value labels/format)
1 = Ja
2 = Nej
3 = Ved ikke

Numerisk variabel
En numerisk variabel kan være et heltal eller et decimaltal. En numerisk variabel indeholder reelle værdier. F.eks. alder, pris, indkomst eller antal.

9.G.3 Indholdet af de enkelte variable skal renses for eventuelle foran- og efterstillede blanktegn.

Indholdet af de enkelte variable skal renses for eventuelle foran- og efterstillede blanktegn (f.eks. mellemrum).

Blanktegn vil altid blive fjernet i arkivets efterfølgende konvertering af afleveringspakken til bevaringsformat. Hvis blanktegnene er betydningsbærende skal de derfor erstattes af andre tilladte tegn. Dette gælder særligt for variable der indgår som nøgler i relationer mellem flere datafiler i afleveringspakken.

9.H. Datatyper

9.H.1 De seks standardiserede datatyper, som skal anvendes i datafilen, fremgår af figur 9.3.

9.H.2 Dataformatnotationer for anvendte dataformater skal angives i metadatafilen, jf. figur 9.3. Dataformatnotationerne er case sensitive.

9.H.2.a Værdier for bogstaverne »w« og »d« i figur 9.3 skal konsekvent angives. »w« angiver datatypens totale bredde. »d« angiver decimaler eller præcision i fraktioner af sekunder.

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.0f	%3.0f
			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**	2019-11-15	**xml**	date	date
		2019-11-15	**Stata**	%tdCCYY-NN-DD	%tdCCYY-NN-DD
		2019-11-15	**SAS**	yymmdd10.	yymmdd10.

	8601:1993 udvidet format: CCYY-MM-DD				
	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: CCYY-MM-DDThh:mm:ss.sss eller CCYY-MM-DD hh:mm:ss.sss**	2019-11-15T08:10:23 Med fraktioner af sekunder: 2019-11-15T08:10:23.123456	**xml**	datetime	datetime
		2019-11-15T08:10:23	**Stata**	%tcCCYY-NN-DD!THH:MM:SS	%tcCCYY-NN-DD!THH:MM:SS
		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

	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	**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-mmm-yyyy hh:mm:ss	15-NOV-2019 08:10:23	**SPSS**	datetime20	datetime20

9.I. Metadatafil

9.I.1 Metadata fra en datafil afleveres som en struktureret tekstfil, der overholder syntaksen beskrevet som EBNF i figur 9.11.

9.I.1.a Metadatafilen udformes som anvist i figur 9.4, hvor otte etiketter opdeler metadata i specifikke kategorier.

9.I.1.b Hver etiket skal forekomme en gang i metadatafilen. Forekomsten af indholdet af etiketten fremgår af kolonnerne »Forekomst« og »Obligatorisk« i figur 9.4.

9.I.1.c Etiketnavne er reserverede ord og må ikke benyttes til navngivning af metadatafilens øvrige indhold.

Eksempel på metadatafil (table1.txt) der overholder EBNF i figur 9.11

table1.txt - Notesblok

Filer Rediger Formater Vis Hjælp

```
SYSTEMNAVN
SPSS

DATAFILNAVN
Børn

DATAFILBESKRIVELSE
Oplysninger om børnene i generationsstudiet

NØGLEVARIABEL
barn_id

REFERENCE
Forældre 'barnid' 'barn_id'
Bedsteforældre 'barnebarnid' 'barn_id'

VARIABEL
barn_id a765
koen f3 koen.
vaegt_idag f3.1
slut_dato sdate10
hobby f3 hobby.
laegebesoeg datetime20
maaling time8
aarsag a100

VARIABELBESKRIVELSE
barn_id 'Barn ID. Unikt ikke-betydningsbærende identifikationsnummer.'
koen 'Er du mand eller kvinde?'
vaegt_idag 'Hvad er din vægt idag? (angivet i kilogram/kg)'
slut_dato 'Gymnasial uddannelse slut dato'
hobby 'Hvad er din hobby?'
laegebesoeg 'Lægebesøg - dato og tidspunkt'
maaling 'Tidspunkt for måling af blodtryk'
aarsag 'Årsag til videre undersøgelse'

KODELISTE
hobby
'1' 'Sport'
'2' 'Musik'
'3' 'Håndværk'
'4' 'Kunst'
'5' 'Bøger'
'9' 'uoplyst'
'10' 'irrelevant'
koen
'1' 'Mand'
'2' 'Kvinde'
'9' 'uoplyst'

BRUGERKODE
hobby '9' '10'
koen '9'
```

Figur 9.4 Metadatafilens struktur

Etiket	Beskrivelse af etiketindhold	Udfaldsrum	Forekomst	Obligatorisk
SYSTEMNAVN	Navn på det program, som data udtrækkes fra, eller datas oprindelige format.	SPSS SAS Stata Excel eller fritekst	1	Ja
DATAFILNAVN	Navnet på datafilen, som den benævnes i brugssammenhæng. Datafilnavn skal overholde krav i punkt 9.I.2.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	1	Ja
DATAFILBESKRIVELSE	Beskrivelse af datafilens indhold.	Fritekst	1	Ja
NØGLEVARIABEL	Datafilens unikke nøglevariabel angivet med navne på den/de variable, nøglevariablen består af.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	0-1	Ja, hvis unik nøglevariabel findes
REFERENCE	Referencer til andre datafiler i afleveringspakken angivet som anvist i punkt 9.I.3.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	0-m	Ja, hvis reference til anden datafil i afleveringspakken findes
VARIABEL	Variable i datafilen. En variabel angives som et sæt bestående af variabelnavn efterfulgt af notation for variablens dataformat, jf. figur 9.3, samt en eventuel kodelistereference, jf. 9.I.5.f. Variable skal overholde krav i punkt 9.I.4.	Navngivning af variable skal overholde ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	1-m	Ja
VARIABELBESKRIVELSE	Beskrivelser af variablenes indhold.	Fritekst	1-m	Ja
KODELISTE	Kodelister angivet med kodelistens navn efterfulgt	Navngivning af kodelister skal	0-m	Ja, hvis der findes kodelister

	af flere sæt bestående af kode og kodebeskrivelse. **Kodelister skal overholde krav i punkt 9.I.5.**	**overholde ISO/IEC 9075:1999 - Database Language SQL (SQL-99)**		
BRUGERKODE	**Brugerdefinerede koder for manglende værdier angivet med navnet på den variabel, hvor koderne anvendes, efterfulgt af de brugerdefinerede koder.** **Brugerdefinerede koder for manglende værdier skal overholde krav i punkt 9.I.6.**	**Navngivning af variable skal overholde ISO/IEC 9075:1999 - Database Language SQL (SQL-99)**	**0-m**	**Ja, hvis der findes brugerdefinerede koder for manglende værdier**

Hvordan skabes metadatafilen?
Metadatafilen kan udtrækkes automatisk fra den oprindelige statistikfil fra statistikprogrammerne SAS, Stata og SPSS ved brug af Rigsarkivets værktøj ASTA.

Laves udtræk fra et regneark, kan metadatafilen udfyldes manuelt i en teksteditor. Se Rigsarkivets vejledning om aflevering af statistikdata fra regneark og csv-filer på www.sa.dk.

SYSTEMNAVN
Under etiketten SYSTEMNAVN angives navn på det program, som data udtrækkes fra, eller datas oprindelige format.
Angiv enten SPSS, SAS, Stata, Excel eller andet i fritekst.

Eksempel:
```
SYSTEMNAVN
SPSS
```

DATAFILNAVN
Under etiketten DATAFILNAVN angives navnet på datafilen, som den benævnes i brugssammenhæng.
Datafilnavne skal være unikke inden for samme afleveringspakke, jf. 9.I.2.
Bemærk at datafilnavn *ikke* skal indeholde filens extension (f.eks. .sav)

Eksempel:
```
DATAFILNAVN
Generationsundersøgelsen
```

DATAFILBEKSRIVELSE
Under etiketten DATAFILBESKRIVELSE angives en beskrivelse af datafilens indhold, f.eks. den fulde titel for statistikfilen, en eksisterende projektbeskrivelse eller et abstract, der præcist dækker statistikfilens indhold.

Bemærk at datafilbeskrivelsen ikke må indeholde tegn for linjeskift.

Eksempel:

```
DATAFILBESKRIVELSE
Generationsundersøgelse som har til formål at undersøge sammenhæng en mellem
sundhed, IQ og arv. Indholdet af dette datasæt er oplysninger om studerende.
```

NØGLEVARIABEL
Under etiketten NØGLEVARIABEL angives datafilens unikke nøglevariabel med navne på den/de variable, nøglevariablen består af. Nøglevariablen skal entydigt identificere en række i datafilen. Dvs. to værdier i variablen må ikke have samme værdi.

En nøglevariabel kan være sammensat af flere variable. Flere variable i en nøglevariabel adskilles af mellemrum.

Eksempel:
```
NØGLEVARIABEL
fødselsdato lbnr
```

REFERENCE
Under etiketten REFERENCE angives referencer til andre datafiler i afleveringspakken, som anvist i punkt 9.I.3.

Hvis der findes flettevariable som kan koble to datasæt (datafiler) i afleveringspakken sammen skal disse angives under etiketten REFERENCE i metadatafilen.

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.

Relationer til eventuelle nøglefiler skal også angives her. Hvis du f.eks. afleverer et datasæt, hvor personfølsomme oplysninger fra datasættet er udtrukket til en nøglefil, skal relationen mellem datasættet og nøglefilen defineres i en reference.

Eksempel:
```
REFERENCE
stata12345 'barn_userid' 'userid'
sas12345 'barnebarn_userid' 'userid'¨
```

VARIABEL
Under etiketten VARIABLE angives alle variable i datafilen.

Variabelnavne skal være unikke inden for samme metadatafil., jf. 9.I.4.

En variabel angives som et sæt bestående af variabelnavn efterfulgt af notation for variablens dataformat, jf. figur 9.3, samt en eventuel kodelistereference, jf. 9.I.5.f.

For hver variabel angives således 2-3 oplysninger.

1) ***Første oplysning*** er angivelse af variabelnavn. Variabelnavne skal overholde SQL99 standarden (ISO/IEC 9075:1999 - Database Language SQL) for angivelse af tabel- og kolonnenavne. Dvs. at variabelnavn må aldrig begynde med et tal, men må godt være en blanding af bogstaver, tal samt underscore på max 128 tegn. Hvis variabelnavnet er et reserveret ord i SQL:1999, skal det enten omdøbes eller omkranses af dobbeltapostrof.
2) ***Anden oplysning*** er angivelse af variablens datatype. Tilladte datatyper for variable fremgår af figur 9.3. Det er dataformatnotationerne angivet i kolonne ”Dataformatnotation i metadatafil”, som skal angives som variablens datatype. Datatypen *f10* i eksemplet nedenfor angiver, at variablen *id* er af datatypen heltal med en længde på 10 cifre.

3) ***Tredje oplysning*** er angivelse af en eventuel kodelistereference. Tilladte notationer for angivelse af kodelistereferencer fremgår af 9.I.5.g og 9.I.5.h.

Eksempel:

```
VARIABEL
id f10
køn f1 køn.
fødselsdato sdate10
fødselstidspunkt time8
ægteskab f2 ægteskab.
nationalitet a2 $nationalitet.
kommentar a200
```

VARIABELBESKRIVELSE

Under etiketten VARIABELBEKSRIVELSE angives beskrivelser af variablenes indhold.

Alle variable skal have en variabelbeskrivelse, der udfyldes så udførligt som muligt, således at en fremtidig bruger har mulighed for at forstå hvilken oplysning der er registreret i variablen. Alle koder i beskrivelsen skal forklares og eventuelle måleenheder skal angives (f.eks. meter, centimeter, kilo). Findes datoer, skal det angives hvad datoen omhandler, f.eks. bryllupsdato eller fødselsdato. Findes oplysninger om flere individer i et datasæt, f.eks. både mor og barn, skal det tydeliget af variabelbeskrivelsen fremgå hvilke variabeloplysninger der er knyttet til hvilke individer, f.eks. '*Barnets køn'* eller *'Respondentens køn'*.

For spørgeskemaundersøgelser anbefaler Rigsarkivet, at alle spørgsmålsteksterne indsættes i variabelbeskrivelserne (variable labels i statistikfilen). Hermed bliver det lettere at søge i og anvende data for fremtidige brugere. Spørgsmålsteksterne bør indsættes i deres oprindelige formulering og fulde længde. Hvis spørgsmålsteksterne er for lange til at blive udtrukket automatisk fra statistikfilen, kan de efterfølgende indsættes direkte i metadatafilen (f.eks. table1.txt) under etiketten VARIABELBESKRIVELSE.

Eksempel:

```
VARIABELBESKRIVELSE
id 'Respondentens cpr-nummer'
køn 'Respondentens køn'
ægteskab 'Er du gift?'
dato 'Hvilken dato blev du gift?'
nationalitet 'Hvilket statsborgerskab har du?'
vægt 'Hvad vejer du? (Angivet i hele kilo)'
kommentar 'Har du yderligere kommentarer kan du angive dem her.'
```

KODELISTE

Under etiketten KODELISTE listes datasættets kodelister angivet med kodelistens navn efterfulgt af flere sæt bestående af kode og kodebeskrivelse.

Kodelister skal desuden overholde kravene i punkt 9.I.5.

I spørgeskemaundersøgelser svarer kodelister til svarkategorier til et spørgsmål.
I statistikprogrammet SPSS svarer kodelister til de *value labels* der er knyttet til variablene.
I statistikprogrammet SAS svarer kodelister til formaterne i format-filen (katalog-filen).

Et kodesæt under etiketten KODELISTE i metadatafilen består af et kodelistenavn hvorunder der angives rækker med koder og tilhørende kodebeskrivelser. Hver række skal bestå af en kode og en kodebeskrivelse. Begge oplysninger skal hver især omkranses af enkeltapostroffer og adskilles med mellemrum.

Eksempel:

```
KODELISTE
Nationalitet
'DA' 'Danmark'
```

```
'GB' 'England'
'SE' 'Sverige'
'9' 'uoplyst'
'10' 'irrelevant'
Køn
'1' 'Mand'
'2' 'Kvinde'
```

BRUGERKODE
Under etiketten BRUGERKODE angives brugerdefinerede koder for manglende værdier med navnet på den variabel, hvor koderne anvendes, efterfulgt af de brugerdefinerede koder omkranset af enkeltapostroffer. Bemærk at betydningen af de brugerdefinerede koder angives i kodelisten knyttet til den variabel, hvor koderne anvendes, under etiketten KODELISTE.

Brugerdefinerede koder for manglende værdier må kun anvendes i kategoriske og numeriske variable med datatyperne heltal, decimaltal, og tekst jf. 9.I.6.

Eksempel:
```
BRUGERKODE
Nationalitet '9' '10'
```

9.I.2 Datafilnavne skal være unikke inden for samme afleveringspakke.

Hvis der udtrækkes data fra flere statistikfiler til en afleveringspakke skal disse navngives forskelligt. Navnet på en datafil angives i metadatafilen til den pågældende datafil under etiketten DATAFILNAVN

9.I.3 Reference

9.I.3.a 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.

Eksempel på angivelse af reference i metadatafilen til Datasæt 1

I dette eksempel er angivet en reference over de to flettevariable *nøglefil_id* og *id*.

Flettevariablen *id* findes i hoveddatasættet med datafilnavnet *Generationsundersøgelsen*.

Flettevariablen *nøglefil_id* findes i datasættet med datafilnavnet *Nøglefil*.

Referencen defineres under etiketten REFERENCE i metadatafilen for hoveddatasættet *Generationsundersøgelsen*, som indeholder flettevariablen (referencevariablen) *id*, som peger på den unikke flettevariabel/nøglevariabel *nøglefil_id* (fremmedvariablen) i datasættet *Nøglefil*.

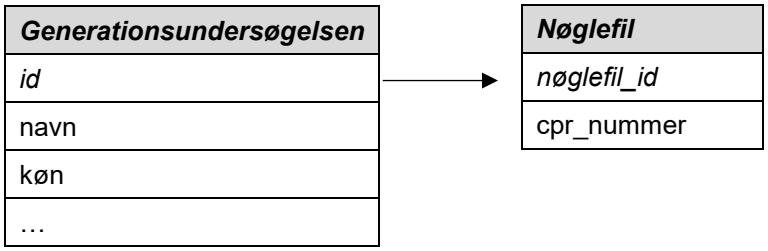
Datasæt 1:
```
…
DATAFILNAVN
Generationsundersøgelsen
…
REFERENCE
Nøglefil 'nøglefil_id' 'id'
```

Datasæt 2:
```
…
DATAFILNAVN
Nøglefil
…
REFERENCE
```

```
VARIABEL
id f8
navn a200
køn f2 køn.
…
```

```
VARIABEL
nøglefil_id f8
cpr_nummer f10
```

Illustration af referencen angivet ovenfor:



9.I.3.b Der skal være fuld overensstemmelse mellem datatype og længde i nøglevariablene, der indgår i referencen.

9.I.4 Variabelnavne skal være unikke inden for samme metadatafil.

To variable under etiketten VARIABEL i metadatafilen knyttet til en datafil må ikke have samme navn.

9.I.5 Kodeliste

9.I.5.a Kun kategoriske og numeriske variable må have henvisninger til en kodeliste.

9.I.5.b Det er kun tilladt at medtage kodelister for variable, der indeholder data af datatypen numerisk heltal, decimaltal eller tekst, jf. datatyper i figur 9.3.

Det er ikke tilladt at anvende koder i og knytte kodelister til variable med datatyperne dato, tidspunkt og tidsstempel.

En kategorisk variabel indeholder værdier, som angiver koder for kategorier. Datatyperne for en kategorisk variabel kan være enten heltal, decimaltal eller tekst.

Eksempler på værdier i kategoriske variable

Heltal
1 = Ja
2 = Nej
3 = Ved ikke

Decimaltal
1.00 = Ingen smerte
2.00 = Lidt smerte
3.00 = Meget smerte

Tekst
DA = Danmark
GB = England
SE = Sverige

En numerisk variabel indeholder reelle værdier. F.eks. alder, pris, indkomst eller antal. Numeriske variable kan have datatyperne heltal eller decimaltal.

Eksempler på værdier i numeriske variable

Heltal
1
2
3
…

Decimaltal
150013,50
12420,67
34880,54
…

9.I.5.c Alle koder i data skal defineres i en kodeliste og forklares med en kodebeskrivelse.

9.I.5.d Hvis koder er defineret som et interval, hvor ikke alle koder har kodebeskrivelser, skal dette dokumenteres i variabelbeskrivelsen.

For variable der anvender koder der er defineret som et interval tilføjes følgende eller lignende dokumentation i variabelbeskrivelsen: ”Skala anvendt. Ikke alle koder har kodebeskrivelse.”

Eksempel på koder defineret som et interval

Et eksempel på koder defineret som et interval er en Likert skala, hvor der f.eks. stilles følgende spørgsmål:

På en skala fra 1-5, hvor enig er du i følgende udsagn: ”Jeg er altid glad”?

Alle besvarelserne 1, 2, 3, 4, 5 findes i datasættet, men der findes kun kodebeskrivelserne 1 = Meget uenig og 5 = Meget enig. Kodelisten angives på følgende måde i metadatafilen:

```
KODELISTE
humør
’1’ ’Meget uenig’
’5’ ’Meget enig’
```

9.I.5.e Koder skal være unikke inden for samme kodeliste.

To eller flere koder i en kodeliste må ikke have samme værdi.

9.I.5.f Referencen mellem en kodeliste og den variabel, som refererer til kodelisten, angives under etiketten VARIABEL, jf. figur 9.4, som anvist i 9.I.5.g og 9.I.5.h.

9.I.5.g Hvis datatypen er numerisk heltal eller decimaltal, angives referencen med det valgte kodelistenavn og et efterstillet punktum ».« (U+002E).

Eksempel på en kodelistereference når variablen er af datatypen heltal eller decimaltal:

```
VARIABEL
ægteskab f2 ægteskab.
```

9.I.5.h Hvis datatypen er tekst, angives referencen med det valgte kodelistenavn med et foranstillet dollartegn »$« (U+0024) og et efterstillet punktum ».« (U+002E).

Eksempel på en kodelistereference når variablen er af datatypen tekst:

```
VARIABEL
nationalitet a2 $nationalitet.
```

9.I.6 Brugerkode

Brugerkode er en kort betegnelse for brugerdefinerede koder for en manglende værdier.

En brugerdefineret kode for en manglende værdi er en kode der angiver hvorfor værdien er manglende. Modsat specialkoder, som skal vælges fra værdier i faste udfaldsrum, kan de brugerdefinerede koder frit defineres af brugeren med bogstaver, tal, specialtegn osv.

Brugerdefinerede koder anvendes typisk i statistikprogrammet SPSS.

Eksempler på brugerdefinerede koder for manglende værdier

9 = irrelevant
10 = uoplyst
11 = deltager ikke
? = uafklaret
fejl = fejl i besvarelse

9.I.6.a En brugerdefineret kode for en manglende værdi er kun tilladt for kategoriske og numeriske variable.

Brugerdefinerede koder kan i statistikprogrammet SPSS anvendes på alle variable, blot de overholder variablens datatype. Jf. bilag 9 er det dog i datafilen kun tilladt at anvende koder for manglende værdier for kategoriske variable med datatyperne heltal, decimaltal eller tekst eller numeriske variable med datatyperne heltal eller decimaltal.

En kategorisk variabel indeholder værdier, som angiver koder for kategorier. Til en kategorisk variabel er skal altid knyttet kodeforklaringer, der tydeliggør betydningen af koderne.

En numerisk variabel indeholder reelle værdier. F.eks. alder, pris, indkomst eller antal.

Eksempler på værdier i kategoriske variable inkl. eksempler på brugerkoder

Heltal
1 = Ja
2 = Nej

3 = Ved ikke
9 = uoplyst
10 = irrelevant

Decimaltal
1.00 = Ingen smerte
2.00 = Lidt smerte
3.00 = Meget smerte
9 = uoplyst
10 = irrelevant

Tekst
DA = Danmark
GB = England
SE = Sverige
9 = uoplyst
10 = irrelevant

Eksempler på værdier i numeriske variable samt eksempel på brugerkoder

Heltal (Alder)
1
2
3
…
999 = uoplyst
1001 = irrelevant

Decimaltal (Indkomst)
150000,00
12420,50
34880,00

Det er ikke tilladt at anvende brugerdefinerede koder for manglende værdier i variable med datatyperne dato, tidspunkt og tidsstempel.

Hvis en dato, fx 9999-12-31, anvendes som en kode for en manglende værdi for en variabel i den statistikfil udtrækket laves fra, bør denne kode omkodes til en manglende værdi (tom), jf. 9.G.2.a, før udtræk til datafil. Er betydningen af denne værdi ikke forklaret fyldestgørende ved en tom værdi eller findes der flere datoer anvendt som koder for manglende værdier med forskellige betydninger i samme variabel, kan disse medtages i udtrækket til datafilen men uden angivelse af brugerkoderne i metadatafilen. Betydningerne af brugerkoderne skal da i stedet dokumenteres i variabelbeskrivelsen for variablen, hvor koderne anvendes, eller i et kontekstdokument.

9.I.6.b En brugerdefineret kode for en manglende værdi angivet i metadatafilen, skal altid fremgå af kodelisten, som den tilhører.

Eksempel på brugerdefinerede koder for manglende værdier angivet i kodelisten

```
KODELISTE
Nationalitet
'DA' 'Danmark'
'GB' 'England'
'SE' 'Sverige'
```

```
'9' 'uoplyst'
'10' 'irrelevant'

BRUGERKODE
Nationalitet '9' '10'
```

Hvad er EBNF?
Figur 9.6 til 9.10 herunder anviser EBNF-syntakser for de tilladte datatyper i datafiler og figur 9.11 og 9.12 for opbygning af en data- og metadatafil i afleveringspakken. EBNF er en notationsform der meget kort og præcist ved brug af særlige tegn definerer hvad der er tilladt. En tegnforklaring til hvordan EBNF-syntaksen skal læses fremgår af figur 9.5.

Et heltal er f.eks. defineret som [FORTEGN] CIFFER {CIFFER}, hvilket betyder at et heltal må bestå af et [FORTEGN], hvor de kantede parenteser angiver at fortegn kan forekomme 0-1 gange, dvs. det er valgt frit. Herefter skal altid komme et CIFFER efterfulgt af {CIFFER}, hvor tuborg-klammerne angiver at dette andet ciffer kan forekommer 0 eller flere gange.

Figur 9.5 EBNF tegnforklaring

::= defineret som
() angiver en gruppering, der skal udføres samlet
[] angiver muligheden for (0 eller 1)
{} angiver mulig gentagelse (0 eller flere)
| angiver et valg (enten eller)
… angiver et fortløbende interval
!! angiver en beskrivende forklaring på alm. dansk
”” omslutter faktiske værdier, der skal skrives uden fortolkning

Figur 9.6 EBNF for heltalstype DS/ISO 6093:1985 (NR1) standard

Nonterminal	**Terminal**	Eksempel
INT ::=	NR1	25
NR1 ::=	[FORTEGN] CIFFER {CIFFER}	+45
FORTEGN ::=	”+” | ”-” *!! Det er valgfrit at anvende fortegn for positive heltal*	-234
CIFFER ::=	”0” | ”1” | … | ”9”	10000

Figur 9.7 EBNF for decimaltalstype DS/ISO 6093:1985 (NR2) standard

Nonterminal	**Terminal**	Eksempel
DECIMAL ::=	NR2	23,75
NR2 ::=	[FORTEGN] CIFFER {CIFFER} | [FORTEGN] CIFFER {CIFFER} DECIMALMÆRKE CIFFER {CIFFER} | DECIMALMÆRKE CIFFER {CIFFER}	.10
FORTEGN ::=	”+” | ”-” *!! Det er valgfrit at anvende fortegn for positive decimaltal !! Det er ikke tilladt at anvende negativt fortegn foran værdien 0*	-123.76
CIFFER ::=	”0” | ”1” | … | ”9”	+123.76
DECIMALMÆRKE ::=	”.” | ”,”	150000.25

Figur 9.8 EBNF for datotyper

Nonterminal	**Terminal**	Eksempel
DATO ::=	ISO-8601-DATE | ALTERNATIV-DATE	
ISO-8601-DATE ::=	CC YY BINDESTREG MM BINDESTREG DD	

CC ::=	CIFFER CIFFER *!! årtusinde angivet med to heltal*	2019-11-15
BINDESTREG ::=	”-” *!! bindestreg (U+002D)*	
YY ::=	CIFFER CIFFER *!! årstal angivet med to heltal*	
MM ::=	CIFFER CIFFER *!! måned angivet med to heltal*	
DD ::=	CIFFER CIFFER *!! dag angivet med to heltal*	
CIFFER ::=	”0” | ”1” | … | ”9”	
ALTERNATIV-DATE ::=	CC YY SKRÅSTREG MM SKRÅSTREG DD	2019/11/15
SKRÅSTREG ::=	”/” *!! skråstreg U+002F*	

Figur 9.9 EBNF for tidstyper

Nonterminal	**Terminal**	Eksempel
TIDSPUNKT ::=	TIME KOLON MINUT KOLON SEKUND	05:10.23 4:22:59
TIME ::=	CIFFER [CIFFER] *!! time angivet med et eller to heltal !! udfaldsrum er 0-23*	
KOLON ::=	”:” *!! kolon (U+003A)*	
MINUT ::=	CIFFER CIFFER *!! minut angivet med to heltal !! udfaldsrum er 00-59*	
SEKUND ::=	CIFFER CIFFER *!! sekund angivet med to heltal !! udfaldsrum er 00-59*	
CIFFER ::=	”0” | ”1” | … | ”9”	

Figur 9.10 EBNF for datetimetyper

Nonterminal	**Terminal**	Eksempel
DATETIME ::=	ISO-8601-DATETIME | ALTERNATIV-DATETIME | IBM-DATETIME	
ISO-8601-DATETIME ::=	CC YY BINDESTREG MM BINDESTREG DD TIDSTEMPELMÆRKE TIME KOLON MINUT KOLON SEKUND [PUNKTUM MILLISEKUNDER]	2019-11-15T08:10:23 2019-11-15T08:10:23.123456 2019-11-15T08:10:23.123 2019-11-15 08:10:23 2019-11-15 08:10:23.123456 2019-11-15 08:10:23.12
CC ::=	CIFFER CIFFER *!! årtusinde angivet med to heltal*	
CIFFER ::=	”0” | ”1” | … | ”9”	
BINDESTREG ::=	”-” *!! bindestreg (U+002D)*	
YY ::=	CIFFER CIFFER *!! årstal angivet med to heltal*	
MM ::=	CIFFER CIFFER *!! måned angivet med to heltal*	
DD ::=	CIFFER CIFFER *!! dag angivet med to heltal*	
TIDSTEMPELMÆRKE ::=	”T” | MELLEMRUM	
MELLEMRUM ::=	” ” *!! mellemrumstegn (U+0020)*	
TIME ::=	CIFFER CIFFER *!! time angivet med to heltal !! udfaldsrum er 00-23*	

KOLON ::=	":" *!! kolon (U+003A)*	
MINUT ::=	CIFFER CIFFER *!! minut angivet med to heltal !! udfaldsrum er 00-59*	
SEKUND ::=	CIFFER CIFFER *!! sekund angivet med to heltal !! udfaldsrum er 00-59*	
PUNKTUM ::=	"." *!! punktum (U+002E)*	
MILLISEKUNDER ::=	CIFFER {CIFFER} *!! millisekunder angivet med op til max 6 cifre*	
IBM-DATETIME ::=	DD BINDESTREG MÅNED BINDESTREG ÅRSTAL MELLEMRUM TIME KOLON MINUT KOLON SEKUND	20-May-2019 11:05:48 15-OCT-2019 08:10:23
MÅNED ::=	"Jan" | "Feb" | "Mar" | "Apr" | "May" | "Jun" | "Jul" | "Aug" | "Sep" | "Oct" | "Nov" | "Dec" *!! der skal anvendes engelske forkortelser for månedsnavn, og de er ikke case sensitive.*	
ÅRSTAL ::=	CIFFER CIFFER CIFFER CIFFER *!! årstal angivet med fire heltal*	

Figur 9.11 EBNF Syntaksregler for metadatafil

Nonterminal	**Terminal**	Eksempel

METADATAFIL ::=	SYSTEMNAVN
	DATAFILNAVN
	DATAFILBESKRIVELSE
	NØGLEVARIABEL
	REFERENCE
	VARIABEL
	VARIABELBESKRIVELSE
	KODELISTE
	BRUGERKODE
		`SYSTEMNAVN`
		`DATAFILNAVN`
		`DATAFILBESKRIVELSE`
		`NØGLEVARIABEL`
		`REFERENCE`
		`VARIABEL`
		`VARIABELBESKRIVELSE`
		`KODELISTE`
		`BRUGERKODE`

SYSTEMNAVN::=	"SYSTEMNAVN" LINJESKIFT "SAS" | "Stata" | "SPSS" | "Excel" | FRITEKST LINJESKIFT LINJESKIFT {LINJESKIFT}	`SYSTEMNAVN`
		`SAS`
		
		`SYSTEMNAVN`
		`R statistikfil`

DATAFILNAVN ::=	"DATAFILNAVN" LINJESKIFT TITEL LINJESKIFT LINJESKIFT {LINJESKIFT}	`DATAFILNAVN` `Generationsundersøgelsen`
LINJESKIFT ::=	(CR LF) | CR | LF	
CR ::=	*!! vognretur (U+000D)*	`DATAFILNAVN`
LF ::=	*!! linjeskift (U+000A)*	

TITEL ::=	(BOGSTAV {BOGSTAV | CIFFER}) | (DOBBELTAPOSTROF BOGSTAV {BOGSTAV | CIFFER} DOBBELTAPOSTROF) *!! En TITEL må aldrig begynde med et tal, men må godt være en blanding af bogstaver og tal på max 128 tegn, og hvis titel er et reserveret ord i SQL:1999, skal titel omkranses af dobbeltapostrof, jf. ISO/IEC 9075:1999 - Database Language SQL (SQL:1999)*	”Aggregate”
DOBBELTAPOSTROF ::=	”” *!! dobbeltapostrof (U+0022)*	
BOGSTAV ::=	”A” |”B” | … | ”Z” | ”a” | ”b” | … | ”z” | *”_”!! Samt andre nationale karakterer, der er tilladte i constraint names i standarden SQL:1999 !! underscore (U+005F)*	
CIFFER ::=	”0” | ”1” | … | ”9”	
DATAFILBESKRIVELSE ::=	”DATAFILBESKRIVELSE” LINJESKIFT FRITEKST LINJESKIFT LINJESKIFT {LINJESKIFT}	DATAFILBESKRIVELSE Generationsundersøgelse som har til formål at undersøge sammenhæng en mellem sundhed, IQ og arv. Indholdet af dette datasæt er oplysninger om studerende.
FRITEKST ::=	BOGSTAV | CIFFER | ANDRETEGN {BOGSTAV | CIFFER | ANDRETEGN}*!! FRITEKST kan være en blanding af bogstaver, tal og andre tegn, så længe det giver semantisk mening*	
ANDRETEGN ::=	*!! alle tilladte tegn i UTF-8, som ikke er bogstaver og tal jf. punkt 9.F*	
NØGLEVARIABEL ::=	”NØGLEVARIABEL” LINJESKIFT {VARIABELNAVN MELLEMRUM} LINJESKIFT LINJESKIFT {LINJESKIFT}	NØGLEVARIABEL fødselsdato lbnr NØGLEVARIABEL cprnummer
VARIABELNAVN ::=	TITEL	
MELLEMRUM ::=	*!! mellemrumstegn (U+0020)*	
REFERENCE ::=	”REFERENCE” LINJESKIFT {FREMMEDDATAFIL MELLEMRUM FREMMEDVARIABEL MELLEMRUM REFERENCEVARIABEL LINJESKIFT} LINJESKIFT	REFERENCE Nøglefil 'nøglefil_id' 'hovedfil_id'
FREMMEDDATAFIL ::=	TITEL *!! navn på den datafil, der refereres til*	
FREMMEDVARIABEL ::=	APOSTROF VARIABELNAVN {MELLEMRUM VARIABELNAVN} APOSTROF *!! navn på nøglevariablen i den datafil, der refereres til*	
REFERENCEVARIABEL ::=	APOSTROF VARIABELNAVN {MELLEMRUM VARIABELNAVN} APOSTROF *!! navn på variablen i datafilen, der refereres fra, som refererer til nøglevariablen i den datafil, der refereres til*	
APOSTROF ::=	”” *!! enkelt apostrof (U+0027)*	
VARIABEL ::=	”VARIABEL” LINESKIFT VARIABELSÆT {VARIABELSÆT} LINJESKIFT {LINESKIFT}	VARIABEL id f10 køn f1 køn. fødselsdato sdate10 fødselstidspunkt time8
VARIABELSÆT ::=	VARIABELNAVN MELLEMRUM DATAFORMATNOTATION MELLEMRUM [KODELISTEREFERENCE] LINJESKIFT	

		ægteskab f2 ægteskab. nationalitet a2 $nationalitet. kommentar a200
KODELISTEREFERENCE ::=	[DOLLAR] TITEL PUNKTUM	køn. $nationalitet.
DATAFORMATNOTATION ::=	INTEGERNOTATION | DECIMALNOTATION | DATONOTATION | DATETIMENOTATION | TIMENOTATION | TEKSTNOTATION	
INTEGERNOTATION ::=	*!! se figur 9.3 for dataformatnotationer for numerisk heltal*	int %3.0f (%w.0f) f3. (fw.) f3 (fw)
DECIMALNOTATION ::=	*!! se figur 9.3 for dataformatnotationer for decimaltal*	decimal %9.2f (%w.df eller %w.dg) f9.2 (fw.d) f9.2 (fw.d)
DATONOTATION ::=	*!! se figur 9.3 for dataformatnotationer for datotyper*	date %tdCCYY-NN-DD yymmdd10. sdate10
DATETIMENOTATION ::=	*!! se figur 9.3 for dataformatnotationer for tidsstempler*	datetime %tcCCYY-NN-DD!THH:MM:SS %tcCCYY-NN-DD!THH:MM:SS.sss %tcCCYY-NN-DD!THH:MM:SS.s e8601dt19. e8601dt25.6 (e8601dtw.d) e8601dt24.5 (e8601dtw.d) ymdhms19 ymdhms26.6 (ymdhmsw.d) ymdhms24.4 (ymdhmsw.d) datetime20
TIMENOTATION ::=	*!! se figur 9.3 for dataformatnotationer for tidspunkter*	time %tcHH:MM:SS time. eller time8. time8
TEKSTNOTATION ::=	*!! se figur 9.3 for dataformatnotationer for tekst*	string

		%20s (%ws) $20. ($w.) a20 (aw)
DOLLAR ::=	”$” *!! dollartegn (U+0024)*	$
PUNKTUM ::=	”.” *!! punktum (U+002E)*	.
VARIABELBESKRIVELSE ::=	”VARIABELBESKRIVELSE” LINJESKIFT BESKRIVELSE {BESKRIVELSE} LINJESKIFT {LINJESKIFT}	VARIABELBESKRIVELSE id ’Respondentens cpr-nummer’ køn ’Respondentens køn’ ægteskab ’Er du gift?’ dato ’Hvilken dato blev du gift?’ nationalitet ’Statsborgerskab’ vægt ’Hvad vejer du? (kilo)’
BESKRIVELSE ::=	VARIABELNAVN MELLEMRUM APOSTROF FRITEKST APOSTROF LINJESKIFT	
KODELISTE ::=	”KODELISTE” LINJESKIFT {KODESÆT} LINJESKIFT {LINJESKIFT}	KODELISTE Nationalitet ’DA’ ’Danmark’ ’GB’ ’England’ ’SE’ ’Sverige’ Køn ’1’ ’Mand’ ’2’ ’Kvinde’ ’.u’ ’Uoplyst’
KODESÆT ::=	KODELISTENAVN LINJESKIFT {KODE MELLEMRUM KODEBESKRIVELSE LINJESKIFT}	
KODELISTENAVN ::=	TITEL	
KODE ::=	APOSTROF HELTAL | DECIMAL | FRITEKST APOSTROF	
KODEBESKRIVELSE ::=	APOSTROF FRITEKST APOSTROF	
HELTAL ::=	*!! se figur 9.6 EBNF for heltalstype*	25 +45 -234 10000
DECIMAL ::=	*!! se figur 9.7 EBNF for decimaltalstype*	23,75 .10 -123.76 +123.76 150000.25
BRUGERKODE ::=	”BRUGERKODE” LINJESKIFT {VÆRDISÆT} LINJESKIFT {LINJESKIFT}	BRUGERKODE Nationalitet ’9’ ’10’
VÆRDISÆT ::=	VARIABELNAVN MELLEMRUM VÆRDI {MELLEMRUM VÆRDI} LINJESKIFT	
VÆRDI ::=	APOSTROF HELTAL | DECIMAL | FRITEKST APOSTROF	

Figur 9.12 EBNF Syntaksregler for datafil

Nonterminal	**Terminal**	Eksempel

DATAFIL ::=	DATAINDHOLD	`id;køn;fødselsdato;fødselstidspunkt`
		`0101013333;2;2016-12-15;17:23:04`
		`0202024444;1;1980-02-21;05:15:42`
		`0303035555;1;1079-11-04;03:33:08`

DATAINDHOLD ::=	OVERSKRIFT LINJESKIFT RÆKKE {RÆKKE}	
OVERSKRIFT ::=	VARIABELNAVN {SEPARATORTEGN VARIABELNAVN}*!! Der anvendes ikke et separatortegn efter det sidste variabelnavn i overskriftslinjen*	
VARIABELNAVN ::=	TITEL	
TITEL ::=	(BOGSTAV {BOGSTAV | CIFFER}) | (DOBBELTAPOSTROF BOGSTAV {BOGSTAV | CIFFER} DOBBELTAPOSTROF) *!! En TITEL må aldrig begynde med et tal, men må godt være en blanding af bogstaver og tal, på max 128 tegn, og hvis titel er et reserveret ord i SQL:1999 skal titel omkranses af dobbeltapostrof, jf. ISO/IEC 9075:1999 - Database Language SQL (SQL:1999)*	
BOGSTAV ::=	”A” |”B” | … | ”Z” | ”a” | ”b” | … | ”z” | *”_”!! Samt andre nationale karakterer, der er tilladte i constraint names i standarden SQL:1999 !! underscore (U+005F)*	
CIFFER ::=	”0” | ”1” | ”…” | ”9”	
SEPARATORTEGN ::=	”;” *!! semikolon (U+003B)*	
LINJESKIFT ::=	(CR LF) | CR | LF	
CR ::=	*!! vognretur (U+000D)*	
LF ::=	*!! linjeskift (U+000A)*	
RÆKKE ::=	VÆRDI | (DOBBELTAPOSTROF VÆRDI DOBBELTAPOSTROF) {SEPARATORTEGN VÆRDI | (DOBBELTAPOSTROF VÆRDI DOBBELTAPOSTROF)} LINJESKIFT *!! Der anvendes ikke et separatortegn efter den sidste værdi i en række. Se 9.G.1.b for omslutning af en værdi, som indeholder separatortegn og/eller dobbeltapostrof*	
VÆRDI ::=	BOGSTAV | CIFFER | ANDRETEGN {BOGSTAV | CIFFER | ANDRETEGN}*!! VÆRDI kan være en blanding af bogstaver tal og andre tegn !!Foran- og efterstillede blanktegn er ikke tilladt i værdier*	
ANDRETEGN ::=	*!! alle tilladte tegn i UTF-8, som ikke er bogstaver og tal, jf. 9.F.*	
DOBBELTAPOSTROF ::=	”” *!! dobbeltapostrof (U+0022)*	

Bilag A – Ændringer i forhold til tidligere bekendtgørelse (bek. 1007)

Større ændringer:

- Nyt format for aflevering af data fra statistikfiler er tilføjet som nyt bilag 9
- Gældende retningslinjer for aflevering af geodata er tilføjet i bilag 5, punkt 5.G.

Mindre ændringer vedrørende forskningsdata:

- Forskningsdata i form af databaser og dokumentsamlinger skal afleveres efter reglerne i bilag 1-8 og vejledningen er tilpasset dette.
- Ny indeksfil der indeholder signifikante egenskaber ved statistikdata afleveret efter bilag 9 er tilføjet i bilag 4, figur 4.5 researchIndex.xml.
- Følgende nye elementer er tilføjet i indeksfilen *archiveIndex.xml* i bilag 6, figur 6.1.
    - researchSIP
- Følgende nye dokumentkategorier er tilføjet i indeksfilen *contextDocumentationIndex.xml* i bilag 6, figur 6.2.
    - 7.a researchProjectDescription
    - 7.b researchQuestionnaire
    - 7.c researchProtocol
    - 7.d researchPublication
    - 7.e researchInformationOther

Mindre ændringer vedrørende administrative data:

- Præcisering af paragraf 2
- Præcisering af vejledning til paragraf 6.
- Præcisering af vejledning vedr. hvordan sager lukkes korrekt i bilag 3, 3.A.3.b
- Præcisering af teksten i bilag 4, 4.A.1
- Præcisering af teksten i bilag 4, 4.D.6
- Præcisering af vejledning til bilag 5, 5.B og figur 5.1
- Præcisering af tekniske krav til TIFF-dokumenter, f.eks. tilladte bitdybder i bilag 5, 5.E.
- Nyt punkt 5.G.1.a om at geodatafiler journaliseret på sager i systemer med dokumenter kan konverteres til TIFF eller JPEG-2000
- Følgende nye elementer er tilføjet i indeksfilen *archiveIndex.xml* i bilag 6, figur 6.1.
    - archiveTypeClosedFiles
    - documentPeriodStart
    - documentPeriodEnd
    - containsGeodata
    - containsResearchData
    - researchSIP
    - documentsDisposal

- Følgende nye dokumentkategorier er tilføjet i indeksfilen *contextDocumentationIndex.xml* i bilag 6, figur 6.2.
    - 1.l systemTaxonomy
    - 1.m systemInstruction