



ARKIVVERKET  
RIKSARKIVET

# Samdok

samla samfunnsdokumentasjon

RAPPORT 2017



Delprosjekt Arkiv i eForvaltning

**PRIORITERT OPPGAVE**

**EFORV 2017\_2**

**Digital langtidsbevaring - Strategi for valg av filformater for bevaring**

Utarbeidet av  
**Anders Alvsåker,**  
**Ingrid Vinje Brustad,**  
**Lars Gaustad,**  
**Jan Tore Jørgensen,**  
**Birgitte Sauge,**  
**Petter Svendsen (leder),**  
**Erik Aaberg**

Rapportdato  
**25.10.2017**



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>1 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

<b>OPPGAVE</b>	<i>EFORV 2017_2</i> <i>Digital langtidsbevaring - Strategi for valg av filformater for bevaring</i>
<b>Ansvarlig delprosjekt</b>	<i>Arkiv i eForvaltning</i>
<b>Arbeidsgruppens leder</b>	Petter Svendsen, Oslo byarkiv
<b>Arbeidsgruppens medlemmer</b>	<ul style="list-style-type: none"><li>- Anders Alvsåker, PixEdit AS</li><li>- Ingrid Vinje Brustad, Statsbygg</li><li>- Lars Gaustad, Nasjonalbiblioteket</li><li>- Jan Tore Jørgensen, Arkivverket</li><li>- Birgitte Sauge, Nasjonalmuseet</li><li>- Petter Svendsen, Oslo byarkiv</li><li>- Erik Aaberg, Arkivverket</li></ul>
<b>Målformulering</b>	Målet for arbeidet er å utrede utfordringene knyttet til dynamikken i utviklingen og forvaltningen av arkivformater, som skissert over, samt å komme med forslag til hvordan dette arbeidet kan gjøres mer tidsmessig.
<b>Relaterte aktiviteter/ prosjekter</b>	<ul style="list-style-type: none"><li>• AV: Forskriftsrevisjon</li><li>• AV: Forvaltning av formater</li><li>• SKATE</li></ul>
<b>Sammendrag</b>	<p><i>Formålet med denne SAMDOK-oppgaven har vært å foreslå tiltak til hvordan valg og bruk av filformater for langtidsbevaring (filformatsforvaltning) kan organiseres. Bakgrunnen har vært en antagelse om at en optimalisert filformatsforvaltning kan bidra til å minske risikoen for informasjonstap og unngå unødig ressursbruk.</i></p> <p><i>Opgaven har blitt utført av en arbeidsgruppe bestående av deltakere fra arkiv-, bibliotek- og museumssektoren, en statlig arkivskaper og en programvareleverandør innen digitalisering og filkonvertering.</i></p> <p><i>I tillegg til å redegjøre for dagens praksis i Norge, har arbeidsgruppen gjennomført en kartlegging av et utvalg av utenlandske depotinstitusjoner. Hensikten med kartleggingen har vært å få et innblikk i sammenlignbare institusjoners strategi og praksis innen filformatsforvaltning. Kartleggingen har også lagt til rette for fremtidig samarbeid med et antall depotinstitusjoner utenfor Norge.</i></p> <p><i>Hovedleveransen fra arbeidsgruppen er fem anbefalinger til tiltak.</i></p>
<b>Oppfølging / anbefalinger om videre arbeid</b>	<p><i>Se <b>Anbefalinger</b> for begrunnelse av anbefalingene</i></p> <ol style="list-style-type: none"><li><i>1. Arkivverket tar ansvaret for forvaltning av filformater for bevaring for en samlet offentlig forvaltning. Det opprettes et råd for filformater i regi av Arkivverket med deltakere på tvers av sektorer. Rådet bistår forvaltningen i spørsmål knyttet til filformater for bevaring og tilgang til digital informasjon over tid.</i></li><li><i>2. Arkivverket prioriterer en høy grad av forpliktende samarbeid med fag-, forsknings- og standardiseringsmiljøer i Norge spesielt, men også i utlandet.</i></li><li><i>3. Arkivverket jobber for økt samarbeid med relevante depotinstitusjoner i inn- og utland.</i></li></ol>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>2 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

	<p>4. <i>Det legges til rette for å lagre samme informasjon i flere filformater for å minske risikoen for informasjonstap og øke muligheten for senere tilgang.</i></p> <p>5. <i>Fokus på fagsystemer økes for å ha bedre kontroll på hvilke typer informasjon (filformater) som forvaltes.</i></p> <p><i>Videre oppfølging</i></p> <ol style="list-style-type: none"><li><i>Kartleggingen</i></li><li><i>Kontakter</i></li></ol>
<b>Vedlegg</b>	<ol style="list-style-type: none"><li><i>Kartlegging – Svar fordelt på spørsmål</i></li></ol>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>3 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Innledning

Det er grunn til å anta at deler av den digitale dokumentasjonen som skapes og forvaltes av offentlig forvaltning og som skulle ha vært bevart, går tapt. Omfanget av hvor mye som går tapt, er ukjent, men basert på kunnskap om de systemene som er i bruk i forvaltningen, regelverket som styrer arkivdanning og -bevaring, og rådende praksis, er det sannsynlig at dokumentasjon tapes daglig. Grunnen til dette, er blant annet liten kontroll med hva som dannes av dokumentasjon, sett i relasjon til kravene om hva som skal bevares og hvordan.

Dokumentasjonstapet spenner mellom arkivdokumenter som går tapt i sin helhet, til arkivdokumenter som blir gjenstand for informasjonstap ved at noe av informasjonen i det enkelte dokumentet blir borte et sted mellom danningen og fremtidig tilgang.

Dette dokumentasjonstapet kan få flere konsekvenser. Dokumentasjon som er avgjørende for saker på alt fra nasjonalt plan til den enkelte borger, vil ikke kunne fremskaffes. Manglende dokumentasjon av installasjoner slik som bygninger og viktig infrastruktur, kan få store økonomiske og sikkerhetsmessige følger, for å nevne noen eksempler.

Tap av dokumentasjon, enten det er snakk om arkivdokumenter i sin helhet eller deler av informasjonen i dem, henger tett sammen med hvilket filformat dokumentasjonen skapes i og i hvilket format den langtidsbevares i. Kjernen i dette er at på den ene siden skapes dokumentasjon i et stort antall filformater (produksjonsformater), mens det er få formater (arkivformater) som egner seg til langtidsbevaring. I tillegg er det i lov- og regelverk ikke spesifisert arkivformater for alle typer dokumentasjon.

I denne rapporten er begrepene dokumentasjon og informasjon brukt om hverandre. Årsaken til det, er at det i noen tilfeller der det er snakk om arkivdokumenter eller dokumentasjonsbegrepet innen offentlig forvaltning, er mer riktig å benytte dokumentasjon. På den annen side er utfordringene knyttet til filformater og filformatsforvaltning, noe som i høy grad berører privat sektor.<sup>1</sup> I noen tilfeller blir det derfor naturlig å benytte det mer generelle begrepet informasjon.

Rapporten er utarbeidet av en arbeidsgruppe med oppgave å foreslå tiltak som skal bidra til å sikre at digital dokumentasjon i enkeltfiler skal være tilgjengelig over tid. Arbeidsgruppen har vært sammensatt av deltakere fra arkiv-, bibliotek- og museumssektoren (ABM-sektoren), en statlig arkivskaper og en programvareleverandør innen digitalisering og filkonvertering. Vi har ikke hatt som oppgave å lage et sett med forslag til konkrete filformater og behandlingen av dem. For å belyse problematikken, har vi tatt tak i bygningsinformasjonsmodellering som et gjennomgående eksempel. I arbeidsgruppen har det vært deltakere fra Arkivverket, Statsbygg og Nasjonalmuseet. Arkivverket har ansvaret for arkivloven som gjennom sine forskrifter angir arkivformater primært for statlige arkivskapere, mens Statsbygg er den største offentlige byggherren og forvalter en stor og viktig bygningsmasse hvor digitale bygningsinformasjonsmodeller har en sentral rolle. Nasjonalmuseet har den største samlingen av privatarkiver fra arkitektbransjen. Arkitekter står for en vesentlig del av denne type arkivdanning. Siden det per i dag ikke er definert noe arkivformat for 3D-modeller i forskrift til arkivloven, har det på mange måter vært et praktisk eksempel for å belyse hvordan en kan ivareta langtidsbevaring av dem. Dette har vært gjennomført som et faglig samarbeid mellom Arkivverket og Statsbygg under arbeidet i arbeidsgruppen. Det å ha et slikt praktisk eksempel har vært svært nyttig for en omforent forståelse av problemområdet og for leveransen fra arbeidsgruppen. I tillegg har det resultert i et konkret arkivformat for bygningsinformasjonsmodeller.

I Norge er det to store offentlige aktører innen digital langtidsbevaring. Dette er Arkivverket og Nasjonalbiblioteket som bevarer og forvalter dokumentasjon fra forskjellige deler av det norske samfunnet, men

<sup>1</sup> Med filformatsforvaltning menes strategi og praksis innen valg og bruk av filformater for langtidsbevaring.



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>4 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

som på mange områder har sammenfallende oppgaver og utfordringer, spesielt når det dreier seg om bevaring av dokumentasjon i enkeltfiler.

En viktig aktivitet i arbeidsgruppens arbeid har vært å gjennomføre en kartlegging av hvordan depotinstitusjoner i andre land griper an utfordringene knyttet til sikring av dokumentasjon gjennom bevaring i enkeltfiler. Det har vært en gjengs antagelse og oppfatning at utfordringene ved misforhold mellom filformater som dokumentasjon skapes i og hvilke formater som er egnet for langtidsbevaring og tilgang over tid, er de samme sett i et globalt perspektiv. Det har derfor vært veldig nyttig å få dannet et bilde av om dette stemmer, og om det i så fall er områder hvor norske depotinstitusjoner og informasjonsforvaltere kan dra veksler på andre lands filformatsforvaltning, og bidra til felles forståelse og hensiktsmessig praksis. Kartleggingen har omfattet et utvalg på rundt førti utenlandske depotinstitusjoner innen arkiv og bibliotek som har fått tilsendt en forespørsel om den enkeltes strategi og praksis på området. Det kom inn i underkant av tretti svar og disse har bidratt til svært nyttig viten og innsikt i hvordan andre depotinstitusjoner forholder seg til digital bevaring. Gjennom svarene har ikke arbeidsgruppen klart å identifisere strategi eller praksis for bevaring av bygningsinformasjonsmodelleringsfiler (BIM).<sup>2</sup>

Det er en utbredt oppfatning at å bevare digital dokumentasjon og holde den tilgjengelig over tid, er en komplisert og ressurskrevende oppgave. Vi vil fremheve at det i tillegg er en *kontinuerlig* oppgave. Den teknologiske utviklingen medfører at nye filformater kommer til og andre forsvinner. Det innføres nye måter å samhandle på, for eksempel gjennom forskjellige former for dialog mellom borger og offentlig forvaltning. Det gjør det nødvendig å ha en dynamikk i filformatsforvaltningen som fanger opp disse endringene.

## Presisering og avgrensning av oppgaven

Rapporten omfatter kun problematisering av logisk bevaring, det vil si at den ikke tar opp lagringsteknologier eller lagringsmodeller, men konsentrerer seg om filformatsforvaltning. Rapporten tar ikke stilling til valg av enkeltformater, selv om det i rapporten nevnes forskjellige konkrete formater.

Rapporten omfatter problemstillinger uavhengig av om det er snakk om filformater for digitalisert dokumentasjon eller digitalt skapt dokumentasjon. Det sistnevnte skaper dog større utfordringer grunnet mer komplekse formater og et større antall formater en må forholde seg til. Rapporten bærer til en viss grad preg av kompleksiteten og mangfoldet knyttet til digitalt skapt dokumentasjon.

## Gjennomføring

### Deltakerne i arbeidsgruppen

Arbeidsgruppen har bestått av 7 medlemmer fra Arkivverket, Nasjonalbiblioteket, Nasjonalmuseet, PixEdit AS, Statsbygg og Oslo byarkiv.

Under følger en kort beskrivelse av de enkelte medlemmene i gruppen.

- Jan Tore Jørgensen har vært ansatt i Arkivverket (Statsarkivet i Kristiansand) siden 2001 og har totalt jobbet med IT i 45 år, primært innenfor systemutvikling og programmering. Han har tung og bred kompetanse innen kartlegging, innsamling og testing av statlige digitale arkiver.

<sup>2</sup> Building Information Modeling



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>5 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

- Erik Aaberg har vært ansatt i Arkivverket (Riksarkivet) siden 2013 og har jobbet innen IT siden tidlig på 1980-tallet. Han arbeider med prosjektledelse av programvareutvikling, digital metodikk, offentlige digitale standarder og bevaring av nettbasert informasjon. Og i tillegg anbefalte metoder for digital langtidsbevaring, hvori inngår arbeid med strategi for valg av filformater.
- Lars Gaustad har vært ansatt ved Nasjonalbiblioteket siden 1992, og har der jobbet med bevaringsproblematikk knyttet til lyd, audiovisuelt materiale og filer. Han leder den tekniske komitéen i International Association of Sound and Audiovisual Archives, og har publisert flere bøker om fagområdet.
- Petter Svendsen har informatikerbakgrunn og jobber med digital bevaring i Oslo byarkiv hvor han har vært ansatt som IKT-arkivar/spesialkonsulent siden 2015. Han har flere års erfaring innen området og har tidligere vært ansatt i Riksarkivet med oppgaver innenfor digitalt depot.
- Anders Alvsåker er daglig leder i PixEdit AS. Hans arbeidsoppgaver er relatert til utvikling av funksjonalitet til PixEdit AS sine produkter for digitalisering og konvertering, blant annet ledelse i forbindelse med et pågående prosjekt, men også R&D til nye moduler.
- Ingrid Vinje Brustad er utdannet Sivilbibliotekar fra Høgskolen i Oslo og har i tillegg en Cand. Mag. fra UiO. Hun er ansatt som seniorrådgiver med fagansvar for arkivtjenesten i Statsbygg, og jobber daglig med bl.a. systemforvaltning og arkivintegrasjoner. Hun har tidligere jobbet som seniorrådgiver for dokumentasjonsforvaltning i Bufdir og som systemkonsulent i De norske bokklubbene.
- Dr. art. Birgitte Sauge er seniorkurator ved Nasjonalmuseet for kunst, arkitektur og design. Etter avlagt doktorgradsavhandling i kunsthistorie med tittelen «Arkitekturtegninger og kontekst», (UiB 2003) har hun fortsatt studiet av historiske og teoretiske aspekter ved bevaring og formidling av arkitekturmateriale. I prosjektet «Architecture Museums and Digital Design Media», støttet av NFR (2015-2019), forsker hun på digital arkitektur.

## Tidsramme

Oppstart: mai 2017

Ferdigstillelse: oktober 2017

## Arbeidsform

Arbeidsgruppen har basert seg på e-post for kommunikasjon internt i gruppen og Google Drive for lagring av relevant dokumentasjon.

Det har blitt avholdt fire møter i arbeidsgruppen. Tre av møtene ble avholdt i Riksarkivet, mens ett ble avholdt hos Statsbygg i forbindelse med en innføring i bygningsinformasjonsmodellering (BIM).

- Oppstartsmøte 31.3.2017
- Arbeidsgruppemøte og BIM-gjennomgang 16.5.2017
- Arbeidsgruppemøte 29.8.2017
- Arbeidsgruppemøte 25.9.2017

Arbeidsgruppen har basert arbeidet på flere kilder. For gruppen som helhet, har egne erfaringer og utfordringer bidratt med nødvendig bakgrunnsinformasjon og til å holde fokus på det aktuelle problemområdet. I tillegg har gjeldende strategier, praksiser og lov- og regelverk vært viktige for å se styrker og svakheter, samt



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>6 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

forbedringsmuligheter. En sentral kilde til forståelsen av andre lands utfordringer og løsninger på filformatfeltet, er kartleggingen som arbeidsgruppen har gjennomført. Søk etter informasjon på nettet og tilgang til et antall artikler og nettsider, har også bidratt til helheten.



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>7 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Filformatstrategi og -praksis i Norge

### Innledning

Denne rapporten har fokus på filformater som egner seg for å lagre informasjon over tid. I arkivsektoren kalles denne informasjonen arkivdokumenter. Et arkivdokument kan her forstås som en mengde digital informasjon som har verdi som dokumentasjon. Når disse arkivdokumentene skal langtidsbevares, er de i mange tilfeller blitt konvertert til filer i formater som betegnes arkivformater. Tidshorizonten innen arkivsektoren for bevaringen er evigheten. Tidshorizonten kan variere sterkt for andre områder enn arkiv, men det er vanlig at det er behov for eller et ønske om at informasjonen skal være tilgjengelig i over ti år. Historisk sett har det innen mange fagområder ikke vært en prioritert oppgave å sørge for at informasjonen forblir tilgjengelig. Filer som er avhengig av spesiell programvare for å kunne vises og behandles, har blitt utilgjengelige når programvaren ikke lenger er aktuell eller mulig å bruke. En annen faktor til at filers innhold blir utilgjengelig, er når det ikke (lenger) er mulig å konvertere filene i det aktuelle formatet til andre formater som opprettholder tilgjengeligheten. Disse faktorene må vi anta har resultert i uønsket tap av informasjon.

For noen tiår siden var det ennå ikke aktuelt med arkivdokumenter som enkeltfiler, men med registerdata hvor eventuelle dokumenter forelå på papir. I tillegg var det ennå ikke vanlig med standardisering av for eksempel bildeformater og video. Etter hvert som teknologien har utviklet seg og blitt allemannseie, har mengden informasjon som skapes, økt enormt. Dette har sammenheng med en bred digitalisering av samfunnet, både innen offentlig forvaltning generelt, ABM-sektoren, næringsliv og for enkeltpersoner. Innen noen områder har digitaliseringen bidratt til økt bevissthet rundt behovet for standardisering og tilgang til informasjonen som skapes. Innen andre områder må en kunne si at det leves mer i nuet og at en ikke har nytte av eller ser nytten av å opprettholde tilgangen til tidligere skapt informasjon. I noen tilfeller vil denne informasjonen allikevel være samfunnsmessig viktig og må tas vare på. Felles for de som skaper og forvalter informasjon som er viktig dokumentasjon for ettertiden, enten som rettighetsdokumentasjon, dokumentasjon av offentlig forvaltning, kulturarv eller tilgjengeliggjøring for fremtidig forskning, er at det er nødvendig å legge til rette for bevaring gjennom valg av hensiktsmessige filformater.

Hvilke filformater som egner seg for å bevare informasjon over tid, er i stadig endring. Dette skjer i stor grad som følge av den teknologiske utviklingen. For noen typer informasjon og innen visse fagfelt, er det relativt opplagte valg av arkivformat. For eksempel kan tabellariske data og innholdet i enkle regneark uten formler, lagres som «kommaseparerte» filer, i praksis i CSV-formatet.<sup>3</sup> Innen arkivsektoren er det mye korrespondanse som skal bevares. I praksis forekommer denne korrespondansen som filer i forskjellige varianter av Word-formatet og som PDF-filer. Slike filer blir i de fleste tilfellene konvertert til PDF/A for langtidsbevaring. For mer komplekse data som 3D-tegninger, er det ikke sikkert det finnes noe egnet filformat for langtidslagring uten informasjonstap.

### Dagens situasjon

Dette kapitlet redegjør for hvordan Arkivverket og Nasjonalbiblioteket forholder seg til valg av arkivformater. I tillegg gis eksempler på hvordan arkivdanningen må forholde seg til et begrenset antall arkivformater. Til slutt kommer det en oversikt over hvilket mangfold av filformater et fagområde forholder seg til, ved at de forvalter dokumentasjon over en relativ lang tidsperiode.

<sup>3</sup> CSV – Comma-separated values





SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>8 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Arkivverket

Arkivverket (Riksarkivet) har siden 1970-tallet mottatt registerdata, dvs. uttrekk fra forskjellige registre hos statlige organer som Statistisk sentralbyrå. Disse uttrekkene har blitt overført til Riksarkivet i form av såkalte flate filer med medfølgende beskrivelse. De eldste registeruttrekkene gjenspeilte struktur og tegnsett i databasene som uttrekkene ble tatt fra. Uttrekkene kom fra hierarkiske databaser, nettverksdatabaser, EBCDIC- og Notis-tegnsett og pakkede numeriske felt.<sup>4</sup>

Det var etter alt å dømme et sted mellom Noark-2 (1984) og Noark-3 (1994) at muligheten for å ha tilgang til en digital versjon av arkivdokumenter, ble spesifisert.<sup>5</sup> Det var først da Noark-4 kom i 1999, at digitale arkivdokumenter med angitte filformater, ble en realitet.<sup>6</sup> Før Noark-4 hadde digitale avleveringer fra Noark-systemer bestått av journaldata. I 1999 kom også Riksarkivarens forskrift som anga krav til digitale uttrekk generelt og til Noark-3 og -4-uttrekk spesielt, i tillegg til å angi filformater for bevaring uavhengig av om det var snakk om fra fag- eller Noark-system.<sup>7</sup>

Etter at det ble vanlig å bevare digitale arkivdokumenter, har det vist seg at det ofte er et misforhold mellom antall forskjellige filformater den enkelte arkivskaper forvalter og antall arkivformater i Riksarkivarens forskrift. I tillegg er det ikke alle formater hos arkivskaper som lar seg konvertere til et arkivformat. Forskriften har gjennomgått én revisjon hva gjelder hvilke filformater som kan benyttes ved bevaring (Kapittel VIII). Denne revisjonen skjedde i 2007 og medførte blant annet at PDF/A-1 ble et arkivformat. I Riksarkivarens forskrift har det hele tiden vært åpnet for at arkivskapere som forvalter informasjon i formater som ikke er definert som arkivformat eller som det ikke finnes et opplagt arkivformat for, kan gå i dialog med Arkivverket. Det er naturlig at arkivskaper gjør dette når avtalen om avlevering eller deponering utformes.<sup>8</sup> Denne muligheten har arkivskapere i varierende grad benyttet seg av. I revisjonen av forskriften vil det etter alt å dømme komme en egen bestemmelse om muligheten for å avtale andre formater enn arkivformatene ved behov.

I februar 2017 gikk det ut et forslag til ny tekst i forskriften, med høringsfrist i juni. Ny forskrift forventes å tre i kraft ved årsskiftet 2017/2018. Det kom inn nærmere 100 svar som alle er gjennomgått, og flere av forslagene som kom inn, er tatt til følge i revidert forslag til ny forskrift. Når det gjelder kapittel VIII og arkivformater, kan det sies at det åpnes opp for **nye** formater blant eksisterende dokumentkategorier (eksempelvis "PNG" for bilder og GML<sup>9</sup> for kart), så vel som nye **kategorier** av dokumenter ("WARC"<sup>10</sup> for nettsider og IFC<sup>11</sup> for objektbaserte informasjonsmodeller). Dette for å være mer i tråd med gjeldende standarder og samtidig harmonere med følgende kriterier for valg av arkivformat som benyttes av Arkivverket:<sup>12</sup>

- Formatet skal følge en internasjonal standard (ISO, ANSI eller lign.)
- Formatet skal være åpent spesifisert
- Formatet skal være i utstrakt bruk

<sup>4</sup> Dagens regelverk i form av Riksarkivarens forskrift bærer ennå preg av disse tidlige uttrekkene og en finner dette også igjen i uttrekkskravene i Noark-3 hvor en har inngående- og utgående dokumenter (pre journalpostbegrepet) i samme datafil på tross av at de har forskjellig feltstruktur.

<sup>5</sup> Noark 91/1 – Justeringer og tillegg til Noark-2

<sup>6</sup> Noark-4, kap. 5.3 Dokumentformater

<sup>7</sup> Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver. (FOR-1999-12-01-1566)

<sup>8</sup> Riksarkivarens forskrift § 8-7. Gjennomføring

<sup>9</sup> <http://www.opengeospatial.org/standards/gml>

<sup>10</sup> <https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml>

<sup>11</sup> Industry Foundation Classes (IFC) er et format for utveksling av bygningsinformasjonsmodeller (BIM).

[https://en.wikipedia.org/wiki/Industry\\_Foundation\\_Classes](https://en.wikipedia.org/wiki/Industry_Foundation_Classes)

<sup>12</sup> Se for øvrig the National Archives (TNA) i Storbritannia sitt dokument om filformatvalg for langtidsbevaring:

<https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>9 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

- Formatet skal ikke være lisensbelagt, i det minste ikke for "arkivbruk", inklusive både bevaring og senere formidling
- 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

PDF/A har i arkivsektoren fått størst oppmerksomhet og utbredelse blant arkivformatene. Derfor kan det være interessant å se på hvordan kravene til PDF-filer har endret seg etter hvert som formatet har endret seg. Arkivverket tillot inntil 1. januar 2009 PDF med visse begrensninger, som arkivformat. Etter denne datoen tok PDF/A-1a/b fra 2005 over som PDF-arkivformat.<sup>13</sup> En av erfaringene med denne versjonen av PDF/A, er at konvertering til PDF/A-1b under visse forutsetninger medfører en 10-20 ganger økning i filstørrelsen. Dette er noe en unngår i PDF/A-2b fra 2011, et format som vil bli vedtatt som et arkivformat i den reviderte versjonen av Riksarkivarens forskrift.<sup>14</sup> Siste versjon av PDF/A er PDF/A-3 som ble vedtatt i 2012.<sup>15</sup> Denne versjonen tillater at filer i vilkårlige formater innkapsles i PDF/A-3-filen. Dette gjør at det ikke uten videre er egnet som arkivformat, men at det må settes krav til hva som kan innkapsles for å sikre tilgangen til informasjonen. Generelt medfører det visse utfordringer med å kontrollere om en PDF-fil som utgir seg for å oppfylle kravene til PDF/A, faktisk gjør det.<sup>16</sup>

Riksarkivarens forskrift er bindende for statlige institusjoner, med mindre det søkes og innvilges et unntak. For kommunale er den kun veiledende, og for private vil den også bare være rådgivende. Historisk har man hatt mest fokus på sakarkiver med tilhørende journalposter og dokumenter. Volumet av "fagsystemer" som ikke passer inn i dette er imidlertid større, og øker ytterligere behovet for en standardisering og kartlegging når det gjelder bruk av filformater i disse systemene.

### Nasjonalbiblioteket

Nasjonalbibliotekets (NB) samlinger inneholder filer som daterer seg tilbake til tidlig 1990-tall. Det digitale materialet i Nasjonalbiblioteket omfatter blant annet aviser, bøker, fotografier, radioopptak og håndskrifter.

NB har hele tiden hatt som utgangspunkt å prioritere bruken av åpne veldokumenterte format med stor utbredelse, og der formatene enten er uten komprimering eller der komprimeringen er tapsfri. Så langt det er mulig, velges det formater som det ikke er knyttet patentrettigheter til. Krypterte filer aksepteres ikke.

Filer som produseres i et komprimert format tas vare på i det formatet, dersom det fortsatt er et format i utstrakt bruk, og konverteres ikke før formatet anses å være ved sitt end-of-life.

Filformatet valideres i ingestprosessen ved mottak, og det genereres en sjekksum for å sikre bitbevaringen av dokumentet.

Når fysiske objekter digitaliseres, eller når det digitale grunnlaget for fysiske publikasjoner avleveres, velges et kvalitetsnivå som gjør det mulig å gjenskape alle relevante aspekter ved det originale objektet.

For digitalt fødte objekter beholdes opprinnelig kvalitet så langt det er mulig.

<sup>13</sup> ISO 19005-1:2005

<sup>14</sup> ISO 19005-2:2011

<sup>15</sup> ISO 19005-3:2012

<sup>16</sup> <https://www.pdfa.org>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>10 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

Når et format nærmer seg end-of-life, eller av andre årsaker ikke bør brukes til digital bevaring, konverteres data til et nytt format som oppfyller kravene til digital bevaring. Så langt det er mulig skal en formatkonvertering være tapsfri, og det skal kunne verifiseres gjennom at det er mulig å konvertere data identisk tilbake til utgangspunktet.

Normalt vil også den originale versjonen av et digitalt objekt beholdes i bevaringsløsningen.

For objekter som ikke kan brukes/vises uten en spesifikk programvare knyttet til en eller flere spesifikke plattformer, baseres fremtidig bruk på emulering av historisk programvare og historiske plattformer. Dette er også en aktuell løsning for formater som ikke lar seg konvertere uten tap av informasjon.

Typiske formater som benyttes av NB i bevaringssammenheng er TIFF og tapsfri JPEG2000 for foto og grafisk materiale, lineær PCM pakket i Wave for lyd, tapsfri JPEG2000 pakket i MXF for video og 16 bit DPX for film.

### Utfordringer ved arkivdanningen

For arkivkaperene er det flere utfordringer knyttet til filformater, i forbindelse med selve arkivdanningen og ved oppfølgingen. Dette gjelder både offentlig og privat sektor. Virksomhetene mottar og skaper dokumentasjon i en rekke programmer og fagsystemer, som benytter de filformatene som best er egnet for å løse de respektive oppgavene. Noen av disse formatene er godkjente som arkivformat, noen kan enkelt konverteres mens andre ikke kan konverteres, eller konverteres på en måte som gjør at innholdet vesentlig forringes.

I tråd med Noark 5 ser det ut til å være en trend at saksbehandlingen flyttes fra det tradisjonelle sakarkivsystemet og over i spesialiserte fagsystemer, og at det gjøres integrasjoner mellom disse. I fagsystemene benyttes en rekke filformater, og det er sjeldent mulig å legge begrensninger på hvilke formater som kan overføres til arkivkjernene. Når disse filene så ikke lar seg konvertere ved journalføring blir arkivkaperene sittende med følgende problem:

Dersom filene ikke har arkivformat kan saker ikke avsluttes i arkivsystemet. De kan heller ikke periodiseres, deponeres eller avleveres til Arkivverket. «Løsningen» i dag vil da være som følger:

- Man velger ikke å importere disse filene eller sletter dem etter import slik at det tekniske problemet er løst, men virksomheten og ettertiden vil gå glipp av viktig dokumentasjon.
- Noen filformater lar seg konvertere til arkivformat, men med stort informasjonstap. Man konverterer kun for konverteringens egen skyld for å unngå problemet med avslutning av saker, på bekostning av at man får en fil med lav informasjonsverdi.

IFC som er et åpent format for objektmodeller, er et eksempel på et format som ikke lar seg konvertere til et av de gjeldende arkivformatene i Riksarkivarens forskrift. IFC er et format som er i bruk i bygge- og eiendomssektoren i Norge og utlandet, fra arkitekt til byggherre og videre til eiendomsforvaltningen. IFC kan forenklet beskrives som lagringsformat for «virtuelle bygninger» med innebygget beskrivelse og dokumentasjon. Det er vanlig at IFC-filer som offentlige virksomheter sender eller mottar, ikke kan registreres inn sakarkivsystem, til tross for at de ofte er å regne som bevaringsverdig dokumentasjon.

Et svært vanlig filformat som kan konverteres til arkivformat, men med fare for stort informasjonstap, er regnearkformatet XLS. Selve ombrekkingen av regnearket til PDFens krav gjør ofte tabellene uleselige, i tillegg til at alle formler, filtreringsmuligheter osv. forsvinner. Saksbehandlere kan som regel benytte det originale produksjonsformatet i sakarkivsystemet, men for ettertiden er mye informasjon gått tapt.



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>11 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

### Arkitektenes bruk av digitale verktøy

For å vise et eksempel på mangfoldet av filformater i bruk innen et spesifikt fagområde, har vi tatt med et resultat fra undersøkelsen "Bruk av digitale medier blant norske arkitekter 2016-17".<sup>17</sup> Innen dette fagområdet som forvalter dokumentasjon som ofte er av stor samfunnsnytte, forekommer det flere filformater og dokumentasjonen spenner over relativt mange år.

«Spørsmålene dekket fire temaer: Generell bakgrunn av firmaet og den digitale statusen; designprosessen og arbeidsflyten; bevaringsstrategier, kommunikasjon og formidling av digital arkitektur. Undersøkelsen ble sendt til de ca. 400 arkitektfirmaene i Norge. Av disse svarte 134. Noen av bedriftene har digitalt arkivmateriale fra 1980-tallet. Flertallet av bedriftene har 15-20 år gamle data.

Nitti-fire prosent svarte at mer enn 50 % av produksjonen deres er digitalt skapt og at mer enn 50 % av arkivbestanden deres også er digitalt skapt. Når det gjelder design og arbeidsflyt, brukes en rekke forskjellige filformater og programvare. Likevel skriver over 60 % av respondentene at de noen ganger tar ut papirkopier. Åttifem prosent svarte at de jobber innenfor et BIM-rammeverk der dette er hensiktsmessig. Formater brukt til datautveksling med eksterne partnere varierer sterkt: Autodesk-format (36 %), standard arkivformater som PDF, PDF/E eller 3D PDF (35 %), og IFC-formatet (19 %) og til slutt 7 % andre som ArchiCad og BIMx. Når det gjelder bevaringsstrategier, er bildet tilsvarende heterogent. Filformatene som brukes er PDF eller PDF / A (32 %), Autodesk's DWG (22 %), Autodesk Revit / 3Ds Max (14 %), IFC (12 %) og ArchiCad (xx).»

<sup>17</sup> Undersøkelsen "Bruk av digitale medier blant norske arkitekter 2016-17", inngår i Birgitte Sauges forskningsprosjekt "Mediascapes – Architecture Museums and Digital Media", som er del av "MEDIASCAPES – Cultural Heritage Mediascapes: Innovation in Knowledge and Mediation Practices", 2015-2019, finansiert av Kulmedia-programmet, Norges Forskningsråd.



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>12 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Kartlegging av andre lands filformatstrategi og -praksis

### Innledning

Arbeidsgruppen gjennomførte sommeren 2017 en kartlegging av et utvalg utenlandske depotinstitusjoner som forvalter digital informasjon. Kartleggingen omfatter arkiv- og bibliotekinstitusjoner på nasjonalt nivå og den enkelte institusjons strategi og praksis innen valg og bruk (forvaltning) av filformater for langtidsbevaring. Utgangspunktet for kartleggingen var en antagelse om at depotinstitusjoner, arkivskapere og andre informasjonsforvaltere i global sammenheng forholder seg til de samme filformatene, ofte de samme store programvareleverandørene, både innenfor det enkelte fagområdet og generelt, og at de derfor har de samme utfordringene knyttet til bevaring og tilgang til informasjonen som norske depotinstitusjoner.

Kartleggingen har med ett unntak omfattet institusjoner utenfor Norden. Årsaken til at Norden er utelatt, er at vi allerede vet mye om nordiske depotinstitusjoner. Innenfor Norden og arkivsektoren er det velkjent hvilke valg det enkelte lands arkivmyndighet har tatt som berører filformater og bevaring. Dette henger blant annet sammen med at riksarkivene i Norge, Sverige, Danmark, Island og Finland siden sent på åttitallet har møttes for å utveksle erfaringer innen metodikk og strategier knyttet til digital bevaring. Det har også blitt gjennomført flere prosjekter i nordisk og europeisk regi med deltakere fra de nordiske landene. Dette er et lite miljø med relativt små utskiftninger og liten vekst i antall personer som arbeider innen feltet, og man kjenner hverandre på tvers av landegrensene. Dette har gjort terskelen for samarbeid og utveksling av erfaring lav også utenfor de etablerte arenaene. Et eksempel på dette, er arkivsektorens valg av modell og system for forvaltning av digital informasjon som skal langtidslagres.

Vi har valgt å ta riksarkivet i Finland med i kartleggingen. Årsaken til dette, er hovedsakelig å supplere informasjonen Arkivverket har om det finske riksarkivet slik at vi kan danne oss et oppdatert bilde av deres formattersforvaltning.

Forespørselen har bestått av følgende fem spørsmål:

1. Hva slags praksis har dere for å forvalte filformater for bevaring og er denne forankret i lov- og regelverk?  
*(What is your current practice and what is it based upon, e.g. what legal provisions and principles apply?)*
2. Hvilke fordeler har dagens praksis?  
*(What are the benefits of the current practice?)*
3. Hvilke ulemper har dagens praksis?  
*(What are the disadvantages of the current practice?)*
4. Hvilke planer har dere når det gjelder forvaltning av filformater for bevaring?  
*(What future plans does your institution have related to choosing file formats for long-term preservation?)*
5. Kan dere tenke dere en form for samarbeid innen forvaltning av filformater med institusjoner i Norge?  
*(Would you be interested in some kind of cooperation with institutions in Norway to share experiences associated with choosing file formats for long-term preservation?)*

I tillegg inneholdt forespørselen et ønske om eventuell tilleggsinformasjon.



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>13 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Metode

Kartleggingen har blitt gjennomført ved å sende forespørselen til de aktuelle depotinstitusjonene på e-post eller ved å fylle ut et skjema der e-postadresse ikke har vært tilgjengelig. E-postene har i de aller fleste tilfellene gått til institusjonens postmottak eller generelle kontaktadresse. Utvalget av depotinstitusjoner er ikke gjort ut i fra noe spesielt mønster. Forespørselen ble sendt til 41 depotinstitusjoner fordelt på 33 land.

## Respons

Arbeidsgruppen fikk tilbakemelding fra 27 av institusjonene, fordelt på 24 land, noe som utgjør en svarprosent på 66. Til tross for et noe uheldig tidspunkt med tanke på noen lands ferieavvikling, vil vi si at responsen har vært god. Kartleggingen kom i gang helt i starten av juli og det har vært åpnet for mottak av svar helt til rapporten er ferdigstilt. Siden forespørselen hovedsakelig ble sendt til postmottak og generelle e-postadresser hos den enkelte institusjon, i tillegg til bruk av skjemaer hos noen, kan svarprosenten ha blitt påvirket i negativ retning. Men med tanke på den ressursbruken det samlet ville ha krevd for å skaffe til veie mer målrettede e-postadresser, og at responsen gjennom bruk av generelle kanaler tross alt må sies å ha vært god, ser det ikke ut til at vi har gått glipp av en signifikant mengde informasjon.

Samtlige svar var positive til henvendelsen. Formen på svarene varierte sterkt, fra presise og omfattende svar på de fem spørsmålene til mer knappe tilbakemeldinger hvor respondentens status ble kort beskrevet. En klar overvekt av respondentene er åpne for samarbeid med norske institusjoner om temaet filformater for langtidsbevaring. Det var flere som ønsket informasjon om prosjektets resultater når de foreligger, og informasjon generelt om filformatsforvaltning i Norge. Noen av respondentene foreslo Open Preservation Foundation (OPF) som samarbeidsarena for norske depotinstitusjoner.<sup>18</sup>

## Funn

Kartleggingen har gitt et viktig innblikk i hvordan depotinstitusjoner i andre land forholder seg til utfordringene ved valg og bruk av filformater for digital bevaring. I tillegg har vi fått et innblikk i hvordan flere forholder seg til digital bevaring generelt. Her presenteres de viktigste funnene.

- Vi har langt på vei fått bekreftet antagelsen om at depotinstitusjoner globalt sett har de samme utfordringene og ofte de samme strategiene innen filformatsforvaltning. I tillegg er det en gjengs oppfatning at digital langtidsbevaring er utfordrende og ressurskrevende, noe som blant annet gjenspeiles i et ønske om å samarbeide om spesifikke formattyper og om generelle prinsipper. For noen innebærer dette å delta i forskjellige prosjekter og nettverkssamarbeid, for eksempel i regi av Open Preservation Foundation.

*“Long term preservation of digital objects is difficult.” – Library and Archives Canada*

- Mens noen ser det som en fordel kun å definere et begrenset antall filformater for langtidsbevaring (arkivformater), godtar andre alle formater som den enkelte informasjonsskaper (arkivskaper) har benyttet og ønsker å overføre til depot. Det forekommer oftere at lov- og -regelverk styrer strategi og praksis for filformatsforvaltning for arkivinstitusjonene enn for bibliotekinstitusjonene. Dette påvirker om det kreves spesifikke formater for langtidsbevaring, eller om institusjonene åpner for de formatene overførende organ benytter. Noen arkivinstitusjoner har dog valgt ikke å legge noen restriksjoner på

<sup>18</sup> OPF er en medlemsorganisasjon som jobber for å opprettholde teknologi og kunnskap innen langtidsbevaring av digital kulturarv. Nasjonalbiblioteket er medlem av OPF. <http://openpreservation.org>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>14 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

hvilke filformater de kan motta.

- For dem som har en nedfelt og etablert praksis for mottak av digital informasjon for langtidsbevaring, er det flere som bevarer både det opprinnelige formatet (produksjonsformat) og en konvertert versjon på et format som oppfyller visse kriterier for langtidsbevaring (arkivformat). For disse depotinstitusjonene varierer praksisen fra å motta både produksjons- og arkivformat, til selv å konvertere til arkivformat ved mottak. De fleste som bevarer den samme informasjonen/dokumentasjonen i flere formatversjoner, oppgir at dette er valgt for å ha en (teoretisk) mulighet for å gå tilbake til den opprinnelige filen hvis noe går galt ved konvertering.
- Majoriteten av depotinstitusjonene har valgt de samme filformatene for langtidsbevaring. Det er også de samme standardene og prinsippene som ligger til grunn for valgene. Dette gjenspeiler et ønske om i størst mulig grad å utelate proprietære filformater. Depotinstitusjoner som er i startgropen når det gjelder å etablere en bevaringsstrategi, følger i noen tilfeller valgene til de mer etablerte institusjonene, for eksempel virker det for noen som nærliggende å basere seg på PRONOM og DROID. Bruken av PDF/A som arkivformat er utbredt. Filformater som TIFF, JPEG, JPEG2000 og PNG er også utbredt. Noen av depotinstitusjonene har valgt formater for levende bilder og 3D-modeller, men jo mer kompliserte formater det blir snakk om, desto færre har en uttalt strategi for bevaring av slik informasjon. Generelt har arkivinstitusjonene flere og i noen tilfeller mer komplekse informasjonstyper å forholde seg til og derav flere filformater de må ta hensyn til enn bibliotekinstitusjonene.
- Vi har ikke klart å identifisere noen depotinstitusjoner blant respondentene som har foretatt et klart valg når det gjelder bevaring av filer for bygningsinformasjonsmodellering (BIM).

I tillegg til funnene som er presentert over, ønsker vi å fremheve følgende funn:

- The National Archives (TNA) i Storbritannia tillater at filformater basert på arkivskaperenes premisser og bruk, overføres til depot. For å kunne håndtere dette, har de opprettet et register over filformater - PRONOM, og utviklet verktøyet DROID for å identifisere filformater. Begge har oppnådd en relativ stor utbredelse hos depotinstitusjoner rundt omkring i verden og det er lagt opp til at alle som ønsker det, bidrar med innhold i PRONOM.

*“... as an institution we are prepared to accept any file format that the depositing body considers to be a record of value for permanent retention and preservation.”* (The National Archives - Storbritannia)

- I Finland ble for få år siden the National Digital Library (NDL) som ivaretar digital langtidsbevaring for arkiv- bibliotek- og museumssektoren i Finland, opprettet. NDL definerer hvilke filformater for langtidsbevaring (arkivformater) som skal benyttes for ABM-sektoren. Det som gjør dette spesielt interessant, er at statsarkivet i Finland i mange år har vært en del av det bevaringsfaglige samarbeidet i Norden. Blant annet kan det nevnes at Finland er med i ADDML-samarbeidet sammen med det norske arkivverket og riksarkivet i Sverige.<sup>19</sup> Hvis vi skal se utenfor Norden, er tilbakemeldingen fra National Library of New Zealand interessant:

*“It is becoming clear that digital preservation has a wider application within New Zealand than just*

<sup>19</sup> ADDML – Archival Data Description Markup Language – standard for tekniske metadatabeskrivelser av datasett.



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>15 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

*fulfilling the National Library and Archives New Zealand collecting mandates. With the vast proliferation of digital information being created, the need for and understanding of digital preservation has also increased and so has its potential application across a range of information sectors.”*

<https://digitalpreservation.natlib.govt.nz/current-projects/dpnz/>

- Nasjonalarkivet i Nederland har basert sin formatforvaltningsstrategi på internasjonale beste praksiser kombinert med at det nå og da lages oversikter over hvilke filformater som er i bruk ute i forvaltningen. Dette har dannet grunnlaget for deres valg av arkivformater. I tillegg til arkivformatene tillates også andre formater etter avtale med arkivskaperen. Det som synes spesielt interessant her, er at Nasjonalarkivet til en viss grad holder oversikt over hvilke filformater som er i bruk, og derav kan legge opp en strategi hvor dette inngår som en faktor.

## Oppfølging

- For å få mer informasjon om depotinstitusjoner i andre land sin strategi og praksis:
  - Følge opp med ytterligere spørsmål til respondentene.
  - Sende forespørselen til flere institusjoner og til dem som ikke har svart.
- Gjennom kartleggingen har det blitt knyttet kontakter som legger til rette for samarbeid og informasjonsutveksling, og det har hos respondentene blitt skapt en forventning om mer kontakt. Dette bør for eksempel følges opp gjennom økt samarbeid eller gjennom eksisterende samarbeidsarenaer.
- Noen av strategidokumentene finnes kun på enkelte lands språk og innholdet har derfor ikke vært direkte tilgjengelig for arbeidsgruppen. For noen av disse dokumentene er det opplyst at oversettelse til engelsk er planlagt. Det kan være en idé å følge med på aktuelle depotinstitusjoners nettsider eller ta kontakt via oppgitte kontaktpersoner hvis innholdet for eksempel kan bidra i en eventuell oppfølging eller videreføring av kartleggingen. Det nederlandske nasjonalarkivet (het Nationaal Archief) er et eksempel på en depotinstitusjon som planlegger å oversette dokumentasjon om filformatstrategi.
- Det er flere depotinstitusjoner som opplyser at de er i ferd med å utarbeide eller revidere sin filformatstrategi. Det kan være interessant å følge opp disse for å få informasjon om hvilken strategi den enkelte har valgt.





SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>16 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Norden

Under følger informasjon om de enkelte nordiske riksarkivers valg av filformater for langtidsbevaring. Innholdet i tabellen er basert på informasjon som er tilgjengelig for Arkivverket, på nettsidene til det enkelte riksarkiv og fra kartleggingssvaret fra det finske riksarkivet.

Land	Strategi og praksis
Sverige	<p>Filformater for langtidsbevaring angis i RA-FS 2009:2.<sup>20</sup></p> <p><i>«Riksarkivets föreskrifter och allmänna råd om tekniska krav för elektroniska handlingar (upptagningar för automatiserad behandling);»</i></p> <p>Når det gjelder valg av filformater, kan RA-FS 2009:2 sammenlignes med Riksarkivarens forskrift. På noen områder er de svenske bestemmelsene mer liberale enn gjeldende norsk forskrift (HTML er et godkjent arkivformat), mens de andre steder er mer restriktive (godtar ikke MP3 for lyd). Med revisjonen av de norske formatbestemmelsene vil det være en enda større grad av samsvar.</p> <p>RA-FS 2009:2 er planlagt revidert.</p>
Danmark	<p>I Danmark angir Bekendtgørelse nr. 1007 af 20. august 2010<sup>21</sup> med supplementet «Dokumentvejledningen»<sup>22</sup> hvilke filformater som skal benyttes ved langtidsbevaring.</p> <p>Bekendtgørelse nr. 1007 af 20. august 2010: <i>«Bekendtgørelse om arkiveringsversioner der fastsætte hvilke formater, struktur mv data og dokumentation skal afleveres i.»</i></p> <p>De danske bestemmelsene samsvarer i stor grad med gjeldende norsk forskrift, men en vesentlig forskjell er at PDF/A ikke er valgt som et arkivformat i Danmark.</p>
Island	Island følger det danske regelverket.
Finland	<p>I Finland har The National Digital Library (NDL) som skal ivareta digital langtidsbevaring for arkiv-, bibliotek- og museumssektoren, blitt opprettet.<sup>23</sup></p> <p>På NDL sine nettsider finnes det en liste over filformater for langtidsbevaring (v1.5.1 datert 15.3.2017). Listen foreligger på engelsk og er basert på en rapport utarbeidet av the Library and Archives Canada.<sup>24</sup></p>

<sup>20</sup> <https://riksarkivet.se/rafs?pdf=rafs/RA-FS%202009-02.pdf>

<sup>21</sup> <https://www.sa.dk/wp-content/uploads/2014/10/Bekendtgørelse-nr-1007-af-20-august-2010.pdf>

<sup>22</sup> [https://www.sa.dk/wp-content/uploads/2015/12/Dokumentvejledning\\_endelig\\_beta.pdf](https://www.sa.dk/wp-content/uploads/2015/12/Dokumentvejledning_endelig_beta.pdf)

<sup>23</sup> <http://www.kdk.fi/en>

<sup>24</sup> <http://www.kdk.fi/images/tiedostot/NDL-File-Formats-v1.5.1-en.pdf>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>17 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Umiddelbare effekter av arbeidsgruppens arbeid

1. IFC-formatet blir tatt med i Riksarkivarens forskrift.<sup>25</sup>
2. Økt viten om andre lands depotinstitusjoner sin strategi og praksis.
3. Identifisert kontaktpersoner ved flere depotinstitusjoner utenfor Norden.
4. Interesse i andre lands depotinstitusjoner for resultatet av kartleggingen og rapporten for øvrig.

---

<sup>25</sup> Industry Foundation Classes (IFC) er et format for utveksling av bygningsinformasjonsmodeller (BIM).



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>18 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

## Anbefalinger

Her presenteres arbeidsgruppens anbefalinger til videre aktiviteter og tiltak. Anbefalingene har kommet frem etter diskusjoner, ofte basert på egen erfaring, møte med fagkompetanse når det gjelder bygningsinformasjonsmodellering (BIM), gjennomgang av strategi og praksis i Arkivverket og Nasjonalbiblioteket, samt resultatene av høringen av Riksarkivarens forskrift. I tillegg har kartleggingen av andre lands strategi og praksis dannet et svært viktig grunnlag for arbeidsgruppens anbefalinger.

Rapporten er en oppgave i SAMDOKs delprosjekt Arkiv i e-forvaltning. Da er det naturlig at Arkivverket gis en sentral plass, både med tanke på Arkivverkets samfunnsoppgave hvor Arkivverket har ansvaret for arkivdanningen, Arkivverkets kompetanse på langtidsbevaring og Arkivverkets deltakelse i SKATE. Selv om rapporten i utgangspunktet berører alle informasjonsforvaltere og arkivskapere som møter utfordringer knyttet til tilgang til digital informasjon over tid, er anbefalingene i hovedsak rettet mot Arkivverket. Det er likevel naturlig at Nasjonalbiblioteket og andre store informasjonsforvaltere i Norge, samt sentrale arkivskapere har en rolle i anbefalingene.

Rapporten inneholder følgende fem anbefalinger til tiltak:

1. Arkivverket tar ansvaret for forvaltning av filformater for bevaring for en samlet offentlig forvaltning. Det opprettes et råd for filformater i regi av Arkivverket med deltakere på tvers av sektorer. Rådet bistår forvaltningen i spørsmål knyttet til filformater for bevaring og tilgang til digital informasjon over tid.
2. Arkivverket prioriterer en høy grad av forpliktende samarbeid med fag-, forsknings- og standardiseringsmiljøer i Norge spesielt, men også i utlandet.
3. Arkivverket jobber for økt samarbeid med relevante depotinstitusjoner i inn- og utland.
4. Det legges til rette for å lagre samme informasjon i flere filformater for å minske risikoen for informasjonstap og øke muligheten for senere tilgang.
5. Fokus på fagsystemer økes for å ha bedre kontroll på hvilke typer informasjon (filformater) som forvaltes.

Under følger en beskrivelse av hva som ligger til grunn for den enkelte anbefaling.

### **1. Arkivverket tar ansvaret for forvaltning av filformater for bevaring for en samlet offentlig forvaltning. Det opprettes et råd for filformater i regi av Arkivverket med deltakere på tvers av sektorer. Rådet bistår forvaltningen i spørsmål knyttet til filformater for bevaring og tilgang til digital informasjon over tid.**

Utfordringene knyttet til valg av egnede filformater for tilgang og bevaring gjelder alle som forvalter informasjon over hele verden, både offentlige og private aktører. I Norge angir Riksarkivarens forskrift filformater for bevaring av noen typer informasjon (arkivformater). For arkivskapere som faller inn under arkivloven, er dette en nødvendig rettesnor selv om det i mange tilfeller innebærer uklarhet om hvordan en skal forholde seg til informasjon som ikke naturlig kan bevares i ett av de angitte formatene. Ofte medfører dette informasjonstap når ikke optimale arkivformater velges for å oppfylle kravene til overføring av arkivmaterialet for langtidsbevaring. Nasjonalbiblioteket har sine krav og retningslinjer til filformater for blant annet pliktavleveringer og digitaliseringsarbeid, mens Difi gjennom Standardiseringsrådets referansekatalog angir formater for utveksling av informasjon mellom offentlige organer, og mellom offentlige organer og private. I referansekatalogen refereres det til formatkravene i Riksarkivarens forskrift for informasjon som skal langtidsbevares. Dette er de mest



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>19 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

utbredte og omfattende kravsettene til filformater i det offentlige som kan relateres til bevaring av informasjon. I tillegg finnes det statlige virksomheter som ikke er underlagt arkivloven, og som forvalter et ukjent antall filformater. For private arkivskapere og i privatarkiver kan det i teorien forekomme et hvilket som helst filformat i arkivmaterialet. For privatarkiver er det derfor spesielt krevende å avgjøre i etterkant hvilke filformater som er egnet for å kunne bevare informasjonen over tid og om mulig få konvertert aktuelle filer til et arkivformat.

Arbeidsgruppen foreslår at Arkivverket, ut fra dets samfunnsoppdrag og posisjon innen arkivdanning, tar ansvaret for forvaltning av filformater for bevaring for en samlet offentlig forvaltning, som inngår som en viktig del av arkivdanningen.

Selv om arkivsektoren kan sies å stå i en særstilling med tanke på underliggende lov- og regelverk, er det allikevel mange flere organer og aktører innen offentlig forvaltning og i privat sektor som må forholde seg til informasjon som befinner seg i enkeltfiler og som det er avgjørende at de har tilgang til både i det daglige og i fremtiden. Noe av det som er felles for alle disse sektorene, organene og aktørene, er at det er mange av de samme formatene som går igjen og ikke minst at det er de samme problemstillingene som dukker opp. I dette ligger det et potensial for enhetlig behandling som kan bidra til redusert bruk av ressurser hos den enkelte i arbeidet med å få avklart hvordan forsvarlig informasjonsforvaltning bør gjennomføres med tanke på valg av egnede filformater. I tillegg vil det også kunne bidra betydelig til å redusere informasjonstapet som forekommer ved valg av uegnede filformater for bevaring og ved manglende filformatstrategier. En annen effekt ved en enhetlig behandling, er at det vil bidra til en mer transparent offentlig forvaltning.

Arbeidsgruppen forslår derfor at det opprettes et råd på permanent basis i regi av Arkivverket bestående av representanter fra offentlig og privat sektor.

Rådets oppgaver vil blant annet omfatte

- å være nasjonalt ressursenter for forvaltning av filformater på tvers av sektorer
- å bygge opp og samle kompetanse om filformater, som vil være fritt tilgjengelig
- å samarbeide med andre råd og institusjoner i utlandet
- å gi råd i konkrete problemstillinger eller henvise til riktig instans

Et slikt råd kan stimulere til digitaliseringen av offentlig forvaltning. Det vil også kunne være en del av Arkivverkets privatarkivstrategi, og bør koordineres med denne. Antagelsen om at mangelfulle filformatstrategier og praksiser innebærer fortløpende informasjonstap, er et klart insentiv til å jobbe mot en enhetlig filformatsforvaltning.

Arbeidet med å få besluttet at et slikt råd skal opprettes, utarbeidelse av mandat og invitasjoner til deltakelse, er noe Arkivverket kan igangsette umiddelbart.

## **2. Arkivverket prioriterer en høy grad av forpliktende samarbeid med fag-, forsknings- og standardiseringsmiljøer i Norge spesielt, men også i utlandet.**

Det er per i dag flere typer informasjon som ikke lar seg langtidsbevare på en god måte uten informasjonstap, for eksempel DAK-modeller, avanserte regneark og GAB-filer (anbud/tilbudsdokumentasjon).<sup>26</sup> Som en del av Arkivverkets samfunnsoppdrag og for å redusere tapet av informasjon, er det naturlig å være kontinuerlig oppdatert på viktige egenskaper ved informasjonen som forvaltes hos arkivskapere. Det er også viktig å være i

<sup>26</sup> Dataassistert konstruksjon (På engelsk: CAD – Computer Aided Design)



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>20 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

forkant når det gjelder å utarbeide strategier for hvordan denne informasjonen kan sikres for ettertiden. En sentral del av denne oppgaven er tett kontakt med aktuelle miljøer i inn- og utland. Det foregår allerede flere samarbeidsaktiviteter hvor både fagmiljøer og kommersielle aktører er involvert. For eksempel er Arkivverket medlem av DLM Forum<sup>27</sup> og involvert i PREFORMA-prosjektet<sup>28</sup>.

Økt samarbeid med fag-, forsknings- og standardiseringsmiljøer vil kunne bidra betydelig til å sikre oppdatert spisskompetanse i Arkivverket og mulighet for å påvirke og fange opp trender og nyvinninger på filformatområdet og andre relevante områder innen digital bevaring og tilgjengeliggjøring.

I tillegg til å øke graden av samarbeid, bør informasjon om slike samarbeidsaktiviteter kommuniseres utad i større grad. Informasjon om pågående og aktuelle aktiviteter vil blant annet kunne bidra til lettere å få med kvalifiserte deltakere fra norsk side.

### 3. Arkivverket jobber for økt samarbeid med relevante depotinstitusjoner i inn- og utland.

Kartleggingen viser at mange depotinstitusjoner utenfor Norge har de samme utfordringene knyttet til filformatsforvaltning som de vi har i Norge. Siden det i svært mange tilfeller er snakk om de samme filformatene for å skape og forvalte korrespondanse, lyd, bilde, 3D-modeller, med mer, er det naturlig at utfordringene når det gjelder bevaring av informasjon og tilgjengelighet, er felles.

Blant annet basert på tilbakemeldinger i kartleggingen, ser vi at det er et ønske hos andre lands depotinstitusjoner om å dele erfaring. I tillegg til økt erfaringsutveksling, er det nærliggende for Arkivverket å delta i felles prosjekter innen standardisering og rutine- og systemutvikling, utover det som nå er tilfelle.

### 4. Det legges til rette for å lagre samme informasjon i flere filformater for å minske risikoen for informasjonstap og øke muligheten for senere tilgang.

Det er ikke uvanlig med systemer hvor informasjonen som skal langtidslagres, er lagret i filformater som hverken er arkivformater etter Riksarkivarens forskrift eller lar seg konvertere til et arkivformat uten at informasjon går tapt. For å øke sjansene for at informasjonen kan leses og behandles på et senere tidspunkt, kan et virkemiddel være å lagre informasjonen i flere formater, for eksempel ved at den overføres til aktuell depotinstitusjon både i opprinnelig format (produksjonsformat) og i det best egnede arkivformatet. Mange produksjonsformater er proprietære og beheftet med lisenser og patenter, eller bryter mot andre prinsipper for valg av formater for langtidsbevaring. Det kan allikevel være verdt å ta vare på filer i slike formater hvis utbredelsen er så høy at det i relativt lang tid er mulig å finne programvare for å tolke og vise informasjonen i filene. Det er i tillegg en mulighet for at det i løpet av denne perioden blir definert et mer egnet arkivformat som filen i produksjonsformat kan konverteres til. Et hensyn som må tas, er behovet for bevaring av informasjonen i flere formater veid opp mot det økte behovet for lagringsplass og en større datamengde som må forvaltes i lang tid.

<sup>27</sup> «The DLM Forum is vibrant community of public archives and interested parties from across government, commercial, academic and voluntary sectors who are active in information governance, including archives, records, document and information lifecycle management. It was founded by the European Commission and first met in 1996. Today it is a not-for-profit foundation providing industry specifications, participating in activities and serving members from all over Europe and the rest of the world.» <http://www.dlmforum.eu/>

<sup>28</sup> PREFORMA (PREservation FORMAts) er et prosjekt som er delfinansiert med EU-midler og som har som oppgave å vurdere dokumentformater. I tillegg utvikler partnerne verktøy for å teste validitet av formatene. Prosjektet koordineres av det svenske riksarkivet. <http://www.preforma-project.eu/>



SAMDOK delprosjekt:	Prioritert oppgave:	SAK (ePhorte):	Dato:	Side: <b>21 av 21</b>
Forfatter:	Tittel: <b>Sluttrapport 2017 – Digital langtidsbevaring - Strategi for valg av filformater for bevaring</b>			

For å være i forkant av slike situasjoner hvor arkivskaper eller depotinstitusjon må ta stilling til om også filer i produksjonsformat skal bevares og holdes tilgjengelig, kan det være rasjonelt både for arkivskaper og depotinstitusjon å ha en filformatstrategi på plass så tidlig som mulig. En slik strategi kan på bakgrunn av kartlegging av hvilke filformater som kan komme til å bli behandlet i det enkelte system, tidlig ta stilling til hvordan informasjonen bør bevares. For eksempel vil det i strategien inngå hvilke filer det vil være nødvendig eller ønskelig å bevare i både produksjons- og arkivformat. Siden det ofte er de samme filformatene som går igjen i offentlig sektor, så vel som i privat, er dette strategier som kan utformes sentralt og kommuniseres bredt.

## 5. Fokus på fagsystemer økes for å ha bedre kontroll på hvilke typer informasjon (filformater) som forvaltes.

Utbredelsen av fagsystemer hvor det utføres saksbehandling øker. Dette bidrar til at det blir vanskelig å ha oversikt over og styre hvilke filformater som lagres via fagsystemene. Statlige organer skal utarbeide bevarings- og kassasjonsplaner hvor det vil fremgå hva slags materiale fra fagsystemene som skal bevares. Det kan legges opp til at bevarings- og kassasjonsplanene, eller Arkivverkets rapporteringsordning for fagsystemer, også kan omfatte en oversikt over hvilke filformater som forekommer i fagsystemet. Hensikten med en slik oversikt er at Arkivverket tidlig kan vurdere om formatene er hensiktsmessige med tanke på bevaring og sammen med arkivskaper planlegge hvordan dokumentasjonen fra fagsystemet kan bevares.

Rapporteringsordningen bør organiseres slik at det er basert på aktiv og løpende kartlegging og oppfølging av Arkivverket eller av det organ/råd som synes hensiktsmessig til oppgaven. Slik vil man i mindre grad basere seg på rapportering fra arkivskaper etter arkivskapers eget initiativ. Arkivverket kan i forbindelse med en justering eller endring av nåværende ordning, være tjent med en evaluering av gjeldende praksis for lettere å identifisere hvilke rapporteringsmetoder for fagsystemer som bør prioriteres.