

Kort og godt om datarensning i arkiveringsversioner

- Vejledning til Bekendtgørelse 1007 af 2010 om arkiveringsversioner (*beta-version, der kun omfatter aflevering til Rigsarkivet*)

Indledning

Det er Rigsarkivets opgave at indsamle og bevare **autentiske** data. Derfor skal arkiveringsversionen af et it-system afspejle data i driftssystemet. Arkiveringsversionen produceres ofte af en ekstern it-leverandør.

Nogle gange ved eller opdager myndigheden eller dennes it-leverandør, at data indeholder forskellige former for fejl.

Ikke alle fejl skal rettes, selvom det er muligt. Dette kan synes overraskende. Det skyldes imidlertid kravet om autentiske data i arkiveringsversionen. Hvis der er truffet en administrativ afgørelse på basis af fejlbehæftede data, skal den eller de fejl også optræde i arkiveringsversionen. Kun således bliver det muligt at forstå grundlaget for den trufne afgørelse.

Det er heller ikke tilladt at forsøge at forbedre dataindhold i forhold til driftssystemet. Fejl må ikke rettes, hvis der kan være tvivl om, hvad fejlen betyder – det er bedre at bevare fejl end at gætte forkert.

Denne vejledning giver eksempler på tilfælde, hvor fejl **skal** rettes, og på tilfælde hvor det **ikke** er tilladt at rette datafejl – at udføre datarensning. Vejledningens eksempler er ikke udtømmende og må ikke opfattes som en facitliste. Der kan opstå mange forskellige former for fejl, og det er næppe muligt at beskrive dem alle. I tvivlstilfælde skal Rigsarkivet kontaktes.

I produktionen af en arkiveringsversion skal der skelnes mellem selve indholdet af de data, som udgør informationsindholdet i driftsversionen af systemet og de strukturer og formater, som data er lagret i. Det er tilladt at ændre strukturer og formater i henhold til bekendtgørelse 1007 om arkiveringsversioner, men ikke at ændre i informationsindholdet af data.

Informationsindholdet af data

Det er ikke tilladt at ændre på informationsindholdet af data uanset formålet.

Det er således ikke tilladt at fjerne data på grund af dataindholdet, heller ikke selv om det sker for at beskytte følsomme oplysninger i systemet. Eksempelvis må myndigheden og leverandøren ikke anonymisere data eller fjerne personfølsomme eller andre fortrolige oplysninger. Hvis systemet indeholder personoplysninger, fortrolige eller klassificerede oplysninger, skal dette oplyses til det Rigsarkivet. Adgangen til følsomme oplysninger reguleres i henhold til arkivlovens tilgængelighedsfrister, som er fastsat i § 22-29.

Det er heller ikke tilladt at forkorte eller ændre på dataindholdet, selvom det vil gøre data nemmere at bruge. Eksempelvis må et timestamp ikke ændres til en dato, og et decimaltal må ikke rundes af til et heltal.

Imidlertid findes der undtagelser, hvor Rigsarkivet kræver, at der foretages ændringer.

Hvis der i driftssystemet er felter, der har foran- eller efterstillede blanktegn, skal disse fjernes. Hvis blanktegn har betydning for nøglerelationer i systemet, skal blanktegnene dog erstattes med et UTF8 tegn, der sikrer, at nøglen fortsat er unik. Ikke alle UTF8-tegn er dog tilladt. Se **Vejledning til bekendtgørelse om arkiveringsversioner**, bilag 5, afsnit D.

Databeskrivelser, strukturer og datatyper

Koder og beskrivelser

Der er ofte koder i data. Det kan være fagkoder eller andet, der anvendes af myndigheden, så som lande-, institutions- eller sygdomskoder. Denne form for koder kaldes i denne vejledning **variant A**. Der kan også være koder, der i modsætning til variant A ikke er synlige for brugeren af systemet. Det er koder, som applikationslaget oversætter til klartekst i brugergrænsefladen. Denne form for kode kaldes i denne vejledning **variant B**.

For begge varianter gælder det, at de skal oversættes i arkiveringsversionen.

- **Variant A:** Den faglige viden, som findes på området og ligger bag koden, skal fastholdes i arkiveringsversionen. Det vil ofte være nødvendigt, at fagfolk hos den afleverende myndighed bidrager med deres viden i forbindelse med dokumentation af data til arkiveringsversionen. Hvis det drejer sig om et gammelt fagsystem, som ikke længere anvendes, kan det være en stor opgave at finde korrekte kodeoversættelser. Myndigheden må ikke gætte på kodernes betydning og anvende disse gæt som oversættelser i arkiveringsversionen. Dokumentation af kodeværdien og den faglige viden skal indgå i arkiveringsversionens kontekstdokumentation. Når der optræder koder af denne type i et system, skal myndigheden kontakte Rigsarkivet, som afgør, hvordan dokumentationen skal udformes.
- **Variant B:** Applikationslaget og brugerfladen ikke indgår i afleveringen. Afkodningen vil her som udgangspunkt fremgå af driftssystemet, fordi oversættelsen ligger i applikationslaget. Det kan dog være en kilde til fejl, hvis koderne ikke er unikke, eksempelvis fordi koder er blevet genbrugt over tid. Det samme kan være tilfældet, hvis opdateringer af it-systemet har medført ændringer i koderne og brugen af disse. Her er det nødvendigt at overveje, hvilke mangler der skal udbedres for at gøre indholdet af arkiveringsversionen forståelig, og hvilke fejl der skal forblive uændret, så arkiveringsversionen bedst afspejler driftsversionen. En sådan problemstilling skal drøftes med Rigsarkivet.

Om koder, se desuden **Vejledning til bekendtgørelse om arkiveringsversioner**, bilag 4, afsnit A.

Datatyper

Det er obligatorisk at konvertere datatyper fra driftsversionen, eksempelvis særlige datoangivelser, som ikke er tilladte i henhold til **Bekendtgørelse 1007 af 2010 om arkiveringsversioner**. Det kan i nogle tilfælde være et vurderingsspørgsmål, hvordan konverteringsopgaven løses bedst. Det er myndighedens leverandør, der skal foretage vurderingen, men Rigsarkivet skal altid kontaktes, hvis der er tvivl om den bedste løsning.

Relationer

Arkiveringsversioner af ikke-relationelle it-systemer skal være relationelle i henhold til **Bekendtgørelse 1007 af 2010 om arkiveringsversioner**. Det betyder, at data omstruktureres, og der er her tale om en nødvendig genopretning af data, der sikrer sammenhæng i arkiveringsversionen, svarende til den sammenhæng, der fandtes i driftssystemet.

Der må kun foretages ændringer i relationer, hvis de afspejler datas sammenhæng i driftssystemet.

- I tilfælde, hvor et applikationslag har håndteret relationer, skal tabelrelationer i etableres i arkiveringsversionen.
- I tilfælde, hvor fejl i relationer har været et problem i driftssystemet, og måske været årsag til et ufuldstændigt beslutningsgrundlag, må fejlen ikke udbedres.

I begge situationer skal det dokumenteres, hvordan relationerne er håndteret. Dokumentationen skal indgå i arkiveringsversionens kontekstdokumentation. Myndigheden skal kontakte Rigsarkivet, som anviser, hvordan dokumentationen skal udformes.

Om relationer, se desuden **Vejledning til bekendtgørelse om arkiveringsversioner**, bilag 3, afsnit B, bilag 4, afsnit C og D.

Overholdelse af afleveringsbestemmelse

Rigsarkivet udsteder en afleveringsbestemmelse, der sammen med **Bekendtgørelse om arkiveringsversioner** definerer arkiveringsversionens indhold. Den skal altid følges nøje, men der kan opstå situationer, hvor dette viser sig vanskeligt. Hvis afleveringsbestemmelsen af en eller anden grund ikke kan følges – hvis eksempelvis et dataudtræk på grund af ferie eller helligdage ikke kan laves på den specificerede slutdato – skal Rigsarkivet kontaktes.

Beta-Version