



Digitale filformater

Forvaltningsforum – 20201029

- Stortingsvedtak
- Sentrale forskrifter
- Lokale forskrifter
- Norsk Lovtidend
- EØS-avtalen
- Norges traktater
- Dommer
- Tariffavtaler
- Statens personalhåndboken
- Oversatte lover / Translated Acts
- Oversatte forskrifter / Translated regulations
- Oversatte avgjørelser / Translated decisions

0 Endret ved forskrift 26 nov 2019 nr. 1582 (i kraft 1 jan 2020).

Del III. Formatkrav m.m. ved uttak fra databasetabeller

§ 5-10. Organisering av arkivuttrekk

(1) Arkivuttrekk fra databaser skal bestå av sekvensielle tekstfiler. Et uttrekk kan bestå av en eller flere filer. Relasjoner mellom filene i uttrekket skal være beskrevet i den medfølgende dokumentasjonen slik det fremgår av § 5-24.

(2) Data skal tas ut på en måte som gjør at hver posttype eksporteres til én enkelt sekvensiell fil. Opprinnelige fysiske pekere skal omgjøres til unike logiske nøkler (primær- og fremmednøkler). Felt som er repetert flere ganger innenfor en post, skal skilles ut fra den opprinnelige posten, og eksporteres som egne poster i en ny fil med tilhørende nøkler. Pakkede og binære tallfelt aksepteres ikke i arkivuttrekk. Ved de nødvendige tilfelle spesifiseres det i § 5-11.

(3) For arkivuttrekk i XML-format kan ovennevnte krav avvikes etter avtale med Arkivverket.

(4) For Noark-systemer gjelder § 5-21 til § 5-23.

§ 5-11. Tegnssett i arkivuttrekk

(1) Arkivuttrekk og medfølgende struktur- og innholdsbeskrivelse skal overføres som tekst i kodert form, og benytte godkjente tegnssett.

(2) Godkjente tegnssett er:

- Unicode UTF-8 (ISO/IEC 10646-1:2000 Annex D)
- ISO 8859-1:1998, Latin 1
- ISO 8859-4:1998, Latin-4 for samiske tegn.

(3) Andre tegnssett aksepteres bare etter avtale med Arkivverket.

§ 5-12. Tekstfilformater i arkivuttrekk

Tekstfiler i et arkivuttrekk skal ha ett av følgende formater:

- Fast format: Alle felt kommer i faste posisjoner innen postene. Lengde og/eller startposisjon for hvert felt skal spesifiseres i struktur- og innholdsbeskrivelsen, som skal følge med uttrekket slik det fremgår av § 5-24. Alle lengder og posisjoner angis ut fra antall tegn.
- Tegnseparert format (kommaseparert format) i henhold til standarden IETF RFC 4180: Poster og felt har variabel lengde og er atskilt med faste, entydige skilletegn. Postskille er påkrevd og skal angis i den medfølgende struktur- og

Hva skjer?

Forskriften skal revideres.

Det gjelder også kapittel 5.

https://lovdata.no/dokument/SF/forskrift/2017-12-19-2286#KAPITTEL_5

(Reprise)

Det **vil** bli enklere å revidere detaljer i anbefalte filformater.

Del 1: Forskriften (i lovs form) – krever en relativt omfattende prosess for endring

Overskrift:

*§ 5-17 Godkjente filformater for arkivdokumenter ved avlevering eller deponering
Arkivdokumenter skal overføres på holdbare formater etter kriterier fastsatt i retningslinjer av Riksarkivaren. For avlevering eller deponering av andre typer formater må det gjøres særskilt avtale med Riksarkivaren jf. § 5-3.*

Kriterier slik de foreligger nå:

- a) *Følgende kriterier skal minimum være oppfylt for at et format er godkjent ved avlevering eller deponering:*
- b) *Formatet skal følge en internasjonal standard (ISO, ANSI eller lign.)*
- c) *Formatet skal være åpent spesifisert*
- d) *Formatet skal være i utstrakt bruk*
- e) *Formatet skal ikke være lisensbelagt*
- f) *Formatet skal ha en lang forventet levetid*
- g) *Formatet skal ikke inneholde eksterne avhengigheter*
- h) *Formatet skal være uavhengig av operativsystem og presentasjonsplattform*

Det vil bli enklere å revidere detaljer i anbefalte filformater.

Del 2: Dirty details: Reglement bestemt av Arkivverket i samarbeid med faginstanser

§ 5-17. Godkjente filformater for arkivdokumenter ved avlevering eller deponering

(1) Følgende formater er godkjente ved avlevering eller deponering:

- a) Arkivdokumenter med ren tekst
 - TXT – ren tekstfil uten innkapslede objekter, ref. spesifikasjon i § 5-11.
 - TIFF – Tag Image File Format versjon 6, med de presiseringer som fremgår av § 5-18, første ledd.
 - PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.
 - XML – Extensible Markup Language versjon 1.0, med de presiseringer som fremgår av § 5-19.
 - For tekstdokumenter i OOXML eller ODF format bør en PDF/A-versjon leveres som tillegg.
- b) Arkivdokumenter inneholdende tekst med objekter:
 - TIFF – Tag Image File Format versjon 6, avbildninger, med de presiseringer som fremgår av § 5-18, første ledd.
 - PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.
- c) For digitale fotografier og bilder aksepteres følgende formater:
 - TIFF versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
 - JPEG (ISO/IEC 10918-1:1994).
 - JPEG 2000 (ISO/IEC 15444-1).
 - PNG (ISO/IEC 15948).
 - PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.
- d) For kart aksepteres følgende formater:
 - TIFF versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
 - SOSI versjon 2.2 (1995) eller nyere.
 - GML (ISO 19136).
- e) For videosekvenser aksepteres følgende formater:
 - MPEG-2 (ISO/IEC 13818-2).
 - MPEG-4/H.264 Part 10: Advanced Video Coding (ISO/IEC 14496-10).
- f) For lydsekvenser aksepteres følgende formater:
 - MP3 (ISO/IEC 13818-3).
 - PCM-basert Wave (ref. Multimedia Programming Interface and Data Specification 1.0 – IBM Corporation og Microsoft Corporation, 1991).
 - FLAC, som spesifisert på xiph.org.
- g) For regneark aksepteres følgende formater:
 - PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.
 - XML, med de presiseringer som fremgår av § 5-19.
 - For regneark i OOXML eller ODF format bør en PDF/A-versjon leveres som tillegg.
- h) For web-sider aksepteres følgende formater:
 - WARC (ISO 28500).
 - HTML (ISO/IEC 15445) basert på tegnsatt som spesifisert i § 5-11.
 - TIFF med de presiseringer som fremgår av § 5-18, første ledd.
 - PDF/A med de presiseringer som fremgår av § 5-18, andre ledd.
- i) For presentasjoner aksepteres følgende formater:
 - PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.
 - XML, med de presiseringer som fremgår av § 5-19.
 - For presentasjoner i OOXML eller ODF format bør en PDF/A-versjon leveres som tillegg.
- j) For objektbaserte informasjonsmodeller for byggverk (BIM) aksepteres følgende format:
 - IFC – Industry Foundation Classes (ISO 16739).

§ 5-18. Spesifiserte krav til arkivdokumenter i TIFF-format eller PDF/A-format

(1) TIFF:

- a) TIFF-dokumenter kan være «single page» eller «multiple page». Ved «single page» må hver side av dokumentet være lagret som en egen fil under en felles katalog.
- b) TIFF-dokumenter med egendefinerte innholdselementer («private tags») aksepteres ikke.
- c) Tapsfri komprimering iht. CCITT/TSS gruppe 4, LZW eller PackBit skal brukes for dokumenter i svart/hvitt, og LZW eller PackBit for dokumenter med farger/gråtoner.
- d) Tapsfri Deflate/LZ76-komprimering generelt og tapsbringende komprimering med JPEG (ISO 10918-1:1994) godtas i TIFF-dokumenter med fotografier eller kart dersom det anvendes følgende endrede komprimeringsverdier som ble fastsatt i TIFF TechNote2 (1995):
 - ved Deflate-komprimering skal komprimeringsverdi 8 brukes i stedet for den opprinnelige komprimeringsverdien 32946 (80b2 HEX) i TIFF versjon 6 (1992),
 - ved JPEG-komprimering (ISO 10918-1:1994) skal komprimeringsverdi 7 brukes i stedet for den opprinnelige komprimeringsverdien 6 i TIFF versjon 6 (1992).
- e) Presisering i forhold til georeferering/stedfesting av TIFF (i form av såkalt GeoTIFF eller annet) aksepteres dersom det er kompatibelt med TIFF versjon 6.

(2) PDF/A:

- a) Følgende varianter aksepteres:
 - PDF/A-1 (ISO 19005-1), versjon 1a, 1b eller 1u
 - PDF/A-2 (ISO 19005-2), versjon 2a, 2b eller 2u
- b) I spesielle tilfeller kan følgende godtas etter søknad:
 - PDF/A-3 (ISO 19005-3), versjon 3a, 3b eller 3u.

§ 5-19. Spesifiserte krav til arkivdokumenter i XML-format

(1) Tegnsatt i XML-dokumenter skal være UTF-8 eller ISO 8859-1/8859-4.

(2) Når XML-dokumenter har DTD (Document Type Definition), alternativt XML Schema, og malfiler, for eksempel for stiloppsett, skal disse ligge ved som egne filer. Tilhørende XML Schema skal også ligge ved i de tilfeller XML-dokumenter bygger på fastsatte standarder som ISO/IEC 26300:2006 – Open Document Format (ODF), Microsøfts og Ecmass Office Open XML-standard – OOXML (ECMA-376 ISO/IEC 29500).

(3) DTD-er, XML Schemas og malfiler som deles av flere dokumenter, kan lagres én gang på overordnet nivå i en hierarkisk filstruktur på vedkommende lagringsenhet, forutsatt at det brukes relative filreferanser. Alle øvrige filer eller objekter som det enkelte XML-dokument refererer til, skal være samlet i en underkatalog, jf. § 5-31 nr. 1 bokstav b. XML filer skal ikke inneholde binære objekter.

(4) Andre og tredje ledd gjelder ikke for arkivuttrekk i XML-format, jf. § 5-13.

Det **kan** bli justeringer i kriterier (vi er åpne for innspill – og det skal på høring)

- a) Følgende kriterier skal minimum være oppfylt for at et format er godkjent ved avlevering eller deponering:
- b) Formatet skal følge en internasjonal standard (ISO, ANSI eller lign.)
- c) Formatet skal være åpent spesifisert
- d) Formatet skal være i utstrakt bruk
- e) Formatet skal ikke være lisensbelagt
- f) Formatet skal ha en lang forventet levetid
- g) Formatet skal ikke inneholde eksterne avhengigheter
- h) Formatet skal være uavhengig av operativsystem og presentasjonsplattform

Vi vil også se litt på hva andre har gjort,
blant annet Rigsarkivet i Danmark

Udbredelse	<i>Formatet er hyppigt anvendt inden for sin indholdstype</i>
Udbredelse II	<i>Formatet er hyppigt anvendt til digital bevaring hos kulturarvsinstitutioner</i>
Levetid	<i>Formatet har eksisteret i mindst 10 år</i>
Levetid II	<i>Formatet har gode fremtidsudsigter</i>
Dokumentation	<i>Formatet er standardiseret</i>
Dokumentation II	<i>Formatets standard er ikke blevet opdateret inden for de seneste 10 år</i>
Dokumentation III	<i>Formatets dokumentation er velbeskrevet men overkommelig at læse</i>
Licensforhold	<i>Formatet er open source</i>
Struktur	<i>Formatet er selv bærende (ikke et containerformat)</i>
Struktur II	<i>Formatet kan læses som ren tekst</i>
Vigtige egenskaber	<i>Formatet understøtter størstedelen af vigtige egenskaber inden for sin indholdstype</i>
Tilgængeliggørelse	<i>Formatet kan genanvendes uden konvertering</i>
Søgbar	<i>Formatet har søgbar information</i>
Interoperabilitet	<i>Formatet er egnet til dataudveksling</i>
Testning	<i>Formatet har værktøjer til identificering og karakterisering</i>
Testning II	<i>Formatet har værktøjer til validering</i>
Kompression	<i>Formatet er ukomprimert eller har tabsfri kompression</i>
Lagring	<i>Formatet optager mindre lagerplads end gennemsnittet for sin indholdstype</i>
Migrering	<i>Formatet kan migreres med acceptabelt tab af vigtige egenskaber til et andet format</i>
Kompatibilitet	<i>Formatet er kompatibelt med flere operativsystemer og programmer</i>

PRONOM mapping vil bli en del av informasjon tilhørende Forskriften

§ 5-17. Godkjente filformater for arkivdokumenter ved avlevering eller deponering

(1) Følgende formater er godkjente ved avlevering eller deponering:

a) Arkivdokumenter med ren tekst:

- TXT⁴⁵ (*x-fmt/111, x-fmt/16*) – ren tekstfil uten innkapslede objekter, ref. spesifikasjon i § 5-11.
- TIFF⁶⁷ (*se § 5-18*); – Tag Image File Format versjon 6, avbildninger av tekst, med de presiseringer som fremgår av § 5-18, første ledd.
- PDF/A (*se § 5-19*) – med de presiseringer som fremgår av § 5-18, andre ledd.
- XML (*se § 5-19*) – Extensible Markup Language versjon 1.0, med de presiseringer som fremgår av § 5-19.
- For tekstdokumenter i OOXML eller ODF format bør en PDF/A-versjon leveres som tillegg.

b) Arkivdokumenter inneholdende tekst med objekter:

- TIFF (*se § 5-18*) – Tag Image File Format versjon 6, avbildninger, med de presiseringer som fremgår av § 5-18, første ledd.
- PDF/A (*se § 5-19*) – med de presiseringer som fremgår av § 5-18, andre ledd.

c) For digitale fotografier og bilder aksepteres følgende formater:

- TIFF (*se § 5-18*) versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
- JPEG⁸ (*fmt/42-44*) (ISO/IEC 10918-1:1994).
- JPEG 2000 (*x-fmt/382*) (ISO/IEC 15444-1).
- PNG⁹ (*fmt/11-13*) (ISO/IEC 15948).
- PDF/A (*se § 5-19*) – med de presiseringer som fremgår av § 5-18, andre ledd.

d) For kart aksepteres følgende formater:

- TIFF (*se § 5-18*) versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
- SOSI¹⁰ (*fmt/1246-1250¹¹*) versjon 2.2 (1995) eller nyere.
- GML¹² (*x-fmt/382*) (ISO 19136).

e) For videosekvenser aksepteres følgende formater:

- MPEG-2¹³ (*fmt/640, fmt/386*) (ISO/IEC 13818-2).
- MPEG-4/H.264Part 10¹⁴ (*fmt/199*): Advanced Video Coding (ISO/IEC 14496-10).

f) For lydsekvenser aksepteres følgende formater:

- MP3¹⁵ (*fmt/134*) (ISO/IEC 13818-3).
- PCM-basert Wave¹⁶ (*fmt/6, fmt/141, fmt/142*) (ref. Multimedia Programming Interface and Data Specification 1.0 – IBM Corporation og Microsoft Corporation, 1991).
- FLAC¹⁷ (*fmt/659*), som spesifisert på xiph.org.

g) For regneark aksepteres følgende formater:

- PDF/A (*se § 5-19*) – med de presiseringer som fremgår av § 5-18, andre ledd.
- XML (*se § 5-19*), med de presiseringer som fremgår av § 5-19.
- For regneark i OOXML eller ODF format (*se § 5-19*) bør en PDF/A-versjon leveres som tillegg.

h) For web-sider aksepteres følgende formater:

- WARC¹⁸ (*x-fmt/265*) (ISO 28500).
- HTML¹⁹ (*fmt/96-100, fmt/471*) (ISO/IEC 15445) basert på tegnssett som spesifisert i § 5-11.

Rimer ikke med § 5-20

Antatt tidsplan

1A: Forslag til Forskrift på høring tidlig i 2021

1B: Forskrift trer i kraft 1. juli 2021

2A: Forslag til detaljer i formater – relevante instanser kontaktes fom. januar 2021

2B: Forslag er mottatt og bearbeidet. Nye detaljer legges ut på nettsider 1/7-2021

Ønsker om tilbakemelding

1. Kommentarer til hovedprinsipper
2. Evt. nye forslag til endringer i formater