

Vedlegg 1A:

Bevaring av frittstående elektroniske dokumenter

Innhold

1	HVA MENES MED FRITTSTÅENDE ELEKTRONISKE DOKUMENTER.....	3
2	UTFORDRINGER KNYTTET TIL FRITTSTÅENDE ELEKTRONISKE DOKUMENTER	3
2.1	DATAMATERIALE FRA FORELDEDE ELLER UKJENTE SYSTEMER.....	3
2.2	ET DOKUMENT KAN FINNES BÅDE PÅ PAPIR OG ELEKTRONISK	3
2.3	MANGLENDE KONTEKST OG METADATA	3
2.4	SKAL EN BEVARINGSINSTITUSJON TA VARE PÅ DEN ORIGINALE KATALOG- OG FILSTRUKTUREN?.....	4
2.5	KAN BEVARINGSINSTITUSJONEN SLETTE ALLE FILER AV EN BESTEMT TYPE?	4
2.6	HVA AVGJØR OMFANGET OG KOMPLEKSITETEN TIL ET ELEKTRONISK ARKIV?	4
3	SKAL EN BEVARINGSINSTITUSJON TA VARE PÅ DE ELEKTRONISKE DOKUMENTENE?	4
4	HVA BØR KREVES AV ARKIVSKAPER/DEN SOM AVLEVERER MATERIALET?	5
4.1	RYDDE OPP I KATALOG- OG FILSTRUKTUR	5
4.2	KONTEKST OG METADATA	5
4.3	AVKLARE OM INFORMASJONEN FINNES PÅ FLERE INFORMASJONSBÆRERE	5
5	FREMANGSMÅTE VED BEVARING.....	6
5.1	MOTTAK	6
5.2	ANALYSE AV ARKIVMATERIALET.....	6
5.3	ARKIVFAGLIG VURDERING.....	6
5.4	KONVERTERING TIL FORMATER FOR LANGTIDSBEVARING	6
5.4.1	<i>Database.....</i>	<i>7</i>
5.4.2	<i>Regneark.....</i>	<i>7</i>
5.4.3	<i>Formaterte tekstdokumenter.....</i>	<i>7</i>
5.4.4	<i>Bildefiler.....</i>	<i>8</i>
5.4.5	<i>Websider.....</i>	<i>8</i>
5.5	STANDARD FOR FILSTRUKTUR	8
6	HVA KREVES FOR Å BEVARE ELEKTRONISKE DOKUMENTER?	9
6.1	TEKNOLOGISKE BETINGELSER	9
6.1.1	<i>Maskinvare</i>	<i>9</i>
6.1.2	<i>Verktøy for analyse av katalog- og filstruktur</i>	<i>9</i>
6.1.3	<i>Programvare for å tolke dokumenter</i>	<i>9</i>
6.1.4	<i>Programvare for å konvertere dokumenter</i>	<i>10</i>
6.1.5	<i>Verktøy for produksjon av metadata</i>	<i>10</i>
6.2	KOMPETANSE	10
6.3	TIDSFORBRUK.....	11

1 Hva menes med frittstående elektroniske dokumenter

Datafiler er skapt av ulike typer programvare, og opptrer som regel i et bestemt filformat. En tekstfil er en sekvens av tegn, ofte organisert i separate linjer med linjeskift. Binære filer blir ofte brukt som betegnelse på alle andre filtyper. I dette dokumentet vil vi diskutere hvordan man kan håndtere informasjon som lagres som frittstående filer, og som ikke er tilknyttet registerdata. Dette kan være tekstfiler, dokumenter, bildefiler og små databaser. Slikt materiale vil være lagret i en mer eller mindre logisk katalog- og filstruktur. Denne typen materiale er ofte uorganisert og kan høre logisk sammen med tilsvarende dokumenter som finnes på papir. I motsetning til registermateriale vil datafiler i en katalog- og filstruktur ikke nødvendigvis kunne omtales som et elektronisk arkiv, men som arkivdokumenter lagret på elektroniske medium.

En katalog- og filstruktur er en metode for å lagre og organisere datafiler slik at det er enkelt å finne og aksessere dem. Strukturen vil ofte være hierarkisk. En katalog- og filstruktur finnes vanligvis på en lagringsenhet, som for eksempel en harddisk eller en CD-ROM. Ulike operativsystemer bruker ulike metoder for å organisere datafiler, men dette dokumentet vil ikke gi en innføring i de ulike metodene. Utgangspunktet er at man har fått tilgang til datafilene, og at disse er tilgjengelige i en katalog- og filstruktur.

2 utfordringer knyttet til frittstående elektroniske dokumenter

2.1 Datamateriale fra foreldede eller ukjente systemer

I mange tilfeller vil bevaringsinstitusjoner stå overfor elektronisk materiale fra systemer som er ukjente eller som ikke lenger er i bruk. Dette kan være fra materiale fra systemer som ikke er tilgjengelig for den som mottar materialet, materiale fra systemer som ikke lenger finnes eller materiale som ikke kan tolkes i nyere versjoner av programvaren som materialet i utgangspunktet ble skapt i. The National Archives i Storbritannia har utviklet et informasjonssystem som kan brukes for å finne informasjon om filformat og hvilken programvare som støtter de ulike formatene. Informasjonssystemet heter PRONOM og er tilgjengelig på nettstedet til The National Archives¹.

2.2 Et dokument kan finnes både på papir og elektronisk

Når bevaringsinstitusjoner mottar et arkiv fra en privat arkivskaper mottar de vanligvis en del arkivmateriale på papir, i tillegg til at de etter hvert får overlevert elektronisk datamateriale på en harddisk, CD eller et annet lagringsmedium. I mange tilfeller kan det samme arkivdokumentet forekomme både elektronisk og på papir. En bevaringsinstitusjon kan også oppleve at dokumentene foreligger i ulike versjoner, noe som kan tyde på at en eller flere av versjonene er utkast, eller for eksempel at versjonene av et brev har blitt endret etter at det ble sendt. utfordringen for bevaringsinstitusjonen er å avgjøre om det elektroniske eller papirbaserte materiale som skal tas vare på, eventuelt begge deler. Se for øvrig neste punkt.

2.3 Manglende kontekst og metadata

Elektroniske dokumenter som ligger lagret i en filstruktur vil svært ofte mangle metadata. Med dette menes informasjon om dokumentets tilblivelse og hvem det henvender seg til, informasjon om mottak, behandling og ekspedering av dokumentet, og informasjon om forholdet mellom dokumenter. Dette er metadata som i mange tilfeller kan avgjøre om elektroniske dokumenter har dokumentasjonsverdi.

¹ <http://www.nationalarchives.gov.uk/pronom/>

2.4 Skal en bevaringsinstitusjon ta vare på den originale katalog- og filstrukturen?

Prinsippet om arkivets indre proveniens innebærer at arkivets indre orden skal bevares, men dette betyr ikke at den fysiske rekkefølgen mellom dokumentene skal beholdes. Arkivkatalogen vil representere den logiske ordenen til arkivet, uavhengig av fysisk organisering og lagringsmedium.

En katalog- og filstruktur kan være opprettet som en del av en bevisst prosess, enten arkivskaperen har brukt arkivnøkkel eller andre prinsipper for å organisere det elektroniske materialet. Dersom strukturen er bygget opp på denne måten bør bevaringsinstitusjoner etterstrebe å bevare den i sin opprinnelige form. I andre tilfeller kan en katalog- og filstruktur ha oppstått som resultat av en rekke tilfeldige valg. Arkivskaper kan ha opprettet filer og kataloger med navn og plassering som virket logisk der og da, men som er vanskelig å bruke for senere gjenfinning, spesielt for utenforstående. I slike tilfeller kan det være nyttig å bygge opp en ny og ryddig struktur etter samme ordningsprinsipper som for ordning av et papirarkiv. Samtidig bør bevaringsinstitusjoner være oppmerksom på at den opprinnelige strukturen i seg selv kan dokumentere hvordan arkivskaperen organiserte arbeidet sitt, og dermed ha en dokumentasjonsverdi i seg selv. En oversikt over den opprinnelige katalog- og filstrukturen kan i slike tilfeller legges ved det elektroniske materialet.

2.5 Kan bevaringsinstitusjonen slette alle filer av en bestemt type?

Når bevaringsinstitusjoner skal langtidsbevare elektroniske dokumenter bør man bare bevare dokumenter på de formater som er aksepterte av Riksarkivaren. Mange filtyper vil med enkle grep kunne konverteres til disse formatene, og vil være bevaringsverdige som arkivdokumenter. Samtidig vil en del filtyper være umulige å konvertere til aksepterte formater, og de vil heller ikke alltid være bevaringsverdige. Et eksempel på slike filer er programfiler og andre filer som brukes av operativsystemet. Det er allikevel vanskelig å anbefale at alle filer av en bestemt type bør slettes. Det kan for eksempel være programfiler som genererer bevaringsverdige dokumenter, og det bør derfor brukes skjønn når filer skal slettes.

2.6 Hva avgjør omfanget og kompleksiteten til et elektronisk arkiv?

I motsetning til med et papirarkiv er det lite hensiktsmessig å regne omfanget av elektronisk arkivmateriale i hyllemeter. Ved bevaringsvurdering av elektronisk arkivmateriale som finnes i en katalog- og filstruktur må man ta hensyn til følgende:

- Antall filer som må konverteres eller manipuleres, og hvor mye tid og arbeid dette vil kreve.
- Et stort antall filer i en komplisert katalog- og filstruktur vil gjøre det vanskelig for en fremtidig bruker av arkivet å finne fram til informasjonen vedkommende er ute etter. Man bør derfor slette kataloger og filer som er tomme, samt kopier eller filer som åpenbart ikke er arkivverdige.

3 Skal en bevaringsinstitusjon ta vare på de elektroniske dokumentene?

I enkelte tilfeller må bevaringsinstitusjoner ta stilling til om det elektroniske materialet skal langtidsbevares, eller om det er tilstrekkelig å ta vare på materialet på papir. Bevaringsinstitusjonen må vurdere nytten av å langtidsbevare det elektroniske materialet opp mot ressursene som kreves for å utføre oppgaven. Selv elektroniske arkiver som kun inneholder én fil vil kreve at materialet overføres til en lagringsenhet i henhold til gjeldende krav, at informasjon om det elektroniske materialet registreres og at materialet plasseres i depot. I tillegg må man vedlikeholde det elektroniske materialet og oppdatere informasjonen om materialet som finnes i de ulike informasjonssystemene.

I enkelte tilfeller vil bevaringsinstitusjoner motta arkiver der det elektroniske materialet kun utgjør noen få dokumenter. I slike tilfeller vil det være naturlig å vurdere om man i det hele tatt skal ta vare på dette materialet i elektronisk form eller om det er tilstrekkelig å ta vare på dokumentene som utskrifter i papirarkivet. Dokumenter kan skrives ut på papir uten at omfanget blir betydelig, og i ordningsarbeidet vil det være enkelt å foreta en slik handling. Likevel vil informasjonen i ett eller flere store elektroniske dokumenter kunne ha større gjenbruksverdi enn dokumenter lagret på papir. Elektroniske dokumenter vil være søkbare og bedre egnet til formidling. Større manuskripter og lignende kan på bakgrunn av dette være eksempler på dokumenter hvor det kan argumenteres for å ta vare på informasjonen i elektronisk form fremfor papir hvis det er mulig. Større filer, eller et stort antall filer, kan tale for elektronisk lagring, mens mindre eller færre filer taler for at informasjon skrives ut på papir.

I denne sammenheng forutsetter vi at materialet er vurdert som bevaringsverdig. Vi argumenterer for at det også bør ligge en arkivfaglig vurdering til grunn for valg av type medium som arkivmaterialet lagres på. De elektroniske versjonenes verdi i forbindelse med gjenbruk og formidling bør tas i betraktning. I en bevaringsinstitusjon bør det derfor eksistere rutiner og metoder som ivaretar en enkel og sikker bevaring av frittstående elektroniske dokumenter.

4 Hva bør kreves av arkivskaper/den som avleverer materialet?

Jo mer arkivskaper kan bidra når det elektroniske materialet skal ordnes, jo bedre kan resultatet av ordningen bli. Dersom det er mulig bør arkivskaper bidra i dette arbeidet ved å avlevere en ryddig og godt beskrevet katalog- og filstruktur.

4.1 Rydde opp i katalog- og filstruktur

En katalog- og filstruktur som ikke er konstruert på en planmessig måte bør ryddes for å gjøre det lettere for utenforstående å navigere i strukturen. Arkivskaper eller andre med kjennskap til det elektroniske materialet bør:

- Slette åpenbart arkivuverdige filer og kataloger, og eventuelle kopier.
- Vurdere om navngiving av filer og kataloger er hensiktsmessig for utenforstående.
- Vurdere om kataloger skal slås sammen, slettes eller deles opp, om filer skal flyttes til andre kataloger og lignende.

4.2 Kontekst og metadata

Som nevnt i kapittel 2.3 vil ofte elektroniske dokumenter lagret i en katalog- og filstruktur mangle metadata og kontekstinformasjon. En del slik informasjon kan fremskaffes ved hjelp av arkivskaper eller andre med kjennskap til materialet. Ved avlevering bør bevaringsinstitusjonen skaffe til veie en oversikt over hvor filstrukturen kommer fra, om det er en enkeltstående arbeidsstasjon eller en nettverksstasjon med flere brukere, hvordan filstrukturen er bygd opp og lignende.

4.3 Avklare om informasjonen finnes på flere informasjonsbærere

Dersom det er mulig bør det komme frem hvorvidt hele eller deler av det elektroniske arkivmaterialet også foreligger på andre informasjonsbærere, som for eksempel papir.

5 Fremgangsmåte ved bevaring

5.1 Mottak

Det første en bevaringsinstitusjon bør gjøre er å foreta en virussjekk og eventuelt fjerne infiserte filer. Dersom det er mulig bør man få arkivskaper til å lage en ny og virusfri avlevering.

Når bevaringsinstitusjonen er sikker på at arkivmediet ikke inneholder virus må innholdet kopieres over på et eget arbeidsområde på den ansvarlige ordningsarbeiderens pc. Det bør deretter lages en sikkerhetskopi før arbeidet med arkivmaterialet begynner.

5.2 Analyse av arkivmaterialet

1. Inspisere det elektroniske arkivmaterialet

Når bevaringsinstitusjonen har kopiert det elektroniske arkivmaterialet over på et arbeidsområde vil det være nyttig å få en oversikt over hva som finnes i arkivet. I enkle tilfeller kan filutforskeren på egen pc benyttes, men i mange tilfeller vil det være nødvendig å bruke programvare som er utviklet for å analysere og beskrive innholdet i filstrukturen. Det finnes verktøy som kan inspisere et filområde, og resultatet av inspeksjonen kan eksporteres over i et regneark for grundigere analyse.

2. Få en grov oversikt

Dersom det elektroniske arkivet består av mange kataloger på flere ulike nivåer bør det skrives ut en oversikt over katalogene, og sammenhengen mellom dem. Videre bør en bevaringsinstitusjon få en oversikt over antall filer, og bruke et verktøy for å finne antall filer av de ulike filtypene. Dersom det er filer som åpenbart mangler eller har feil suffiks (tre-tegns fileternavn) bør dette endres. Det kan være at arkivskaper utilsiktet har endret suffikset. Etter å ha gått gjennom listen med ulike filtyper og endret manglende eller feil suffiks bør man gå tilbake til trinn 1 i denne beskrivelsen og foreta en ny inspeksjon av det elektroniske arkivet.

3. Slette åpenbart arkivverdige filer og kataloger

Mange filtyper kan som regel slettes uten at det elektroniske arkivet mister sin informasjonsverdi. Eksempel på slike filtyper vil være *.bat, *.bin, *.dll, *.exe, andre systemfiler og midlertidige filer. Men i enkelte tilfeller kan slike filer inneholde funksjoner som genererer arkivverdige dokumenter.

5.3 Arkivfaglig vurdering

Trinn 1-3 kan utføres av elark-konsulent. Etter at denne har fått en oversikt over arkivet og slettet åpenbart arkivverdige filer og kataloger bør en arkivar foreta arkivfaglige vurderinger. Det kan være tilfeller der det elektroniske materialet foreligger på papir, og da må det foretas en bevaringsvurdering av det elektroniske materialet. I samarbeid bør arkivar og elark-konsulent rydde opp i filstrukturen.

5.4 Konvertering til formater for langtidsbevaring

Ifølge retningslinjene for privatarkiver skal en bevaringsinstitusjon sikre at arbeidet med bevaring av privatarkiver utføres planmessig og systematisk og følger allment godtatt arkivfaglige prinsipper og rutiner. Dette innebærer at avleveringsbestemmelsenes² krav til formater for langtidsbevaring bør følges.

² FOR 1999-12-01 nr 1566: Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver.

Dokumentformatene som aksepteres er:

- Ren tekst – ISO 8859-1: 1998, Latin-1 og ISO 8859-4: 1998, Latin-4 for samiske tegn
- TIFF – Tag Image File Format, versjon 6
- PNG – Portable Network Graphics, versjon 1.0
- XML – Extensible Markup Language, versjon 1.0
- PDF – Portable Document Format.

For fotografier, lyd- og video-sekvenser er følgende formater godkjent:

- For digitale fotografier samt for store digitale bilder og kart: JPEG (ISO 10918-1).
- For digital lyd: MP3 (ISO 11172-3).
- For digital video: MPEG2 (ISO 13818-2).

5.4.1 Database

En database kan ikke bevares i sin opprinnelige form. For å tilfredsstille Arkivverket sine krav for langtidsbevaring finnes to alternativer:

- Uttrekk til flate, sekvensielle tekstfiler. Ved uttrekk til flate, sekvensielle tekstfiler er det nødvendig å beskrive databasens tekniske metadata i ADDMML.
- Utdrag i form av rapporter, konvertert til PDF. Dette kan blant annet gjøres ved å eksportere databasen inn i et regneark. Se for øvrig problemer knyttet til regneark i kapittel 5.4.2

Hvilke av de to alternativene som bør velges avhenger av databasens størrelse og kompleksitet, samt hvilken gjenbruksverdi materialet har. Dersom det genereres rapporter og disse konverteres til PDF-dokumenter, vil det ikke lenger være mulig å rekonstruere databasen for å bruke datamaterialet til søk i bestemte felt og lignende. Dersom en database er lite kompleks, og inneholder lite informasjon kan det være tilstrekkelig å ta ut rapporter og bevare disse. En database bestående av en tabell med under tusen poster, der datamaterialet er blitt brukt til å skrive en rapport hvor dataene blir sammenfattet, vil typisk være gjenstand for konvertering til PDF. Materialet vil da bli mindre fleksibelt, men arbeidsmengden ved å generere rapporter vil sannsynligvis være mindre enn ved å trekke ut data til flate sekvensielle tekstfiler og beskrive tilhørende metadata. Det er ikke hensiktsmessig å komme med konkrete anbefalinger, og de to alternativene må vurderes i hvert enkelt tilfelle.

5.4.2 Regneark

Regneark er i likhet med databaser data organisert i tabeller. Videre kan regneark bestå av grafiske elementer som for eksempel diagrammer. I likhet med databaser kan data fra regneark tas ut som flate sekvensielle tekstfiler, eller konverteres til et PDF-dokument. Dersom regnearket konverteres til PDF-dokument må man være oppmerksom på at det redigeres på samme måte som når det blir tatt en utskrift på papir, dvs. at det settes hensiktsmessige kolonnebredder, radhøyder, sidebredde og -høyde, samt at tabelloverskrifter og sidetall gjør datamaterialet leselig. Det kan ikke settes i gang et automatisk konvertering til PDF uten en oversikt over hvordan regnearkene ser ut, og hvordan resultatet av konverteringen vil bli. Videre må bevaringsinstitusjonen vurdere om det er ønskelig å bevare data i form av resultater, formler eller begge deler. I enkelte tilfeller kan et regneark beregne en mengde verdier, beregnet på grunnlag av en bestemt modell eller formel. I slike tilfeller kan det være selve modellen, mer enn verdiene som er lagt inn, som dokumenterer grunnlaget for en saksbehandlings- eller beslutningsprosess.

5.4.3 Formaterte tekstdokumenter

Formaterte tekstdokumenter vil vanligvis være produsert i et eller annet tekstbehandlingsverktøy. For å beholde utseendet til slike dokumenter bør de konverteres til PDF.

5.4.4 Bildefiler

Bildefiler blir vanligvis langtidsbevart som TIFF, JPEG eller PNG. Valg av filtype avhenger av hvilke behov og muligheter som ligger til grunn.

TIFF

Tagged Image File Format (TIFF) er et utbredt filformat for lagring av bitmap-grafikk. TIFF-bilder kan være i sort/hvitt, gråtoner eller farger, og de kan ha en hvilken som helst oppløsning. Formatet er "lossless" (tapsfritt) og fører ikke til tap av data. TIFF er derfor godt egnet til trykking på papir og til nærmere granskning av detaljer via skjerm. Dersom en fil i utgangspunktet er TIFF bør den ikke konverteres til annet format for langtidsbevaring. TIFF støtter multiple page, dvs. at en fil kan inneholde flere sider med bilder, noe som gjør det mulig å lagre mangesidige elektroniske dokumenter som en enkelt TIFF-fil.

TIFF har med tiden blitt tilført proprietære tagger, informasjon som støtter bestemte typer funksjonalitet for bestemte typer programmer. Det har ført til at programmer i enkelte tilfeller kan ha problemer med noen typer TIFF-filer. TIFF-filer skal ved langtidsbevaring ikke inneholde egendefinerte tagger.

JPEG

Joint Photographic Experts Group (JPEG) er et filformat laget for å komprimere fotografiske bilder, som f.eks. malerier, fotografier, teksturer etc. Formatet gir svært bra kompresjon (10:1-100:1) og gir langt mindre lagringskostnader enn TIFF, men bildefilene blir kvalitativt dårligere. Komprimeringen som brukes i JPEG er optimalisert for fotografier eller bilder med et stort antall farger. Dette vil ofte føre til at ikoner og detaljerte småbilder med store kontraster i mange tilfeller vil se bedre ut hvis de lagres som PNG.

PNG

Portable Network Graphics (PNG) gir, i likhet med JPEG, komprimerte filer, men dette foregår uten tap av data.

5.4.5 Websider

HTML er ikke lenger et godkjent langtidslagringsformat. For å bevare websider må de enten lagres som XML-dokumenter, eller sidene kan konverteres til PDF. Dersom en PDF-skriver blir benyttet til å konvertere websider til PDF, må man være oppmerksom på at sider med en dårlig grafisk layout vil kunne bli gjengitt på en måte som ikke samsvarer med det som vises på skjerm. Dette gjelder sider som består av rammer (frames) eller av tabeller som er for brede for en vanlig utskrift. I slike tilfeller kan det være et alternativ å endre til liggende retning på utskriften til PDF.

5.5 Standard for filstruktur

I følge avleveringsbestemmelsene skal elektroniske dokumenter organiseres i filer og kataloger etter ISO 9660 Level 1. Alternativt kan fil- og katalogstruktur følge Microsofts Joliet-standard, som aksepteres som en utvidet ISO 9660-spesifikasjon i Noark-4 for å muliggjøre håndtering av fil- og katalognavn med større lengde enn 8 tegn.

ISO 9660 Level 1 er et hierarkisk filformat på samme måte som MS-DOS. Det bruker det samme 8.3 systemet for filnavn med opptil 8 bokstaver i filnavnet, et punktum, og opptil 3 bokstaver etter punktum. Det støttes bare 8 nivåer i hierarkiet, og godkjente tegn er de engelske bokstavene fra A-Z, tallene 0-9 og understrek "_".

Joliet er Microsofts utvidelse av ISO 9660. Her er alle Unicode-tegn lovlig og filnavn kan være på 64 tegns lengde. De færreste operativsystemer vil ha tilgang til de lange filnavnene, men en CD-R der filene er organisert etter Joliet filstrukturformat er lesbar fra alle ISO 9660-kompatible operativsystemer fordi et ISO 9660-formatert filnavn vil bli lagret. For eksempel vil filnavnet *langt filnavn.txt* være *LANGTF~1.TXT* som ISO 9660. I utgangspunktet anbefales valg av ISO 9660 Level 1 som filstrukturingsformat, da dette vil være lesbart på flest operativsystemer. Men av praktiske grunner vil det bli svært vanskelig å følge opp dette. Dersom et avlevert elektronisk arkiv inneholder et stort antall filer vil det ikke være praktisk mulig å endre navn på samtlige filer. En løsning vil være å bruke Joliet som filstrukturingsformat for elektroniske arkiv. Ved hjelp av et dertil egnet verktøy er det mulig å lage en oversikt som viser filnavnene og de tilhørende ISO 9660-formaterte filnavnene. Å benytte Joliet kan være nyttig dersom brukeren av arkivet bare kan lese de korte navnene.

6 Hva kreves for å bevare elektroniske dokumenter?

6.1 Teknologiske betingelser

6.1.1 Maskinvare

En bevaringsinstitusjon må ha maskinvare som kan få tilgang til informasjon lagret på elektroniske medier. I de fleste tilfellene er det tilstrekkelig med en ordinær pc med 3,5" diskettstasjon og CD-leser, men det kan i enkelte tilfeller være nødvendig med annen maskinvare.

6.1.2 Verktøy for analyse av katalog- og filstruktur

Inspector F

Inspector F er et verktøy som kan brukes til å få oversikt over alle filer og kataloger som befinner seg på et gitt område på en lagringsenhet. Resultatet kan sees i programmet eller eksporteres til tekstdokument, regneark, database eller XML for videre bearbeiding.

I prosjektet ble Inspector F 1.1 brukt til å:

- Generere en oversikt over katalogene, og den hierarkiske strukturen på disse.
- Generere en total oversikt over alle filer, med filnavn, plassering og filtype. Dette ble eksportert til Microsoft Excel for videre bearbeiding.

Inspector F er lisensiert, men det gis mulighet for 90 dagers gratis test av programmet. Inspector F 1.2 skal gi bedre muligheter for analyse av materialet enn det som er skissert i dette dokumentet. Programmet er tilgjengelig på produsenten Basu Technology Inc's nettsted³.

Regneark

Regneark kan benyttes til å filtrere og sortere datamaterialet for å finne frem til hvor mange filer det er av hver ulik filtype, hvor filene er plassert og lignende.

Søkefunksjonen i operativsystemet

Søkefunksjonen i operativsystemet er svært nyttig for å finne filer som kan slettes, filer som må få nytt suffiks og lignende.

6.1.3 Programvare for å tolke dokumenter

En viktig forutsetning for å kunne konvertere dokumenter er at bevaringsinstitusjonen har tilgang til verktøy som er i stand til å få tilgang til og tolke den digitale representasjonen på det

³ <http://www.basutech.com/>

elektroniske mediet. Selv om bevaringsinstitusjonen ikke har umiddelbar tilgang til slike verktøyer finnes løsninger som gir et tilfredsstillende resultat. Internett gir tilgang til en rekke verktøyer med begrenset funksjonalitet, men som kan fungere tilfredsstillende som leseverktøy.

6.1.4 Programvare for å konvertere dokumenter

Konvertering av dokumenter vil ofte kreve en god del improvisasjon, og ved å bruke ulike verktøyer vil det være mulig å oppnå ønsket resultat.

Adobe Acrobat

Adobe Acrobat er en programvare som er utviklet av Adobe Systems Inc. Med Adobe Acrobat Standard er det mulig å konvertere alle elektroniske dokumenter til PDF-filer for langtidsbevaring. Adobe Acrobat er lisensiert og kan kjøpes via leverandørens nettsted⁴.

OpenOffice

OpenOffice er en gratis kontorpakke utviklet ved hjelp av åpen kildekode, som inneholder programvare for tekstbehandling, regneark, presentasjoner, tegning og lignende. OpenOffice tilbyr funksjonalitet for å konvertere dokumenter til PDF. Programvaren kan ikke konvertere flere dokumenter i samme operasjon, men det er utviklet makroer som utfører dette. OpenOffice⁵ og tilhørende makroer⁶ kan fritt lastes ned fra Internett.

PDF-skriver

Det finnes flere enkle programvarer for konvertering til PDF. Disse blir installert som skrivere, og gir mulighet for konvertering til PDF gjennom leseverktøyenes utskriftsfunksjonalitet. Ved konvertering fra HTML til PDF kan det være nødvendig å bruke denne typen verktøy. Et eksempel på slikt verktøy er CutePDF Writer som kan lastes ned fra leverandørens nettsted⁷.

6.1.5 Verktøy for produksjon av metadata

Arkadukt

Arkadukt er et åpent tilgjengelig verktøy som er utviklet og vedlikeholdt av Riksarkivet, og blir brukt til å beskrive en databases tekniske metadata. Verktøyet genererer en komplett strukturbeskrivelse som følger ADDMML (Archives Data Description and Manipulation Markup Language), et markeringsspråk basert på XML. ADDMML.DTD er spesifisert i avleveringsbestemmelsene.

Teksteditor

Siden strukturbeskrivelser som følger ADDMML.DTD i utgangspunktet er uformaterte tekstfiler er det tilstrekkelig å benytte en hvilken som helst teksteditor til å lage en slik beskrivelse. En slik fremgangsmåte vil imidlertid være arbeidskrevende. I tillegg er XML av en slik natur at kun én feil i forhold til ADDMML gjør beskrivelsen ugyldig.

6.2 Kompetanse

Hvilken kompetanse som kreves i arbeidet med å bevare elektroniske dokumenter vil kunne variere fra arkiv til arkiv. Den største utfordringen vil som regel være at disponibel maskin- og/eller programvare ikke alltid gir tilgang til materialet. I slike tilfeller vil det være en utfordring å finne ut hva slags verktøy som kreves, og om disse skal tas i bruk. Videre vil arbeidet kreve at vedkommende som arbeider med materialet har forutsetninger for å bruke verktøyet, enten ved at

⁴ <http://www.adobe.no/>

⁵ <http://no.openoffice.org>

⁶ Prøv å søke etter en eller flere av ordene *openoffice batch converter macro* i en søkemotor.

⁷ <http://www.cutepdf.com/>

vedkommende har kjennskap til virkemåte eller evner til å lære seg å bruke det. Slike avvikende situasjoner vil som regel kreve en stor grad av improvisasjon og derfor en del IT-kompetanse.

6.3 Tidsforbruk

Det er ingen enkel oppgave å estimere tiden som kreves for å klargjøre elektroniske dokumenter lagret i katalog- og filstruktur for langtidsbevaring.

- Å konvertere et ryddig og enkelt arkiv kan utføres relativt raskt, mens klargjøring av en kompleks struktur med mange filer, som ikke uten videre lar seg åpne og konvertere, kan kreve mye arbeid.
- Avleveringer av eldre materiale vil ofte kreve mer arbeid enn yngre materiale som er tilgjengelig ved hjelp av dagens programvare.

Tidsforbruket i bevaringsarbeidet er en faktor som bør tas i betraktning ved klargjøring av elektroniske dokumenter (jf. kapittel 3).