



Folloarkivets dagskonferanse 2014

Filer som varer
Tony Lærdahl,
tony.lardahl@ciber.com

ciber[®]

Filer som varer

- Elektroniske dokumenter
- Filformater
- Standarder

- utfordringer på lang sikt

Skal ikke snakke om...

Lagringsmedier



ciber[®]

Hva er et dokument?

- Wikipedia:

Et dokument er en lesbar meddelelse som inneholder informasjon.

- Arkivloven sier:

”Dokument: medium som lagrar informasjon for seinare lesing, lyding, framsyning eller overføring.” (Denne definisjonen brukes også i offentlighetsloven.)

- Nå teknologinøytral:

Dokument: en logisk avgrenset informasjonsmengde som er lagret på et medium for senere lesing, lytting, framføring, overføring eller lignende.

Hva er et dokumentformat?

Hva er et filformat?

Hva er et filformat?

- Nesten det samme i denne sammenhengen

Hva er et filformat?

doc

Hva er et filformat?

doc

docx

Hva er et filformat?

doc

docx

html

Hva er et filformat?

doc

docx

html

xml

Hva er et filformat?

doc

docx

html

xml

PDF

Hva er et filformat?

doc

docx

html

xml

PDF

TIFF

Hva er et filformat?

doc

JPEG

docx

html

xml

PDF

TIFF

ciber[®]

Hva er et filformat?

doc

docx

html

xml

PDF

TIFF

JPEG

mp3

Hva er et filformat?

doc

docx

html

xml

PDF

TIFF

JPEG

mp3

MPEG-2

Hva er et filformat?

doc

docx

html

xml

PDF

TIFF

JPEG

mp3

MPEG-2

MPEG-4

Hva er et filformat egentlig?

- Et sett med syntaktiske og semantiske regler for mapping mellom abstrakt informasjonsinnhold og bit-strømmer
- Hvis vi er interessert i å bevare innhold, og ikke bare bit-strømmer, er det fundamentalt viktig å ha kontroll på formater

```
ffd8ffe000104a46  
4946000102010083  
00830000ffed0fb0  
50686f746f73686f  
7020332e30003842  
494d03e90a507269  
6e7420496e666f00  
0000007800000000  
0048004800000000  
02f40240ffeeffee  
0306025203470...
```

Hva er et filformat egentlig?

- Et sett med syntaktiske og semantiske regler for mapping mellom abstrakt informasjonsinnhold og bit-strømmer
- Hvis vi er interessert i å bevare innhold, og ikke bare bit-strømmer, er det fundamentalt viktig å ha kontroll på formater

ffd8ffe000104a46	SOI
4946000102010083	APP0 JFIF 1.2
00830000ffed0fb0	APP13 IPTC
50686f746f73686f	APP2 ICC
7020332e30003842	DQT
494d03e90a507269	SOF0 183x512
6e7420496e666f00	DRI
0000007800000000	DHT
0048004800000000	SOS
02f40240ffeeffee	ECS0
0306025203470...	...

Hva er et filformat egentlig?

- Et sett med syntaktiske og semantiske regler for mapping mellom abstrakt informasjonsinnhold og bit-strømmer
- Hvis vi er interessert i å bevare innhold, og ikke bare bit-strømmer, er det fundamentalt viktig å ha kontroll på formater

ffd8ffe000104a46	SOI
4946000102010083	APP0 JFIF 1.2
00830000ffed0fb0	APP13 IPTC
50686f746f73686f	APP2 ICC
7020332e30003842	DQT
494d03e90a507269	SOF0 183x512
6e7420496e666f00	DRI
0000007800000000	DHT
0048004800000000	SOS
02f40240ffeeffee	ECS0
0306025203470...	...



ciber[®]



 [Log in / create account](#)

Page [Discussion](#)

[Read](#)

[View source](#)

[View history](#)

File Formats

This is the top of the Ontology.

This is naturally an in-flux directory of categorizations and names - it should not be considered in stone.

The goal is to be inclusive of all things File Format throughout all of human existence and material.

Electronic File Formats

The heart of this project: the myriad electronic information containers of the past century.

Examples: [3D Models](#) · [Audio](#) · [Camera/Digital Image Sensor Files](#) · [Interactive Fiction](#) · [Video](#)

Physical File Formats

Containers for information in the physical realm.

Examples: [Floppy Disks](#) · [Photographic Film](#) · [Piano Rolls](#) · [Videotape](#)

Organic File Formats

File Formats found in nature or encoded in living beings.

Examples: [Tree Rings](#) · [DNA](#) · [RNA](#)

Languages

Spoken, gestural or written encoding intended for beings to communicate with other beings or machines.

Examples: [Spoken Languages](#) · [Written Languages](#) · [Programming Languages](#)

Navigation

[Main page](#)

[File formats](#)

[Formats by extension](#)

[Still more extensions](#)

[Software](#)

[Glossary](#)

[Library](#)

[Sources](#)

[Categories](#)

[Community portal](#)

[Recent changes](#)

[Random page](#)

Toolbox

[What links here](#)

[Related changes](#)

[Special pages](#)

[Printable version](#)

[Permanent link](#)

Overordnet mål:

- Dokumentene skal være brukbare for all framtid!

(dokumentene blir utrolig fort gamle)



Avleveringsbestemmelsene eller det mer formelle

Riksarkivarens forskrift, kapittel VIII:

Bestemmelser om elektronisk arkivmateriale som avleveres eller overføres som depositum til Arkivverket

<http://www.lovdata.no/cgi-wift/lcles?doc=/sf/sf/sf-19991201-1566.html#map058>

§ 8-17. Godkjente dokumentformater ved avlevering og deponering

Elektroniske dokumenter med tekst og bilde eller grafikkobjekter kan være lagret i følgende dokumentformater ved avlevering og deponering:

- a) Som ren tekst: UTF-8 (ISO/IEC 10646-1:2000 Annex D) eller ISO 8859-1:1998, Latin 1. ISO 8859-1:1998, Latin 1 kan erstattes med ISO 8859-4:1998, Latin 4 for samiske tegn.
- b) TIFF - Tag Image File Format versjon 6, med de presiseringer som fremgår av § 8-18.
- c) XML - Extensible Markup Language versjon 1.0, med de presiseringer som fremgår av § 8-19.
- d) PDF/A - ISO 19005-1:2005, versjon 1a eller 1b («Conformance Level» A eller B). PDF/A erstatter Adobe PDF, jf. § 8-20 tredje ledd.

For digitale fotografier, kart, video-, og lydsekvenser aksepteres følgende formater:

- a) fotografier og bilder: TIFF versjon 6 og JPEG (ISO 10918-1:1994).
- b) kart: TIFF versjon 6 og SOSI versjon 2.2 (1995) eller nyere.
- c) video: MPEG-2 (ISO 13818-2.)
- d) lyd: MP3 (ISO 11172-3), PCM eller PCM-basert Wave. Valget mellom disse lydformatene skal i hvert tilfelle være avtalt med Arkivverket før deponering eller avlevering.

§ 8-17. *Godkjente dokumentformater ved avlevering og deponering*

Elektroniske dokumenter med tekst og bilde eller grafikkobjekter kan være lagret i følgende dokumentformater ved avlevering og deponering:

- a) Som ren tekst: UTF-8 (ISO/IEC 10646-1:2000 Annex D) eller ISO 8859-1:1998, Latin 1. ISO 8859-1:1998, Latin 1 kan erstatte med ISO 8859-4:1998, Latin 4 for samiske tegn.
- b) TIFF - Tag Image File Format versjon 6, med de presiseringer som fremgår av § 8-18.
- c) XML - Extensible Markup Language versjon 1.0, med de presiseringer som fremgår av § 8-19.
- d) PDF/A - ISO 19005-1:2005, versjon 1a eller 1b («Conformance Level» A eller B). PDF/A erstatter Adobe PDF, jf. § 8-20 tredje ledd.

For digitale fotografier, kart, video-, og lydsekvenser aksepteres følgende formater:

- a) fotografier og bilder: TIFF versjon 6 og JPEG (ISO 10918-1:1994).
- b) kart: TIFF versjon 6 og SOSI versjon 2.2 (1995) eller nyere.
- c) video: MPEG-2 (ISO 13818-2.)
- d) lyd: MP3 (ISO 11172-3), PCM eller PCM-basert Wave. Valget mellom disse lydformatene skal i hvert tilfelle være avtalt med Arkivverket før deponering eller avlevering.

Fremstillingen av arkivversjoner med komprimerte JPEG-filer, MP3 lydfiler og MPEG-2 videofiler skal skje slik at det ikke medfører ytterligere komprimering og svekket bilde- eller lyd kvalitet.

§ 8-18. *Spesifiserte krav til elektroniske dokumenter i TIFF-format*

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.

TIFF-dokumenter med egendefinerte innholdselementer («private tags») aksepteres ikke.

Tapsfri komprimering iht. CCITT/TSS gruppe 4, LZW eller PackBit skal brukes for dokumenter i svart/hvitt, og LZW eller PackBi for dokumenter med farger/gråtoner.

§ 8-17. Godkjente dokumentformater ved avlevering og deponering

- Bestemmelsene inneholder konkrete angivelser av hvilke format som er gyldige for gitte situasjoner

§ 8-17. Godkjente dokumentformater ved avlevering og deponering

- Dokumenter med tekst og bilde
- Fotografier og bilder
- Lyd
- Video
- Kart

§ 8-17. Godkjente dokumentformater ved avlevering og deponering

Typer ikke dekket av bestemmelsene:

- Regneark
- Websider
- 3D
- Mappestrukturer
- E-post

- og mange fler

Dokumenter med tekst og bilde

Godkjente format:

- UTF-8 (kun ren tekst)
- ISO 8859-1 evt ISO 8859-4 (kun ren tekst)
- TIFF, versjon 6
- XML
 - Må oppdateres!
- PDF/A
 - Må oppdateres!

Bilder

Godkjente format:

- TIFF, versjon 6
- JPEG

Lyd

Godkjente format:

- MP3
- PCM
- PCM-basert Wave

Video

Godkjente format:

- MPEG-2
 - Må oppdateres!

Kart

Godkjente format:

- SOSI versjon 2.2 eller nyere
- TIFF, versjon 6

Det viktigste formatet er nok... PDF/A

PDF/A er en ISO-standard for langtidsarkivering av digitale dokumenter.

PDF/A *in a Nutshell*
PDF for long-term archiving

Hvilken PDF/A?

PDF/A-1 ?

PDF/A-2 ?

PDF/A-3 ?

Kom i dialog med Riksarkivet!

- Riksarkivet skal gi råd
– men det kan vi også!

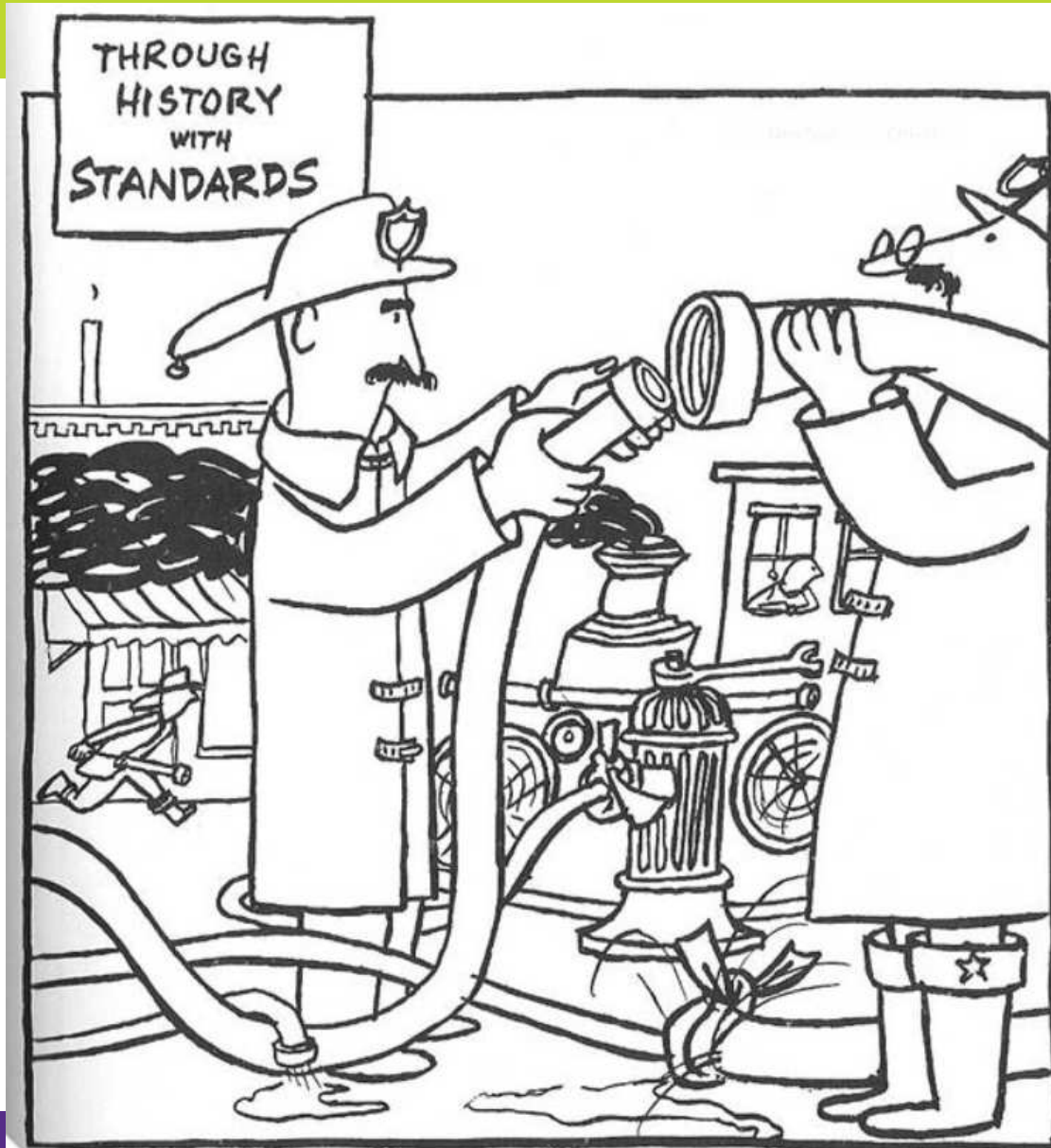


Riksarkivet kontrollerer filene

Testing av dokumentformater

- Er dokumentet virkelig hva det gir seg ut for å være?
 - Validering
 - Ikke stol for mye på fil-ender. De kan lyve og samme fil-ender kan bli brukt på mange typer filer
- Kontrollér filene **før** de sendes Riksarkivet!

Hva er en standard?



Hva er en standard?



ciber[®]

Hva er en god standard?

En fri og åpen standard

- Den er immun for leverandørinnlåsing.
- Den er adoptert og vil bli vedlikeholdt av en ikke-kommersiell organisasjon - ISO
- Den er publisert og spesifikasjonsdokumentet er fritt tilgjengelige
- Det er tillatt for alle å kopiere, distribuere og bruke den uten begresninger

Åpenhet

- Åpne data
- Åpen standard
- Åpen programvare



Difis referanse katalog for IT-standarder i offentlig sektor

Formål

- Referanse katalogen for IT-standarder i offentlig sektor skal legge grunnlaget for en velfungerende IT-samhandling i offentlig sektor.
- Internt i en offentlig virksomhet, mellom offentlige virksomheter eller med innbyggere og næringsliv

Standardiseringsportalen

IKT-standarder i offentlig sektor



Om standardiseringsar

Standardar

Høringer

Forslag og utredninger

Vanlige spørsmål

Møter i rådet

Slik jobbar råd


Forsiden > Standardar

Standardar

23.10.2013

Her finner du oversikt over alle anvendelsesområder og alle standarder som er vedtatt obligatorisk eller anbefalt for offentlig sektor. Den enkleste tilgangen er gjennom de ulike anvendelsesområdene, alternativt kan du laste ned hele katalogen gjennom en av dokumentene under.

[Forskrift om IT-standarder i offentlig forvaltning](#)

 [Referansekatalogen versjon 4.0 \(PDF\)](#)

 [Referansekatalog for IT-standarder i offentlig sektor \(HTML\)](#)

Anvendelsesområder

» [Arkivstandarder](#)

--

Standarder

» [AAC - Advanced Audio Coding](#)

»

2.2.2 Bruk av standarder på området

Det er **obligatorisk** for offentlige virksomheter å **sende** dokumenter beregnet for lesing på PDF 1.4 – 1.6, PDF 1.7 (ISO 32000-1:2008) eller PDF/A (ISO 19005-1:2005) i kommunikasjon med innbyggere og næringsliv. Det er anbefalt å benytte samme format ved intern utveksling mellom offentlige virksomheter.

Det er tillatt å parallellsende i flere format. Dette anbefales, siden enkelte løsninger for svaksynte har dårlig støtte for PDF-formatet.

Ved produksjon av PDF dokumenter anbefales det at fonter innkapsles (embed) i PDF dokumentene. Dette for i større grad å sikre at dokumentet blir gjengitt korrekt, uavhengig av plattform som benyttes, ved visning av dokumentet. Dokumenter som følger PDF/A (ISO 19005-1:2005) har alltid fontene innkapslet i dokumentet.

Det er **obligatorisk** å kunne **motta** ferdigstilte dokumenter og dokumenter for videre bearbeidelse i e-post fra innbyggere/ næringsliv/ andre offentlige virksomheter på vanlige dokumentformater, som for eksempel Open document format (ODF), Office Open XML (OOXML), PDF (alle versjoner), Portable Network Graphics (PNG ISO/IEC 15948:2003) og Joint Photographic Experts Group (JPEG IDO/IEC 10918-1:1994).

Det er anbefalt at dokumentene som sendes gjøres tilgjengelig. Her vises det til veileder for utforming av tilgjengelige dokumenter.

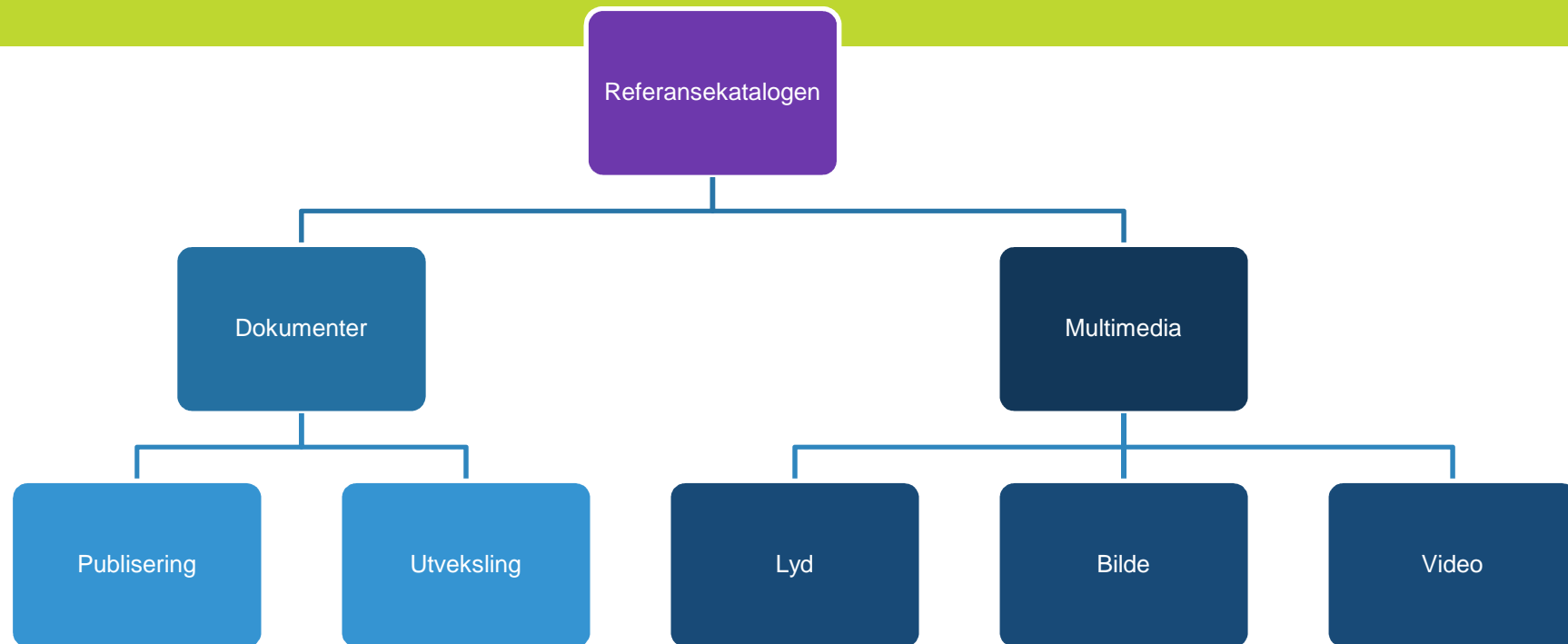
2.2.3 Frist for å tilpasse seg spesifiserte standarder for bruksområdet

Kravene er trådt i kraft.

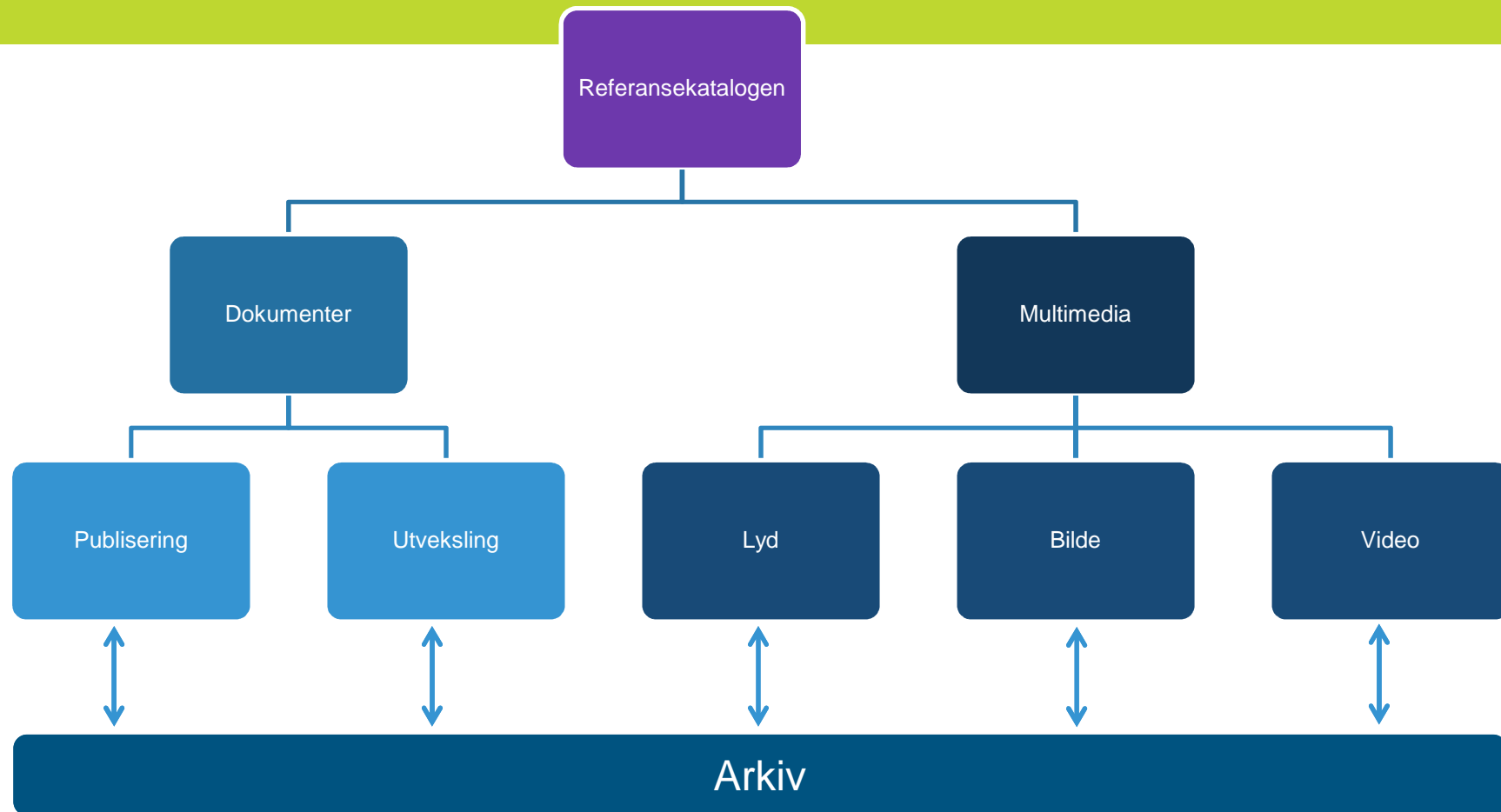
Difis referanse katalog

- Forskjellige standarder for forskjellige filformater
- Dokumentstandarder
- Multimediastandarder

Difis referanse katalog



Difis referanse katalog



Difis referanse katalog

Bruk difis referanse katalog!

For eksempel

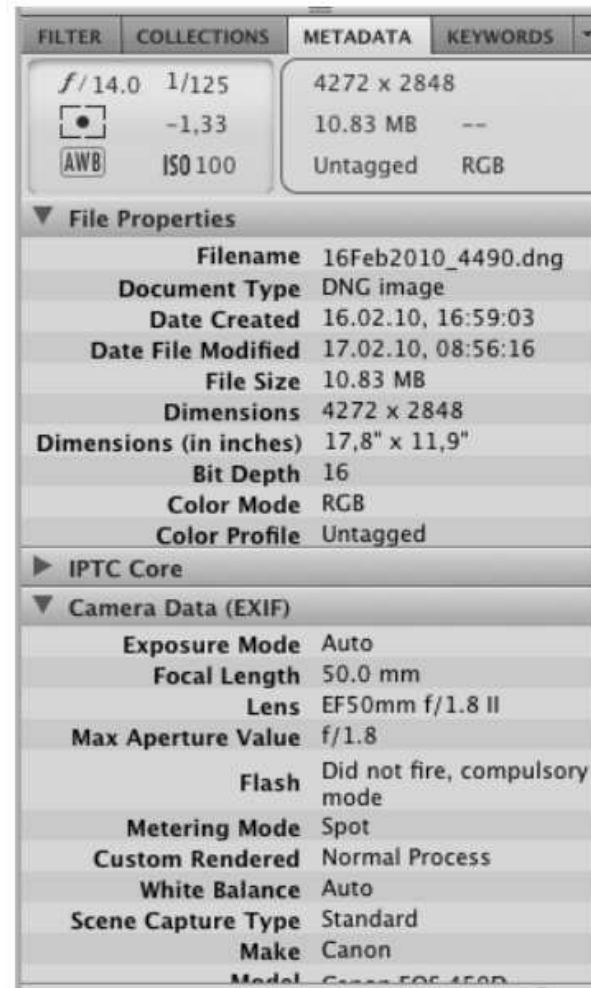
Ved offentlige anskaffelser

- Statens standardavtale
 - Sett krav til leverandørene



Metadata

- ...er lurt. 😊
- Eks. fra foto.
- Når ble bildet tatt?
- Hvor ble bildet tatt?
- Alt lagres i bildefilen!
- Automatiske metadata
- Påførte metadata



ciber®

Hva bør du huske?

- Problemene fikser seg ikke selv
- Problemene går ikke over av selv
- Gjør noe med det i dag, før det er for sent

Get started now not later

Hva bør du huske?

- Standarder i difis referanse katalog
- Avleveringsbestemmelsene til Riksarkivet
- Gå i dialog med Riksarkivet
- Vi kan hjelpe dere