



Strategi for arkivering af digitalt skabte arkivalier

Dette dokument beskriver Rigsarkivets strategi for modtagelse og bevaring af digitalt skabte arkivalier. Ved digitalt skabte arkivalier forstås upubliceret information, som er skabt eller tilvejebragt i digital form i forbindelse med en myndigheds, en organisations, en forskers eller en privatpersons virksomhed. Strategien har til formål at understøtte Rigsarkivets overordnede formål, som iht. arkivloven er at sikre bevaringen af arkivalier, der har historisk værdi eller tjener til dokumentation af forhold af væsentlig administrativ eller retlig betydning for borgere og myndigheder. Visionen er at sikre:

At digitalt skabte arkivalier, der har historisk værdi, bliver bevaret på en måde der sikrer, at den værdi som de repræsenterer, bliver realiseret både nu og i fremtiden

På den baggrund er de strategiske mål:

1. At bevare digitalt skabte arkivalier, så deres autenticitet opretholdes, og så de kan findes og genanvendes, både på kort og på langt sigt
2. At bevare digitalt skabte arkivalier på en måde, der er så omkostningseffektiv som muligt for både arkivskaberne og Statens Arkiver

Udfordringer

Digitalt skabte arkivalier har andre sårbarheder og byder på andre udfordringer end arkivalier, som er skabt i fysisk form. De medier, som data er lagret på, kan nedbrydes, og den teknologi, som er nødvendig for at læse medierne, kan blive forældet. På samme måde kan de formater, som den digitale information findes i, blive forældede, og viden om fortolkning af data samt deres kontekst kan gå tabt. Bevaring af digital information på langt sigt i en form, som gør, at den kan genanvendes, forudsætter bevidste valg i forhold til metoder, teknologier og dokumentation.

Digital bevaring må nødvendigvis ske under hensyntagen til samfundsøkonomien, herunder både myndighedernes omkostninger til produktion og aflevering af arkiveringsversioner og arkivets omkostninger til bevaring og tilgængeliggørelse.

Grundlæggende strategivalg

Arkiver, som modtager digitalt skabte arkivalier, kan vælge mellem at modtage og bevare det digitalt skabte materiale i originalformater (produktionsformater) og evt. migrere (normalisere) det efter behov i forbindelse med tilgængeliggørelse, eller migrere materialet til bevaringsformat enten før eller i forbindelse med modtagelse. Ved migrering i forbindelse med modtagelse er hensigten at sikre, at bevaringen sker i et format, der er mere egnet til langtidsbevaring end produktionsformatet.

Det grundlæggende strategivalg for bevaringsinstitutioner er, hvorvidt man vil følge en *emuleringsstrategi* eller en *migreringsstrategi*. Dette valg får afgørende betydning for, hvorledes man i øvrigt tilrettelægger sit arbejde med digital bevaring.

Emuleringsstrategien baserer sig på, at man til stadighed kan udvikle eksempelvis programmer og operativsystemer, som efterligner (emulerer) og fungerer som de programmer, der blev anvendt, da de digitale arkivalier blev skabt. På den måde kan man bevare den originale bitstrøm intakt i forventning om, at den til stadighed kan bevares og stilles til rådighed på alle fremtidige platforme, idet man kan afvikle gamle programmer på moderne maskiner.

Emuleringsstrategien er derfor tiltrækkende ud fra et autenticitetshensyn, idet man ikke foretager ændringer i det oprindelige digitale materiale. Den er imidlertid risikabel, fordi der endnu ikke er dokumentation for, at den vil kunne fungere i stor skala, dvs. som bevaringsmetode for samtlige typer af digitale dokumenter og databaser, som anvendes i det danske samfund generelt. Samtidig vil der være store omkostninger



forbundet med en emuleringsstrategi og metoden egner sig bedre til bevaring af andre typer af digitale materialer end dem, som Rigsarkivet modtager, f.eks. computerspil.

Rigsarkivet har derfor valgt at basere den digitale bevaring på en migreringsstrategi. Strategien indebærer, at arkivskaberne skal migrere digitalt skabte arkivalier til få, veldefinerede standardformater, som Rigsarkivet har udpeget til langtidsbevaring¹. Denne migrering kan være temmelig indgribende for materialet, og det er derfor bevaringsmæssigt en fordel at sikre, at denne sker så tæt på skabelsestidspunktet som muligt. Derved forøges muligheden for at finde hensigtsmæssige værktøjer til eksport og konvertering, ligesom det vil være muligt at kontrollere konverteringsresultatet mod data i det system, hvor de er skabt.

Migreringsstrategien betyder endvidere, at data, dokumenter og de strukturer, som anvendes i Rigsarkivets bevaring, med mellemrum må migreres til nye formater og strukturer. Enhver migrering medfører en risiko for forvanskning af data, og bør derfor kun ske, når det er nødvendigt ud fra en afvejning af omkostninger og risici ved at migrere sat over for omkostninger og risici ved at bevare data i det hidtidige format.

Systemuafhængig bevaring

Bevaring af data efter Rigsarkivets strategi skal ske systemuafhængigt. Det vil sige, at genanvendelsen og fortolkningen af data ikke må afhænge af, at man fortsat har adgang til det system, hvor data oprindeligt blev skabt. Data skal kunne fortolkes og genanvendes i andre systemer.

Systemuafhængig betyder i sagens natur ikke, at data i de formater og strukturer, vi i dag vil betegne som systemuafhængige, for tid og evighed kan fastholdes i netop denne form. Det betyder heller ikke, at man kan opfatte sine arkivalier som uafhængige af bestemte teknologier, som er bredt anvendt inden for en række af systemer. Informationsteknologien udvikler sig og før eller siden vil de digitalt skabte arkivalier, der er indsamlet i en bestemt længerevarende periode, opfattes som så afhængige af et bestemt teknologisk miljø, at det er nødvendigt at migrere data til et nyt bevaringsformat.

Genstanden for digital bevaring

I et traditionelt arkiv bevarer man arkivalierne i den form, de er skabt, dvs. at man fysisk bevarer det originale stykke papir el. lign., som rent faktisk blev skabt i den administrative proces. Når det gælder digital arkivering, giver det ikke på samme måde mening at tale om en "original". Uanset om der er tale om fx udtræk af data fra tabeller i en database, data fra en surveyundersøgelse eller digitale dokumenter, bevarer man en *repræsentation* af indholdet i bevaringsformat. Man er i den forbindelse nødt til at sikre, at der sker en vurdering af, hvilke data, der skal bevares; at der træffes beslutning om, hvilke formater, data skal findes i, samt hvilken dokumentation, der er nødvendig, for at data kan genanvendes.

Genstanden, som bevares i et digitalt arkiv, er således først og fremmest *data* eller *information*. I forbindelse med aflevering til arkiv er det Rigsarkivets opgave at sikre, at den relevante information, som er skabt i forbindelse med en bestemt proces, bliver bevaret på en måde, så den troværdigt repræsenterer den information, som arkivskaberen tilvejebragte og havde til rådighed. Det er bevaringen af autentisk information, der står i centrum, også selv om den til arkivformål må finde en anden teknisk repræsentation (dvs. migreres til arkivformater). Det er i den forbindelse væsentligt, at informationen er tilstrækkeligt dokumenteret til at kunne fortolkes og genanvendes uafhængigt af det system, hvor data er skabt.

Implementering

Den konkrete implementering af strategien i Rigsarkivet hviler på tre søjler²:

- Tidlig identifikation og godkendelse af systemer med henblik på aflevering af data og eventuelt digitale dokumenter

¹ Rigsarkivet migrerer selv visse private arkivskaberes data og digitale dokumenter.

² Denne implementering retter sig mod statslige myndigheder. For kommunale og regionale myndigheder såvel som private arkivskabere gælder samme bevaringsprincipper, men den konkrete implementering af bl.a. identifikation og afleveringsforløb sker med en række variationer



- Tidlige og hyppige afleveringer i systemuafhængigt format
- Løbende bevaringsplanlægning og periodiske migreringer til nyt bevaringsformat

Identifikation

Statslige myndigheder skal anmelde nye it-systemer til Rigsarkivet. I forbindelse med anmeldelsen foretager Rigsarkivet en bevarings- og kassationsvurdering, således at det ved systemernes ibrugtagning bliver klart, om data skal afleveres eller ej. Systemer med bevaringsværdige data skal godkendes, således at der er en sikkerhed for, at myndighederne på et tidligt tidspunkt i systemernes levetid har sikret sig, at data kan eksporteres til Rigsarkivets afleveringsformat.

Bevaringsvurderinger foretages af Rigsarkivets arkivarer. Der er ingen lovbestemte krav til, at bestemte typer af data skal afleveres til offentligt arkiv, og arkivarerne træffer ikke deres beslutninger på baggrund af faste regler, men en fagligt begrundet vurdering af datas værdi.

Ved vurderingen tages udgangspunkt i datas

- genanvendelighed, dvs. om data kan anvendes til at belyse konkrete problemstillinger i et historisk perspektiv. Det er ikke nødvendigvis det samme som, at data er troværdige eller retsgyldige.
- repræsentativitet, dvs. om data dækker en tilstrækkelig stor population, enten i tid eller i antal individer.
- singularitet, dvs. om data er unikke eller redundante i forhold til andre data.

Rigsarkivet udfører med mellemrum tilsyn med myndighedernes anvendelse af it-systemer, således at der findes et godt overblik over, hvilke systemer den offentlige forvaltning anvender og hvilke systemer, der indeholder bevaringsværdige data. Dette overblik er en nødvendig forudsætning for at planlægge de regelmæssige afleveringer af data.

På det kommunale og regionale område identificeres bevaringsværdige data og dokumenter gennem de tilsyn, som Rigsarkivet gennemfører med regelmæssige mellemrum.

På det private område sker identifikationen af bevaringsværdige data igennem samarbejdet med de arkivskabere, hvis arkivalier generelt anses for bevaringsværdige.

På forskningsområdet suppleres ovenstående med kontakt til offentlige og private fonde.

Tidlige og hyppige afleveringer

Offentlige data afleveres med jævne mellemrum, normalt hvert femte år. I modsætning til papirarkivalier, som først afleveres, efter at de er gået ud af administrativ brug, afleveres data hyppigt og uden skelnen til administrativ brug og status. Det vigtigste hensyn i forbindelse med digital bevaring er at sikre, at data og dokumenter bliver afleveret til arkiv, inden de bliver udsat for teknologisk forældelse og inden der sker sletning af visse data iht. anden lovgivning. Det er også nødvendigt for at sikre den nødvendige dokumentation, mens viden stadig er til stede hos den person eller organisation, som har skabt data. Man kan derfor ikke vente med aflevering af data til den administrative proces er afsluttet, og der ikke længere er brug for data hos myndigheden.

Denne tilgang betyder, at data på samme tid kan findes både i myndighedens administrative systemer og hos Rigsarkivet. Det betyder ligeledes, at data kan findes i flere kopier hos Rigsarkivet, idet afleveringer fra et system, der har eksisteret i mere end 5 år, ofte vil indeholde data, som også findes i tidligere afleveringer. Disse problemstillinger kan håndteres pragmatisk, men det understreger, at den traditionelle arkivistiks koncepter og modeller ikke altid kan anvendes uforandret i en digital verden.

Det tilstræbes, at data og eventuelt digitale dokumenter fra private arkivskabere også afleveres med jævne mellemrum, men i realiteten sker det typisk først, når det it-system, som de er skabt i, går ud af brug.



Data skabt eller tilvejebragt i forbindelse med et forskningsprojekt afleveres som udgangspunkt, når projektet er afsluttet.

Bevaring af data sker altid i en relationel databasestruktur, idet evt. dokumenter dog lagres i en filstruktur. En arkiveringsversion af en database kan bestå af

- et udtræk af samtlige data fra samtlige tabeller i databasen
- et udtræk, hvor data fra rent systemtekniske tabeller eller andre tabeller, der ikke har nogen værdi i en systemuafhængig kontekst, er udeladt.
- et bestemt udtræk af kerneoplysninger fra en database eller et løst koblet system

Findes den bevaringsværdige information hos arkivskaber ikke i en relationel database, men f.eks. i en hierarkisk mappestruktur, skal informationen omdannes til relationel databasestruktur i forbindelse med aflevering.

Data afleveres i systemuafhængige arkiveringsversioner, dvs. informationspakker, hvor data, dokumenter og dokumentation er samlet til et hele, der i så høj grad som muligt er selvdokumenterende.

Statslige myndigheder har pligt til at aflevere bevaringsværdige arkivalier til Rigsarkivet. Kommunale og regionale myndigheder kan som et alternativ vælge at oprette et eget offentligt arkiv. Den af Rigsarkivet fastsatte standarden for aflevering af digitale arkivalier er en national standard, som anvendes ved aflevering af alle offentlige digitale arkivalier, uanset om de afleveres til Rigsarkivet eller et andet offentligt arkiv.

Fysisk bevaring

De digitale arkivaliers bits og bytes skal bevares intakt. For at sikre arkivalierne så godt som muligt anvender Rigsarkivet en distribueret digital bevaring, hvor data findes i flere identiske eksemplarer på flere forskellige medietyper, både optiske og magnetiske, på flere forskellige geografiske placeringer.

OAIS-standard

I bevaringsinstitutioner verden over har OAIS-standard (Standard for an Open Archival Information System) stigende betydning som fælles referenceramme for bevaring af digitalt skabte arkivalier. Statens Arkivers strategi for digital langtidsbevaring er udviklet før OAIS-standard og er således som udgangspunkt uafhængig heraf. Standardens funktionelle entiteter er dog velegnede til at beskrive Statens Arkivers implementering af digital bevaring.

OAIS -standardens funktionelle model giver et overordnet billede af et data flow fra arkivskaber til arkiv og fra arkiv til bruger.

I OAIS-standard kaldes en arkiveringsversion af data, som afleveres til arkiv, for en SIP (*Submission Information Package*). I bevaringssammenhæng kaldes arkiveringsversionen for *bevaringsversion* eller AIP (*Archival Information Package*) og til udlevering til arkivets brugere anvendes en *tilgængeliggørelsesversion* eller DIP (*Dissemination Information Package*).

Arkiveringsversionen skal indeholde den information, som er nødvendig for at kunne fortolke og formidle indholdet. For at sikre, at data til stadighed er læsbare, herunder at der sker de nødvendige migreringer til rette tid, skal et digitalt arkiv have en løbende bevaringsplanlægning.

Bevaringsplanlægning

I overensstemmelse med OAIS-standard udfører Rigsarkivet løbende en bevaringsplanlægning, som bl.a. definerer, hvornår skift af bevaringsstandard skal ske, samt hvornår der skal ske en migrering til nye formater.

Skift af bevaringsstandard skal som udgangspunkt ske så sjældent som muligt. Samtidig er det dog væsentligt, at standarden er tilpasset de aktuelle behov hos både arkivskabere og arkiv. Behov for revision af bevaringsstandard kan bl.a. opstå, når arkivskaberne begynder at skabe og opbevare bevaringsværdig



digital information i formater, der ikke tilfredsstillende kan migreres til de hidtidige bevaringsformater. Behovet kan også opstå ud fra skift i teknologier til processering af data, således at aflevering til stadighed kan ske så effektivt som muligt og basere sig på det aktuelle teknologiske stade.

Det tilstræbes, at migrering af eksisterende samling til nyt bevaringsformat sker inden for en kortere årrække efter skift af bevaringsformat, således at data migreres, mens organisationens viden om og praktiske erfaring med formater og tilhørende programmer fortsat er intakt. Det er Rigsarkivets erfaring, at det er omkostningskrævende at vedligeholde viden og systemer til håndtering af bevaringsformater, som ikke længere er i daglig brug til modtagelse og test. Der vil derfor være en stigende risiko for at bringe datas integritet i fare, hvis de opbevares i formater, der ikke længere aktivt anvendes af organisationen. Af hensyn til bevaringen af alle offentlige arkivalier, vil Rigsarkivet ved skift af bevaringsformat aktivt fremme migreringen af de arkiveringsversioner, som er afleveret til andre offentlige arkiver.

En afgørende forudsætning for bevaringsplanlægningen er, at Rigsarkivet løbende opsamlar viden om, hvordan offentlige myndigheder og andre arkivskabere udnytter it-teknologien til at fastholde dokumentation af deres virksomhed, samt hvilke bevaringsløsninger andre bevaringsinstitutioner og andre udvikler. På baggrund heraf justerer Rigsarkivet til stadighed implementeringen af sin strategi, så visionen forbliver indenfor rækkevidde.