

**Vedlegg 2A:**  
**Bevaring av e-post**

## Innhold

<b>1</b>	<b>INNLEDNING</b>	<b>4</b>
1.1	HVA ER E-POST?	4
1.2	ENDRING I KOMMUNIKASjonsFORM	4
<b>2</b>	<b>TEKNISK BESKRIVELSE AV E-POST</b>	<b>4</b>
2.1	KLIENT-TJENER	4
2.1.1	<i>E-post via Internett</i>	4
2.1.2	<i>E-post via lokalt nettverk</i>	5
2.1.3	<i>E-postkonto</i>	5
2.2	SORTERING AV E-POST I MAPPER	5
2.3	MELDINGSFORMAT	5
2.4	KODING AV INNHOLDET I E-POSTMELDINGER	6
2.5	ÅPNE ELLER PROPRIETÆRE SYSTEMER	7
2.5.1	<i>Åpne systemer</i>	7
2.5.2	<i>Proprietære systemer</i>	7
2.6	HVORDAN E-POSTEN BLIR LAGRET	7
2.6.1	<i>Åpne formater</i>	7
2.6.2	<i>Proprietære formater</i>	8
<b>3</b>	<b>E-POST SOM ARKIVMATERIALE</b>	<b>8</b>
3.1	E-POST KAN OPPFATTES SOM MINDRE ARKIVVERDIG ENN TRADISJONELL KORRESPONDANSE	8
3.2	UFULLSTENDIGE E-POSTSAMLINGER	8
3.2.1	<i>Begrenset lagringskapasitet</i>	8
3.2.2	<i>Lokal lagring av e-post</i>	9
3.3	E-POST SOM ARKIVSYSTEM	9
3.4	SAMMENHENG MELLOM E-POST OG ØVRIG ELEKTRONISK ARKIVMATERIALE	9
3.5	E-POST SOM FENOMEN	9
3.6	ER E-POST ARKIVVERDIG?	9
<b>4</b>	<b>UTTREKKSFORMATER- OG METODER</b>	<b>10</b>
4.1	DOKUMENTFORMATER	10
4.2	UTVEKSLING AV E-POST MELLOM ULIKE KLIENTER	10
4.3	EKSPORT FRA ULIKE E-POSTKLIENTER	10
4.3.1	<i>Microsoft Outlook</i>	10
4.3.2	<i>Microsoft Outlook Express</i>	11
4.3.3	<i>Mozilla Thunderbird</i>	11
4.3.4	<i>Eudora</i>	11
<b>5</b>	<b>ULIKE ALTERNATIVER FOR BEVARING AV E-POST</b>	<b>11</b>
5.1	E-POSTMELDINGER SOM ENKELTSTÅENDE DOKUMENTER	11
5.1.1	<i>Dokumentformat og filstruktur</i>	11
5.1.2	<i>Journal</i>	12
5.1.3	<i>Verktøy</i>	12
5.1.4	<i>Fordeler</i>	12
5.1.5	<i>Ulemper</i>	13
5.2	XML	13
5.2.1	<i>Dokumentformat og filstruktur</i>	13
5.2.2	<i>Verktøy</i>	14
5.2.3	<i>Fordeler</i>	14
5.2.4	<i>Ulemper</i>	15
5.3	FLATE SEKVENSIELLE TEKSTFILER MED TILHØRENDE METADATA	15
5.3.1	<i>Dokumentformat og filstruktur</i>	15
5.3.2	<i>Verktøy</i>	15
5.3.3	<i>Fordeler</i>	15
5.3.4	<i>Ulemper</i>	15
5.4	MBOX	15

5.4.1	<i>Dokumentformat og filstruktur</i> .....	15
5.4.2	<i>Fordeler</i> .....	15
5.4.3	<i>Ulemper</i> .....	16
5.5	MAILDIR .....	16
<b>6</b>	<b>HVA KREVES FOR Å KUNNE BEVARE E-POST</b> .....	<b>16</b>
6.1	AV ARKIVSKAPER .....	16
6.1.1	<i>Sortering</i> .....	16
6.1.2	<i>Avlevering</i> .....	16
6.2	TEKNOLOGISKE BETINGELSER .....	16
6.2.1	<i>Uttrekk</i> .....	16
6.2.2	<i>Overføring til langtidslagringsformat</i> .....	16
6.3	KOMPETANSE .....	17
6.4	TIDSFORBRUK.....	17
<b>7</b>	<b>AVSLUTNING</b> .....	<b>17</b>

## 1 Innledning

### 1.1 Hva er e-post?

E-post er en forkortelse for elektronisk post, og er en metode for å lage, sende og motta meldinger via elektroniske kommunikasjonssystemer. De fleste e-postsystemer i dag bruker Internett, og e-post er en av de mest populære bruksmåtene for Internett.

Elektronisk post har mange fellestrekk med papirbasert post. Brukeren skriver brev, sender, mottar og organiserer posten etter behov. Elektronisk post gjør det også svært lett å sende vedlegg til e-postmeldingene. En bruker kan sende med dokumenter, programfiler og multimediafiler. Dette gjør e-post til et velegnet verktøy for utveksling av elektronisk materiell til de fleste formål.

### 1.2 Endring i kommunikasjonsform

E-postmeldinger kan være mer muntlige i stilen enn papirbaserte brev. Hurtigheten og de lave utgiftene i meldingsutvekslingen gjør at en bruker gjerne benytter mindre tid på å skrive en melding, og ofte vil det skrives flere meldinger frem og tilbake. Selv om e-postutveksling er *asynkron*, dvs. at sender og mottaker ikke trenger være tilstede samtidig, vil en slik utveksling i mange tilfeller være preget av flere korte meldinger som blir sendt med kort tids mellomrom. Dette gjør at e-postutveksling på mange måter er å betrakte som en kombinasjon av tradisjonell brevveksling og en telefonsamtale.

## 2 Teknisk beskrivelse av e-post

### 2.1 Klient-tjener

For å kunne bruke e-post er en bruker avhengig av å ha tilgang til en *posttjener*, en maskin som sender og mottar elektronisk post. Nye elektroniske brev som kommer til brukerens adresse tas hånd om av posttjeneren, og blir plassert i brukerens *postkasse*. For å få tilgang til sin elektroniske post må brukeren koble seg til posttjeneren ved hjelp av en *postklient*, et brukergrensesnitt for å få tilgang til og behandle e-post. En e-postklient er ofte egen programvare på brukerens arbeidsstasjon.

Posttjeneren kan befinne seg på forskjellige steder, alt etter hvilken tilkobling brukeren har til Internett. Hos større organisasjoner/firmaer vil posttjeneren ofte være plassert i et lokalnett, mens privatpersoner og mindre organisasjoner/firmaer ofte vil kjøpe tjenester hos kommersielle leverandører som tilbyr en postkasse på sin posttjener. Hvor posttjeneren er plassert vil ofte avgjøre hvilke *protokoller* som blir brukt i meldingsutvekslingen. En protokoll er en oppskrift for hvordan forskjellige datasystemer kommuniserer, og regler for hvordan meldinger mellom systemene skal bygges opp og utveksles. Sammen kalles gjerne en posttjener og en postklient for *e-postsystem*, og vi vil i stor grad bruke denne betegnelsen i rapporten.

#### 2.1.1 E-post via Internett

For å kunne sende og motta e-post via Internett må klienten kommunisere med posttjeneren ved hjelp av protokollene *POP* eller *IMAP* for mottak av post, og *SMTP* for sending av post. I arkivsammenheng er det protokollene for innkommende post som er av interesse. Eksempel på programvare som blir brukt til å hente ned e-post via Internett er Microsoft Outlook Express, Eudora og Mozilla Thunderbird.

### 2.1.1.1 POP

POP står for Post Service Protocol. Ved bruk av POP blir hver e-postmelding lastet ned på brukerens maskin, før den blir slettet fra postkassen. Dersom brukeren laster ned e-postmeldinger på flere ulike datamaskiner vil han ikke få en samling med meldinger på en maskin, men en rekke meldinger spredd på ulike maskiner. Disse meldingene kan bli midlertidig lagret, for så å bli slettet når brukeren logger av maskinen, eller de kan bli lagret på en datamaskin med manglende rutiner for sikkerhetskopiering. E-postmeldingene vil heller ikke være tilgjengelige fra andre maskiner enn den de er lastet ned på. Brukeren kan imidlertid velge å bevare en kopi av e-postmeldingen i postkassen, men dette er vanligvis ikke en standard innstilling.

### 2.1.1.2 IMAP

IMAP står for Internet Message Access Protocol. Ved bruk av IMAP blir ikke e-postmeldingen lastet ned på klientmaskinen. Det eneste som ligger på klientmaskinen er en liste over hvilke e-postmeldinger som finnes. Brukeren er nødt til å være tilkoblet internett for å få muligheten til å lese innholdet i meldingene. Fordelen med IMAP er at e-postmeldingene ligger lagret på posttjeneren. Eneste måten de blir fjernet på er at brukeren sletter dem selv.

## 2.1.2 E-post via lokalt nettverk

Dersom brukeren har tilgang til internett via et lokalt nettverk vil han ofte ha postkassen sin på en posttjener i nettverket. Også for kommunikasjon med slike postkasser kan brukeren benytte POP eller IMAP for å hente e-postmeldinger ned til klienten. I lokale nettverk vil brukerne ofte ha tilgang til mer avanserte kommunikasjonssystemer som integrerer e-post med gruppevarefunksjoner, som for eksempel kalenderfunksjon og muligheter for å reservere møterom. Eksempel på slike systemer er Microsoft Exchange, Lotus Domino og Novell Groupwise. Kommunikasjonen mellom klient og tjener vil ofte kreve andre protokoller enn POP/IMAP og SMTP, som for eksempel *MAPI* (Messaging Application Programming Interface) som er protokollen som blir benyttet for kommunikasjon mellom Microsoft Outlook og Microsoft Exchange Server.

## 2.1.3 E-postkonto

En brukers tilgang til en posttjener blir gjerne omtalt som *e-postkonto*. Dersom all e-post blir lagret på posttjeneren vil den totale mengden av e-post utgjøre e-postkontoen. Dersom brukeren benytter en protokoll der alle e-postmeldingene blir overført til klientmaskinen, vil kontoen kun bestå av postkassen på posttjeneren. Etter hvert som brukeren åpner e-postmeldinger vil de bli fjernet fra posttjeneren, og dermed fra e-postkontoen.

## 2.2 Sortering av e-post i mapper

De fleste e-postsystemer gir brukeren mulighet til å sortere meldinger i *e-postmapper*. Alle meldinger blir vanligvis lagt i en mappe kalt *innboks*, og brukeren kan deretter fordele e-postmeldingene i et hierarki av e-postmapper. Sendte e-postmeldinger blir lagret i en egen mappe, men blir av mange systemer slettet etter en viss tid. Brukeren kan selv velge å overføre de sendte meldingene til andre e-postmapper.

## 2.3 Meldingsformat

Formatet for e-post via Internett er definert i RFC 2822<sup>1</sup>. En e-postmelding består av to hovedkomponenter, *meldingshode* (header) og *meldingskropp* (body).

---

<sup>1</sup> Request for Comments. RFC kan ansees som grunnloven for Internett.

- Meldingshode - I meldingshodet, finnes vanligvis fire felt:
  - From – E-postadressen til senderen av meldingen.
  - To – E-postadressen til mottakeren av meldingen.
  - Subject – En emnetekst.
  - Date – Det lokale klokkeslettet og datoen for når meldingen ble sendt.

Andre vanlige felt i meldingshodet er:

- Cc – Carbon copy. Dette er mottakere av kopier av meldingen.
- Bcc – Blind Carbon Copy. Mottakerene i To og Cc feltet ser ikke hvem som er definert som mottakere i Bcc-feltet.
- Received – Sporinginformasjon generert av posttjenere som har behandlet meldingen.
- Content-type – Informasjon om hvordan meldingen skal vises, vanligvis en MIME-type.

I tillegg finnes en rekke andre felter som blir generert av posttjeneren eller -klienten.

- Meldingskropp – Selve meldingsteksten.

## 2.4 Koding av innholdet i e-postmeldinger

E-postprotokollen SMTP er kun definert til å overføre meldinger kodet i 7-bits ASCII. Dette er effektivt med på å begrense e-postmeldinger til bare å inneholde 128 forskjellige tegn, noe som for eksempel ekskluderer de norske bokstavene æ, ø og å. Derfor har e-post blitt utvidet ved hjelp av *MIME-standard*, en standard som definerer mekanismer for å sende tekst som ikke bruker 7-bits ASCII. I tillegg tillater bruk av MIME at det blir sendt binære vedlegg som bilde, lyd og andre typer filer. Konverteringen inn og ut av MIME-formatet blir utført av posttjeneren eller -klienten når brukeren sender og mottar e-post.

Det fremkommer av meldingshodefeltet *MIME-Versjon* at en melding er MIME-formatert. Vanlig versjonsnummer er 1.0:

*MIME-Version: 1.0*

Videre følger et felt for innholdstype og undertype. Dette feltet, kalt MIME-type vil kunne se slik ut:

*Content-type: text/plain*

En rekke filformater har registrerte MIME-typer, og dette blir brukt til å identifisere hvilken filtype et vedlegg er. Videre har alle teksttyper i tillegg et tegnsett-parameter.

MIME definerer et sett av metoder for å representere binære data i ASCII tekstformat. Meldingshodefeltet *content-transfer-encoding* indikerer hvilken metode som blir brukt.

I kapittel 2.6.1.1 viser vi et eksempel på hvordan de ulike meldingshodefeltene kan se ut i en e-postmelding.

## 2.5 *Åpne eller proprietære systemer*

### 2.5.1 Åpne systemer

Tjenersystemer som bygger på standarder for meldingsutveksling over Internett (f.eks. MIME for koding av meldingsinnhold, SMTP som protokoll for postutveksling og POP/IMAP for tilgang til egen post) kalles åpne systemer. De åpne standardene er dokumentert i allment tilgjengelige RFC'er, og tilgangen til systemer som benytter protokollene er god. Standardene er vanligvis definert flere år tilbake i tid (med nyere versjoner som tilfører ny funksjonalitet), og applikasjonsbasen har ofte fått flere år på å stabilisere seg før dagens versjoner.

### 2.5.2 Proprietære systemer

Tjenersystemer som ikke baserer seg på åpne, veldokumenterte protokoller blir kalt proprietære systemer. Slike systemer baserer seg ofte på gateways som skal oversette det interne formatet til formatene som brukes eksternt. E-postmeldinger som er lagret på proprietære formater vil kreve at systemer som kjenner formatet fortsatt er i drift.

## 2.6 *Hvordan e-posten blir lagret*

E-postmeldinger vil foreligge i tekstfiler eller binære filer, alt etter hvilke type systemer som er brukt. Vi vil her redegjøre for hva som kjennetegner de ulike formatene.

### 2.6.1 Åpne formater

#### 2.6.1.1 Mbox

Mbox er et format der e-postmeldingene blir lagret på tekstfiler. Alle e-postmapper er representert ved hver sin mbox fil. Hver ny melding blir lagret fortløpende i den samme filen, og filen blir større for hver ny e-postmelding som blir lagt til.

En melding som er kodet i mbox-format begynner med en **From\_**-linje, fortsetter med en serie ikke-**From\_** linjer, og blir avsluttet med en blank linje. En **From\_**-linje er en linje som begynner med bokstavene F, r ,o, m og mellomrom.

En e-postmelding på mbox-format kan se sånn ut:

```
From avsender@adresse.no Mon Oct 4 12:22:34 2004
Return-Path: <avsender@adresse.no>
Delivered-To: mottaker@adresse.no Date: 4 Oct 2004 12:22:34 +0100
From: Avsender <avsender@adresse.no>
To: mottaker@adresse.no
Subject: Her er din e-postmelding

MIME-Version: 1.0
Content-Type: text/plain;
  charset="iso-8859-1"

Her er e-postmeldingen din.

Mvh
Avsender
```

I motsetning til andre protokoller for utveksling av e-post, har ikke mbox blitt formelt definert gjennom RFC's standardiseringsmekanismer. Implementeringen av formatet er derfor opp til den som utvikler e-postklienten. Dette gjør at man kan se ulike varianter av mbox-filene. Ulike e-postklienter vil bruke ulike suffiks på mbox-filene. I Mozilla Thunderbird har ikke filene suffiks i det hele tatt, mens de i Eudora har suffikset .mbx.

Siden alle meldingene blir lagret i én mbox-fil kan brukeren oppleve problemer når vedkommende for eksempel sletter en e-postmelding samtidig som det kommer en ny e-postmelding som skal skrives til filen. Dette gjør at mbox er et format som har mistet noe av sin popularitet til fordel for maildir-formatet.

#### **2.6.1.2 Maildir**

Maildir er i likhet med mbox et format der e-postmeldingene blir lagret som tekstfiler, men maildir lagrer meldingene i separate filer. Filene får et unikt filnavn basert på bl.a. tidsstempel.

### **2.6.2 Proprietære formater**

#### **2.6.2.1 Microsoft Exchange Server**

E-postmeldinger som er lagret på en Microsoft Exchange Server er lagret i en proprietær database. For å få tilgang til e-postmeldingene må man benytte seg av protokollen MAPI.

#### **2.6.2.2 Pst**

En pst-fil er en fil med personlige mapper. Formatet blir brukt av Microsoft Outlook til å arkivere filer for ikke å ta opp plass på posttjeneren dersom brukeren benytter Microsoft Exchange Server, eller som lagringsformat dersom brukeren benytter Outlook som klient mot en POP-server. Outlook vil ofte være satt opp til å slette e-postmeldinger fra Exchange Server og lagre dem i en pst-fil med jevne mellomrom.

#### **2.6.2.3 Dbx**

E-postmeldinger i Microsoft Outlook Express ligger i databasefiler med etternavn .dbx. Det finnes en dbx-fil for hver e-postmappe i Outlook Express. For eksempel har innboksen navnet inbox.dbx.

## **3 E-post som arkivmateriale**

### **3.1 E-post kan oppfattes som mindre arkivverdig enn tradisjonell korrespondanse**

Kombinasjonen av papirbasert kommunikasjon og telefon gjør at mange e-postmeldinger kan inneholde en blanding av privat konversasjon og mer saksrelaterte opplysninger. Denne uformelle og muntlige stilen som til dels preger en e-postutveksling kan gjøre at mange velger å ikke betrakte e-post som fullverdig med papirbaserte brev, og ikke tillegger den noen arkivverdi. Terskelen for å slette e-postmeldinger kan dermed være lavere enn for å kassere papirbaserte brev.

### **3.2 Ufullstendige e-postsamlinger**

#### **3.2.1 Begrenset lagringskapasitet**

I organisasjoner med felles posttjener lar man vanligvis hver e-postbruker ha sin egen postkasse på tjeneren. På grunn av begrenset lagringskapasitet lar man hver bruker disponere en gitt lagringsplass som ikke skal overskrides. E-postmeldinger med store vedlegg tar mye plass, og for mange e-postbrukere vil den tildelte lagringsplassen fort bli brukt opp. I mange systemer vil brukeren bli varslet om å slette e-postmeldinger dersom den tildelte lagringsplassen overskrides.

Dersom brukerne ikke selv sletter e-postmeldinger vil systemet automatisk slette meldinger etter en viss tid. Enkelte e-postsystemer inneholder også funksjonalitet for å overføre e-postmeldinger til såkalte arkivmapper, dvs. proprietære filer på brukerens pc eller andre nettverksstasjoner. Slike filer kan senere importeres til brukerens postkasse ved behov. Denne begrensningen av lagringsplass kan føre til at mange brukeres e-postutveksling ikke er komplett dokumentert.

### 3.2.2 Lokal lagring av e-post

Dersom brukeren har tilgang til e-post via POP vil som regel e-posten bli lastet ned til klienten, og lagret enten på brukeren sin pc eller på en nettverksstasjon. Lokal lagring kan i mange tilfeller medføre dårlige rutiner på sikkerhetskopiering av e-posten, eller at en samling e-postmeldinger er fordelt på flere maskiner.

### 3.3 *E-post som arkivsystem*

E-postmeldinger kan foreligge med eller uten innhold i form av tekst i meldingskroppen, og med eller uten vedlagte dokumenter. Meldinger med innhold vil kunne betraktes som selvstendige dokumenter, mens meldinger med vedlegg kan betraktes som kontekstinformasjon til vedlagte dokumenter. Også i tilfeller der bevaringsinstitusjonen ikke har mulighet til å ta vare på vedlegg, vil e-postmeldinger kunne inneholde opplysninger som kan betraktes som metadata til dokumenter lagret i en katalog- og filstruktur på en pc eller server. Sendte e-postmeldinger med vedlegg vil som regel inneholde dokumenter som finnes i en slik struktur, og vedlegg i innkomne e-postmeldinger blir ofte lagret i strukturen. Dersom disse dokumentene fremdeles finnes i katalog- og filstrukturen, kan e-postmeldingene gi en brukbar, men ikke absolutt og sikker kontekstinformasjon.

### 3.4 *Sammenheng mellom e-post og øvrig elektronisk arkivmateriale*

For at avlevering av e-post til en bevaringsinstitusjon skal være mulig, må det eksistere teknologi og rutiner som sikrer lagring av e-postmeldingene over tid hos arkivskaperen. Stilt overfor et e-postsystem som inneholder dokumentasjon og informasjon fra en lengre tidsperiode, bør bevaringsinstitusjonen se dette i sammenheng med annet arkivmateriale hos arkivskaper. Har virksomheten betydelige mengder arkivmateriale utenfor e-postsystemet som gjør dette til et parallelt arkiv, så reduseres viktigheten av å bevare e-postsystemet. Arkivskapere kan ha lagt opp rutiner for å overføre e-post til det ordinære arkivsystemet. Dette vil tale for å ta vare på ordinært arkivsystem fremfor e-postsystem.

Flere arkivskapere bruker i dag e-postsystemet som sitt primære arkivsystem. Det er ikke usannsynlig at e-postsystemet vil bli mer og mer viktig hos en rekke mindre arkivskapere.

### 3.5 *E-post som fenomen*

E-post som fenomen kan være av interesse. En komplett e-postsamling vil kunne ha dokumentasjonsverdi ved at den dokumenterer en arkivskapers funksjon, selv om vedleggene ikke blir tatt vare på og systemet dermed vil mangle en del informasjon.

### 3.6 *Er e-post arkivverdig?*

Det er først og fremst i tilfeller hvor det ordinære arkivsystemet ikke i tilstrekkelig grad dokumenterer virksomheten, eller hvor e-postsystemet dokumenterer vesentlig mye mer, at det vil være aktuelt å vurdere avlevering og bevaring. En e-postmelding og tilhørende vedlegg vil ha dokumentasjons- og informasjonsverdi og kan betraktes som bevaringsverdig. Selv uten vedlegg vil e-postmeldingene kunne dokumentere en aktivitet hos arkivskaper som i enkelte tilfeller ikke finnes andre steder, er unik, og verdt å vurdere for bevaring.

Men en ufullstendig samling e-postmeldinger vil i mange tilfeller tale for at en bevaringsinstitusjon ikke tar vare på data fra e-postsystemet.

## 4 Uttreksformater- og metoder

### 4.1 Dokumentformater

Avleveringer av elektronisk post bør følge avleveringsbestemmelsenes<sup>2</sup> krav til dokumentformat. 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)

Av dette følger at e-postmeldinger ikke kan langtidslagres på proprietære formater. Slike e-postmeldinger må trekkes ut eller konverteres til godkjente formater, og tekstfiler som bruker et annet tegnsett enn ISO 8859 må konverteres til akseptert tegnsett.

Vedlegg bør håndteres i tråd med anbefalinger som er gitt i vedlegg 1A: Bevaring av frittstående elektroniske dokumenter.

### 4.2 Utvexling av e-post mellom ulike klienter

Flere e-postsystemer har funksjonalitet for å importere e-post som er lagret i andre e-postsystemer. Denne funksjonen kan benyttes til å konvertere data fra et format til et annet og mer ønskelig format. Dette kan igjen brukes til å importere e-postmeldingene til et verktøy utviklet for å produsere avleveringer i samsvar med et av alternativene som er skissert i kapittel 5. Dette er ikke helt uproblematisk. Erfaringer viser at ved konvertering kan det være fare for å miste eller få endret informasjon, noe som er uheldig med tanke på det elektroniske materialets autentisitet. Deler av det importerte materialet bør derfor kontrolleres for å se om det har oppstått feil.

Dersom e-postmeldinger blir importert fra Microsoft Outlook til Mozilla Thunderbird vil adressene til avsender og mottakere av e-postmeldinger som er sendt internt i e-postsystemet mangle, og importen vil inneholde feil navn.

### 4.3 Eksport fra ulike e-postklienter

Vi har ikke hatt kapasitet til å undersøke eksportfunksjoner i alle e-postklienter som finnes på markedet. Punktene under vil derfor beskrive funksjonalitet knyttet til noen sentrale klienter.

#### 4.3.1 Microsoft Outlook

Microsoft Outlook har en innebygget eksportfunksjon. Her kan man blant annet eksportere e-postmeldinger til kommaseparerte eller tabulatorseparerte tekstfiler. Uttrekket blir veldig dårlig,

---

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

og linjeskift i meldingsteksten gir linjeskift i uttrekket. Man har ikke mulighet til å få med vedlegg i uttrekket. Man kan også eksportere e-postmeldinger til en .pst-fil, men dette er et proprietært format og er ikke brukbart for langtidslagring.

#### 4.3.2 Microsoft Outlook Express

Microsoft Outlook Express kan kun eksportere til Microsoft Outlook eller Microsoft Exchange.

#### 4.3.3 Mozilla Thunderbird

E-postmeldinger i Mozilla Thunderbird blir lagret som mbox-filer (Se kapittel 2.6.1.1). Disse filene kan brukes i behandling av e-postmeldingene i andre systemer. Filene har ikke noe suffiks, og er lagret på e-postklienten.

#### 4.3.4 Eudora

E-postmeldinger fra Eudora vil, i likhet med fra Mozilla Thunderbird, bli lagret i mbox-filer. Filene har suffikset mbx, og er lagret på brukerens pc eller et nettverksområde.

## 5 Ulike alternativer for bevaring av e-post

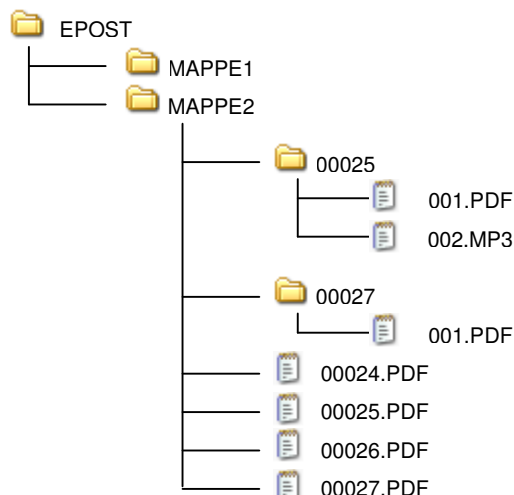
Dette kapitlet skisserer ulike alternativer for bevaring av e-postmeldinger. Ingen av alternativene er endelige løsninger, og det kan være formålstjenlig å kombinere eller videreutvikle alternativene.

### 5.1 E-postmeldinger som enkeltstående dokumenter

#### 5.1.1 Dokumentformat og filstruktur

E-postmeldinger kan konverteres til enkeltstående dokumenter. Dersom det er ønskelig å ta vare på utseendet til e-postmeldingene slik de forelå for brukeren, kan langtidslagringsformatet PDF benyttes. En annen mulighet er å langtidslagre e-postmeldinger som rene tekstfiler, men da vil ikke e-postmeldingenes visuelle uttrykk bli tatt vare på.

Dokumentene bør navngis med et nummer som blir generert fortløpende, for eksempel 00001.PDF. Tilhørende vedlegg bør legges i en katalog med samme navnet som dokumentet, for eksempel 00001. Hvert vedleggsdokument bør også nummereres fortløpende innenfor hver katalog. Dokumentene lagres i en katalogstruktur som gjenspeiler den opprinnelige strukturen i e-postklienten.



Figuren over viser hvordan en avlevering med e-postmeldinger lagret som enkeltstående dokumenter kan organiseres. Eksempelet er sterkt forenklet og består bare av to e-postmapper, MAPPE1 som inneholder 23 e-postmeldinger, og MAPPE2 som inneholder 4 e-postmeldinger. Meldingene i MAPPE1 er ikke synlige i eksempelet. E-postmeldingene 00025.PDF og 00027.PDF har vedlegg, og disse er plassert i katalogene 00025 og 00027.

### 5.1.2 Journal

Som fremfinningsmiddel er det mulig å benytte en journal som bl.a. inneholder data fra meldingshodet til hver e-postmelding. Slike data vil være:

Løpenummer	Automatisk generert løpenummer som knytter sammen postene i journalen med e-postmeldinger og tilhørende vedlegg.
Mottaker(e)	Mottaker(e) av e-postmeldingen.
Mottaker(e) av kopi	Mottaker(e) av kopi av e-postmeldingen.
Avsender	Avsender av e-postmeldingen.
Emne	Emnet som følger e-postmeldingen.
Sendetidspunkt	Sendetidspunkt for e-postmeldingen.
Meldingstekst	Meldingsteksten inneholder av og til bare informasjon som fungerer som kontekstinformasjon for vedlegget.
Vedlegg	Antall vedlegg. (Evt. egen tabell for vedlegg som tar vare på det originale filnavnet til vedlegget).
Katalog	Sti til katalog med e-postmelding og vedlegg.

Journalen bør langtidslagres som en flat sekvensiell tekstfil eller som XML<sup>3</sup> Dette kan brukes til å rekonstruere en tabell for søk, sorteringer og lignende. Alternativt kan det genereres rapporter som lagres som PDF-dokumenter, men da med begrensede muligheter for søking og sortering. Se for øvrig *Vedlegg 3A: Bevaring av fagsystemer* for fremgangsmåte.

### 5.1.3 Verktøy

Vi har ikke lyktes i å finne et verktøy som kan produsere en avlevering av den typen som er skissert over. Vi vil ikke avskrive muligheten for å utvikle et slikt verktøy ved å benytte seg av programvare basert på åpen kildekode, som for eksempel Mozilla Thunderbird. Ved å modifisere kildekode til denne e-postklienten vil det være mulig å lage funksjonalitet for å eksportere e-postmeldinger som enkeltstående dokumenter og metadata i en egen fil.

### 5.1.4 Fordeler

Ved å lagre e-postmeldinger som enkeltstående dokumenter og ha en journal/oversikt over e-postmeldinger vil man raskt få en oversikt over innholdet, uten å være avhengig av spesiell programvare for å rekonstruere datamaterialet til brukervennlig form. E-postmeldinger lagret som enkeltstående dokumenter er et alternativ som det sannsynligvis er mulig å utvikle løsninger for.

<sup>3</sup> Dersom det er en XML-fil bør det utvikles et stilsett som gjør det mulig å vise de feltene man selv ønsker, for eksempel ved at man tar med eller tar bort meldingsteksten.

### 5.1.5 Ulemper

Det er ikke sikkert det er mulig å finne/utvikle verktøy som håndterer avleveringer fra alle typer e-postsystemer/-formater.

## 5.2 XML

### 5.2.1 Dokumentformat og filstruktur

Bevaringsinstitusjoner kan bruke XML som langtidslagringsformat for e-postmeldinger. Hver e-postmelding blir gitt et løpenummer, og vedleggene lagres i kataloger som angitt i kapittel 5.1.1. Man kan lage en DTD (Document Type Definition) for å spesifisere regler for hvordan XML-elementene, attributtene og dataene skal struktureres og relateres.

En avlevering kan se ut som i eksempelet under. Eksempelet viser samme avlevering som i kapittel 5.1.1. Elementene som har + foran har innhold som ikke vises i eksempelet. Eksempelet må ikke sees på som en ferdig løsning, men som et forslag til hvordan en avlevering av e-post som XML kan bevares.

```

<?xml version='1.0' encoding='ISO-8859-1'?>
<!DOCTYPE emdmmml SYSTEM 'emdmmml.dtd'>
- <email>
+ <email_folder name="MAPPE1">
- <email_folder name="MAPPE2">
  <path>EPOST/</path>
  - <email id="00024">
    <from>avsender@adresse.no</from>
    <to>mottaker@adresse.no</to>
    <cc>mottaker2@adresse.no</cc>
    <subject>Invitasjon til foredrag</subject>
    <date>Fri, 17 Sep 2004, 10:01:11</date>
    <message>
    - <![CDATA[
      Hei mottaker
      Du er velkommen til vårt foredrag i morgen.

      Hilsen Avsender
    ]]>
    </message>
  </email>
  - <email id="00025">
    <from>avsender2@adresse.no</from>
    <to>mottaker@adresse.no</to>
    <cc>mottaker2@adresse.no</cc>
    <subject>Lyd og tekst</subject>
    <date>Fri, 17 Sep 2004, 21:55:20</date>
    <message>
    <![CDATA[
      Hei mottaker
      Sender filene du ba om til møtet med partiledelsen.
      Hilsen Avsender2
    ]]>
    </message>
    - <attachments>
      - <attachment>
        <name>001.PDF</name>
        <path>EPOST/MAPPE2/00025</path>
      </attachment>
      - <attachment>
        <name>002.MP3</name>
        <path>EPOST/MAPPE2/00025</path>
      </attachment>
    </attachments>
  </email>
+ <email id="00026">
+ <email id="00027">
</email_folder>
</email>

```

## 5.2.2 Verktøy

Vi har ikke lyktes i å finne et verktøy som kan produsere en avlevering av den typen som er skissert over.

## 5.2.3 Fordeler

Det finnes store fordeler knyttet til søkbarhet og strukturingsmuligheter ved valg av denne løsningen.

#### 5.2.4 Ulemper

Eksportverktøyer må kunne kjenne til XML-syntaksen som er spesifisert for denne type avleveringer. Det er muligens mer realistisk med komma-/tabulatorseparert fil, og metadata som XML.

### **5.3 Flate sekvensielle tekstfiler med tilhørende metadata**

#### 5.3.1 Dokumentformat og filstruktur

E-postmeldinger og vedlegg kan bevares i flate sekvensielle tekstfiler. Hver e-postmelding lagres som en post, med fast feltlengde eller en form for feltskilletegn. Posttypen vedlegg inneholder et felt som refererer til e-postmeldingen, og selve vedleggene lagres i kataloger som angitt i kapittel 5.1.1. Metadata beskrives i verktøy som genererer en addmml-fil, for eksempel Arkadukt.

#### 5.3.2 Verktøy

Vi har ikke lyktes i å finne et verktøy som kan produsere en avlevering av den typen som er skissert over.

#### 5.3.3 Fordeler

Uttrekk av e-post på flate sekvensielle tekstfiler krever ikke at e-postklienten, eller verktøyet som benyttes til å eksportere dataene, kjenner til en bestemt XML-struktur. Det går an å bruke eksportresultatet slik det er, og beskrive metadataene i et dertil egnet verktøy.

#### 5.3.4 Ulemper

Produksjon av metadata kan være en komplisert oppgave, men det er ellers få ulemper knyttet til dette alternativet.

### **5.4 Mbox**

#### 5.4.1 Dokumentformat og filstruktur

Avleveringer på mbox vil være en av de enkleste måtene å ta imot e-postmeldinger. Filene befinner seg hos brukeren og det er tilstrekkelig at filene blir kopiert for klargjøring til langtidslagring. En mbox-fil vil inneholde både meldingshodefeltet og selve meldingsteksten. Vedleggene ligger kodet i henhold til MIME-standard. Dersom e-postmeldingene er sortert i e-postmapper vil man kunne bevare denne strukturen i form av en fil- og katalogstruktur på lagringsenheten.

#### 5.4.2 Fordeler

Mbox er en standard som foreløpig støttes av flere e-postklienter. Så lenge bevaringsinstitusjoner har tilgang til e-postklienter som bruker mbox som lagringsformat vil de være i stand til å presentere e-postmeldingene på en akseptabel måte. Dersom det utvikles et verktøy med grunnlag i programvare basert på åpen kildekode, som for eksempel Mozilla Thunderbird, vil det være muligheter for å konvertere avleveringen til enkeltstående dokumenter som beskrevet i kapittel 5.1.1, til XML som beskrevet i kapittel 5.2.1, eller som flate sekvensielle tekstfiler som beskrevet i kapittel 5.3.1

Siden mbox-filer er rene tekstfiler, vil arkivbrukeren, uansett om bevaringsinstitusjonen ikke har tilgang til verktøyer som nevnt over, kunne lese innholdet i e-postmeldingene.

### 5.4.3 Ulemper

Dersom bevaringsinstitusjoner tar imot en samling med e-postmeldinger som mbox-filer vil de få en rekke filer som er mulige å lese, men i utgangspunktet lite brukervennlige for presentasjon. Mbox-filer skal inneholde en del elementer som for eksempel **From**\_linjen, men siden mbox ikke har blitt formelt definert gjennom RFC's standardiseringsmekanismer kan man se ulike varianter av mbox-filene. Dersom e-postmeldinger blir importert fra en klient til en annen vil man også oppleve at mbox-filene blir ulike. Dette kan gjøre det vanskelig å utvikle løsninger for å presentere e-postmeldingene for brukere av arkivet.

Vedlegg vil i mbox-format være kodet i henhold til MIME. Dersom man bevarer e-post på mbox-formatet vil vedleggene ikke bli konvertert til langtidslagringsformater.

## 5.5 Maildir

Maildir lagrer, i likhet med mbox, e-postmeldinger som ren tekst, men hver melding utgjør en egen fil. Derfor kan også e-postmeldinger lagret i maildir oppbevares på det formatet de foreligger hos arkivskaper.

## 6 Hva kreves for å kunne bevare e-post

### 6.1 Av arkivskaper

#### 6.1.1 Sortering

Arkivskaper bør foreta en sortering av e-postmeldingene i e-postmapper. Også sendte meldinger bør sorteres i e-postmapper, aller helst i samme mapper som logisk tilhørende mottatte meldinger.

#### 6.1.2 Avlevering

Arkivskaper bør sørge for å overføre e-postmeldingene til et format og lagringsmedium som kan brukes til videre behandling i depotinstitusjonen. Dette gir to alternativer, der det første er mest ønskelig:

- Avleveringen er i samsvar med et av alternativene i kapittel 5. E-posten vil da være klar til langtidslagring.
- E-postmeldingene er lagret på et proprietært format. Dette forutsetter at bevaringsinstitusjonen har tilgang til e-postsystemer som kan tolke e-post lagret på slike formater, og er i stand til å konvertere e-postmeldingene til et av alternativene i kapittel 5.

### 6.2 Teknologiske betingelser

#### 6.2.1 Uttrekk

E-postsystemet må være integrert med funksjonalitet for eksport av e-postmeldinger for at det skal være mulig med avlevering til en bevaringsinstitusjon. Dette lar seg gjøre i de aller fleste e-postsystemer. Bevaringsinstitusjonen må ha tilgang til verktøy (ulike e-postklienter) som kan importere/tolke e-post lagret på det avleverte formatet.

#### 6.2.2 Overføring til langtidslagringsformat

Eventuelle verktøy for overføring til langtidslagringsformat er beskrevet i kapittel 5.

### **6.3 Kompetanse**

Hvilken kompetanse som kreves er avhengig av hvilket e-postsystem meldingene er lagret i, og hvilket alternativ som velges for å langtidslagre e-posten. Som et minimum bør bevaringsinstitusjonen vite hvor e-postmeldingene er lagret, hvilket format de er lagret i, og hvordan e-postmeldingene kan overføres mellom ulike formater og lagringsmedier.

### **6.4 Tidsforbruk**

Hvor mye tid som kreves er også avhengig av hvilket e-postsystem meldingene er lagret i, og hvilket alternativ som velges for langtidslagring av e-posten. Uttrekk av e-postmeldingene vil som regel kreve et visst tidsforbruk, uavhengig om man har 200 eller 2000 meldinger. Det som derimot vil kreve betydelig mer arbeid er en eventuell bearbeiding og konvertering av vedlegg.

## **7 Avslutning**

Diskusjonen i kapittel 3 viser at e-post bør vurderes for bevaring, også i tilfeller der de ikke lar seg gjøre å ta vare på vedleggene. E-postsystemet kan brukes som en innfallsvinkel til det øvrige elektroniske arkivmaterialet, eller det kan dokumentere arkivskapers funksjon. Dersom bevaringsinstitusjonen står overfor et e-postsystem hvor det mangler store deler av e-postsamlingen, eller arkivskaper har betydelige mengder arkivmateriale utenfor e-postsystemet, så vil verdien av å ta vare på materialet fra e-postsystemet være liten.

Langtidsbevaring av e-post er en problematisk oppgave. Det finnes per i dag ingen verktøy som kan konvertere en samling e-postmeldinger med tilhørende vedlegg på en tilfredsstillende måte. Men det er ikke en umulig oppgave å utvikle verktøy for dette formålet. Dersom bevaringsinstitusjonen velger å ta vare på e-postmeldinger som ren tekst, og vedlegg formatert i henhold til MIME-standarden, vil de ha mulighet for å bruke det elektroniske materialet i form av lesbar tekst i meldingshode og –kropp, og de har mulighet å konvertere materialet til mer brukervennlige formater dersom verktøy blir utviklet.