



Rigsarkivet

Danmarks
hukommelse

Vejledning i brug af erstatningsdokumenter og Parent <pID> i docIndex.xml

Rigsarkivet maj 2025
Version 1.0



Indledning

Dette er en vejledning i opmærkning af parent-relationer i tabeldata (table[n].xml filer) og i docIndex.xml filen (<pID>) samt i registrering af kassationsårsag i forbindelse med kassation og eventuel anvendelse af erstatningsdokumenter.

Det modtagende arkiv kan vælge mellem de nedenstående løsninger afhængigt af de specifikke arkivmæssige hensyn, arkivet ønsker at varetage.

I denne vejledning anvendes begrebet *fil* når der henvises til et filformat (fx zip-fil, msg-fil eller pakket fil) eller hvis begrebet *fil* anvendes i definitionen af et krav i Bekendtgørelse nr. 128 af 12/02/2020 om arkiveringsversioner (fx elementet <oFn> *OrigFilename* i docIndex.xml). Begrebet *dokument* anvendes i bekendtgørelsens betydning, dvs. som et logisk sammenhørende indhold med en tilhørende dokumentregistrering. Bemærk at et dokument kan bestå af flere single-page TIFF-filer placeret i samme dokumentmappe i arkiveringsversionen. Bemærk endvidere at begrebet dokument også dækker lyd, video og geodata.

Krav til og formål med docIndex.xml

Alle dokumenter placeret i arkiveringsversionens mappe *Documents* skal registreres i docIndex.xml. Indeksfilen docIndex.xml skal indeholde oplysninger om hvert dokumentets placering i arkiveringsversionen, oprindeligt filnavn inkl. ekstension (*OrigFilename*), filtype i arkiveringsversionen (*ArchivalFormat*) samt dets eventuelle overordnede dokument (*Parent*).

docIndex.xml danner forbindelsen mellem hvert dokument registreret i tabeldata og dokumentets placering i dokumentmappen i en dokumentssamlingsmappe i en *Documents* mappe på et lagringsmedie (*MediaID*). Denne forbindelse sker via kolonnen i tabeldata opmærket med *functionalDescription* "Dokumentidentifikation" i tableIndex.xml, svarende til *DocID* (<dID>) i docIndex.xml, svarende til nummeret på dokumentmappen i dokumentssamlingsmappen i mappen *Documents* (se **Fejl! Henvissningskilde ikke fundet.1**). Indeksfilen docIndex.xml danner denne forbindelse, da den indeholder oplysning om dokumentssamlingsmappen (*SubFolder* i elementet <dCf>), hvori den pågældende dokumentmappe, navngivet efter *DocID* (<dID>), med dens fil(er) er placeret. Bemærk at et dokument kan bestå af flere single page TIFF-filer. docIndex.xml indeholder også oplysning om lagringsmediet (*MediaID*) i elementet <mID>, hvori den pågældende dokumentssamlingsmappe er placeret.

Ved definitionen af docIndex.xml i 2010 i bek. 1007 valgte Rigsarkivet bevidst, at i docIndex.xml måtte et dokument kun have ét overordnet dokument (*Parent*) jf "Forekomst pr. dokument" lig 0-1 (et valg af enkelt frem for udviklet, på bekostning af funktionalitet). Det var desuden på daværende tidspunkt antagelsen, at oplysninger i docIndex.xml allerede ville forefindes i tabeldata, og at docIndex.xml således skulle fungere som et indeks, hvor man ét sted kunne få oplysningerne, og dermed lette udvikling af løsninger til fremfinding, validering, migrering og tilgængeliggørelse af dokumenter, uden at skulle gennemløbe alle arkiveringsversionens medier og dokumentssamlingsmapper for at finde et givent dokument.





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

Tabel viewer ER Diagram					
select * from dokumenter					
c1	c2	c3	c4 (FK)	c5 (FK)	c6 (FK)
dokumentid	dato	dokumenttitel	lagringsform	dokumentregistreringsID	relation
INTEGER	DATE	CHARACTER VARYING(81)	SMALLINT	INTEGER	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	Foredragsholder 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	aaaben_kilde_multimedia	1	13	
11	2018-07-31	LibreOffice, Winter 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	Rægnemark 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>
  </columns>
</table>
```

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>
```

Figur 1 docIndex.xml og DocID

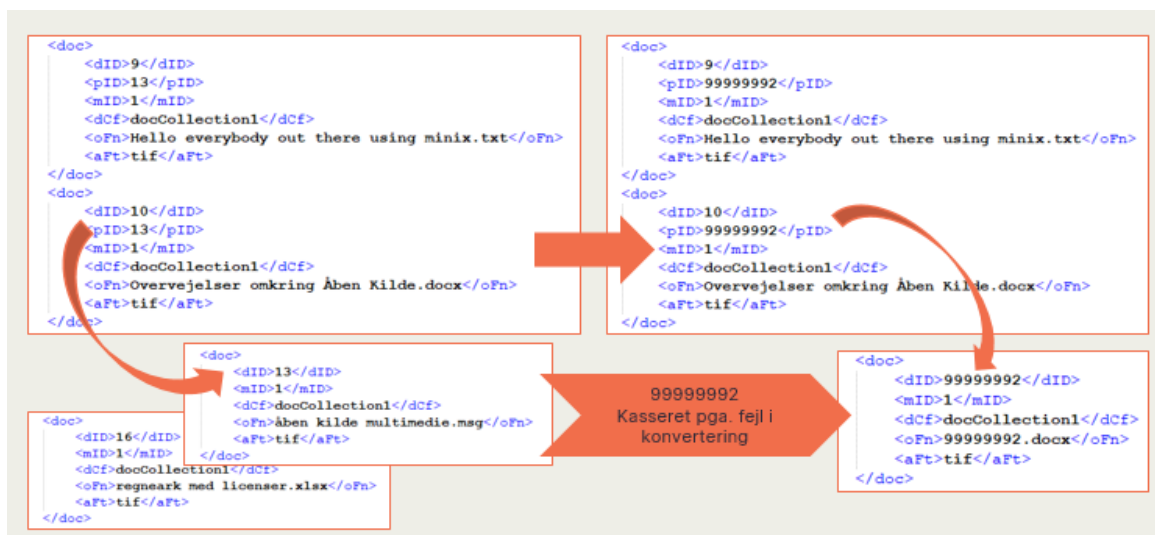
Problemstilling

I visse tilfælde (angivet herunder) er det ikke muligt at lave en korrekt registrering af relationer mellem dokumenter i docIndex.xml (dvs. med brug af "ID på overordnet dokument" Parent <pID> pegende på et



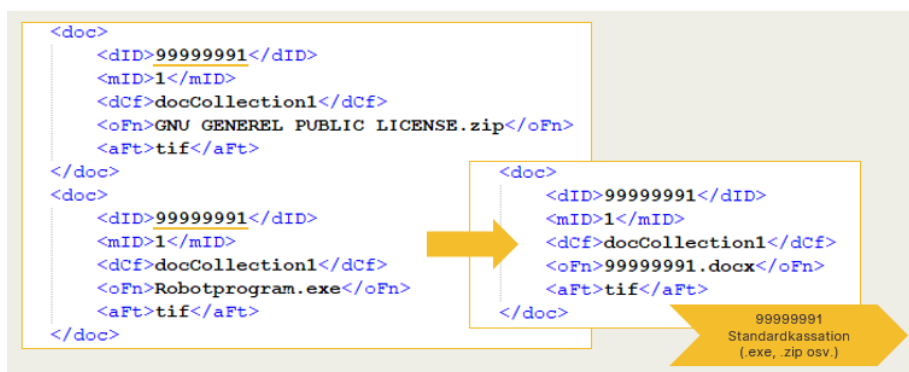
andet *DocID* <dID>). Disse tilfælde forekommer, hvis dokumenter udelukkende registreres i docIndex.xml, og hvis der er krav om brug af erstatningsdokumenter med betydningsbærende *DocID* <dID> (99999991, 99999992 eller 99999993):

- **Kassation af overordnede dokumenter (parent).** Hvis flere dokumenter <dID> peger på forskellige kasserede overordnede dokumenter <pID> (fx når man standardkasserer en zip-fil). Her skal de forskellige kasserede overordnede dokumenter have samme <dID>. Dette er ikke muligt i docIndex.xml. Anvendelsen af *DocID* for erstatningsdokumenter (fx 99999992) medfører også, at originalt filnavn og ekstension angivet i elementet <oFn> for de kasserede dokumenter ikke kan bevares i docIndex.xml, hvorved informationer om de kasserede overordnede dokumenter mistes (se figur 2).



Figur 2 Originalt filnavn og ekstension mistes i docIndex.xml når flere overordnede dokumenter kasseres

- **Kassation af underordnede dokumenter (child).** Hvis flere forskellige dokumenter <dID> udpakket fra en pakket fil <pID> kasseres (fx når man udpakker fra en msg-fil, beholder meddelelseteksten og kasserer nogle vedhæftninger). Her skal de kasserede underordnede dokumenter have samme <dID>. Dette er ikke muligt i docIndex.xml. Anvendelsen af *DocID* for erstatningsdokument (fx 99999991) medfører også, at originalt filnavn og ekstension angivet i elementet <oFn> for de kasserede dokumenter ikke kan bevares i docIndex.xml (se figur 3).



Figur 3 Originalt filnavn og ekstension mistes i docIndex.xml når flere dokumenter kasseres

- **Kasseret dokument (child) med flere overordnede dokumenter (parents).** Hvis et kasseret dokument `<dID>` har flere overordnede dokumenter (*Parent* `<pID>`) (fx fordi et dokument var bilag til to forskellige møder og dermed journaliseret på to forskellige sager) skal flere `<pID>` angives i docIndex.xml for det samme dokument. Dette er ikke muligt i docIndex.xml, jf. bek. 128, punkt 4.C.6.b, Figur 4.4. Alle parent-relation `<pID>` kan i disse tilfælde ikke angives i docIndex.xml.

Rigsarkivets anviste løsning på problemstillingen

Løsningen på ovenstående problemstilling er:

- at omdanne driftssystemets struktur, herunder registrering af dokumenter og relationer mellem disse, til en eller flere tabeller (table[n].xml filer)¹. Relationer mellem dokumenter kan fx være mellem en e-post og dens vedhæftninger.
- at kassationsårsag for et dokument angives i kolonne opmærket med *functionalDescription* "Kassation" `<functionalDescription>Kassation</functionalDescription>`. Derved undgås brugen af betydningsbærende *DocID* `<dID>`.
- at strukturen (dvs. opmærkning af *Parent* `<pID>`) angives i docIndex.xml², når dette er muligt. Dette er et krav angivet i bekendtgørelse 1007 af 20/08/2010 og bekendtgørelse 128 af 12/02/2020, og hvis der dispenseres fra kravet vil det have konsekvenser for de offentlige arkiver, som har baseret deres systemunderstøttelse på dette krav i bekendtgørelsen.
- Det modtagende arkiv kan jf. bekendtgørelse nr. 128, § 6 vurdere om det er nødvendigt at anvende erstatningsdokumenter for kasserede dokumenter og vurdere om deduplikering skal anvendes af hensyn til pladsreduktion i forbindelse med en aflevering eller om årsag til kassation udelukkende skal registreres i kolonne opmærket med *functionalDescription* "Kassation". Med deduplikering menes, at det samme *DocID* `<dID>` anvendes for alle kasserede dokumenter med samme kassationsårsag.
Den valgte løsning til håndtering af kasserede dokumenter skal fremgå af afleveringsbestemmelsen.

Registrering af dokumenter og relationer i tabeller

Alle dokumenter i arkiveringsversionen (placeret i dokumentmapper i en eller flere dokumentsamlingsmapper i en eller flere *Documents* mapper) skal registreres med sigende dokumenttitler samt eventuel anden relevant metadata i arkiveringsversionens tabeller, såsom dokumentdato, afsender/modtager, forbindelse til sag m.m. Dette gør dokumenterne søgbare for arkivbrugerne.

¹ jf. bek 128, punkt 4.A.3

² jf. bek 128, punkt 4.C.6.a



Når pakkede filer (fx en zip-fil eller en msg-fil) pakkes ud i forbindelse med en migrering til arkiveringsversion, skal de udpakkede dokumenter derfor altid registreres i en eller flere tabeller i tabeldata i arkiveringsversionen og relationer mellem dokumenterne opmærkes med brug af fremmednøgler.

De udpakkede dokumenter kan registreres i en ny tabel skabt til arkiveringsversionen. I tabelbeskrivelsen i tableIndex.xml for en tilføjet tabel skal det fremgå tydeligt, at tabellen er tilføjet under aflevering til arkiv, af hvilken årsag den er tilføjet, samt hvad den indeholder.

En pakket fil (fx en zip-fil eller en msg-fil indeholdende de udpakkede dokumenter) antages at være registreret i IT-systemet i en eksisterende tabel, der har en relation til sager.

Man kan dermed nøjes med at tilføje en ny tabel til registrering af de udpakkede dokumenter med følgende 6 kolonner, som illustreret i Figur 4:

1. **kolonne med dokumentregistreringsID opmærket som primærnøgle** (typisk et ikke-betydningsbærende fortløbende tal)
2. **kolonne med dokumentets titel** (optimalt en sigende dokumenttitel og som minimum filnavnet, som også skal angives i docIndex.xml i elementet <oFn>)
3. **kolonne med lagringsform** indeholdende værdien "1" (Helt eller delvis digitalt), hvis dokumentet findes digitalt i arkiveringsversionen eller der er tilføjet et erstatningsdokument for et kasseret dokument.
Kolonnen skal i tableIndex.xml være opmærket med *functionalDescription* "Lagringsform",
<functionalDescription>Lagringsform</functionalDescription>.
4. **kolonne med DocID** (svarende til elementet <dID> i docIndex.xml) indeholdende nummeret på den dokumentmappe, hvori det digitale dokument er placeret i dokumentsamlingsmappen i *Documents* mappen.
Kolonnen skal i tableIndex.xml være opmærket med *functionalDescription* "Dokumentidentifikation",
<functionalDescription>Dokumentidentifikation</functionalDescription>.
5. **kolonne med relation** til tabellen, der indeholder dokumentregistrering af den pakkede fil (fx en zip-fil eller en msg-fil, som dokumentet er udpakket fra) opmærket som fremmednøgle. Denne kolonne kan indeholde angivelse af den parent-relation, som ikke kan angives som <pID> i docIndex.xml (da denne kun tillader én *Parent* per *DocID*).
I Figur 4 refererer fremmednøglen "parentrelation" (c5) i tabel "udpakkede_dokumenter" til primærnøglen "dokumentregistreringsID" (c5) i tabel "dokumenter" i Figur 5.
6. **kolonne med kassation**. I denne kolonne angives en af kassationskoderne for erstatningsdokumenter (enten 99999991, 99999992 eller 99999993) afhængigt af årsagen til kassation af dokumentet, jf. figur 4 i afsnit 4.E i Rigsarkivets dokumentvejledning. Bemærk at modtagende arkiv også kan anviser andre kassationskoder.
Kolonnen skal i tableIndex.xml være opmærket med *functionalDescription* "Kassation",
<functionalDescription>Kassation</functionalDescription>





dokumenter

udpakke dokumenter

dokumentregistreringsID (PK)

dokumenttitel

lagringsform (FK)

SMALLINT

(FK) av_kodetabel_lagringsform (lagringsform)

docID

parentrelation (FK)

kassation (FK)

INTEGER

(FK) av_kodetabel_kassation (kassation)

select * from udpakke dokumenter

c1 (PK)	c2	c3 (FK)	c4	c5 (FK)	c6 (FK)
dokumentregistreringsID	dokumenttitel	lagringsform	docID	parentrelation	kassation
INTEGER	CHARACTER VARYING(81)	SMALLINT	INTEGER	INTEGER	INTEGER
7	Licensing conditions - creative	1	5	12	
8	Creative Commons - What's New in 4.0	1	6	12	
9	Open Source på arbejde - Mød Kjeld Jensen - robotforsker ved Syddansk Universitet	1	7	13	
10	Podcast om spekulation 1 på robotdag	1	8	13	
18	GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - multipage TIFF	1	1	1	
19	GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - singlepage TIFF	1	14	1	

Figur 4 Tabel "udpakke dokumenter" med registrering af de udpakke dokumenter

dokumenter

▶ dokumentid

▶ dato

▶ dokumenttitel

▶ lagringsform (FK)

SMALLINT

(FK) av_kodetabel_lagringsform (lagringsform)

▶ dokumentregistreringsID (PK)

▶ kassation (FK)

INTEGER

(FK) av_kodetabel_kassation (kassation)

▶ udpakke dokumenter

select * from dokumenter

c1	c2	c3	c4 (FK)	c5 (PK)	c6 (FK)
dokumentid	dato	dokumenttitel	lagringsform	dokumentregistreringsID	kassation
INTEGER	DATE	CHARACTER VARYING(81)	SMALLINT	INTEGER	INTEGER
15	2007-06-29	GNU GENERAL PUBLIC LICENSE zip	1	1	99999991
15	2009-02-13	robotprogram.exe	1	2	99999991
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	Foredragsholder til konference	1	6	
9	1991-08-25	Hello everybody out there using minix	1	11	
10	2019-02-02	Overvejelser omkring Åben Kilde	1	12	
16	2019-02-02	aaben_kilde_multimedie	1	13	99999992
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	
16	2019-02-02	Regneark med licenser	1	17	99999992

Figur 5 Tabel "dokumenter" med bestående registreringer, herunder af pakke filer, gentaget DocID (15 og 16) ³

select * from av_kodetabel_kassation	
c1 (PK)	c2
kassationsårsagskode	kassationsårsag
SMALLINT	CHARACTER VARYING(26)
99999991	Standardkassation, jf. Rigsarkivets retningslinjer
99999992	Dokument kasseret på grund af fejl i konverteringen
99999993	Dokument kasseret efter aftale med Rigsarkivet (statslige myndigheder) eller dokument kasseret efter bekendtgørelse 183/2018 med senere ændringsbekendtgørelser (kommuner).

Figur 6 Tabel "av_kodetabel_kassation" med kassationsårsager

Eksempelarkiveringsversion AVID.SA.18011 på Rigsarkivets hjemmeside viser eksemplet i figur 4-6.

Da en kolonne i tabeldata opmærket med *functionalDescription* "Dokumentidentifikation" (jævnfør punkt 4 ovenfor) modsat docIndex.xml godt kan indeholde samme DocID'er <did>, er det her muligt at forskellige kasserede dokumenter refererer det samme erstatningsdokument i arkiveringsversionen, fx angivet med "docID" lig 16 i kolonne "dokumentid" (c1) i Figur 5⁴.

I tabellen "dokumenter" i Figur 5 ovenfor ses fx at både dokumentet "åben_kilde_multimedie" med "dokumentregistreringsID" (c5) lig 13 og dokumentet "Regneark med licenser" med "dokumentregistreringsID" lig 17 er kasseret på grund af fejl i konverteringen. Begge disse dokumentregistreringer refererer derfor erstatningsdokumentet i dokumentmappen med "dokumentid" lig 16 (c1) og kassationsårsagen er angivet i kolonne "kassation" med værdien 99999992⁵. Oplysninger om disse kasserede dokumenter fremgår stadig af kolonnen "dato" (c2) og

³ Bemærk, at for at flere dokumentregistreringer kan benytte samme docid kræves, at OrigFilename er det samme.

⁴ forudsat at denne kolonne ikke opmærkes som en primærnøgle.

⁵ Jf. Dokumentvejledningen, afsnit 4. E, figur 4

"dokumenttitel" (c3). Bemærk at det også er muligt ikke at anvende erstatningsdokumenter, men blot registrere kassationsårsagen i tabeldata og efterlade kolonne med *DocID* med værdien NULL for de kasserede dokumenter eller bibeholde et i forvejen eksisterende *DocID* i kolonnen og angive værdien 3 (= ikke relevant, dvs. ikke har tilknyttet en fil) i kolonne opmærket med *functionalDescription* "Lagringsform".

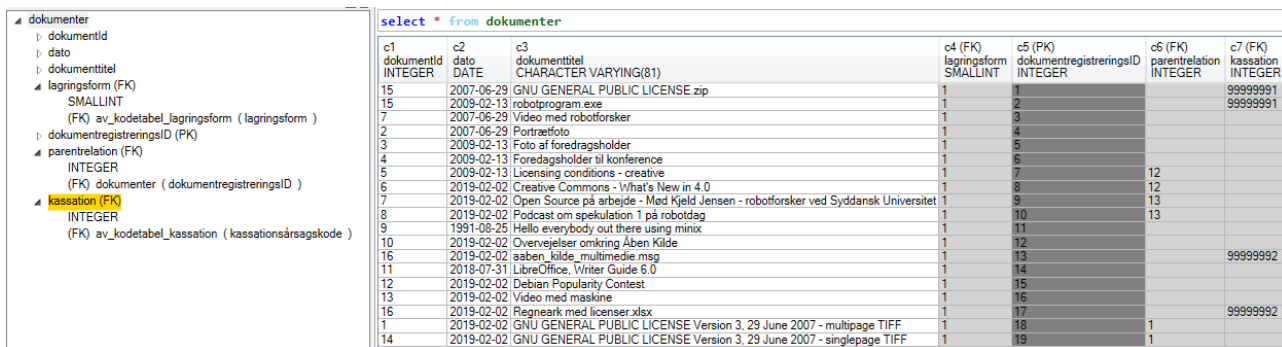
I tabellen "udpakkede dokumenter" i Figur 4 er angivet en relation (fremmednøgle) i kolonnen "parentrelation" (c5) til det dokument, som de udpakkede dokumenter er udpakket fra. Her ses fx at dokumenterne "Open source på arbejde – Mød Kjeld Jensen – robotforsker ved Syddansk Universitet" med "dokumentregistreringsID" (c1) lig 9 og "Podcast om spekulation 1 på robotdag" med "dokumentregistreringsID" lig 10 har en "parentrelation" (c5) med værdien 13, som refererer dokumentet "aaben_kilde_multimedie" registreret i tabellen "dokumenter" i Figur 5 med "dokumentregistreringsID" (c1) lig 13. Dette dokument er dog efter udpakning kasseret grundet fejl i konverteringen, hvorfor kolonnen "kassation" (c1) er udfyldt med værdien 99999992.

Bemærk at ovenstående eksempel illustreret i Figur 4 blot viser det minimum af kolonner med oplysninger, som skal registreres for de udpakkede filer. Hvis der er behov for at registrere andre oplysninger om udpakkede dokumenter tilføjes blot de nødvendige kolonner, fremmednøgler eller tabeller til dette. Ved anvendelse af afleveringsmodel B er kolonnen "Afleveret" fx obligatorisk. Ved udpakning af pakkede filer kan der også forekomme multible hierarkier (fx en mail med en mail vedhæftet, der også har en mail vedhæftet). Parent-relationer kan i disse tilfælde sikres ved, at tilføje en ekstra kolonne til registrering af en fremmednøgle, der refererer kolonnen "dokumentregistreringsID" i selvsamme tabel, her tabellen "udpakkede_dokumenter".



Alternativt til en ny tabel kan de udpakkede dokumenter registreres i en eller flere eksisterende tabeller med dokumentregistreringer og relationerne mellem dokumenterne etableres med de eksisterende fremmednøgler, fx som illustreret i Figur 7.

Eksempelarkiveringsversion AVID.SA.18009 på Rigsarkivets hjemmeside viser eksemplet i figur 7.



c1	c2	c3	c4 (FK)	c5 (PK)	c6 (FK)	c7 (FK)
dokumentid	dato	dokumenttitel	lagringsform	dokumentregistreringsID	parentrelation	kassation
INTEGER	DATE	CHARACTER VARYING(81)	SMALLINT	INTEGER	INTEGER	INTEGER
15	2007-06-29	GNU GENERAL PUBLIC LICENSE zip	1	1		
15	2009-02-13	robotprogram.exe	1	2		
7	2007-06-29	Video med robotforsker	1	3		99999991
2	2007-06-29	Portrætfoto	1	4		99999991
3	2009-02-13	Foto af foredragsholder	1	5		
4	2009-02-13	Foredragsholder 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 - Mad 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		
16	2019-02-02	aaben_kilde_multimediae.msg	1	13		99999992
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		
16	2019-02-02	Regneark med licenser.xlsx	1	17		99999992
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	

Figur 7 Tabel "dokumenter" med udpakkede dokumenter indsat, gentaget DocID (15 og 16)

I Figur 7 er dokumenterne med "dokumentregistreringsID" (c5) lig 7, 8, 9, 10, 18 og 19 udpakket i forbindelse med skabelse af arkiveringsversion og indsat i den bestående dokumentregistreringstabel "dokumenter".

Relationen til dokumentregistreringen for en pakket fil, som disse udpakkede dokumenter (fx 9 og 10) er udpakket fra fremgår af kolonnen "parentrelation" (c6) (værdien 13), der er angivet som en fremmednøgle til kolonnen "dokumentregistreringsID" (c5) i samme tabel.

Rækkerne med "dokumentregistreringsID" (c5) lig 18 og 19 er eksempler på to udpakkede dokumenter, der via kolonnen "parentrelation" (c6) refererer en kasseret zip-fil (dokumentregistreringsID 1), der peger på en dokumentmappe navngivet med "dokumentid" 15 (c1) i arkiveringsversionen, som indeholder et erstatningsdokument for standardkasserede dokumenter, jf. Bilag 1 i Rigsarkivets dokumentvejledning⁶. Af kolonnen "kassation" (c7) er angivet værdien "99999991" som årsagen til kassation. Forklaringen af koden "99999991" findes i kodelisten av_kodetabel_kassation (Figur 6).

⁶ Jf. Dokumentvejledningen, afsnit 4. E, figur 4



Behov for deduplikering grundet størrelse eller antal af erstatningsdokumenter

Deduplikering opnås, når dokumenter, der alle er kasseret af samme grund, peger på det samme erstatningsdokument, dvs. registreres med samme *DocID* i kolonne opmærket med *functionalDescription* "Dokumentidentifikation".

Deduplikering kan anvendes, hvis arkivet ønsker at reducere arkiveringsversionens størrelse.

En anden måde at reducere arkiveringsversionens størrelse på er, at registrere kassationsårsagen i tabeldata og efterlade kolonnen opmærket med *functionalDescription* "Dokumentidentifikation" med værdien NULL for de kasserede dokumenter. Det er også muligt, at bibeholde et i forvejen eksisterende *DocID* i kolonnen og angive værdien 3 (= ikke relevant, dvs. ikke har tilknyttet en fil) i kolonne opmærket med *functionalDescription* "Lagringsform".

Ikke behov for deduplikering, men behov for registrering af alle parents i docIndex.xml

Hvis arkiveringsversionens størrelse, inklusiv dubletter af erstatningsdokumenter, ikke er et problem, er det tilladt at anvende unikke *DocID*'er (dvs. en given *DocID* anvendes kun til én dokumentregistrering). Derved bliver det muligt at registrere alle parent-relationer <pID> i docIndex.xml. Denne løsning er eksemplificeret i Figur 8, Figur 9 og Figur 10.

dokumenter

udpakkede_dokumenter

dokumentregistreringsID (PK)

dokumenttitel

lagringsform (FK)

SMALLINT

(FK) av_kodetabel_lagringsform (lagringsform)

docID

parentrelation (FK)

INTEGER

(FK) dokumenter (dokumentregistreringsID)

kassation (FK)

INTEGER

(FK) av_kodetabel_kassation (kassation)

select * from udpakkede_dokumenter

c1 (PK)	c2	c3 (FK)	c4	c5 (FK)	c6 (FK)
dokumentregistreringsID	dokumenttitel	lagringsform	docID	parentrelation	kassation
INTEGER	CHARACTER VARYING(81)	SMALLINT	INTEGER	INTEGER	INTEGER
7	Licensing conditions - creative	1	5	12	
8	Creative Commons - What's New in 4.0	1	6	12	
9	Open Source på arbejde - Med Kjeld Jensen - robotforsker ved Syddansk Universitet	1	7	13	
10	Podcast om spekulation 1 på robotdag	1	8	13	
18	GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - multipage TIFF	1	1	1	
19	GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007 - singlepage TIFF	1	14	1	

Figur 8 Tabel "udpakkede_dokumenter" med registrering af de udpakkede dokumenter

dokumenter

dokumentid

dato

dokumenttitel

lagringsform (FK)

SMALLINT

(FK) av_kodetabel_lagringsform (lagringsform)

dokumentregistreringsID (PK)

kassation (FK)

INTEGER

(FK) av_kodetabel_kassation (kassation)

udpakkede_dokumenter

select * from dokumenter

c1	c2	c3	c4 (FK)	c5 (PK)	c6 (FK)
dokumentid	dato	dokumenttitel	lagringsform	dokumentregistreringsID	kassation
INTEGER	DATE	CHARACTER VARYING(81)	SMALLINT	INTEGER	INTEGER
15	2007-06-29	GNU GENERAL PUBLIC LICENSE zip	1	1	99999991
17	2009-02-13	robotprogram.exe	1	2	99999991
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	Foredragsholder til konference	1	6	
9	1991-08-25	Hello everybody out there using minix	1	11	
10	2019-02-02	Overvejelser omkring Åben Kilde	1	12	
16	2019-02-02	aaben_kilde_multimedia	1	13	99999992
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	
18	2019-02-02	Regneark med licenser	1	17	99999992

Figur 9 Tabel "dokumenter" med bestående registreringer, herunder af pakkede filer, unik DocID





▲ dokumenter ▷ dokumentId ▷ dato ▷ dokumenttitel ▲ lagringsform (FK) SMALLINT (FK) av_kodetabel_lagringsform (lagringsform) ▷ dokumentregistreringsID (PK) ▲ parentrelation (FK) INTEGER (FK) dokumenter (dokumentregistreringsID) ▲ kassation (FK) INTEGER (FK) av_kodetabel_kassation (kassationsårsagskode)		select * from dokumenter				
c1	c2	c3	c4 (FK)	c5 (PK)	c6 (FK)	c7 (FK)
dokumentId	dato	dokumenttitel	lagringsform	dokumentregistreringsID	parentrelation	kassation
INTEGER	DATE	CHARACTER VARYING(81)	SMALLINT	INTEGER	INTEGER	INTEGER
15	2007-06-29	GNU GENERAL PUBLIC LICENSE.zip	1	1		99999991
17	2009-02-13	robotprogram.exe	1	2		99999991
7	2007-06-29	Video med robotforskere	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 - Mød 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		
16	2019-02-02	aaben_kilde_multimediae.msg	1	13		99999992
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		
18	2019-02-02	Regneark med licenser.xlsx	1	17		99999992
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	

Figur 10 Tabel "dokumenter" med udpakkede dokumenter indsat, unik DocID

Af Figur 9 fremgår det, at de fire kasserede dokumenter med angivelse af kassationsårsag 99999991 eller 99999992 i kolonne "kassation" (c6) alle har unikke *DocID*'er (15, 17, 16 og 18) angivet i kolonnen "dokumentId" (c1).

Eksempelarkiveringsversion AVID.SA.18010 på Rigsarkivets hjemmeside viser eksemplet i Figur 10.



Bilag – krav og hjemmel

Krav til registrering af dokumenter i tabeldata

Af bekendtgørelse 128 af 12/02/2020, bilag 4, punkt 4.A.3 fremgår det at alle dokumenter i IT-systemet skal registreres i tabeldata:

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.

Selvom dokumenterne oprindeligt ikke er lagret i en relationel databasestruktur, skal den betydningsbærende struktur således under aflevering til arkiv migreres til en relationel database. Betydningsbærende strukturer er således også dokumenter lagret og ordnet i mapper på mappedrev, relationer mellem mails og deres vedhæftninger i e-post-systemer, gruppering af dokumenter i zip-filer, relationer defineret i applikationslaget osv.

Krav til brug af indeksfilen docIndex.xml

Bekendtgørelse 128 af 12/02/2020, bilag 4, punkt 4. C. 1.b angiver krav om brug af indeksfilen docIndex.xml, som skal indeholde præcist ét *DocID* <dID> per dokument, og højst ét *Parent* <pID> per dokument.

Brug af functionalDescription "Kassation"

Bekendtgørelse 128 af 12/02/2020, bilag 6, figur 6.6 giver mulighed for i tabeldata at angive kassationsårsag for alle kasserede dokumenter ved brug af *functionalDescription* "Kassation".

