Rigsarkivet

Håndtering af dokumenter ved produktion og aflevering af arkiveringsversioner til Rigsarkivet

Retningslinjer efter reglerne i bekendtgørelse nr. 128 om arkiveringsversioner

Rigsarkivet

Indhold

1. Læsevejledning til retningslinjerne 2

A. Vejledningens målgruppe og anvendelse 2
B. Henvisning til øvrig vejledning 3
C. Lovgivning og retsforskrifter 3
D. Definitioner 3

2. Håndtering af digitale dokumenter 5

A. Kvalitetskontrol af dokumenter i driftssystemet 5
B. Flere versioner af samme dokument 5
C. Indlejrede dokumenter og vedhæftede filer 5
D. Pakkede og komprimerede filer (zip-filer) 6
E. Kassation af dokumenter 7

3. Dokumentformater 9

A. Hvad er et dokument? 9
B. Tekst og billeder 9
C. Lyd og video 10
D. Geografiske data 10
E. Databaser 11
F. Indlejrede elementer i dokumenter 11
G. Særlige forhold ved regneark 12
H. Binary Large Object (dokumenter i BLOB-felter) 12
I. XML og anden opmærket tekst 13
J. Øvrige dokumentformater 14

4. Konvertering af dokumenter til bevaringsformat 16

A. Ressourceforbrug ved konvertering 16
B. Navngivning og placering af dokumenter i arkiveringsversionen 16
C. Konvertering af dokumenter til TIFF 19
D. Konverteringsfejl 20
E. Erstatningsdokumenter 20
F. Manglende dokumenter i arkiveringsversionen 22
G. Dokumenter med adgangskoder 22
H. Tomme dokumenter 22

5. Bilag 1: Dokumenter, der altid kan kasseres 24
6. Bilag 2: Fejlliste 25

Rigsarkivet

1. Læsevejledning til retningslinjerne

Offentlige myndigheder er forpligtede til at aflevere en kopi i form af bl.a. tabelindhold og af digitale dokumenter fra bevaringsværdige it-systemer til et offentligt arkiv i form af en arkiveringsversion. Statslige myndigheder skal aflevere til Rigsarkivet. Kommuner og regioner kan vælge, om de vil aflevere til Rigsarkivet, eller om de vil oprette deres eget arkiv efter arkivlovens § 7 (kommunearkiv/stadsarkiv/§ 7-arkiv).

Af hensyn til bevaring og fremtidig brug af data har Rigsarkivet fastsat regler for produktion af arkiveringsversioner, som alle myndigheder skal overholde. Disse bestemmelser fremgår af Rigsarkivets bekendtgørelse om arkiveringsversioner (bekendtgørelse nr. 128 af 28. februar 2020, herefter benævnt bekendtgørelse 128/2020).

Statslige myndigheder skal anmelde deres it-systemer til Rigsarkivet før ibrugtagning. Bevaringsværdige systemer skal endvidere godkendes af Rigsarkivet, for at det kan sikres, at data og dokumenter kan konverteres til en arkiveringsversion. Bestemmelserne er fastsat i Rigsarkivets cirkulære om anmeldelse og godkendelse (cirkulære nr. 9290 af 21. juni 2013, herefter benævnt cirkulære 9290/2013).

Kommuner og regioner skal ikke anmelde deres it-systemer. Systemerne skal heller ikke godkendes, men de skal leve op til bestemmelserne i bekendtgørelse 128/2020.

A. Målgruppe og anvendelse

Retningslinjerne henvender sig til de statslige, kommunale og regionale myndigheder, som skal producere og aflevere arkiveringsversioner af it-systemer med dokumenter til Rigsarkivet. Dvs. til de myndigheder, som har Rigsarkivet som modtagende arkiv, og hvor Rigsarkivet tolker Vejledning til bekendtgørelse 128/2020.

Retningslinjerne beskriver enkelte steder de situationer, hvor myndigheden skal kontakte Rigsarkivet for at drøfte konkrete problemstillinger. It-leverandører bør gøre deres kunder opmærksomme på situationer, hvor der skal laves aftaler med eller indhentes tilladelser fra Rigsarkivet, hvis myndigheden ikke selv er opmærksom på forholdet.

Rigsarkivet har ikke hjemmel til at fastsætte, at Retningslinjerne også skal gælde for § 7-arkiverne. Dvs. Rigsarkivet kan ikke fastsætte regler for, hvordan § 7-arkiverne skal tolke Vejledning til bekendtgørelse 128/2020 over for deres kommuner. Derfor er § 7-arkiverne ikke omfattet af Retningslinjerne. Rigsarkivet anbefaler dog, at § 7-arkiverne også anvender Retningslinjerne ved aflevering af it-systemer med dokumenter.

B. Henvisning til øvrig vejledning

Foruden disse retningslinjer har Rigsarkivet udarbejdet andet vejledningsmateriale, der har betydning for produktion og aflevering af arkiveringsversioner:

- Vejledning til bekendtgørelse 128/2020 om arkiveringsversioner
- Vejledning til cirkulære 9290/2013 om anmeldelse og godkendelse af statslige it-systemer
- Anvisning i aflevering af bitemporale data
- Kort og godt om aflevering af it-systemer
- Kort og godt om kassation af dokumenter
- Kort og godt om test af arkiveringsversioner
- Kort og godt om views og SQL-forespørgsler
- Oversigt over de mest almindelige fejl ved test
- Eksempel på arkiveringsversion af et system med dokumenter (AVID.SA.18000)
- Introduktion til Ada, Rigsarkivets program til test af arkiveringsversioner
- Brug af erstatningsdokumenter og Parent <pID> i docIndex.xml
- Vejledning til deduplikering af dokumenter

Vejledningsmaterialet findes på Rigsarkivets hjemmeside

- Statslige myndigheder
- Regionale og kommunale myndigheder

C. Lovgivning og retsforskrifter

Information om lovgivning og retsforskrifter findes på Rigsarkivets hjemmeside:

- Statslige myndigheder
- Regionale og kommunale myndigheder

D. Definitioner

Arkiveringsversioner af it-systemer med dokumenter består overordnet set af

1) Tabeldata der afleveres i en nærmere defineret xml-struktur (se bekendtgørelse 128/2020),
2) Dokumenter, der skal afleveres i Rigsarkivets bevaringsformater.

Bevaringsformater: Dokumenter i it-systemer skal afleveres i et af følgende bevaringsformater: TIFF, JPEG2000, MP3, WAVE, MPEG2 og MPEG4. Derudover lagres eventuelle geodata som GML-filer efter nærmere fastsatte retningslinjer. Yderligere om formater: se afsnit 3.

Dokumenter er filer fra kontorprogrampakker som tekstbehandling, præsentationer og regneark, foruden billeder, lyd og video. Det kan fx være: ind- og udgående breve, mødereferater, notater, indstillinger, digitale billeder, forskningsrapporter m.v. Lyd, video og geodata kan også indgå i

ESDH- og EDH-systemer. Databaser er ikke omfattet af Rigsarkivets og denne vejlednings dokumentbegreb. Om databaser i it-systemer med dokumenter, se afsnit *3.E Databaser*.

Myndigheder er alle offentlige myndigheder på alle administrative niveauer: kommuner, regioner og stat, herunder domstolene, særlige forvaltningsenheder, offentligt ejede virksomheder samt de selvejende institutioner, der har pligt til at aflevere til et offentligt arkiv. Hvis der er tvivl om, hvorvidt en institution falder ind under dette myndighedsbegreb, kan Rigsarkivet hjælpe med afklaring.

Standardkassation er kassation af filer, der optræder i Retningslinjernes bilag 1.

2. Håndtering af digitale dokumenter

A. Kvalitetskontrol af dokumenter i driftssystemet

Statslige myndigheder skal løbende foretage kvalitetssikring af it-systemets dokumenter og dokumentregistrering (dokumenternes metadata). Når en statslig myndighed får godkendt et system efter cirkulære 9290/2013, stiller Rigsarkivet krav om, at der kun gemmes dokumenter i filformater, der kan konverteres eller udpakkes til et bevaringsformat. Det er myndighedens ansvar at kontrollere, at dette overholdes i praksis, også selvom systemet teknisk set kan håndtere de ikke tilladte formater.

Rigsarkivet anbefaler, at også kommuner og regioner kontrollerer, at deres dokumenter kan konverteres til et bevaringsformat, da deres arkiveringsversioner skal leve op til bekendtgørelse 128/2020.

Myndighedens kvalitetskontrol skal sikre, at dokumenterne er læsbare, og at de er korrekt registreret i systemet (har de rigtige metadata), og at dokumenterne kan konverteres til bevaringsformaterne (Se vejledning til bekendtgørelse 128/2020, bilag 5.E).

B. Flere versioner af samme dokument

I mange it-systemer med dokumenter bliver der lagret flere versioner af samme dokument, fx hver gang der trykkes på "Gem". Dette kaldes versionering. Versionering kan dermed være udtryk for, at et dokument udarbejdes over en periode. I disse tilfælde skal kun det færdige dokument indgå i arkiveringsversionen. Håndteringen af dette aftales med Rigsarkivet.

Versionering kan også anvendes i forbindelse med forelæggelse eller godkendelse af et dokument, så den administrative og/eller politiske beslutningsgang dokumenteres. Ved aflevering af et system, hvor versionering har været brugt på denne måde, skal myndigheden kontakte Rigsarkivet for at lave en aftale om håndtering af de versionerede dokumenter.

C. Indlejrede dokumenter og vedhæftede filer

I mange programmer kan man indlejre en fil i et andet dokument. Det kan fx være:

- Et billede, animation (gif-fil) eller video, der indlejres i et tekstdokument, et regneark eller en præsentation
- Et billede, tekstdokument, regneark, der er indlejret i stedet for at være vedhæftet i en e-mail

Et billede vises normalt i det dokument, det er indlejret i. I sådanne tilfælde kommer billedet normalt med ved konvertering til bevaringsformat, men myndigheden skal alligevel være opmærksom ved

produktion af arkiveringsversionen og sikre sig, at de indlejrede billeder af synlige. Indlejrede regneark og mediefiler skal oftest eksekveres (startes, åbnes), for at indholdet bliver synligt. Det betyder, at indholdet ikke automatisk kommer med i konvertering til bevaringsformat. Det kan i stedet være nødvendigt at håndtere dokumenterne manuelt ligesom dokumenter med konverteringsfejl eller få programmeret særlige rutiner til dette. Se afsnit *4.D Konverteringsfejl*.

I mange mails er der vedhæftede eller indlejrede dokumenter, som skal medtages i arkiveringsversionen.

Dokumentoplysninger om bevaringsværdige vedhæftede eller indlejrede dokumenter, der pakkes ud under konvertering til arkiveringsversion, skal registreres i tabeldata, hvis dokumenterne ikke allerede er registreret i it-systemets database, jf. bekendtgørelse 128, punkt 4.A.3. I mange mails indgår der logo-filer, visitkort og digitalsignaturer som vedhæftede eller indlejrede dokumenter. Denne slags dokumenter kan som udgangspunkt standardkasseres, jf. bilag 1. Det vil derfor være en fordel at frasortere filerne i det omfang, det er muligt, forud for konvertering og aflevering.

Standardkasserede dokumenter i henhold til bilag 1 er således ikke bevaringsværdige. Det er derimod dokumenter, som er kasseret fordi de ikke kunne konverteres under aflevering til arkiv, svarende til de dokumenter der fremgår af fejllisten, jf. bilag 2.

Dokumenterne skal registreres med sigende titler og så mange dokumentoplysninger som muligt i tabeldata, så de kan fremsøges af en arkivbruger. Som minimum kan angives de oplysninger, der også registreres i docIndex.xml filen, dvs. filens originaltitel. Dokumentoplysningerne kan tilføjes i en ny tabel skabt til arkiveringsversionen eller i en eksisterende tabel med dokumentregistreringer. Referencer mellem de udpakkede dokumenter og de hoveddokumenter, de har været indlejret i eller vedhæftet, skal etableres med brug af fremmednøgler.

Referencer mellem dokumenterne skal derudover også registreres ved at tilføje et *ParentID* <pID> til child-dokumentet i docIndex.xml, når det er muligt. Dette er fx ikke muligt, når der gøres brug af erstatningsdokumenter til kasserede dokumenter, for at reducere arkiveringsversionens størrelse (deduplikering). Se vejledning til bekendtgørelse 128/2020, afsnit 4.C.6.b, særligt figur 4.4.

Se Rigsarkivets vejledning *Brug af erstatningsdokumenter og Parent <pID> i docIndex.xml* for en mere præcis anvisning i, hvordan indlejrede filer og erstatningsdokumenter skal registreres i arkiveringsversionen.

D. Pakkede og komprimerede filer (zip-filer)

Pakkede og komprimerede filer skal pakkes ud og registreres enkeltvis med særskilt dokument ID i tabeldata og dokumenternes relation til den oprindelige zip-fil skal etableres ved brug af fremmednøgler.

Dokumenternes relation til den oprindelige pakkede fil (zip-fil) skal også opmærkes som en parent/child-relation i docIndex.xml filen, når det er muligt.

Når dokumenterne er pakket ud og konverteret til deres respektive bevaringsformat, kan den oprindelige zip-fil kasseres efter samme procedure som ved standardkassation. I fald det standardkasserede dokument, fx den pakkede fil, er registreret i tabeldata, skal kassationskoden 99999991 for standardkassation angives i tabeldata for dokumentet i en kolonne opmærket med *functionalDescription* "Kassation".

Ved brug af erstatningsdokumenter for kasserede filer, eller anden deduplikering af ens filer, er det ikke altid muligt at registrere parent/child-relation i docIndex.xml filen, fordi der kun kan angives én parent <pID> per dokument i docIndex.xml. Se afsnit *4.E Erstatningsdokumenter*.

Pakkede og komprimerede filer kan håndteres, inden der laves dataudtræk til arkiveringsversionen. I denne forbindelse skal myndigheden tage stilling til, om de udpakkede filer kan standardkasseres i henhold til Bilag 1.

Se Rigsarkivets vejledning *Brug af erstatningsdokumenter og Parent <pID> i docIndex.xml* for en mere præcis anvisning i, hvordan pakkede filer og erstatningsdokumenter skal registreres i arkiveringsversionen.

E. Kassation af dokumenter

Som udgangspunkt udsteder Rigsarkivet bestemmelser om bevaring eller kassation af hele it-systemer hos statslige myndigheder. I statslige it-systemer med dokumenter kan Rigsarkivet dog i nogle tilfælde give tilladelse til kassation af større, afgrænsede grupper af dokumenter, der falder i en af disse tre kategorier:

- Standardbreve eller andre standardiserede (ens) dokumenter
- Dokumenter lagret i forbindelse med rutinesagsbehandling
- Dokumenter, der udelukkende indeholder oplysninger, som også findes som tabelindhold (registerdata) i systemet

Dokumenterne skal kunne udskilles på en klar og entydig måde. I et ESDH-system eller andet system med dokumenter kan det fx dreje sig om alle dokumenter i en specifik saggruppe. Et eksempel på det kasserede dokument kan indgå i kontekstdokumentationen. De kasserede dokumenters dokumentoplysninger i tabeldata må ikke slettes og kassationsårsag skal angives i tabeldata i en kolonne opmærket med *functionalDescription* "Kassation", jf. figur 4 i afsnit *4.E Erstatningsdokumenter*.

Hvis en statslig myndighed ønsker at kassere dokumenter i en eller flere af de ovenstående kategorier, skal den ansøge Rigsarkivet herom. I ansøgningen skal myndigheden kunne

- Udpege klart og entydigt afgrænsede grupper af dokumenter
- Redegøre for udsøgningen og udskilningen af de pågældende dokumenter
- Argumentere for, hvorfor disse kan kasseres

Kommunale myndigheder skal følge kassationsreglerne i bekendtgørelse nr. 183 af 26. januar 2018 med senere ændringsbekendtgørelser.

Hvis en statslig myndighed har fået tilladelse til at kassere dokumenter i et it-system, skal godkendelsen indgå i kontekstdokumentationen.

For kommunale it-systemer skal der være en angivelse i arkivelndex.xml af, om der er kasseret i henhold til bekendtgørelse 183/2018 med senere ændringsbekendtgørelser, hvis der er foretaget kassation af dokumenter. Jf. Vejledning til bekendtgørelse 183/2018, s. 78.

For dokumenter kasseret efter aftale med Rigsarkivet eller i henhold til bekendtgørelse 183/2018 med senere ændringsbekendtgørelser, skal kassationskoden 99999993 angives.

Der er i øvrigt en række dokumenttyper, som umiddelbart kan standardkasseres (kassationskoden 99999991). Se bilag 1.

3. Dokumentformater

A. Hvad er et dokument?

Dokumenter er som udgangspunkt alle former for dokumentation, der har været anvendt af myndigheden, og som ikke er lagret som rå data i en database. Det er tekstfiler, herunder e-mails, regneark og præsentationer, men kan også være lyd og billedgengivelser, som er skabt eller har været anvendt hos myndigheden. Rigsarkivets brug af begrebet "dokument" er dermed bredere end den almene forståelse.

Hvis myndigheden er i tvivl om, hvorvidt en dokumenttype skal bevares, skal myndigheden kontakte Rigsarkivet for at få en afklaring og eventuel løsning på udfordringer ved konvertering til bevaringsformat. Foruden dokumenttypen lægger Rigsarkivet især vægt på dokumenternes indhold ved vurdering af bevaring.

B. Tekst og billeder

Der er to tilladte bevaringsformater til tekstdokumenter, regneark, præsentationer og billeder: TIFF og JPEG2000. Det er det digitale dokuments indhold, der er afgørende for valget af bevaringsformat, og det er derfor også tilladt at benytte begge formater inden for samme arkiveringsversion.

TIFF er velegnet til tekstdokumenter samt billeder, hvor den eksakte bevaring er af betydning. Ud over filformat kan der vælges forskellige kombinationer af bitdybde (antallet af farver) og komprimeringsmetode alt efter indholdet.
JPEG2000 er velegnet til store billeder med mange farver, hvor den eksakte farve ikke er afgørende.

Dokumenttype	**Egnet format, bitdybde og kompression**
Sort/hvide tekstdokumenter	TIFF 1 bit med gruppe 4 eller LZW kompression
Tekstdokumenter med gråtoner eller få farver eller billeder, hvor farverne ikke er betydningsbærende	TIFF 2, 4 og 8 bit med PackBit eller LZW kompression
Billeder med mange farver, hvor den eksakte farve er af betydning, fx satellitfotos af afgrøder, hvor tolkningen af farven på de enkelte pixels kan være af betydning	TIFF 24 bit med LZW kompression

Dokumenttype	Egnet format, bitdybde og kompression
Store billeder med mange farver, hvor den eksakte farve er af mindre betydning, fx grafisk formidling af kort og tegninger, skannede byggesager eller affotograferede skibstegninger.	JPEG 2000

Kvaliteten skal svare til den kvalitet, som myndigheden har anvendt i driftsversionen, hvilket for digitalt skabte dokumenter svarer til en læsbar udskriftskvalitet. For indskannet eller affotograferet materiale skal kvaliteten modsvare kvaliteten af det oprindelige materiale, der dannede grundlag for digitaliseringen.

Imidlertid afspejler kvaliteten af billeder, hvor formålet er grafisk formidling, ikke nødvendigvis den administrative anvendelse, og der kan være tilfælde, hvor det er hensigtsmæssigt at fortage en reduktion af kvaliteten for disse billeder i forbindelse med konvertering til bevaringsformat. Hvis it-systemet indeholder sådanne billeder, skal myndigheden kontakte Rigsarkivet for at træffe aftale herom.

C. Lyd og video

Hvis myndigheden har anvendt lyd og video, skal det i udgangspunktet afleveres i en kvalitet, der svarer til den, som myndigheden har anvendt. Lyd skal konverteres til MP3 eller WAVE. Video skal konverteres til MPEG-2, hvor lyden er indkodet som MP3, eller som MPEG-4, hvor lyden er indkodet som AAC. Konverteringen skal foretages, så den oprindelige kvalitet i størst muligt omfang beholdes.

Kvaliteten af optagelser afspejler ikke nødvendigvis den administrative anvendelse, og der kan være tilfælde, hvor det er hensigtsmæssigt at foretage en reduktion af kvaliteten i forbindelse med konvertering til bevaringsformat. Hvis it-systemet indeholder lyd eller video, skal myndigheden kontakte Rigsarkivet for at træffe aftale herom.

D. Geografiske data

Geodata er stedbestemte data og kan være lagret som tabelindhold i en (spatial) database, en eller flere filer i særlige geodataformater – fx shapefiler – eller i georefererede grafiske formater.

Strukturerede geodata skal som udgangspunkt konverteres til udvekslingsformatet GML som beskrevet i bekendtgørelse 128/2020, pkt. 5.G., for at undgå informationstab. Hvis it-systemet indeholder geodata, skal myndigheden kontakte Rigsarkivet for at træffe nærmere aftale herom.

Hvis der i et it-system findes enkelte dokumenter i geodataformater (dvs. ustrukturerede geodata), eksempelvis i et ESDH-system, skal dokumenterne konverteres til TIFF. Hvis dokumenterne ikke kan konverteres korrekt til GML eller TIFF, skal myndigheden kontakte Rigsarkivet.

E. Databaser

Hvis der identificeres mindre databaser i et it-system i forbindelse med produktion af arkiveringsversionen, og data derfra har haft betydning for sagsbehandlingen, skal data fra databasen konverteres til et godkendt bevaringsformat. Det kan eksempelvis være rapporter eller tabeludskrifter, der derefter konverteres til TIFF. Hvis ovenstående løsning giver myndigheden problemer, skal myndigheden kontakte Rigsarkivet, der vil vurdere, om der skal udarbejdes en selvstændig arkiveringsversion af data fra databasen, eller om data kan kasseres.

Kommunale og regionale myndigheder skal aflevere eller kassere databaser i henhold til bekendtgørelse nr. 183/2018. Hvis der er tvivl om behandlingsmetoden, skal Rigsarkivet træffe beslutning om den videre proces.

F. Indlejrede elementer i dokumenter

Et dokument, der er konverteret til TIFF, svarer kort fortalt til en udskrift. Hvis et dokument indeholder et indlejret element – fx skjult tekst i form af noter i et præsentationsprogram – vil elementet ikke blive synligt i bevaringsformat.

Rigsarkivet ser løbende problemer i forbindelse med konverteringer af dokumenter fra kontorprogrammer som Office-pakken og lignede programmer. Da disse dokumenter udgør en væsentlig del af de dokumenter, der afleveres, er der her en række udfordringer, hvor myndigheden skal være særlig opmærksom på konverteringsfejl. Det drejer sig primært om tekstbehandling, præsentationer og regneark (om regneark, se også afsnit *3.G Særlige forhold ved regneark*).

Kontorprogrammer giver i mange tilfælde brugeren mulighed for at indlejre elementer i dokumenter. Det drejer sig primært om følgende:

- Skjult tekst (fx noter i præsentationsprogrammer)
- Kommentar
- Korrektur

Hvis indlejring er anvendt, skal myndigheden sikre, at elementerne er synlige, inden konvertering til bevaringsformat, så oplysningerne bevares ved konvertering.

G. Særlige forhold ved regneark

Et regneark, der er konverteret til TIFF, svarer til en udskrift. Hvis et regneark har skjult indhold, fx kommentarer (”gule sedler”) eller skjulte faneblade, vil elementet ikke blive synligt i bevaringsformat. Man bør derfor være opmærksom på følgende:

- Skjulte faneblade
- Skjulte rækker og kolonner
- Kommentarer
- Brugerdefinerede visninger
- Formelark
- Labels
- Kæder mellem faneblade eller andre regneark

Alle ark, rækker og kolonner af betydning for en sag skal konverteres til bevaringsformat. Kolonnernes bredde skal svare til længden af den indeholdte tekst, for ellers vil man ikke kunne se hele teksten i det konverterede dokument. Hvis der har været brugerdefinerede visninger, beregninger eller formler af betydning for en sag, skal de også konverteres til bevaringsformat.

Hvis konverteringen af de brugerdefinerede visninger, beregninger eller formler resulterer i et meget stort antal TIFF-sider, kan Rigsarkivet træffe beslutning om, hvor stor en del af visningerne, beregningerne og formlerne der skal medtages.

Hvis der er anvendt betydningsbærende farvekoder, må regnearket ikke konverteres til sort/hvid.

Når regneark konverteres til TIFF, vil det ofte fylde mange sider. For at sikre overblik anbefales det, at der tilføjes kolonne- og rækkeangivelser på samtlige sider. En hyppigt forekommende fejl er, at brugeren har skrevet et enkelt ubetydeligt tegn langt nede i regnearket, eller at en kolonne er autoudfyldt. Det betyder, at der i konverteringen af arket vil blive medtaget mange kolonner og rækker som er uden indhold. Det tager uforholdsmæssigt megen plads og vil vise sig ved test af arkiveringsversionen. Dette kan imødegås ved at tjekke dokumentmængden for meget store regneark inden konvertering. Det vil spare tid og plads i den sidste ende.

Har myndigheden anvendt store, komplekse regneark, eller hvis myndigheden har regneark, der indeholder forskningsdata, skal konvertering af disse drøftes med Rigsarkivet.

H. Binary Large Object (dokumenter i BLOB-felter)

I nogle it-systemer lagres dokumenter i databasen som Binary Large Objects (BLOB). Hvor dette er tilfældet, skal dokumenterne konverteres til bevaringsformat og registreres som dokumenter. Det skal ske således, at de pågældende dokumenter fremstår i den samme visning, som har været tilgængeligt i systemets driftsversion.

I. XML og anden opmærket tekst

Arkiveringsversioner af it-systemer må ikke indeholde XML-filer. De skal konverteres eller kasseres, alt efter deres formål i driftsversionen. XML-filerne skal håndteres som vist her i figur 1:

XML-filens funktion i driftsversionen	Håndtering	Bemærkning
XML-følgesedler til e-mails, der er journaliseret sammen med den pågældende mail	Standardkasseres, jf. Bilag 1	
XML-filer, der er vedhæftet e-mail, men som har en selvstændig betydning som bilag til e-mailen, dvs. ikke er en følgeseddel	Konverteres almindeligvis til TIFF	Hvis it-systemet indeholder mange filer af denne type, skal myndigheden kontakte Rigsarkivet og aftale, hvordan de skal håndteres
XML-filer, der bruges til at lagre tekstdokumenter i enkelte felter i databasen som Binary Large Object (BLOB)	Dokumenterne i BLOB-felterne skal trækkes ud, konverteres til bevaringsformat og lagres som dokumenter.	Se afsnit *3.H Binary Large Object (dokumenter i BLOB-felter)*
XML-filer, der er brugt som bærende datastruktur	Myndigheden skal kontakte Rigsarkivet og aftale, hvordan XML-data skal håndteres.	Fx XML-databaser, spørgeskemaer, indberetningsskemaer, logfiler

Figur 1

J. Øvrige dokumentformater

Andre dokumentformater kan vise sig umulige at konvertere uden væsentligt datatab. Hvis et it-system indeholder dokumenter af denne type, skal myndigheden kontakte Rigsarkivet for at drøfte mulige løsninger. Figur 2 viser eksempler på sådanne filtyper og på mulige løsninger:

Filtype	Håndtering
Microsoft Project-filer med mange visningsmuligheder såsom gant- diagrammer, ressourcer og kalender	Microsoft Project-filer skal udskrives som dokumenter og medtages i bevaringsformat i arkiveringsversionen, hvis dokumenterne har haft betydning for pågældende sag/projekt
Vektorgrafiske formater som CAD- formater og SVG	2D-formater skal afleveres i TIFF eller JPEG2000 3D-formater skal afleveres i formatet TIFF eller JPEG2000 som et vertikalt og et horisontalt billede

Figur 2

Hvis der i systemet i øvrigt er dokumenter, som ikke kan konverteres, skal Rigsarkivet kontaktes.

4. Konvertering af dokumenter til bevaringsformat

A. Ressourceforbrug ved konvertering

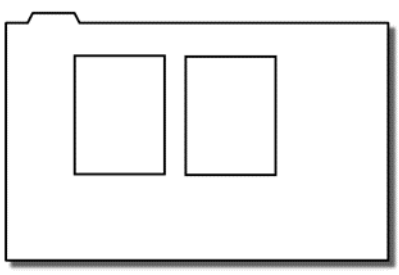
Når der skal produceres en arkiveringsversion af et it-system med dokumenter, skal dokumenterne konverteres til de tilladte bevaringsformater. Arkiveringsversionen skal være færdig cirka 6-8 måneder efter afslutningen af journalperioden (systemets driftsperiode) eller af arkivperioden. Afleveringsfristen fastsættes af Rigsarkivet i samråd med myndigheden.

Myndigheden skal sikre sig – eventuelt i samarbejde med leverandøren – at samtlige dokumenter i systemet kan konverteres til et bevaringsformat inden for ovenstående tidsramme. Ved større it-systemer vil det i en række tilfælde betyde, at forberedelserne til konvertering af dokumenter skal påbegyndes i god tid inden afslutningen af journalperioden/arkivperioden.

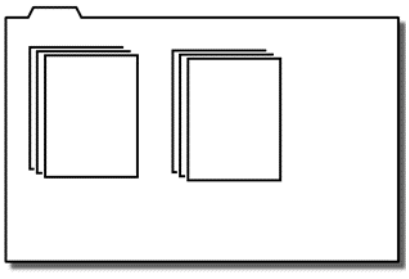
B. Navngivning og placering af dokumenter i arkiveringsversionen

Hvert dokument i arkiveringsversionen består af en eller flere filer i samme bevaringsformat, der navngivnes fortløbende begyndende fra 1 samt formatets filendelse og placeres i en særskilt dokumentmappe, der navngives med dokumentets ID uden foranstillede nuller. Dokumentmappen placeres i en dokumentsamlingsmappe, som navngives "docCollection[fortløbende nummer]" begyndende med 1 og må indeholde op til 10.000 dokumentmapper. Se i øvrigt vejledningen til bekendtgørelse 128/2020, afsnit 4.G.

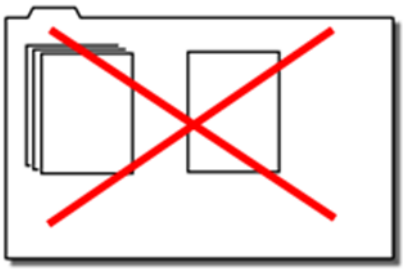
Dokumenters filer i arkiveringsversionen skal navngives 1.[ekstension] (eksempelvis 1.tif eller 1.mp3). Det enkelte TIFF-dokument skal være enten singlepage eller multipage. TIFF-dokumenter, der er multipage, må gerne bestå af flere multipage-filer, se figur 3.



Billede 3: Tilladt: Flere singlepage-filer



Billede 2: Tilladt: Flere multipage-filer



Billede 1: Ikke tilladt: Multipages- og singlepage- filer i samme mappe

Figur 3

Hvis et dokument består af flere filer, skal de placeres i samme dokumentmappe og navngives med et fortløbende nummer.

For alle dokumenter i *Documents* mappen skal der i docIndex.xml angives oplysninger om dokumentets oprindelige filnavn og filtype, samt parent/child-relationer mellem dokumenter når muligt, eksempelvis mails (parent) og vedhæftninger (child). Se også afsnit *2.C Indlejrede dokumenter og vedhæftede filer* og *2.D Pakkede og komprimerede dokumenter*. Se i øvrigt Vejledningen til bekendtgørelse 128/2020, afsnit 4.C.6.a.

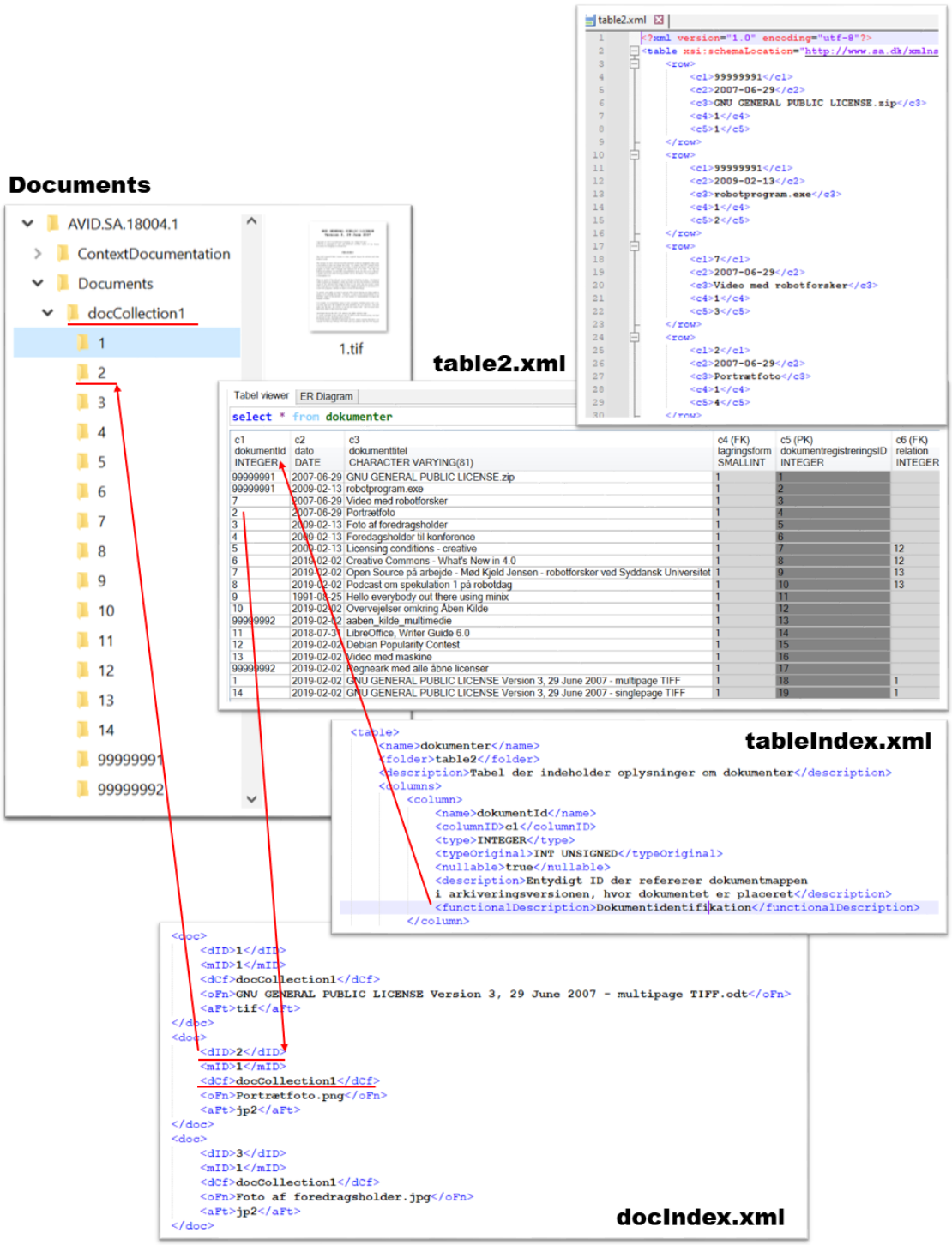
Alle dokumenter i *Documents* mappen skal registreres med oplysninger om dokumenterne i tabeldata (table(n).xml filerne i *Tables* mappen). Relationer mellem dokumenterne, fx mellem dagsorden og bilag eller mailen og dennes vedhæftninger, skal angives i tabeldata ved brug af fremmednøgler. I tabeldata skal også angives hvilken lagringsform dokumentet har i arkiveringsversionen.

Bemærk at dokumentets ID skal angives tre steder i arkiveringsversionen (Billede 4):

1. I navngivning af dokumentmappen i *Documents* mappen
2. I en dokumentregistreringstabel (table[n].xml) i kolonne opmærket med *functionalDescription* "Dokumentidentifikation"
3. I docIndex.xml filen i elementet <dID>

Kolonner opmærket med *functionalDescription* "Dokumentidentifikation" i tabeldata, indeholder således via dokumentets id en reference, der udpeger den dokumentmappe i arkiveringsversionen, hvor det/de digitale dokumenter, der hører til dokumentregistreringen, er placeret. Stien til denne dokumentmappe kan udledes af docIndex.xml filens oplysninger om lagringsmedie <mID>, dokumentsamlingsmappe <dCf> og dokumentets id <dID> (fx AVID.SA.18004.1\docCollection1\2).

På denne måde skabes der i en arkiveringsversion af et it-system med digitale dokumenter en forbindelse mellem registreringerne om dokumentet i tabeldata og selve dokumentet i *Documents* mappen som illustreret i Billede 4:


Documents
AVID.SA.18004.1
ContextDocumentation
Documents
docCollection1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
99999991
99999992
1.tif
table2.xml
table2.xml
<?xml version="1.0" encoding="utf-8"?>
<table xsi:schemaLocation="http://www.sa.dk/xmlns
<row>
<c1>99999991</c1>
<c2>2007-06-29</c2>
<c3>GNU GENERAL PUBLIC LICENSE.zip</c3>
<c4>1</c4>
<c5>1</c5>
</row>
<row>
<c1>99999991</c1>
<c2>2009-02-13</c2>
<c3>robotprogram.exe</c3>
<c4>1</c4>
<c5>2</c5>
</row>
<row>
<c1>7</c1>
<c2>2007-06-29</c2>
<c3>Video med robotforsker</c3>
<c4>1</c4>
<c5>3</c5>
</row>
<row>
<c1>2</c1>
<c2>2007-06-29</c2>
<c3>Portrætfoto</c3>
<c4>1</c4>
<c5>4</c5>
Tabel viewer
ER Diagram
select * from dokumenter
c1 dokumentId INTEGER
c2 dato DATE
c3 dokumenttitel CHARACTER VARYING(81)
c4 (FK) lagringsform SMALLINT
c5 (PK) dokumentregistreringsID INTEGER
c6 (FK) relation INTEGER
99999991 2007-06-29 GNU GENERAL PUBLIC LICENSE.zip 1 1
99999991 2009-02-13 robotprogram.exe 1 2
7 2007-06-29 Video med robotforsker 1 3
2 2007-06-29 Portrætfoto 1 4
3 2009-02-13 Foto af foredragsholder 1 5
4 2009-02-13 Foredagsholder til konference 1 6
5 2009-02-13 Licensing conditions - creative 1 7 12
6 2019-02-02 Creative Commons - What's New in 4.0 1 8 12
7 2019-02-02 Open Source på arbejde - Med Kjeld Jensen - robotforsker ved Syddansk Universitet 1 9 13
8 2019-02-02 Podcast om spekulation 1 på robotdag 1 10 13
9 1991-08-25 Hello everybody out there using minix 1 11
10 2019-02-02 Overvejelser omkring Åben Kilde 1 12
99999992 2019-02-02 aaben_kilde_multimedie 1 13
11 2018-07-31 LibreOffice, Writer Guide 6.0 1 14
12 2019-02-02 Debian Popularity Contest 1 15
13 2019-02-02 Video med maskine 1 16
99999992 2019-02-02 Regneark med alle åbne licenser 1 17
1 2019-02-02 GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - multipage TIFF 1 18 1
14 2019-02-02 GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - singlepage TIFF 1 19 1
tableIndex.xml
<table>
<name>dokumenter</name>
<folder>table2</folder>
<description>Tabel der indeholder oplysninger om dokumenter</description>
<columns>
<column>
<name>dokumentId</name>
<columnID>c1</columnID>
<type>INTEGER</type>
<typeOriginal>INT UNSIGNED</typeOriginal>
<nullable>true</nullable>
<description>Entydigt ID der refererer dokumentmappen
i arkiveringsversionen, hvor dokumentet er placeret</description>
<functionalDescription>Dokumentidentifikation</functionalDescription>
</column>
docIndex.xml
<doc>
<dID>1</dID>
<mID>1</mID>
<dCf>docCollection1</dCf>
<oFn>GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - multipage TIFF.odt</oFn>
<aFt>tif</aFt>
</doc>
<doc>
<dID>2</dID>
<mID>1</mID>
<dCf>docCollection1</dCf>
<oFn>Portrætfoto.png</oFn>
<aFt>jp2</aFt>
</doc>
<doc>
<dID>3</dID>
<mID>1</mID>
<dCf>docCollection1</dCf>
<oFn>Foto af foredragsholder.jpg</oFn>
<aFt>jp2</aFt>
</doc>

Billede 4

Rigsarkivet

C. Konvertering af dokumenter til TIFF

Komprimering

Inden for de overordnede regler for komprimering af dokumenter ved konvertering til bevaringsformat er det muligt at nedbringe pladsforbruget i arkiveringsversionen fx ved anvendelse af optimal bitdybde. Rigsarkivet anviser overordnede regler for komprimering af dokumenter ved konvertering til bevaringsformat jf. bekendtgørelse nr. 128/2020.

Dokumenter i TIFF RGB må udelukkende benytte bitdybderne 1, 2, 4, 8, 24 og 32, mens dokumenter i TIFF CMYK udelukkende må benytte bitdybderne 1, 2, 4, 8, 32 og 40.

Man kan optælle antallet af unikke farver og dermed bestemme den bedste bitdybde. En bitmap indeholder et antal pixels, som hver især repræsenterer en farve. Hver pixel optager plads afhængig af, hvordan man gemmer. En pixel kan gemmes i form af index til palette (1 til 8 bit) eller direkte som en binær værdi (typisk 16 eller 24 bit). En side med 15 unikke farver behøver kun 4 bits (en halv byte), hvorimod en side med 34.243 unikke farver har brug for 16 bit (2 bytes) og kan gemmes i 24 bit. Hvis man gemmer som 24 bit, vil der blive anvendt 3 bytes, uanset om siden indeholder mange, få eller ingen data. Hvis man optæller antallet af unikke farver anvendt på en side, kan man bestemme, hvor mange bits der er nødvendige for at udtrykke netop dette antal farver, og derefter gemme siden i denne bitdybde.

Dokumenter i TIFF RGB må maksimalt benytte 3 farvekanaler med en maksimal bitdybde på 24 bit (8x8x8 bit) og evt. suppleret med én alfakanal på 8 bit. Den samlede bitdybde må ikke overstige 32 bit.

Dokumenter i TIFF CMYK må maksimalt benytte 4 farvekanaler med en maksimal bitdybde på 32 bit (8x8x8x8 bit) og evt. suppleret med én alfakanal på 8 bit. Den samlede bitdybde må ikke overstige 40 bit.

Anvendelse af ekstrakanaler (alpha-kanaler) bør undlades, med mindre der er tale om specifik anvendelse (fx afmaskning af bitmap).

Hvis man kender filformatet, kan man ofte bestemme den optimale bitdybde. Tekstdokumenter (txt) kan gemmes som TIFF grp. 4, og JPG-filer kan gemmes som TIFF 24 bit LZW uden yderligere analyse. Dokumentsider med bitmap af meget begrænset størrelse må antages at være uden betydning for forståelsen af sidens indhold (logo, streger, ikoner). De kan derfor gemmes som 8 bit uanset antallet af unikke farver. Rigsarkivet anbefaler, at der anvendes grp. 4- og LZW-kompression, hvor det er velegnet, eftersom dokumenter komprimeret med disse algoritmer fylder mindre, end hvis der benyttes andre kompressioner.

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

TIFF-angivelse af X-resolution og Y-resolution (DPI/opløsning) er betydningsbærende og vil blive benyttet til skalering i forbindelse med visning/konvertering. Angivelse af X-resolution og Y-resolution bør som udgangspunkt være identisk (fx 300x300 DPI), men kan være forskellig for TIFF-filer, som benytter grp. 3- og grp. 4-kompression. Myndigheden skal sikre sig, at dokumenterne også efter konvertering til bevaringsformatet er læsbare fx 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.

D. Konverteringsfejl

Ved konvertering viser det sig ofte, at dokumenter, der skal indgå i arkiveringsversionen, ikke kan konverteres automatisk. Producenten skal efter automatisk konvertering gennemgå alle de dokumenter, der ikke er konverteret korrekt, og konvertere disse manuelt i det omfang, det er muligt.

Myndigheden skal sende en liste over de dokumenter, som det hverken er muligt at konvertere automatisk eller manuelt (fejllisten), til Rigsarkivet. I fejllisten angives hvilke fejl, der knytter sig til det enkelte dokument. Fejllisten skal udformes som beskrevet i bilag 2. Rigsarkivet godkender fejllisten, hvorefter fejllisten og godkendelsen deraf (fejlnotatet) skal indgå som kontekstdokumentation i arkiveringsversionen.

De fejlbehæftede dokumenter medtages ikke i arkiveringsversionen, men kan erstattes af ét erstatningsdokument og registreres med kassationskode 99999992 i tabeldata, se afsnit *4.E Erstatningsdokumenter*.

E. Erstatningsdokumenter

I fald arkivet beslutter, at der skal tilføjes erstatningsdokumenter for kasserede dokumenter i arkiveringsversionen, skal der oprettes et eller flere erstatningsdokumenter i arkiveringsversionen for:

- Dokumenter der standardkasseres i henhold til *Bilag 1: Dokumenter, der altid kan kasseres* (kassationskode 99999991).
- Dokumenter kasseret på grund af konverteringsfejl, jf. afsnit *4.D Konverteringsfejl* (kassationskode 99999992).
- Dokumenter kasseret efter aftale med Rigsarkivet, jf. afsnit *2.E Kassation* af større grupper af dokumenter (statslige myndigheder) (kassationskode 99999993)
- Dokumenter kasseret efter bekendtgørelse 183/2018 med senere ændringsbekendtgørelser (kommunale myndigheder) (kassationskode 99999993)

Erstatningsdokumentet skal forklare årsagen til kassation. Der oprettes mindst ét erstatningsdokument per kassationsårsag.

Dokumentreferencen (dokumentets ID), angivet i kolonne opmærket med *functionalDescription* "Dokumentidentifikation" i tabeldata for hvert kasseret dokument, skal pege på det erstatningsdokument, der forklarer dokumentets kassationsårsag.

Hvis flere end ét dokument er kasseret af samme årsag, og Rigsarkivet vurderer at deduplikering skal anvendes for at reducere arkiveringsversionens størrelse, skal deres dokumentreferencer pege på et og samme erstatningsdokument, nemlig det, der forklarer årsagen til kassation.

For kasserede dokumenter registreret med dokumentoplysninger i tabeldata skal kassationskode angives i en kolonne opmærket med *functionalDescription* "Kassation" i tableIndex.xml. Kassationskoderne angivet i figur 4 skal anvendes.

Bemærk at det også er muligt ikke at anvende erstatningsdokumenter, men blot registrere kassationsårsagen i tabeldata og angive værdien i kolonnen opmærket med *functionalDescription* "Dokumentidentifikation" som NULL for de kasserede dokumenter. Det er også muligt, at bibeholde et i forvejen eksisterende ID i kolonnen og angive værdien 3 (= ikke relevant, dvs. ikke har tilknyttet en fil) i kolonne opmærket med *functionalDescription* "Lagringsform".

Arkivets beslutning om håndtering af kasserede dokumenter vil fremgå af afleveringsbestemmelsen.

	Årsag til kassation	Kassationskode
	Standardkassation, jf. Rigsarkivets retningslinjer	[99999991]
	Dokument kasseret på grund af fejl i konverteringen (enten i drift eller under aflevering til arkiv)	[99999992]
Statslige myndigheder	Dokument kasseret efter aftale med Rigsarkivet	[99999993]
Kommuner	Dokument kasseret efter bekendtgørelse 183/2018 med senere ændringsbekendtgørelser.	[99999993]

Figur 4

Hvis der i et system er yderligere årsager til kassation, skal producenten kontakte Rigsarkivet.

Se også afsnit *4.D Konverteringsfejl*.

F. Manglende dokumenter i arkiveringsversionen

Hvis der ved konvertering af systemet til arkiveringsversionen opstår problemer med dokumenter, der ikke kan konverteres til bevaringsformat, eller hvis der er dokumenter, som helt mangler i systemet, fx grundet tab ved tidligere konverteringer af driftssystemet, skal myndigheden forsøge at konvertere dokumenterne manuelt eller genskabe dem.

Ved genskabelse menes der, at dokumentet eftersøges i mailprogram, på personlige drev, fællesdrev og lignende og lægges i arkiveringsversionen på det fejlkonverterede dokuments plads. Arbejdet kan foretages i dialog med Rigsarkivet.

Nogle it-systemer tillader, at brugerne kan henvise til dokumenter uden for systemet. Eksempelvis kan der være mulighed for at henvise til dokumenter på lokale fildrev, myndighedens intranet eller i dokumentsiloer uden for selve systemet. Hvis ikke det er muligt at medtage dokumenterne i afleveringen, så de lægges ind og erstatter det henvisende link, skal myndigheden kontakte Rigsarkivet for at finde en løsning. Hvis der har været sådanne eksternt henviste dokumenter, skal dette dokumenteres i arkiveringsversionen, så en fremtidig bruger kan orientere sig om, hvordan systemet har været anvendt. Rigsarkivet beslutter dokumentationens udformning.

G. Dokumenter med adgangskoder

Alle dokumenter i systemet skal indgå i arkiveringsversionen, medmindre der er tale om de dokumenter, der altid kan standardkasseres, eller hvis Rigsarkivet giver tilladelse til kassation, se afsnit 2.E. Afleveringspligten omfatter også dokumenter, der er beskyttet med adgangskode (password).

Rigsarkivet anbefaler alle myndigheder at forsøge at forhindre, at adgangskodebeskyttede dokumenter gemmes i afleveringspligtige it-systemer. Dette gøres almindeligvis ved at skrive det ind i journalinstruksen og ved løbende kvalitetskontrol.

Hvis der alligevel optræder adgangskodebeskyttede dokumenter i it-systemet, når arkiveringsversionen skal produceres, skal myndigheden i første omgang forsøge at få adgang til dokumentets indhold gennem den konkrete medarbejder. Hvis myndigheden har ret til indholdet i dokumentet, kan myndigheden desuden forsøge at knække adgangskoden (bryde passwordet). Hvis ingen af disse løsninger giver positivt resultat, skal myndigheden anmode Rigsarkivet om tilladelse til at kassere dokumentet, jf. afsnit *2.E Kassation af dokumenter*.

H. Tomme dokumenter

Når dokumenter konverteres til bevaringsformatet, kan der opstå konverteringsfejl, så en eller flere sider bliver blanke. Rigsarkivets testværktøj ADA kan undersøge, om der i arkiveringsversionen er blanke dokumenter eller dokumenter hvor første side er blank.

Hvis der er i forbindelse med test af arkiveringsversionen findes flere blanke dokumenter, tolkes dette som en konverteringsfejl. Er der i driftssystemet blanke dokumenter eller dokumenter, hvor første side er blank, skal dette dokumenteres i kontekstdokumentationen, for at arkiveringsversionen kan godkendes. Rigsarkivet anbefaler derfor, at denne test foretages før afleveringen, så eventuelle problemer kan håndteres på forhånd. Hvis der er helt tomme eller blanke dokumenter i driftssystemet på grund af fejljournalisering eller lignende, skal de medtages på listen over fejlede dokumenter og fremgå i kontekstdokumentationen.

5. Bilag 1: Dokumenter, der altid kan kasseres

Mange it-systemer indeholder dokumenter, som ikke har et betydningsbærende indhold med relevans for arkiveringsversionen.

Dokumenter i følgende filformater kan altid kasseres og skal ikke indgå i arkiveringsversioner, jf. for kommunernes vedkommende bekendtgørelse 183/2018, § 4.

- XML-følgesedler og GML-følgesedler. Se afsnit *3.I XML-dokumenter*.
- Pakkede filer – fx zip- eller rar-filer – hvor indholdet er pakket ud, og hver enkelt dokument er registreret med selvstændigt dokument-id i både tabeldata og docIndex.xml filen. Se afsnit *2.D Pakkede og komprimerede dokumenter*.
- VCF-filer eller tilsvarende "visitkort", hvor oplysningerne om afsender/modtager fremgår af
- den mail, visitkortet har været vedhæftet på
- E-mailsignaturer med dertilhørende logoer.
- Digitale signaturer.
- Thumbnails af digitale dokumenter, hvor originalerne indgår i arkiveringsversionen.
- Programfiler, fx msi-, dll-, ini-, exe-, class- eller sys-filer
- Interne og eksterne genvejsfiler, fx til internettet eller brugerens skrivebord. Se afsnit *4.F Manglende dokumenter i arkiveringsversionen*.

For kassation af ovenstående dokumenter gælder det, at de skal kunne identificeres sikkert i it-systemet. Myndigheden skal optælle antallet af dokumenter fordelt på filformat, der kasseres i henhold til bilag 1, og sende opgørelsen til Rigsarkivet.

Rigsarkivet

6. Bilag 2: Fejlliste

Listen over dokumenter, der ikke kunne konverteres, skal indeholde følgende informationer - i det omfang, felterne er relevante for systemet:

- Journal- eller sagsnummer
- Saggruppe, facet eller emneord
- Sagstitel
- Aktnummer
- Akttitel
- Dokumentnummer
- Dokumenttitel
- Dokumentdato
- Filnavn
- Filformat
- Sagsbehandler eller anden kontaktperson
- Kort fejlbeskrivelse. Hvor der er flere ens fejl, skal teksten være identisk

Dokumenter, der er omtalt i bilag 1 og derfor standardkasseres, skal ikke indgå i fejllisten.

Statslige myndigheder

Dokumenter, der kasseres efter bestemmelser fastsat af Rigsarkivet, skal ikke indgå i fejllisten.

Kommunale myndigheder

Dokumenter, der kasseres i henhold til bekendtgørelse nr. 183/2018 med senere ændringsbekendtgørelse, skal ikke indgå i fejllisten.