

Digital bevaring – status og viden 2016

Rigsarkivet december 2016

Indledning

Langt størstedelen af den dokumentation, der i dag skabes af det danske samfund, bliver skabt i digital form. Det gælder både i den offentlige og i den private sektor. Det giver særlige udfordringer at bevare denne dokumentation, både på kort og på langt sigt. Det skyldes, at digitale medier ikke er langtidsholdbare på samme måde som papir, og at den teknologi, som man skal bruge til at anvende data og digitale dokumenter, meget hurtigt forældes.

Rigsarkivet er Danmarks fælles hukommelse og har til opgave at sikre dokumentation af det danske samfund, både så den kan anvendes i den nære fremtid, og så kommende generationer kan skrive og forstå deres historie. Vi indsamler derfor løbende viden om, hvordan de nye tekniske muligheder, der kommer på markedet, bliver anvendt, og hvordan andre kulturbevaringsinstitutioner løfter bevaringsopgaven både nationalt og internationalt. Resultatet publiceres i en årlig rapport, som udkommer for sidste gang i år. Arbejdet med opsamle viden på området og stille den til rådighed vil dog fortsætte.

Rapporten er disponeret efter Rigsarkivets strategi for arkivering af digitalt skabte arkivalier, som kan findes på www.sa.dk. Strategien bygger på følgende principper:

- Tidlig identifikation og godkendelse af nye statslige it-systemer med henblik på aflevering
- Hyppige afleveringer i systemuafhængige formater
- Løbende bevaringsplanlægning og periodiske migreringer til nyt bevaringsformat

Rapporten afspejler, at rammerne for Rigsarkivets arbejde med digital bevaring er i disse år under forandring, først og fremmest fordi implementeringen af EU's nye persondataforordning i dansk lovgivning kommer til at medføre en revision af arkivloven. Men også den fællesoffentlige digitaliseringsstrategi for perioden 2016-2020 sætter retning for Rigsarkivets virksomhed i de kommende år. Det skyldes både, at strategien giver et fingerpeg om, hvor de offentlige myndigheders fokus på digitalisering vil være, men også at strategien har haft indflydelse på Rigsarkivets nye strategi, Vores fælles hukommelse, som dækker perioden frem til 2025. Her er de strategiske mål:

- Arkiverne bringes i spil til flere,
- Bidrag til effektivisering af den offentlige sektor
- Udvikling og fremtidssikring af samlingerne

Det er vores håb, at alle, der arbejder med digitalt forvaltning, digital dokumentation af deres virksomhed og digital bevaring, kan have glæde af rapporten.

God fornøjelse med læsningen.

Indholdsfortegnelse

Digital bevaring – status og viden 2016	1
Indledning	2
Indholdsfortegnelse.....	3
1 Overvågning af digital forvaltning i Danmark.....	4
1.1 Ny persondataforordning - hvad betyder den for arkiverne?.....	4
2 Tidlig identifikation og godkendelse af systemerne med henblik på aflevering.....	4
2.1 Tilsyn med kommuner og regioner 2015 - afsluttende bemærkninger	4
2.2 Revision af bk-bestemmelser for kommuner og regioner	5
2.3 Et kig ind i myndighedernes journaliseringspraksis - tilsyn med statslige myndigheders ESDH-systemer	5
2.4 Datamanagement og anmeldelse af forskningsdata.....	6
2.5 Virksomhedsprojektet	6
2.6 Bevaringsaftaler – status for den private sektor	6
3 Hyppige afleveringer i systemuafhængigt format.....	7
3.1 Geodata – ny anvisning	7
3.2 Handlingsplan for effektivisering af afleveringsprocesser	8
3.3 E-ARK – status.....	9
4 Formater	9
4.1 Formater for forskningsdata.....	9
4.2 Dokumentformater	10
5 Løbende bevaringsplanlægning og periodiske migreringer til nyt bevaringsformat	12
5.1 Infrastrukturrapport	12
6 Tilgængeliggørelse af data i arkiveringsversioner	12
7 Samarbejdsrelationer og internationale indtryk	13
7.1 Samarbejdsrelationer – OPF, DLM mv.....	13
7.2 Samarbejdet i Det Nationale DM-Forum.....	14
7.3 Nordisk ekspertseminar om bevaring og kassation	14

1 Overvågning af digital forvaltning i Danmark

1.1 Ny persondataforordning - hvad betyder den for arkiverne?

I 2016 blev databeskyttelsesforordningen vedtaget i EU efter flere års drøftelser. Forordningen finder anvendelse fra den 25. maj 2018. I Danmark har Justitsministeriet oprettet et særligt databeskyttelseskontor, som vil være hovedansvarlig for, at forordningen bliver implementeret i Danmark, og at øvrig lovgivning bliver tilpasset i fornødent omfang. Rigsarkivet deltager som observatør i en af de arbejdsgrupper, som Databeskyttelseskontoret har nedsat. Implementeringen af forordningen i arkivlovgivningen skal i øvrigt drøftes i arkivlovsudvalget, der er blevet nedsat i efteråret 2016.

”Retten til at blive glemt” har været et nøglebegreb i arbejdet med at udforme forordningen, hvilket gav arkiverne anledning til betydelig bekymring for, om forordningen ville betyde en forringelse af mulighederne for at modtage, bevare og stille autentiske data med personoplysninger til rådighed.

Forordningens artikel 89 indeholder dog en række undtagelsesmuligheder i forhold til den registreredes muligheder for eksempelvis at kræve berigtigelse eller sletning af personoplysninger, såfremt der er tale om en ”behandling til arkivformål i samfundets interesse, til videnskabelige eller historiske forskningsformål eller til statistiske formål”.

Den konkrete implementering af forordningens undtagelsesmuligheder i henhold til artikel 89 skal afklares nærmere i arkivlovsudvalget.

2 Tidlig identifikation og godkendelse af systemerne med henblik på aflevering

2.1 Tilsyn med kommuner og regioner 2015 - afsluttende bemærkninger

I 2015-2016 gennemførte Rigsarkivet det 4. tilsyn med kommunerne og det 2. tilsyn med regionerne, jf. vidensrapport 2015, s. 8. Begge tilsyn skulle følge op på tidligere tilsyn, men der skulle også indsamles oplysninger til brug for udarbejdelse af reviderede bevaringsbestemmelser for kommuner og regioner.

Hovedresultatet for kommunernes vedkommende er, at de generelt er gode til at sikre, at it-systemer med bevaringspligtige data og dokumenter ikke bliver slettet inden arkivering. Tilsynet har også vist, at de 125 km bevaringspligtige papirarkiver bliver opbevaret under betryggende forhold i de fleste kommuner. Rigsarkivet har i 2016 udarbejdet en afrapportering af kommunetilsynet. Den indeholder en række anbefalinger dels til kommunernes varetagelse af arkivmæssige hensyn, dels til Rigsarkivets tilsynsvirksomhed over for kommunerne:

[Rapport: Rigsarkivets tilsyn med kommunerne 2015](#)

For regionernes vedkommende var det overordnede resultat, at regionerne har lagt et stort arbejde i at få konsolideret oversigterne over it-systemer med bevaringspligtige data og dokumenter, herunder de cirka 60 kliniske kvalitetsdatabaser, som rummer vigtige data til sundhedsforskning. Regionerne skal nu i gang med at få afleveret data fra deres bevaringspligtige it-systemer. En del af systemerne har data tilbage til 1990'erne, som ikke tidligere er blevet afleveret. Afrapportering af det regionale tilsyn indeholder en række anbefalinger til regionernes varetagelse af arkivmæssige hensyn:

2.2 Revision af bk-bestemmelser for kommuner og regioner

Der blev i 2010 fastsat bestemmelser for kommunale arkivalier fra perioden efter 2007. I 2016 har bestemmelserne været i kraft i fem år, og kommunerne har gjort sig erfaringer med dem. Samtidig er der sket ændringer i ressortfordelingen mellem staten og kommunerne, Udbetaling Danmark har overtaget mange opgaver fra kommunerne, og det kommunale it-systemlandskab er under forandring efter monopolbruddet med KMD. Rigsarkivet nedsatte derfor i foråret 2016 en arbejdsgruppe, som skal give den nuværende bestemmelse et serviceeftersyn. Arbejdsgruppen består af en repræsentant fra KL, to repræsentanter fra offentlige arkiver nedsat i henhold til arkivlovens § 7 (§ 7-arkiverne) og tre repræsentanter fra Rigsarkivet, herunder projektlederen. Rigsarkivets direktion er styregruppe, og Udvalget vedrørende bevaring, aflevering og tilgængeliggørelse af kommunale og regionale arkivalier (K-udvalget) er referencegruppe. Arbejdsgruppen har bl.a. arbejdet med at analysere dataflowet mellem stat og kommune, forældreintranet på skolerne, bevaring af oplysninger om flygtninge og klagesager. Arbejdsgruppen er på nuværende tidspunkt ved at lægge sidste hånd på en indstilling, som skal videre til K-udvalget. Det forventes, at udkast til ny bekendtgørelse kommer i offentlig høring i begyndelsen af 2017.

2.3 Et kig ind i myndighedernes journaliseringspraksis - tilsyn med statslige myndigheders ESDH-systemer

Rigsarkivet har i mange år godkendt statslige myndigheders ESDH-systemer og har i tidens løb gennemgået talrige journalplaner, facetoversigter, instrukser og anvisninger. Men hvordan arbejder myndighederne egentlig med journalisering, bliver instrukserne anvendt, eller lander de i skuffen, indtil systemet skal afleveres efter fem år? Kan stringent journalisering gennem hele perioden gøre det billigere at aflevere en arkiveringsversion af systemet, når den tid kommer? Hvilken kvalitetssikring af sager og dokumenter foretager myndighederne i deres systemer? Dette var nogle af de spørgsmål, Rigsarkivet stillede sig selv, da der skulle udarbejdes et nyt anmeldelsescirkulære i 2013. Med det nye cirkulære blev der derfor indført et tilsyn med myndighedernes ESDH-systemer 2 år efter ibrugtagning.

I 2016 blev de første otte tilsyn gennemført og har givet et første kig ind i myndighedernes journaliseringspraksis. Tilsynene viser, at mange myndigheder har indført regelmæssig kvalitetssikring af deres sager og dokumenter. Ved kontrollen opdager de fx, hvis der kommer filer ind i deres systemer, som ikke kan arkiveres, sager, der ikke lukkes, dokumenter, som fejlagtigt ikke er tilknyttet en sag, samt titler på dokumenter, som gør det svært at finde det relevante dokument i en sag. Forhold, som tidligere først blev adresseret i forbindelse med arkiveringen, og som fordyrede arkiveringsprocessen. Nu får både myndigheden og arkivet glæde af rettelserne, da de sker i det "levende system".

Det kan imidlertid være udfordring for myndighederne at få de tilstrækkelige ressourcer til kontrollen. På tilsynsmøder vejleder Rigsarkivet derfor i, hvordan kontrollen kan udføres med færrest mulige omkostninger.

I mange systemer kan kvalitetssikringen understøttes bedre via forskellige rapporter og/eller automatiserede processer, så den manuelle del af kvalitetssikringen kan mindskes. Langt de

fleste systemleverandører har allerede fokus på dette område, og det er Rigsarkivets forventning, at kvaliteten i myndighedernes registreringer vil stige over de kommende år.

2.4 Datamanagement og anmeldelse af forskningsdata

Rigsarkivet har i indeværende år arbejdet videre med en bekendtgørelse om anmeldelse af statslige forskningsdata. I begyndelsen af året blev arbejdet præsenteret for en referencegruppe nedsat af Danske Universiteter. Tilbage meldingen var, at det daværende udkast til bekendtgørelse var for upræcist i forhold til, hvilke datasæt der skal anmeldes, og hvornår de skal anmeldes. Rigsarkivet har derfor i løbet af året arbejdet videre med denne problematik. Det er bl.a. sket ved deltagelse i et projekt under Det Nationale Datamanagement Forum, som arbejder med at udvikle en guide til datamanagement til forskerne, og hvor man opstiller en kategorisering af forskningsdata. Denne kategorisering forventes også at kunne anvendes i den kommende bekendtgørelse, hvorved der sikres et ensartet begrebsapparat til beskrivelse af danske forskningsdata. Det er Rigsarkivets forventning, at bekendtgørelsen kan sendes i høring i begyndelsen af 2017.

2.5 Virksomhedsprojektet

Erhvervs livet er en vigtig del af det danske samfund. Bevaring af dokumentation fra private virksomheder, herunder i digital form, er derfor et væsentligt element i Rigsarkivets indsamling, der har til formål at sikre en helhedsdækkende samfunds dokumentation.

Rigsarkivet har i 2016 påbegyndt et udredningsarbejde om virksomheders arkivforhold med særligt fokus på deres kendskab til Rigsarkivets service, hvad der hindrer virksomhederne i at aflevere data til Rigsarkivet, samt hvad der kunne få dem til at aflevere. Virksomhedsprojektet skal skabe et evidensbaseret grundlag for en ny strategi for indsamling og bevaring af digital dokumentation og it-systemer fra danske erhvervs virksomheder.

Projektet består af en række afgrænsede delundersøgelser, herunder bl.a. af virksomhedernes syn på bevaringssamarbejde med offentlige arkiver. Endelig skal undersøgelsen kortlægge brugernes – og det er helt overvejende forskeres – efterspørgsel efter og forventninger til den dokumentation, der bevares fra virksomheder.

Første fase består i en survey-undersøgelse, der udsendes til ca. 100 danske virksomheder. Undersøgelsen skal afdække virksomhedernes kendskab til Rigsarkivet og muligheden for at indgå digitale bevaringssamarbejder med Rigsarkivet. Samtidig skal undersøgelsen skabe klarhed om de faktiske barrierer for aflevering.

2.6 Bevaringsaftaler – status for den private sektor

Rigsarkivet har siden 2009 lavet en opsøgende indsats for at sikre en tidlig identifikation af private arkivskaberes it-systemer for at sikre deres bevaring og senere aflevering. Det sker ved indgåelse af en bevaringsaftale, der sikrer, at den private arkivskaber ikke uden forudgående kontakt til Rigsarkivet sletter oplysninger i it-systemet, men i stedet afleverer oplysningerne til Rigsarkivet via en afleveringsaftale.

Den opsøgende indsats er fortsat i 2016 med et særligt fokus på foreninger og organisationer inden for klima- og miljøområdet.

I de første tre kvartaler er der indgået 17 bevaringsaftaler og fem afleveringsaftaler med private arkivskabere – fra virksomheden A/S Hvide Sande Fiskernes Fremtid til Økologisk Landsforening og Skole og Forældre. Der er endvidere indgået afleveringsaftaler med 14 folketingsmedlemmer, hvor partiet Alternativet er godt repræsenteret med otte aftaler.

Status for aftaler er, at der siden 2009 er indgået bevaringsaftaler med 51 private aktører samt 28 afleveringsaftaler; endvidere er der modtaget data fra 23 aktører.

Herudover er der indgået afleveringsaftaler med 75 folketingsmedlemmer og modtaget data fra 16.

3 Hyppige afleveringer i systemuafhængigt format

3.1 Geodata – ny anvisning

I juni 2016 udstedte Rigsarkivet en ny anvisning til aflevering af geodata. Denne anvisning erstatter de krav til aflevering af geodata, der i dag fremgår af bekendtgørelse nr. 1007 om arkiveringsversioner. Den nye anvisning vil blive implementeret som krav i bekendtgørelsen, næste gang den revideres.

Rigsarkivets nye format til aflevering af geodata er GML version 3.1.1. GML står for Geography Markup Language. Med denne standard går Rigsarkivet væk fra kravet om anvendelse af den specifikke danske standard GML 3.v.3.DK, som erfaringer har vist ikke er tidssvarende og operationel, og mod en mere generel og internationalt anvendt standard. Det europæiske E-ARK-projekt, der arbejder på at udvikle fælles europæiske afleveringsformater og tilgængelighedsløsninger til digitalt skabte data, opstiller identiske retningslinjer for brug af GML. Formatet anvendes også i den danske implementering af Inspire-direktivet samt i FOT-Danmark, der er et samarbejde mellem staten og kommunerne om etablering af et fælles geografisk administrationsgrundlag målrettet offentlig forvaltning.

Ud over anvendelse af dette nye GML-format stiller Rigsarkivet nu også krav om, at GML-dokumenter og deres lokale skemaer udarbejdes i en bestemt obligatorisk struktur med enkelte obligatoriske elementer. Desuden skal geodata være veldokumenterede ved hjælp af lokale features. Hvis geografien fx er et punkt, skal der være et antal lokale features, der specificerer, hvad punktet angiver. Er punktet en angivelse af placeringen af et skibsvrag, en bygning eller et træ? Endelig skal det entydigt dokumenteres, hvordan geodata i GML-dokumentet relateres til øvrige data i arkiveringsversionen, fx hvordan geodata i et GML-dokument er koblet til en sag registreret i arkiveringsversionens tabeller.

De nye regler for aflevering af geodata udgør således et domænespecifikt Application Schema, hvor Rigsarkivet har defineret, hvilke elementer og beskrivelser der er relevante for aflevering af geodata inden for domænet "Arkiv". Disse regler har til hensigt at gøre det muligt at teste afleveringerne ens og øge brugervenligheden ved anvendelse af de afleverede geodata.

Testværktøjet ADA kan nu teste afleveringer, der indeholder geodata.

Anvisning til aflevering af geodata til Rigsarkivet samt et eksempel på en arkiveringsversion, der indeholder GML-data, kan findes på <https://www.sa.dk/aflevering-arkivet/statslige/aflevering-it-systemer>.

3.2 Handlingsplan for effektivisering af afleveringsprocesser

Det er Rigsarkivets mål, at det skal være så nemt og billigt som muligt at aflevere bevaringsværdige data. For et par år siden blev der derfor indsamlet forslag og ønsker til afleveringsprocessen fra myndigheder og leverandører. På baggrund af disse inputs blev der udarbejdet en handlingsplan for effektivisering, som kan ses her:

[Handlingsplan for effektivisering af processerne knyttet til aflevering af data til Statens Arkiver](#)

Overordnet set har arbejdet med handlingsplanen fulgt tre hovedspor:

- Udarbejdelse og revision af vejledningsmateriale, der skal gøre det lettere for myndigheder og leverandører at efterleve gældende regler
- En analyse af omkostningerne ved at ændre på gældende formater samt metoder og procedurer knyttet hertil
- En analyse af hvorvidt arbejdet med produktion af arkiveringsversioner kan organiseres anderledes, fx ved at Rigsarkivet overtager flere opgaver

3.2.1 Vejledning

Rigsarkivet har i forbindelse med handlingsplanen opdateret sin vejledningsstrategi og på baggrund heraf udarbejdet en lang række vejledningsprodukter samt revideret og opdateret eksisterende materiale.

Vejledningsmaterialet omfatter bl.a.

- nye eksempelarkiveringsversioner
- en række "Kort og godt"-vejledninger om bl.a. aflevering, test, SQL-forespørgsler og kassation
- en ny dokumentvejledning
- en video med introduktion til, hvad en arkiveringsversion er
- en liste over leverandører, der har fået godkendt mindst én arkiveringsversion efter reglerne i bekendtgørelse 1007.

Ud over det skriftlige vejledningsmateriale er Rigsarkivets kurser for myndigheder blevet opdateret, så de inkluderer de emner, som myndighederne har efterspurgt.

3.2.2 Processer ved aflevering

Hvad angår processerne omkring aflevering, har Rigsarkivet arbejdet målrettet på at nedbringe tidsforbruget til test. I 2013 var den gennemsnitlige sagsbehandlingstid fra modtagelse til svar på test på 3,8 måneder. I 2015 var dette tal nedbragt til 2,2 måneder i gennemsnit. Det vil også fremover være et fokusområde for Rigsarkivet at holde sagsbehandlingstiderne i testværkstedet så lave som muligt.

Rigsarkivet har konstateret, at det for mange leverandører og myndigheder vil være enklere og mere effektivt at aflevere data via en sikker FTP-forbindelse. En sådan løsning indebærer såvel sikkerheds- som kapacitetsmæssige udfordringer, men Rigsarkivet er indstillet på snarest muligt at gennemføre et pilotprojekt, som kan afklare mulighederne for at tilbyde aflevering ad denne vej.

3.2.3 Nye modeller for produktion af arkiveringsversioner

Eftersom udfordringerne med at få leverandørerne til at løse konverteringsopgaven inden for en rimelig tid og ikke mindst til en rimelig pris var det mest påtrængende set fra myndighedernes side, var det et udbredt

ønske i indhentningen af forslag til effektivisering, at Rigsarkivet i stedet skulle overtage produktion af arkiveringsversioner eller i det mindste stå for dokumentkonverteringen.

Rigsarkivet har på baggrund af dette ønske gennemført en analyse af mulighederne herfor. Analysen viser, at det ikke er muligt for Rigsarkivet at overtage hverken produktionen af arkiveringsversioner eller konverteringsopgaven inden for de nuværende juridiske og økonomiske rammer.

Som alternativ vil Rigsarkivet i stedet fortsat have fokus på at understøtte konkurrenceudsættelsen af arkiveringsopgaven. I den forbindelse har Rigsarkivet og SKI drøftet, hvorvidt og hvordan SKI-aftaler kan spille en rolle i forhold til reduktion af myndighedernes arkiveringsudgifter.

3.3 E-ARK – status

Det fælleseuropæiske projekt E-Ark, som Rigsarkivet deltager i, er godt i gang med sit tredje år. Projektets formål er at identificere, videreudvikle og standardisere metoder til digital bevaring på europæisk plan. Projektet har udarbejdet specifikationer for såvel afleveringsformater som arkiverings- og tilgængeliggørelsesformater og udviklet værktøjer, der understøtter arbejdet hermed.

Af særlig interesse for Rigsarkivet har været arbejdet med databasearkivering, hvor der er udarbejdet en ny udgave af databasearkiveringsformatet SIARD, som også ligger til grund for de eksisterende danske regler for aflevering af data til Rigsarkivet. Projektet har udviklet værktøjet Database Preservation Toolkit til udtræk af data fra relationelle databaser, som kan anvendes ved produktion af arkiveringsversioner efter såvel de eksisterende danske regler som det nye SIARD 2.0-format. Rigsarkivet har gennemført en række pilotforsøg med værktøjet, og de foreløbige resultater er lovende. Ligeledes har projektet udviklet programmet Database Visualization Toolkit til søgning i de arkiverede databaser.

Projektet slutter 1. februar 2017, men der vil blive arbejdet videre med resultaterne i relevante europæiske fora. DLM-Forum står for vedligeholdelse af standarderne og har i den forbindelse nedsat en ”Digital Archiving Standards Board”, mens Open Preservation Foundation står for vedligehold af de udviklede programmer. Der er således et stort potentiale for, at arkiverne i Europa kan begynde at anvende ensartede formater og programmer, hvilket vil indebære en betydelig gevinst for vores fælles hukommelse.

Projektet kan følges på www.eark-project.com

4 Formater

4.1 Formater for forskningsdata

Der har ikke tidligere eksisteret ensartede regler for modtagelse af forskningsdata. Derfor har Rigsarkivet traditionelt lagt en stor arbejdsindsats i at konvertere og oparbejde forskningsdata fra samfunds- og sundhedsvidenskaberne. Rigsarkivet har i 2016 arbejdet videre med at afdække, hvilke formater der fremover kan sikre en mere effektiv modtagelse og behandling af forskningsdata. I første omgang har der været fokus på surveydata o.lign. Efter flere overvejelser og forsøg er den foreløbige vurdering, at det vil være mest hensigtsmæssigt at skelne imellem et egentligt afleveringsformat og et langtidsbevaringsformat.

Det betyder i praksis, at forskningsdata skal afleveres i et bestemt afleveringsformat. Når der er gennemført en succesfuld test af afleveringen, skal data efterfølgende konverteres til Rigsarkivets

langtidsbevaringsformat i relationel struktur, sådan som det fremgår af bekendtgørelse 1007 om arkiveringsversioner.

De overordnede krav til det afleveringsformat for forskningsdata, som p.t. er på tegnebrættet, er følgende:

1. Udtræksformatet skal være systemuafhængigt, dvs. være uafhængigt af specifikke programmer
2. Udtræksformatet skal være tekstbaseret, både hvad angår metadata og egentlige data
3. Udtræksformatet skal bygge på en stringent logisk repræsentation af data og metadata, beskrevet i en "EBNF"¹, som skal udmønte sig i egentligt regelsæt på området
4. Udtræksformatet skal kunne udtrækkes ved brug af den software, der normalt anvendes til statistikbaseret forskning, så udtrækket af data kan foretages af den personkreds, der udfører forskningen
5. Udtræksformatet skal med et minimum af omkostninger kunne transformeres tabsfrit til Rigsarkivets langtidsbevaringsformat

Et afleveringsformat, der understøtter de ovenstående fem krav, er i skrivende stund under afprøvning hos eksterne parter. De foreløbige resultater/tilbagemeldinger ser meget lovende ud, og de vil komme til danne en del af grundlaget for en endelig beslutning om formatvalg. Et nyt afleveringsformat kan dog ikke stå alene, og den eksterne afprøvning skal ligeledes give inputs til den vejledning og eventuelt hjælpemidler, som Rigsarkivet skal stille til rådighed for forskere og forskningsinstitutioner. Undersøgelsen har samtidigt vist, at udlevering af forskningsdata til interessenter nemt vil kunne fortages ved hjælp af en efterfølgende konvertering fra Rigsarkivets langtidsbevaringsformat til udlevering i gængse statistikformater som SPSS eller SAS. Det eneste, der kræves i den forbindelse, er udvikling af et mindre program, der understøtter en sådan transformation.

4.2 Dokumentformater

Da Rigsarkivets handlingsplan for effektivisering blev udarbejdet, var mange leverandører og myndigheder i gang med at gøre de første erfaringer med arkivering af store dokumentmængder fra forvaltningens ESDH-systemer. Indkøringsfasen blev oplevet som ganske tung, og der var derfor stort fokus på muligheden for at bruge andre formater end TIFF til dokumenter for om muligt at nedbringe konverteringsomkostningerne.

Rigsarkivet har derfor undersøgt en række muligheder, herunder bl.a. XML-baserede dokumentformater, rene tekstdokumenter (TXT), WKT (well-known text) til geodata og PDF/A.

I forhold til OOXML, ODF eller PDF/A er der ikke umiddelbart nogle ændringer i omverdenens håndtering eller understøttelse af formaterne, som ændrer på tidligere eller nuværende vurdering af dem som bevaringsformater. Den største udfordring er og bliver manglende eller utilstrækkelig mulighed for validering. Selv hvis/når det bliver muligt at validere, tyder alt på, at problematikken omkring performance i forbindelse med test af dokumenter ikke vil være løst.

Hvad angår rene tekstdokumenter, er der ingen tvivl om, at tekstdokumenter gemt som TXT fylder væsentlig mindre end TIFF-dokumenter. Der vil typisk kunne opnås en besparelse på en faktor 10 eller mere for de dokumenter, som eventuelt kunne afleveres som TXT i stedet for TIFF. TXT vil dog kun være en reel mulighed for et mindre udsnit af de dokumenter, som Rigsarkivet modtager (skønsmæssigt 3-6 %).

¹ EBNF Extended Backus Naur Form. En notation, der kan anvendes til at beskrive kontekstfri formelle sprogs syntaks

Samtidig er der en række variationer i forhold til bl.a. tegnsæt, og en anvendelse af ren tekst som afleverings- og bevaringsformat forudsætter derfor, at Rigsarkivet specificerer det præcise tekstformat, som i givet fald skal anvendes, og definerer valideringsrutiner herfor. Såvel specifikation som implementering hos arkiv, myndigheder og leverandører vil være ganske tidskrævende, hvilket bekræftes af en mini-høring hos leverandørerne, der ganske klart konkluderede, at indførelsen af TXT som ekstra afleveringsformat ville forøge udgifterne til produktion af arkiveringsversioner.

Konklusionen er derfor, at det ud fra et omkostnings- og effektiviseringsperspektiv vil være tilrådeligt at ændre så lidt som muligt på de eksisterende regler. I al fald på nuværende tidspunkt. Enhver ændring vil medføre implementeringsomkostninger hos såvel myndigheder og leverandører som Rigsarkivet, og eventuelt nye afleveringsformater skal derfor betyde en markant reduktion af konverteringsomkostningerne for at være relevante.

Formatet PDF/A vil muligvis kunne betyde en besparelse for myndighederne, hvis det tillades som afleveringsformat. PDF/A kan imidlertid ikke anvendes som bevaringsformat, så længe der ikke findes mulighed for en pålidelig validering af, at et PDF-dokument overholder arkivstandarden PDF/A. Rigsarkivet følger dog med interesse de aktuelle bestræbelser på at udvikle en metode til validering af PDF/A-filer, som sker i det EU-finansierede Performa-projekt. Projektet finansieres under rammeprogrammet FP7, hvor OpenPreservationFoundation, som Rigsarkivet er medlem af, deltager i samarbejde med firmaet veraPDF. Performa-projektet er undervejs med en open source validator, som, ud over at teste for overholdelse af selve PDF/A-standarden, tester for nogle af de mere basale og dermed mere udbredte fejltyper, som optræder i PDF som overordnet format.

Projektet søger at definere de mange regler, som skal overholdes, ved hjælp af eksempelfiler med tilhørende testprofiler. Validatorprogrammet skal altså kende (indlæse) en profil (regel) og teste reglen op mod den valgte PDF.

veraPDF har frigivet de første prototyper af Performa-valideringsværktøjet, men funktionaliteten er indtil videre begrænset til test af, hvorvidt et dokument er en valid PDF/A, og ikke hvorvidt den underliggende PDF ligeledes er i orden.

Det er på nuværende tidspunkt uvist, hvor omfattende en endelig validator vil kunne teste, og det er ikke muligt at sige noget om, hvor lang tid en fuldstændig test af et dokument vil komme til at tage. Rigsarkivets foreløbige tests af de prototyper, som er releaset, viser, at de performancemæssigt er langt fra produktionsklare i forhold til afleveringer med så mange dokumenter, som Rigsarkivet modtager.

Rigsarkivet vil dog naturligvis følge udviklingen og vil nøje vurdere, om formatet i de kommende år kan blive modent til bevaringsformål.

5 Løbende bevaringsplanlægning og periodiske migreringer til nyt bevaringsformat

5.1 Infrastrukturrapport

Hvor meget stiger vandstanden i havet de kommende år, og hvilke konsekvenser har det for behov for diger og kystsikring? Hvor mange børn skal daginstitutioner og skoler være dimensioneret til i 2025? Hvordan er holdbarheden i statens finanser fra 2025 og fremad?

Spørgsmål som disse er ofte til debat. Vi ved naturligvis godt, at dagens ansatte i stat, regioner og kommuner ikke er i stand til at forudsige fremtiden præcist. For at vi kan være så godt rustet som muligt til fremtidens udfordringer, er vi dog nødt til at forsøge at opstille de mest realistiske scenarier, baseret på dagens viden.

Det gælder også den digitale udvikling og de krav, den fremover vil stille til Rigsarkivets digitale infrastruktur, lagerkapacitet og organisation. Vi er nødt til at forsøge at vurdere fremtidens udfordringer, således at vi løbende kan tilpasse os hertil. Som et led i dette arbejde har Rigsarkivet udarbejdet en rapport, som beskriver de udfordringer, udviklingen medfører, i relation til hele infrastrukturen for modtagelse, test og bevaring.

Allerede nu ser vi, at de stigende datamængder sætter Rigsarkivets infrastruktur under pres. Baseret på den aktuelle digitaliseringsgrad i forvaltningen forventes det, at mængden af digitalt skabte data, som Rigsarkivet modtager, vil nærme sig 100 terabyte årligt, mod de ca. 25 terabyte, der er blevet modtaget årligt de sidste to år. Dette repræsenterer således en ganske voldsom stigning. Hertil kommer den omfattende digitalisering af papirarkivalier, som gennemføres internt, for at kunne stille arkivalierne til rådighed via internettet. I sagens natur betyder dette, at Rigsarkivets udgifter til digital lagring vil stige betragteligt år for år. Prognoserne tager ikke højde for en situation, hvor Rigsarkivets bevaringsområde eventuelt udvides til også at omfatte bevaring af større mængder overvågningsdata. I givet fald vil der skulle tilføres ekstra ressourcer til at etablere og vedligeholde en passende infrastruktur hertil.

Rapporten peger på nødvendigheden af, at der opretholdes specialistkompetencer på lagringsområdet internt i Rigsarkivet.

Forventningen er, at den nuværende grundlæggende infrastruktur for digital bevaring i Rigsarkivet kan skaleres gennem parallelisering og optimering, således at der nu og her ikke er behov for en grundlæggende omvæltning. Det er imidlertid også klart, at behovet for et større teknologiskifte er lige om hjørnet. Rigsarkivet skal derfor inden for det næste par år afsætte tilstrækkeligt med ressourcer til at gennemføre en grundig markedsanalyse af mulighederne for at skifte til en mere standardiseret lagringsløsning. En vigtig komponent i den forbindelse er de standardiserede modeller for lagringsstruktur, som udarbejdes i det fælleseuropæiske E-ARK-projekt, og som muligvis vil kunne understøtte en eventuel overgang til nye lagringssystemer. Dette falder godt i tråd med rapportens bud på strategiske retningslinjer for Rigsarkivets bevaringssystem, som bl.a. anbefaler et tæt samarbejde med andre bevaringsinstitutioner omkring teknologiovervågning samt anvendelse af standardiserede lagringsløsninger. Omkostninger til teknologiskifte og implementering i den forbindelse må dog ikke undervurderes.

Overordnet er målsætningen, at Rigsarkivets bevaringssystem også fremover skal være dimensioneret til omkostningseffektivt og sikkert at kunne modtage og bevare de datamængder, som Rigsarkivet skal håndtere. Bevaringen skal ske på en måde, som sikrer, at data kan hentes og genanvendes efter behov.

6 Tilgængeliggørelse af data i arkiveringsversioner

6.1 Myndighedernes adgang til egne arkiverede data

Rigsarkivet oplever en stadig stigende interesse fra myndighederne om at kunne tilgå data i arkiveringsversioner, som de selv eller deres forgængermyndigheder har afleveret til arkivet. Interessen skyldes, at myndighederne ofte holder "liv" i systemer, som er gået ud af daglig brug, men som der fortsat kan være behov for at slå op i fra tid til anden. Endvidere har en del myndigheder ønske om at kunne tilgå data fra netop afsluttede journalperioder hos arkivet. Rigsarkivets kan tilgængeliggøre arkiveringsversionerne af myndighedernes systemer med it-systemet Sofia. Sofia er på nuværende tidspunkt kun til rådighed på arkivets læsesale. I Sofia kan myndighederne slå op i arkiveringsversionerne (hvis der er følsomme persondata i arkiveringsversionen efter høring af Datatilsynet). Rigsarkivet kan tilbyde enkelte opslag i arkiveringsversioner som gratis ydelse; hvis behovet er større, tilbydes opslag som indtægtsdækket virksomhed.

I de kommende år skal det endvidere undersøges, hvilke muligheder/udfordringer den nye persondataforordning giver i forhold til myndighedernes adgang til "egne" data hos arkiverne.

7 Samarbejdsrelationer og internationale indtryk

7.1 Samarbejdsrelationer – OPF, DLM mv.

Rigsarkivet indgår i en række internationale samarbejdsrelationer, såvel formelle som uformelle. I de senere år har hovedvægten i Rigsarkivets internationale engagement ligget i EU-projektet E-ARK (se afsnit 3.3).

I Open Preservation Foundation (OPF) er Rigsarkivet i Danmark sammen med nationalarkiverne i Nederlandene og Estland i færd med at etablere en "Archives Interest Group", som skal identificere og arbejde videre med de problemstillinger, som er de væsentligste set ud fra et arkivperspektiv. Blandt de områder, som gruppen vil arbejde videre med, er spørgsmålet om, hvilke filformater der er egnede til langtidsbevaring, samt risikohåndtering i den digitale bevaring. Mange spørgsmål inden for digital bevaring er så komplekse, at det er af afgørende betydning, at der samarbejdes internationalt.

I DLM Forum, som er et europæisk forum for nationalarkiver, it-leverandører og konsulenter med fokus på informationsforvaltning i bred forstand, har arbejdet i E-ARK-projektet de senere år fået en væsentlig plads.

Det er efter Rigsarkivets vurdering en særdeles glædelig udvikling, eftersom projektet er med til at øge fokus på, at digital forvaltning ikke kun er et spørgsmål om at bruge informationen nu og her, men også om at sikre, at den bevaringsværdige digitale information kan bevares på langt sigt.

7.2 Samarbejdet i Det Nationale DM-Forum

Nationalt Forum for Forskningsdata Management (DM-Forum), som fremmer fag- og forskernære initiativer inden for forskningsdatamanagement, har i 2016 arbejdet på en lang række projekter, som skal understøtte universitetsforskernes håndtering af forskningsdata, fra de bliver skabt, til de skal bevares/genbruges eller slettes. Rigsarkivet deltager i forummet for at sikre godt dokumenterede data og viden om arkivering og genbrug af forskningsdata. Rigsarkivet indgår derfor i projekter om en Datamanagement Guide, som skal vejlede forskeren i håndtering af data gennem hele forskningsprocessen samt i et projekt om datamanagement undervisning for ph.d.-studerende på de humanistiske og sundhedsvidenskabelige fakulteter. Endelig deltager Rigsarkivet også i en gruppe, som har fokus på de tekniske aspekter af datahåndteringen, og i et netværk om de juridiske udfordringer på området. Arbejdet med de forskellige aktiviteter forventes færdiggjort i 2017.

[Læs mere om DM-Forum og aktiviteterne](#)

7.3 Nordisk ekspertseminar om bevaring og kassation

Nordisk Arkivakademi afholdt i september 2016 seminar om bevaring og kassation. Seminaret blev afholdt i Island og havde deltagere fra Finland, Island, Norge, Sverige og Danmark. Danmark var repræsenteret af fire arkivarer fra Rigsarkivets sølje for Indsamling og Bevaring.

Temaet for seminaret var bevaring og kassation. De enkelte lande præsenterede dels, hvad status er for arkivarbejdet, og dels hvad der sker inden for bevarings- og kassationsområdet. Norge og Finland står over for de største ændringer på arkivområdet. I Norge er man i gang med en større omorganisering, hvor man går fra Rigsarkiv med tilknyttede landsarkiver til en enhedsorganisation med opgaver på tværs af lokationer. Desuden er man i gang med at implementere nye kommunale bevarings- og kassationsbestemmelser samtidig med, at landet gennemfører en kommunalreform. I Finland arbejder man på en lignende omorganisering af arkivområdet. Desuden skal landet have ny arkivlov. I Finland har man hidtil udskrevet digitale arkivalier på papir og bevaret dem sådan. Nu vender man 180 grader og vil bevare digitalt. Det betyder, at man igangsætter et stort skanningsprojekt, hvor papirdokumenter skannes og bevares ad denne vej.

I Sverige nytænker man bevarings- og kassationsarbejdet, idet der er lavet forsøg med risikoanalyse som metode ved bevaringsvurdering af arkivalier. Indtil videre er risikoanalysen afprøvet på allerede gennemførte bevarings- og kassationsbestemmelser for at vurdere effekten af metoden.

I Island har arkivområdet været præget af en konflikt mellem regeringen og Nationalarkivet om tolkningen af arkivlovgivningen. Er en e-post et arkivalie, før det ligger i et dertil indrettet it-system? Sagen er ikke afsluttet, men jurister har erklæret sig enige med arkivet i, at der er tale om arkivalier.

Akademiet sluttede med en hensigtserklæring om at arbejde for, at bevaring og kassation bliver et emne på De Nordiske Arkivdage i Island i 2018.