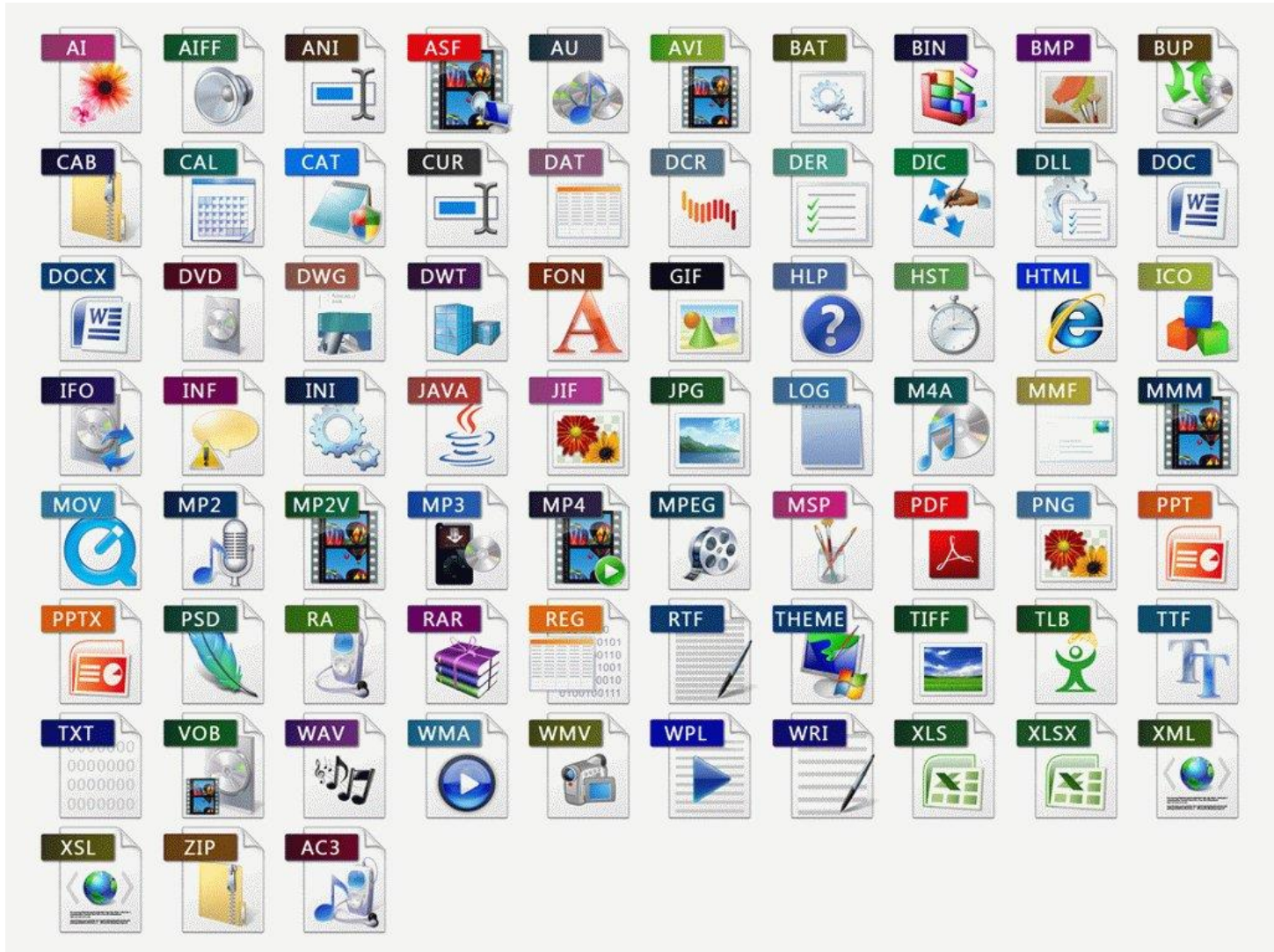


# Digitale dokumentformater

Forvaltningsforum 12.09.2019



## Filformater





## Filformater – Agenda

1. Kriterier for valg
2. Om forskriften
3. Om formater, kategorier
4. Om unntak
5. Om PRONOM
6. Om prosess videre



## Kriterier for valg av digitale formater – vi tenker langtids holdbarhet/forståelse

- Formatet skal følge en internasjonal standard (ISO, ANSI eller lign.)
- Formatet skal være åpent spesifisert
- Formatet skal være i utstrakt bruk
- Formatet skal ikke være lisens belagt
- Formatet skal ha en lang forventet levetid
- Formatet skal ikke inneholde eksterne avhengigheter. Det vil si at all informasjon som er nødvendig for fremtidig presentasjon skal være inneholdt i selve dokumentet.
- Formatet skal være uavhengig av operativsystem og presentasjonsplattform



## Om Forskriftens Kapittel 5, formatkrav

[https://lovdata.no/dokument/SF/forskrift/2017-12-19-2286#KAPITTEL\\_5](https://lovdata.no/dokument/SF/forskrift/2017-12-19-2286#KAPITTEL_5)

(3) Statlige organer skal deponere en arkivversjon hos Arkivverket straks perioden er avsluttet, jf. § 9-5 femte ledd.

### Kapittel 5. Elektronisk arkivmateriale. Lagringsmedier, format og avlevering

#### Del I. Generelle bestemmelser

##### § 5-1. Virkeområde

(1) Bestemmelsene i dette kapitlet gjelder arkivmateriale fra statlige organers elektroniske systemer som skal bevares for ettertiden henhold til arkivloven, arkivforskriften og forskriften her.

(2) For kommuner og fylkeskommuner gjelder bestemmelser i dette kapitlet i den utstrekning det er fastsatt i kapittel 3 og 8. For øvrig fastsetter kommuner og fylkeskommuner selv spesifiserte krav til arkivmateriale som skal avleveres eller deponeres til arkivdepot.

- I kraft fra 1. januar 2018
- Mange nye formater tillatt
- Litt tungvint prosess å revidere
- Kun bindende for statlige – ellers rådgivende
- Definerer formater - og tilleggskrav (tegnsett, ikke kryptert, ikke passord beskyttet - - )

##### § 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 innkapstede objekter, ref. spesifikasjon i § 5-11.
- TIFF – Tag Image File Format versjon 6, avbildninger av tekst, 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:

- H.264 (ISO/IEC 14496-10).

f)

- Application Programming Interface and Data Specification 1.0 – IBM Corporation og Microsoft

g) For digitale fotografier og bilder aksepteres følgende formater:

- TIFF versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
- PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.

- For tekstdokumenter i OOXML eller ODF format bør en PDF/A-versjon leveres som tillegg.

h) For digitale fotografier og bilder aksepteres følgende formater:

- TIFF versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
- PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.

i) For digitale fotografier og bilder aksepteres følgende formater:

- TIFF versjon 6 med de presiseringer som fremgår av § 5-18, første ledd.
- PDF/A – med de presiseringer som fremgår av § 5-18, andre ledd.

j) For objektoaserte informasjonsmodeller for byggverk (BIM) aksepteres følgende format:

- IFC – Industry Foundation Classes (ISO 16739).



## Kategorier av formater som er dekket av forskriften

- Papir/Scan/Text/Bilder – fortsatt krav til PDF/A (A-1/A-2)  
Nytt her: Office formater – 2007 eller senere – tillatt.  
Også nytt: Digitale signaturer kan inkluderes i PDF-filer
- Lyd
- Bilde
- Film
- GeoData
- CAD/CAM/3D data (ny kategori 2018)
- Database dump, XML (JSON?)
- HTML og Nettsider (nytt fra 2018)



## Unntak

1. Det er fortsatt mulig å søke om unntak fra Forskriften
2. Det blir en vurdering av de «10 kriterier» - veiet opp mot muligheten for arkivskaper til å etterleve kravene
3. En annen faktor er mulig informasjonstap ved konvertering  
I slike tilfelle kan man oppfordres til å levere både «produksjonsformat» og konvertert «arkivformat»



## Pronom – hva er det?

- Den definitivt mest brukte klassifikasjon av filformater i arkivverdenen. Vedlikeholdt av UK Archives.

The screenshot shows the website for The National Archives. At the top left is the logo 'The National Archives' with a large red 'A'. To the right is a search bar with the text 'Search our website' and a magnifying glass icon. Below the logo is a navigation menu with five items: 'About us', 'Education', 'Records', 'Information management', and 'Archives sector'. Below the menu is a breadcrumb trail: 'You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > PRONOM'. The main content area features an image of a CD-ROM and the text 'The technical registry PRONOM'. To the right of this is a vertical menu with icons and links: 'Welcome', 'Search', 'Add an entry', 'Help', and 'Information resources'. Below the main heading is a sub-heading 'Welcome to PRONOM' and a link to 'PRONOM changes and DROID signature file release notes.'. Further down is another link 'DROID signature files.'. The main text block contains two paragraphs: the first mentions the removal of the Linked Data PRONOM version due to a reorganisation, and the second describes PRONOM as a resource for technical information about file formats and software products.





## DROID - programvare

- Assosiert med PRONOM – DROID – som er:  
«Signaturfiler» for gjenkjenning, og programvare for gjenkjenning

The screenshot shows the website interface for The National Archives. At the top left is the logo 'The National Archives' with a large red 'A'. To the right is a search bar with the text 'Search our website' and a magnifying glass icon. Below the logo is a navigation menu with tabs: 'About us', 'Education', 'Records', 'Information management', and 'Archives sector'. The main content area shows a breadcrumb trail: 'You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > [PRONOM](#) > DROID signature files'. Below the breadcrumb is a section titled 'The technical registry PRONOM' with an image of a CD-ROM. To the right of this section are links: 'Welcome', 'About', 'Add an entry', 'Search', 'Help', and 'Information resources'. The main heading is 'DROID signature files (archive)'. The text below explains that the page contains an archive of all previously released signature files, generated by PRONOM and used by DROID for file format identification. It notes that signatures are presented in XML form and that there are gaps in the numbering of files (e.g., V27 is missing between V26 and V28). At the bottom, there is a list of links for 'DROID Signature File V1' through 'DROID Signature File V6'.



## 3 parallelle Pronom aktiviteter

# Arkivverket lager en mapping – vil bli lagt ut på Arkivverkets web

### § 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<sup>45</sup> (*x-fmt/111, x-fmt/16*) – ren tekstfil uter
- TIFF<sup>67</sup> (*se § 5-18*) – Tag Image File Format ve 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 OOX

**KDRS kartlegger**

b) Arkivdokumenter inneholdende

- TIFF (*se § 5-18*) – Tag Imag fremgår av § 5-18, første led
- PDF/A (*se § 5-19*) – med de

c) For digitale fotografier og bilder a

- TIFF (*se § 5-18*) versjon 6 n
- JPEG<sup>3</sup> (*fmt/42-44*) (ISO/IEC
- JPEG 2000 (*x-fmt/382*) (ISC
- PNG<sup>9</sup> (*fmt/42-44*) (ISO/IEC
- PDF/A (*se § 5-19*) – med de

1	A	B	C	D	E	F
1	PUID	EXT	FORMAT NAME	VERSION	MIME TYPE	
23	x-fmt/28	cal	CALS Compressed Bitmap			
24	x-fmt/29	cdr	CorelDraw Drawing	6.0		
25	x-fmt/30	cdt	CorelDraw Template			
26	x-fmt/31	cdx	CorelDraw Compressed Drawing			
27	x-fmt/32	ch3	Harvard Graphics Chart	3.0		
28	x-fmt/33	clk	Corel R.A.V.E.	1		
29	x-fmt/34	cmx	Corel Presentation Exchange File	5.0		
30	x-fmt/35	cmx	Corel Presentation Exchange File	6 and Later		
31	x-fmt/36	cpx	Corel CMX Compressed			
32	x-fmt/37	ctb	AutoCAD Colour-Dependant Plot Style Table			
33	x-fmt/38	cus	AutoCAD Custom Dictionary			
34	x-fmt/39	dbq	AutoCAD dbConnect Query Set			
35	x-fmt/40	dbt	AutoCAD dbConnect Template Set			
36	x-fmt/41	dif	Data Interchange Format			
37	x-fmt/42	doc	WordPerfect Secondary File	5.0		
38	x-fmt/43	doc	WordPerfect Secondary File	5.1/5.2		
39	x-fmt/44	doc	WordPerfect for MS-DOS/Windows Document	6.0		
40	x-fmt/44	w60	WordPerfect for MS-DOS/Windows Document	6.0		
41	x-fmt/44	wp	WordPerfect for MS-DOS/Windows Document	6.0		
42	x-fmt/44	wpp6	WordPerfect for MS-DOS/Windows Document	6.0		
43	x-fmt/44	wpsd	WordPerfect for MS-DOS/Windows Document	6.0		
44	x-fmt/45	dot	Microsoft Word Document Template	97-2003		
45	x-fmt/46	dqy	Microsoft Excel ODBC Query			
46	x-fmt/47	dwn	Micrografx Draw	2.0		
47	x-fmt/48	dvb	Visual Basic Macro			
48	x-fmt/49	dwr	AutoCAD Design Web Format	6.0		
49	x-fmt/50	dws	AutoCAD Drawing Standards File			
50	x-fmt/51	dwt	AutoCAD Drawing Template			
51	x-fmt/52	dcx	Drawing Interchange Format Standard Extract			
52	fmt/122	eps	Encapsulated PostScript file format	1.2		
53	fmt/122	epsf	Encapsulated PostScript File Format	1.2		
54	x-fmt/53	fhs	Macromedia Freehand	5		

## Noark 5 tjenestegrensesnitt

### Katalog over M701-verdier til bruk med Noark 5 Tjenestegrensesnitt

Kode	PRONOM	MIME-type	Navn	Merknad
RA-CSV	x-fmt/18	text/csv	Kommaseparerte verdier	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 innholdsbeskrivelsen. Feltskilletegn, feltnavn og eventuelle overskriftslinjer skal spesifiseres i den medfølgende struktur- og innholdsbeskrivelsen, jf. § 5-24. Merk at PRONOM-formatet ikke henviser til IETF RFC 4180.
RA-DOCX	fmt/412 fmt/494	application/vnd.openxmlformats-officedocument.wordprocessingml.document	ISO OOXML docx	ECMA-376 ISO/IEC 29500. Tilhørende XML Schema skal legges ved i arkivkapselen som egne filer. For tekstdokumenter i OOXML-format bør en PDF/A-versjon leveres som tillegg.
RA-FLAC	fmt/279	audio/flac	FLAC	FLAC, som spesifisert på xiph.org.
RA-GML	fmt/1047 x-fmt/227	application/gml+xml	GML	Geography Markup Language is defined in ISO 19136.
RA-HTML	fmt/96 fmt/97 fmt/98 fmt/99 fmt/100 fmt/103 fmt/471	text/html	HTML	(ISO/IEC 15445) basert på tegnsnitt som spesifisert i § 5-11.
RA-IFC	fmt/659 fmt/699 fmt/700 fmt/664		IFC	Industry Foundation Classes (ISO 16739).
RA-JP2	x-fmt/392	image/jp2	JPEG 2000	JPEG 2000 (ISO/IEC 15444-1).
RA-JPEG	fmt/41 fmt/42 fmt/43 fmt/44	image/jpeg	JPEG	JPEG (ISO 10918-1:1994).

- <sup>4</sup> <http://www.nationalarchives.gov.uk/PRONOM/Forma>
- <sup>5</sup> <http://www.nationalarchives.gov.uk/PRONOM/Forma>
- <sup>6</sup> <http://www.nationalarchives.gov.uk/PRONOM/Forma>
- <sup>7</sup> <http://www.nationalarchives.gov.uk/PRONOM/Forma>
- <sup>8</sup> <http://www.nationalarchives.gov.uk/PRONOM/Forma>
- <sup>9</sup> <http://www.nationalarchives.gov.uk/PRONOM/Forma>

Nr	M701	Navn	format	Forekomster	En
		Obligatorisk/valgfri	Obligatorisk		En
		Definisjon	Dokumentets format		
		Arkivenhet	dokumentobjekt		
		Kilde	Registreres automatisk når dokumentet arkiveres		
		Arv	Nei		
		Betingelser	Kan ikke endres		
		Kommentarer	Faste verdier bestemmes senere		
		Referansekrav			

Eksempel på nytt innhold: RA-FLAC -> fmt/659



## Oppfølging/tilbakemelding

1. Kommentarer til Forskriftens utforming
2. Ønsker om endringer/tillegg til Forskriften  
- spesielt: mangler vi noe
3. Hvor strenge bør man være – når unntak skal innvilges
4. Innspill til PRONOM prosess