

# Omkostninger ved digital bevaring

## Projektrapport for fase 2



DET KONGELIGE BIBLIOTEK



STATENS ARKIVER

Ulla Bøgvad Kejser (Det Kongelige Bibliotek),  
Anders Bo Nielsen og Alex Thirifays (Statens Arkiver)

Februar 2011

# Indholdsfortegnelse

<b>1</b>	<b>Resumé (Executive summary)</b> .....	<b>3</b>
1.1	<i>Baggrund</i> .....	3
1.2	<i>Formål</i> .....	4
1.3	<i>Hovedaktiviteter</i> .....	4
1.4	<i>Ressourcer</i> .....	5
1.5	<i>Produkter</i> .....	5
1.6	<i>International orientering</i> .....	5
1.7	<i>Fremadrettede aktiviteter</i> .....	5
1.8	<i>Konklusioner</i> .....	6
<b>2</b>	<b>Indledning</b> .....	<b>7</b>
2.1	<i>Læsevejledning</i> .....	7
<b>3</b>	<b>Indtagning af digitale data i arkivet (Ingest)</b> .....	<b>8</b>
3.1	<i>Udarbejdelse af afleveringsaftale</i> .....	9
<b>4</b>	<b>Omkostninger ved Ingest</b> .....	<b>12</b>
<b>5</b>	<b>Logisk bevaring (Preservation Planning)</b> .....	<b>13</b>
5.1	<i>Kort status for emulering</i> .....	13
5.2	<i>Kort status for migrering</i> .....	14
<b>6</b>	<b>Omkostninger ved migrering</b> .....	<b>15</b>
6.1	<i>Valg af kvalitet af migrering</i> .....	15
6.2	<i>Hvornår skal man migrere?</i> .....	16
6.3	<i>Hvor ofte skal man migrere?</i> .....	18
6.4	<i>Omkostninger til systemer til overvågning og migrering</i> .....	19
<b>7</b>	<b>Omkostninger ved bitbevaring (Archival Storage)</b> .....	<b>23</b>
<b>8</b>	<b>Indhentning af omkostningsdata for Ingest</b> .....	<b>24</b>
<b>9</b>	<b>Vejledning til CMDP model (regneark)</b> .....	<b>26</b>
9.1	<i>Opbygning</i> .....	26
9.2	<i>Arket Migrate</i> .....	26
9.3	<i>Arket PAIMAS</i> .....	30
<b>10</b>	<b>Konklusioner</b> .....	<b>31</b>
<b>11</b>	<b>Ordlister</b> .....	<b>36</b>
<b>12</b>	<b>Bilag</b> .....	<b>48</b>
12.1	<i>File format support in Microsoft Excel 2010</i> .....	48
12.2	<i>Omkostning i Statens Arkiver vedr. Ingest</i> .....	49

# 1 Resumé (Executive summary)

## 1.1 Baggrund

Projektet blev dannet som følge af et forslag til Kulturministeriets "Udvalg for Digital Bevaring" fra Det Kongelige Bibliotek (KB), Statsbiblioteket (SB) og Statens Arkiver (SA) om at udarbejde en omkostningsmodel for bevaring af digitalt materiale.

Projektet består foreløbigt af 2 faser:

I første fase af projektet (2009) har vi udviklet en omkostningsmodel, der betegnes "Cost Model for Digital Preservation" (CMDP). Den tilgrundliggende metode bygger på to elementer: Arkivstandard Open Archival Information System (OAIS) Reference Model<sup>1</sup> og den aktivitetsbaserede omkostningsmodel, ABC<sup>2</sup>.

OAIS modellen beskriver de forskellige funktioner i et digitalt langtidsbevaringssystem fra data indtages i bevaringsinstitutionens arkiv til de bevares og gøres tilgængelige for institutionens brugere. Vi har brugt OAIS modellen til at identificere funktioner og opdele dem i afgrænsede omkostningskritiske aktiviteter. Kun aktiviteter, der minimum tager 1 personuge at gennemføre, er medtaget i beregningerne. Herefter har vi identificeret de omkostningsparametre (variable), der er forbundet med de enkelte aktiviteter, og udtrykt dem i formler i et regneark.

Omkostninger opgøres som den tid det tager at gennemføre en aktivitet gange timeprisen. Hertil kommer udgifter til evt. anskaffelser. CMDP skal i princippet medregne alle direkte omkostninger, både udviklings-, etablerings- og driftsudgifter, samt indirekte administrationsomkostninger (overhead). Indtil videre er overhead dog kun implementeret indirekte i modellen i form af lønniveau. Andre udeståender i forhold til fordeling af omkostninger er indarbejdning af afskrivning i modellen samt afklaring af, om udgifter til almindelig systemunderstøttelse, f.eks. udgifter til etablering af infrastruktur, skal kategoriseres under OAIS arkivfunktionerne eller under støttefunktionen Common Services. Modellen skal på sigt også kunne tage højde for finansielle justeringer som inflation.

Pga. begrænsede ressourcer har vi været nødt til at operationalisere omkostningsmodellen etapevis.

I første fase af projektet har vi opgjort omkostninger for logisk bevaring og fokuseret på omkostninger ved den bevaringsstrategi, som bygger på migrering. Første fase af projektet er afrapporteret i "The Cost of Digital Preservation, November 2009"<sup>3</sup>. Rapporten indeholder en introduktion til emnet omkostninger for digital bevaring og gennemgår eksisterende omkostningsmodeller. Desuden beskriver rapporten den metode, vi har brugt til udviklingen af omkostningsmodellen. Endelig beskriver den hvordan vi har operationaliseret modellen for logisk bevaring ved migrering og præsenterer resultaterne fra vores test af modellen på eksisterende omkostningsdata fra migreringsprojekter gennemført af Statens Arkiver.

I den anden fase af projektet, som er gennemført i 2010, har vi fokuseret på omkostninger forbundet med indtagning af digitale samlinger i arkivet.

---

<sup>1</sup> <http://public.ccsds.org/publications/archive/650x0b1.pdf>

<sup>2</sup> [http://en.wikipedia.org/wiki/Activity-based\\_costing](http://en.wikipedia.org/wiki/Activity-based_costing)

<sup>3</sup> CMDP regneark, projektrapport og anden dokumentation fra første fase af projektet kan downloades fra projektets hjemmeside: [www.costmodelfordigitalpreservation.dk](http://www.costmodelfordigitalpreservation.dk).

Denne rapport beskriver projektets anden fase. Rapporten er målrettet projektets styregruppe og andre med interesse i bestemmelse af omkostninger for digital langtidsbevaring og kortlægning af bevaringsaktiviteter.

Hvor det er relevant er OAIS modellens funktionsnavne anført på engelsk og med stort begyndelsesbogstav. I kapitel 11 findes en ordliste.

## **1.2 Formål**

Projektets overordnede mål er at udarbejde en generisk model, der kan opgøre faktiske afholdte omkostninger og estimere de forventede omkostninger forbundet med bevaring af digitalt materiale på ABM institutioner.

Denne fases mål var først og fremmest at afdække omkostninger ved indtagning (Ingest) af materialer i arkivet og at forbedre CMDP modellens evne til at estimere omkostninger forbundet med logisk bevaring ved migreringsstrategien. Det var også målet at indhente omkostningsdata for, hvad det faktisk koster at indtage digitale materialer samt at teste og justere modellen i forhold til disse input. Desuden skulle modellen udvikles med området bitbevaring (Archival Storage). Endelig var det hensigten at udarbejde en brugervenlig grænseflade til CMDP regnearket og en vejledning i, hvordan man bruger modellen.

## **1.3 Hovedaktiviteter**

Projektet består af en række hovedaktiviteter som ses herunder. Aktivitetsnumrene henviser til projektinitieringsdokumentet.

- 5.3.1 Udvidelse af modellen med "aflevering" og "modtagelse" (Ingest)
- 5.3.2 Indhentning af oplysninger om Ingest omkostninger
- 5.3.3 Videreudvikling af modellen på området "migrering" (Preservation Planning)
- 5.3.4 Anvendelse af tilpasset model på indhentede oplysninger
- 5.3.5 Udvidelse af modellen med "bitbevaring" (Archival Storage)
- 5.3.6 Samlet rapport og regneark for modellen
- 5.3.7 Brugervenlig grænseflade til regneark
- 5.3.8 Oversættelse af slutrapport til engelsk

På nuværende tidspunkt har vi udvidet modellen med Ingest samt indhentet oplysninger om faktisk afholdte omkostninger forbundet med Ingest.

De resultater vi har fået fra vores spørgeskemaer til offentlige myndigheder (Producer) vedr. deres aflevering til offentlige arkiver, går i mange retninger. De viser dog, at ca. 2/3 af omkostningen for myndigheden går til selve skabelsen af informationspakken (Information Package) til arkivet, altså til migreringen (normaliseringen) af dokumenter.

Vi har også fået mere viden om, hvad der har betydning for omkostningerne forbundet med logisk bevaring og forbedret modellens evne til at anslå disse omkostninger. Dog må vi konstatere, at modellen fortsat bygger på en række skøn, og at der er et stort spænd i skønnet af omkostninger, fordi disse i sagens natur afhænger af, hvilken bevaringskvalitet der vælges: Der er eksempelvis stor forskel på, om man prioriterer valg af bevaringsformater, som skønnes at holde i 8 eller 20 år, om man prioriterer simple eller komplekse bevaringsformater samt hvor fejlbehæftede migreringer man vil acceptere.

Skønnene bygger primært på erfaringer fra arkivsektoren, eftersom bibliotekerne endnu ikke har megen erfaring med gennemførelse af migreringer og dokumentation for omkostningerne. Af samme grund må man forvente, at skønnene er mest relevante for arkiver.

## **1.4 Ressourcer**

Projektet fik tildelt 12 mandmåneder (mm.), i alt 600.000 kr. fra Kulturministeriet. Ressourcerne blev fordelt med 3 mm. til 1 projektdeltager fra Det Kongelige Bibliotek og med 8 mm. til 3 projektdeltagere fra Statens Arkiver og en fra Københavns Stadsarkiv. Endvidere fik Statsbiblioteket tildelt 1 mm. til review af projektrapport og CMDP modellen (regneark). De tildelte ressourcer er anvendt.

## **1.5 Produkter**

Projektet har leveret følgende produkter:

- Omkostningsmodel (regneark) operationaliseret for områderne Ingest, Preservation Planning og udvalgte funktioner under Administration.
- Projektrapport på dansk

## **1.6 International orientering**

Projektet har desuden holdt sig ajour med den internationale udvikling af omkostningsmodeller for digital bevaring:

Vi har løbende haft kontakt med det britiske LIFE projekt (<http://www.life.ac.uk/>). I august var CMDP projektet vært for et møde på KB med projektlederen og en projektdeltager fra LIFE projektet. Mødets formål var at evaluere LIFE3 beta modellen (regneark). I forhold til den tidligere udgave af LIFE projektets regneark, er der sket betydelige ændringer, og modellens fokus ligger nu på omkostninger for digitalisering og lagring. Alle beregninger for bevaringsplanlægning og migrering, den såkaldte Generic Preservation Model (GPM), er taget ud af modellen, fordi det var for komplekst et område at operationalisere. På mødet fortalte projektlederen for LIFE, at hensigten var lave en webbaseret løsning til regnearket. LIFE3 projektet sluttede med udgangen af august 2010 med håb om at aktiviteterne kunne videreføres i et LIFE4 projekt eller under Open Planets Foundation<sup>4</sup>. Status er i dag at hverken regneark eller webløsning er tilgængelige.

Et andet omkostningsprojekt, vi har fulgt, er Keeping Research Data Safe (KRDS). Projektet er udarbejdet af det private konsulentfirma Charles Beagrie (<http://www.beagrie.com/>). KRDS projekterne har udgivet forskellige rapporter og guidelines, men den skitserede model er endnu ikke operationaliseret.

CMDP har også bidraget med et oplæg på et internationalt ekspertmøde om omkostninger ved digital bevaring afholdt af Nederlandenes Koalition for Digital Bevaring på National Biblioteket i Haag<sup>5</sup>.

## **1.7 Fremadrettede aktiviteter**

Styregruppen har lagt op til, at der i marts 2011 afholdes en workshop om omkostninger ved digital bevaring på KB med deltagelse af en lidt bredere kreds fra institutionerne. Workshoppen skal fokusere på selve CMDP modellen, som skal testes på en række cases.

En udvidelse af modellen, så den omfatter omkostninger for bitbevaring, forventes foretaget under Bitmagasin-projektet. Denne opgave vil være relativt enklere at løse, da SB, SA og KB har lang erfaring med bitbevaring, og der findes tilgængelige omkostningsdata for området.

Derudover vil mulighederne for at promovere CMDP modellen og projektets øvrige resultater internationalt blive undersøgt, herunder at få projektrapporten oversat til engelsk og publicere den på projektets hjemme-

---

<sup>4</sup> <http://www.openplanetsfoundation.org/>

<sup>5</sup> <http://www.ncdd.nl/en/documents/20100916PriceTagsConferenceReportfinal.pdf>

side. Vi har også diskuteret muligheden for at lave en brochure (A4 side trykt på begge sider) med faktuelle oplysninger om CMDP modellen.

Endnu større opmærksomhed ville kunne opnås ved at formidle projektets resultater i et internationalt tidsskrift eller på en international conference. Præsentationen af første fase af projektet på iPRES 2009 konferencen medførte således, at vi fik en invitation til at publicere en artikel i International Journal of Digital Curation. Desuden blev vi som nævnt inviteret til at præsentere vores arbejde på et ekspertmøde på det nederlandske National Bibliotek i Haag.

Formidling på nettet, deltagelse i konferencer og ekspertmøder er ikke kun en måde at promovere institutionernes arbejde på; det åbner også op for at hente ny viden til institutionerne og skaber muligheder for at indgå i samarbejde med andre projekter om det videre arbejde med omkostningsmodellen.

## ***1.8 Konklusioner***

Overordnet fungerer metoden til at fremfinde omkostninger. Modellen som helhed er dog endnu ikke detaljeret nok til at give retvisende resultater for alle typer af materialer. Al empiri stammer fra offentlige arkivalier, hvilket medfører at modellen p.t. er bedst egnet til at skønne omkostninger netop for den type materiale.

Generelt set viser vores arbejde med omkostninger ved digital bevaring at bevaringsinstitutioner er stærkt afhængige af at kunne bruge standardiserede løsninger, da det bliver meget dyrt, hvis de selv skal udvikle alle løsninger.

En vigtig konklusion i arbejdet med Ingest, som spørgeskemaundersøgelsen bekræfter, er, at normalisering af formater udgør den største omkostning i Ingest, nemlig omtrent 2/3 af de samlede omkostninger.

En anden vigtig konklusion er, at valget af det digitale objekt (formatet), dets kompleksitet og mængde som beregningsmæssigt udgangspunkt, gør modellen potentielt generisk og dermed i stand til at beregne omkostningerne for forskellige digitale samlinger.

Såfremt CMDP skal fuldstændiggøres både metodisk og angående præcision samt udvides til at inkludere ikke blot alle OAIS funktionerne, men ligeledes indeholde mulighed for f.eks. beregning af forskellige bevaringsstrategier og forskellige samlinger, kræves flere ressourcer.

## 2 Indledning

Denne rapport beskriver anden fase af projektet ”Omkostninger ved digital bevaring”.

Omkostninger ved digital bevaring omfatter alle omkostninger forbundet med etablering og drift af et OAIS baseret digitalt langtidsbevaringssystem, dvs. fra digitale materialer bringes ind i arkivet, bevares og til de gøres tilgængelige for arkivets brugere.

Ifølge OAIS standarden er et arkiv (Archive) en organisation bestående af systemer og mennesker, som er ansvarlige for at modtage og bevare arkivets beholdning og stille materialerne til rådighed for arkivets brugergruppe (Designated Community). Ordet arkiv relaterer til alle typer bevaringsinstitutioner, og altså ikke specielt til arkivsektoren. OAIS standarden indeholder en funktionel model, der beskriver et arkiv med syv hovedfunktioner (Ingest, Archival Storage, Data Management, Administration, Preservation Planning, Access og Common Services) og tre roller (Producer, Consumer, Management). OAIS standarden definerer også tre typer informationspakker: Submission Information Package (SIP), Archival Information Package (AIP) og Dissemination Information Package (DIP).

I første fase af projektet har vi udarbejdet formler for aktiviteter under OAIS hovedfunktionen Preservation Planning og for de funktioner under Administration, der indgår i og understøtter migreringsprocesser. I anden fase har vi udarbejdet formler for aktiviteter under Ingest samt udvalgte funktioner under Administration, der også er en del af indtagningsprocessen. Vi har også brugt PAIMAS<sup>6</sup> standarden til identificering af aktiviteter i forbindelse med Ingest. PAIMAS, der bygger på OAIS standarden, beskriver aktiviteter forbundet med udformning af en afleveringsaftale (Submission Agreement) mellem Producer og Archive, samt selve overførslen af data (Transfer) og valideringen af overførslen (Validation).

### 2.1 Læsevejledning

I kapitel 3 beskriver vi, hvordan vi ved brug af OAIS og PAIMAS standarderne har udvidet CMDP modellen med Ingest,. Herefter beskriver vi i kapitel 4, hvad Ingest omkostningerne består af, og hvad de afhænger af. I kapitel 5 gør vi status over de to bevaringsstrategier emulering og migrering, og forklarer, hvorfor migreringsstrategien er den mest udbredte. Herefter beskriver vi i kapitel 6 omkostningerne ved migreringsstrategien fordelt på løbende overvågning og migrering af data, og herunder diskuterer vi migreringskvalitet, formatlevetid og migreringshyppighed. Denne diskussion har også dannet baggrund for vores videreudvikling af formlerne til logisk bevaring. I kapitel 7 lister vi de omkostningselementer, som indgår i bitbevaring. Analysen af disse omkostninger er endnu ikke gennemført i forhold til OAIS. I kapitel 8 præsenterer vi resultaterne af en spørgeskemaundersøgelse til indhentning af faktiske omkostningsdata fra afleveringer. Kapitel 9 indeholder en kort vejledning i brug af omkostningsmodellen, der endnu ikke har en brugervenlig grænseflade. Kapitel 10 konkluderer på projektet og modellens anvendelighed. Endelig findes der i kapitel 11 en litteraturliste, og i kapitel 12 er en ordliste over de begreber, vi anvender i rapporten og i regnearket.

---

<sup>6</sup> <http://public.ccsds.org/publications/archive/651x0m1.pdf>

### 3 Indtagning af digitale data i arkivet (Ingest)

Ingest er beskrevet i store træk i OAIS, men skaberne af OAIS standarden har fundet det nødvendigt at lave en nøjere beskrivelse af den interaktion, der finder sted mellem producent og arkiv i forbindelse med aflevering af data. Dette er gjort i standarden Producer-Archive Interface Methodology Abstract Standard (PAIMAS). PAIMAS har sit fokus på, hvad der kræves for at kunne udfærdige en afleveringsaftale (Submission Agreement), og beskriver en foreløbig og en endelig fase for at kunne udforme en aftale, samt en overførselsfase og godkendelsesfase.

Som første skridt i identificeringen af aktiviteter forbundet med indtagning af digitale materialer, udarbejdede vi på baggrund af OAIS standarden et flowdiagram over processen. I det følgende beskriver vi først dette flow, så man kan se hvilke aktiviteter processen involverer. Bemærk at aktiviteten Generate SIP ikke findes i OAIS. Den har vi tilføjet i CMDP, for at have et sted at placere evt. omkostninger forbundet med normalisering af informationspakkerne, inden de afleveres til arkivet.

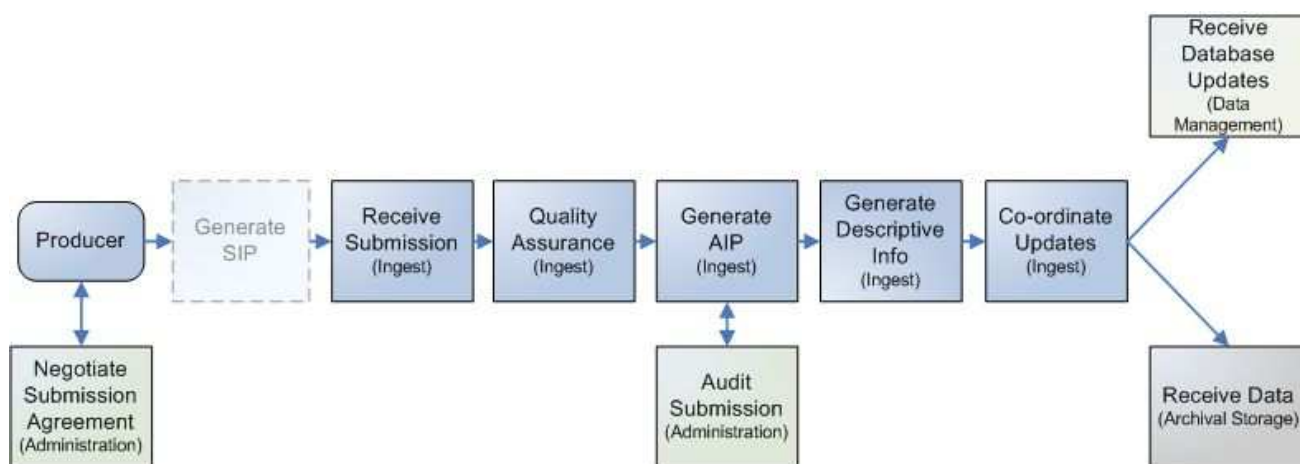


Diagram 1 viser flowet mellem OAIS funktioner fra producent til arkiv ved aflevering af digitale materialer.

#### Negotiate Submission Agreement

Inden data overføres fra producent til arkiv, kan en afleveringsaftale udarbejdes mellem parterne. I OAIS modellen sker forhandlingen af afleveringsaftalen mellem producenten og den funktion under Administration, som hedder Negotiate Submission Agreement.

#### Generate SIP

Hvis arkivet kan stille krav til producenten om, hvordan informationspakkerne skal afleveres, kan det være nødvendigt for producenten at normalisere data til godkendte afleveringsformater og/eller berige pakkerne med metadata, før pakkerne (SIP) overføres til arkivet. Denne forberedelse af pakkerne er omkostningskrævende, og da aktiviteten ikke indgår specifikt i OAIS modellen, har vi tilføjet dette element til CMDP og kaldt det Generate SIP.

#### Receive Submission

Når afleveringsaftalen er indgået, sender producenten SIP pakkerne til arkivet (Transfer), hvor de modtages af Receive Submission og placeres på et midlertidigt lager. Overførslen kan ske via flytbare lagringsmedier, som dvd eller harddiske, eller via netværk.

#### Quality Assurance

Herefter kontrolleres afleveringen for fejl af Quality Assurance funktionen. Det kan f.eks. ske vha. en checksumkontrol. Hvis pakkerne er i orden, sendes en bekræftelse på modtagelse til producenten. Hvis der er fejl eller mangler, gives der også besked til producenten, som herefter overfører pakkerne igen.



## **Generate AIP**

Dernæst migreres SIP pakkerne til AIP pakker i funktionen Generate AIP under Ingest. Denne transformation fra SIP til AIP kan indebære, at formater normaliseres til godkendte bevaringsformater, at repræsentationsinformationen (Representation Information) ændres og/eller at informationen i pakkerne reorganiseres. Funktionen kan også bede funktionen Data Management om at sende information, der er nødvendig for beskrive pakken fuldstændigt.

## **Audit Submission**

Audit Submission funktionen kontrollerer herefter om pakkerne lever op til bestemmelserne i afleveringsaftalen og sender en audit rapport tilbage til Generate AIP. Hvis der er fejl eller mangler, gives der endvidere besked til producenten, som herefter udbedrer fejlen og overfører pakkerne igen.

## **Generate Descriptive Information**

Funktionen Generate Descriptive Information trækker så den beskrivende information (Descriptive Information, dvs. metadata, som bruges til at søge og fremfinde pakkerne) ud af AIP pakken, og sender informationen til Co-ordinate Updates. Der findes ikke en funktion i OAIS, der genererer metadata og anden dokumentation. Metadata trækkes kun fra eksisterende kilder. Da tilvejebringelse af metadata kan være en betydelig omkostning, har vi i CMDP valgt at inkludere denne aktivitet i Generate SIP.

## **Co-ordinate Updates**

Co-ordinate Updates sender AIP pakken/pakkerne videre til Archival Storage, som kvitterer med en bekræftelse og allokering af et ID for AIP pakken, når lagringen er udført og verificeret. Co-ordinate Updates inkluderer denne ID i den beskrivende information og sender den videre til Data Management.

### ***3.1 Udarbejdelse af afleveringsaftale***

PAIMAS standarden beskriver, hvordan man udformer en afleveringsaftale (Submission Agreement) mellem producent og arkiv, gennemfører dataoverførslen (Transfer) og godkender at overførslen er gået som planlagt (Validation). Standarden opdeler et afleveringsprojekt i fire faser:

- Afklaringsfasen skal vise om et afleveringsprojekt er gennemførligt og rentabelt. Fasen består af den første kontakt mellem producent og arkiv, den foreløbige definition af projektets formål og kontekst, udkast til beskrivelse af de digitale data og deres struktur samt udkast til selve afleveringsaftalen.
- Definitionsfasen skal resultere i en formel afleveringsaftale mellem producent og arkiv. Fasen skal beskrive SIP designet samt præcist, hvilke digitale data, der skal afleveres. Den skal også fastlægge forhold omkring lovmæssige og kontraktuelle forhold, sikkerhed samt beskrive, hvordan overførslen og valideringen af overførslen skal foregå. Endelig skal den lægge en tidsplan for projektet.
- Overførselsfasen skal sørge for, at de digitale materialer (SIP) overføres fra producent til arkivet, og at arkivets indledende processering af data sker efter de retningslinjer, som er bestemt i afleveringsaftalen.
- Godkendelsesfasen skal sikre at overførslen af de digitale data valideres efter de retningslinjer, som er bestemt i afleveringsaftalen.

PAIMAS er en checkliste, der beskriver de enkelte aktiviteter i hver af faserne i tilstræbt kronologisk orden. De to første faser er dem, der i OAIS varetages af funktionen Negotiate Submission Agreement under Administration. Overførselsfasen svarer til Receive Submission under Ingest, mens godkendelsesfasen svarer til funktionerne Quality Assurance under Ingest samt Audit Submission under Administration. PAIMAS standardens styrke ligger i den detaljerede beskrivelse af, hvad en afleveringsaftale skal indeholde, hvordan data skal overføres og hvordan overførslen skal godkendes.

Vi startede med at analysere PAIMAS for at identificere omkostningskritiske aktiviteter og deres afhængigheder. Denne analyse viste, at graden af detaljering i PAIMAS overstiger den grænse for omkostningskritiske aktiviteter, vi opererer med i CMDP (1 personuge). Vi samlede derfor en række af aktiviteterne. Diagram 2 viser de omkostningskritiske aktiviteter i forbindelse med udarbejdelse af en afleveringsaftale. De største kasser repræsenterer hovedaktiviteterne og de mindre kasser, som er forbundet med hovedaktiviteterne, viser hvilke delaktiviteter de indeholder.

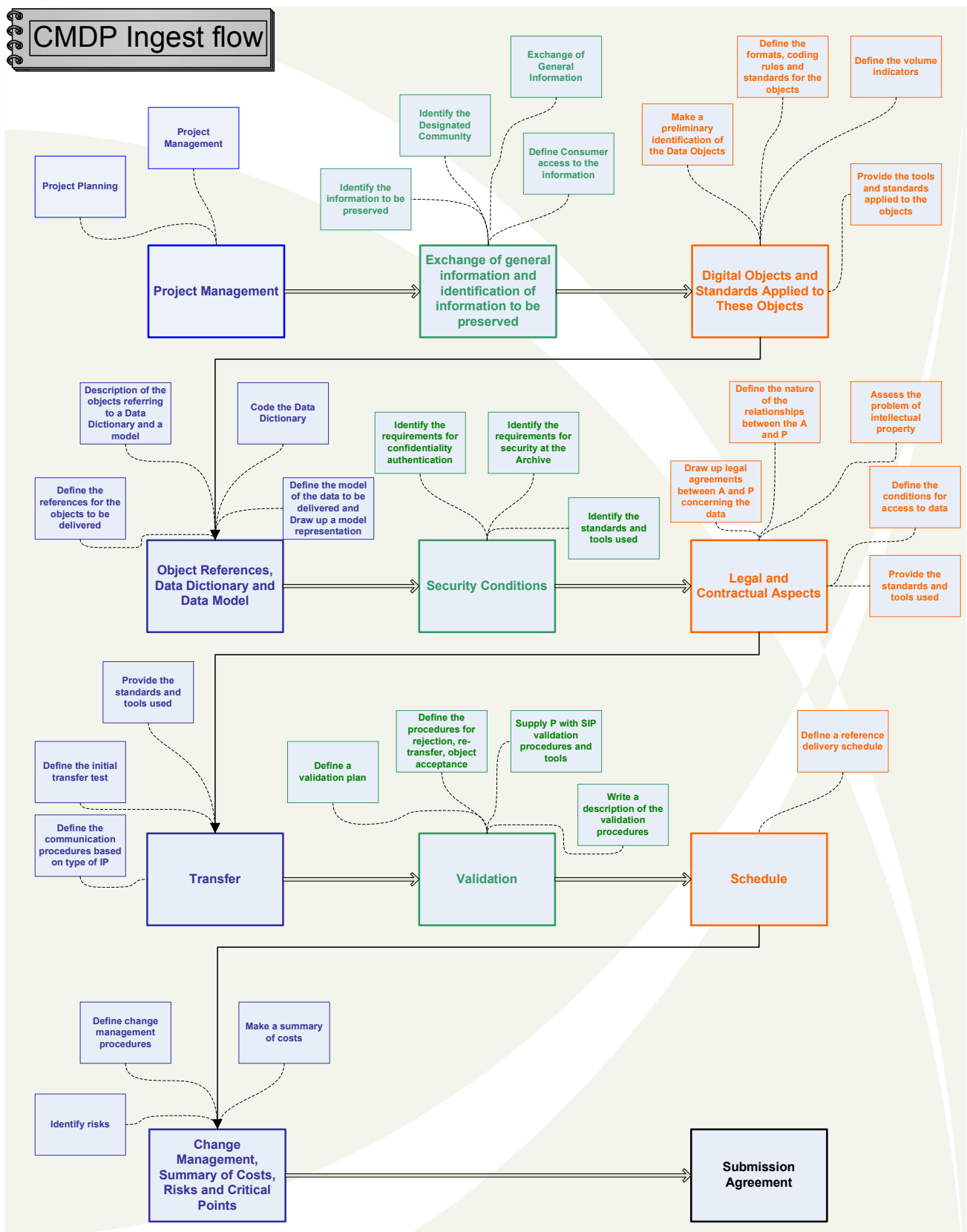


Diagram 2. Omkostningskritiske aktiviteter i forbindelse med udarbejdning af en afleveringsaftale mellem producent og arkiv.

## 4 Omkostninger ved Ingest

Indtagning af bevaringsværdige data fra producenten kan være en betydelig omkostning ved digital bevaring. I det følgende gennemgår vi de omkostningskritiske aktiviteter ved Ingest og deres afhængigheder.

Omkostningerne ved udarbejdelse af en afleveringsaftale afhænger først og fremmest af styrkeforholdet mellem producent og arkiv, af hvor forskelligartede og komplekse data er, af hvor godt de er dokumenteret samt af størrelsen af afleveringsprojektet.

Hvis Ingest omfatter normalisering af data fra de formater og strukturer, som blev anvendt hos producenten, til de, som anvendes i arkivet, medfører det en betydelig omkostning, og det uanset om det er producenten eller arkivet, der afholder denne omkostning.

I OAIS nævnes normalisering under beskrivelsen af Ingest funktionen Generate AIP, hvor der står at migreringen fra SIP til AIP bl.a. kan indebære en normalisering af formater (s. 4-6). PAIMAS nævner også kort under beskrivelsen af afleveringsaftalen (s. 3-21), at hvis det eksisterende format ikke svarer til det specificerede, skal producent og arkiv finde en løsning, f.eks. i form af normalisering. Det er dermed styrkeforholdet mellem arkiv og producent, der afgør, hvor omkostningen afholdes. Omkostningerne ved normalisering er beskrevet i kapitel 6 "Omkostninger ved migrering".

Omkostningerne forbundet med overførslen af data fra producent til arkiv består af etablering af infrastruktur og et midlertidigt lager. Vi har endnu ikke afklaret om disse og lignende udgifter til systemunderstøttelse skal kategoriseres som en del af Ingest omkostningerne eller sættes under OAIS hovedfunktionen Common Services.

Efter overførslen skal data kontrolleres for bitfejl, men denne proces er automatisk og udgør kun initialomkostninger for etablering af kvalitetssikringssystem.

Den efterfølgende generering af AIP pakken/pakkerne, kan indebære normalisering. Vi har beskrevet denne omkostning i kapitel 6. Tilførslen af beskrivende metadata til brug for søgning og fremfinding udtrækkes normalt automatisk fra pakken og er derfor ikke omkostningskritisk. Derimod kan godkendelsen af AIP pakkerne, som foretages af Audit Submission funktionen under Administration, være omkostningskrævende. Det afhænger først og fremmest af eventuelle krav stillet i afleveringsaftalen, der igen afhænger af styrkeforholdet mellem bevaringsinstitution (Archive) og arkivsskaber (Producer). Godkendelsen omfatter typisk at informationspakkerne er fuldstændige, at integriteten er bevaret, og at de passer ind i datamodellen., herunder at de aftalte dataformater er anvendt og deres syntaks overholdt.

Den efterfølgende koordinering med Data Management og overførsel til Archival Storage sker normalt også automatisk og er derfor heller ikke omkostningskritisk.

## 5 Logisk bevaring (Preservation Planning)

I den første projektrapport har vi beskrevet identificeringen af omkostningskritiske aktiviteter og afhængigheder for logisk bevaring ved migreringsstrategien.

Omkostninger ved logisk bevaring omfatter omkostninger til at gengive indholdet af data i en forståelig form. De mest realistiske bevaringsstrategier for logisk bevaring ud fra såvel en teknisk som en økonomisk tilgang skønnes at være:

- emulering (efterligning) af de oprindelige systemer, der kan anvende data i uændret form
- migrering af data, så de kan anvendes af tidssvarende systemer

Begge strategier medfører risiko for at indholdet af data ikke bevares fuldstændigt, enten ved at man ikke fuldstændigt kan efterligne de oprindelige systemer, eller ved at man ikke fuldstændigt kan migrere data så de gengives korrekt i et tidssvarende system.

### 5.1 Kort status for emulering

I de sidste 10 år er emulering blevet stadig bedre. Inden for bevaringsinstitutionernes område har det nederlandske nationalbibliotek anvendt et UVC (Universal Virtual Computer) system til tilgængeliggørelse af billeder og bl.a. via EU-projektet Planets udviklet emuleringssystemet Dioscuri, som kan emulere x86 maskinarkitekturen.

Projektet Preserving Virtual Worlds, der fokuserer på bevaring af spil og interaktiv fiktion, og EU-projektet KEEP (Keeping Emulation Environments Portable), arbejder også med emulering. Inden for denne forskning er status, at der findes lovende prototyper, men at der er langt til produkter, der kan fungere i et bevaringssystem. Hertil er opgaven forkrævende i forhold til bevillingerne.

Computerspilbranchen har også bidraget til udvikling af emulering. Spilkonsolproducenter har således udviklet emuleringssystemer til deres spilkonsoller, som spilproducenter bruger på pc maskiner. Desuden findes brugergrupper inden for computerspil, som udvikler emulatorer med henblik på at gamle spil, f.eks. til Commodore 64, kan afvikles på tidssvarende maskiner.

Emulering af de oprindelige systemer kræver, at man foruden de oprindelige data gemmer de tilsvarende oprindelige systemer og underliggende styresystemer, samt dokumentation for deres opsætning og anvendelse.

Ved gengivelse af indholdet af data mange år efter dets skabelse, må man forudse, at det i sagens natur sker på det oprindelige systems præmisser. Brugere skal således kunne betjene ældre systemer og styresystemer for at kunne tilgå data. Udviklingen inden for betjening af systemer peger i disse årtier i retning af forholdsvis ensartede brugergrænseflader. Dog er det bemærkelsesværdigt, at systemer, der blot er fra midten af 1990'erne, og som ikke har en grafisk brugergrænseflade, men er baseret på funktionstaster og kommandoer, synes at være en stor barriere for unge brugere, der ikke selv har anvendt sådanne systemer.

Status for emulering som bevaringsstrategi er således p.t., at emulering må anses for at være krævende at indføre, og relativt usikkert at vedligeholde. Det kan desuden være svært at forudsige om gengivelsen af indhold og funktionalitet vil være anvendelig på lang sigt. De bevaringsinstitutioner, som har en aktiv bevaringsindsats, undgår ofte emuleringstrategien, hvis deres data på nogen måde kan bevares via migreringsstrategien.

Der findes dog materialetyper, f.eks. interaktive materialer, hvis funktionalitet ikke kan bevares ved migreringsstrategien. Her satser nogle bevaringsinstitutioner på bitbevaring af alle de komponenter, der er nødvendige, hvis det på et senere tidspunkt bliver muligt at give adgang til materialerne via emulering.

## ***5.2 Kort status for migrering***

Muligheden for migrering af data i form af migrering af data fra et dataformat til et andet har altid været til stede, og i de sidste 10 år, særligt de seneste, er det blevet en mindre omkostningskrævende opgave. Det er sket som følge af, at der i dag anvendes færre formater end førhen, at de relativt anvendes i længere tid og at de relativt set er bedre dokumenterede (og bl.a. derfor kan anvendes af flere systemer).

Selv om migrering relativt set er blevet mindre krævende, er det dog målt absolut stadig en ganske omkostningskrævende opgave. Den omfatter bl.a. anskaffelse, drift og videreudvikling af systemer til løbende at overvåge, hvilke formater, der er forældede, og hvis indhold skal migreres til andre formater, samt ikke mindst selve migreringen af data fra et format til et andet og kontrol af, at data ikke forandres utilsigtet, altså ud over hvad det valgte format medfører.

Data i sig selv er normalt ikke anvendelige, men kræver et system, der kan fortolke dataformatet for at kunne gengive indholdet af data. Desto mere komplekst indholdet af data og dataformatet bliver, desto mere afhængig bliver man af systemer til at fortolke dataformatet. Migreringsstrategien skiller data fra dets system, og dermed risikerer man ikke at kunne gengive indholdet af data fuldstændigt i et andet system, end det, som det var skabt i. Det gør sig særligt gældende for dataformater, hvor producenten ikke har nogen interesse i, at de fortolkes af andre systemer end det oprindelige.

Migreringsstrategien er således på ingen måde fuldkommen, men p.t. den teknisk og økonomisk set mest enkle og sikre bevaringsstrategi, og vel af samme grund den som i praksis anvendes af de fleste bevaringsinstitutioner.

I det følgende vil omkostninger ved logisk bevaring udelukkende blive behandlet for migreringsstrategien.

## 6 Omkostninger ved migrering

Der er en lang række af omkostninger ved migrering, hvoraf de vigtigste er anskaffelse, drift og videreudvikling af systemer til

- løbende at overvåge, hvilke formater, der er forældede, og hvis indhold skal migreres til andre formater.
- selve migreringen af data fra et format til et andet, herunder kontrol af, at data ikke forandres utilsigtet.

Skønnene over omkostninger er baseret på egne erfaringer og på gennemgang af kildemateriale om emnet. I forhold til videreudviklingen af formlerne for migreringsomkostninger har vi særligt været inspirerede af vejledningen: Software Development Cost Estimating Guidebook<sup>7</sup> (se også litteraturlisten).

### 6.1 Valg af kvalitet af migrering

Som for mange andre omkostninger er valg af kvalitet af migreringen afgørende for omkostningsniveauet. Migreringens kvalitet bestemmes primært af valg af det format, der migreres til (herefter udformat), og af fejltolerancen ved migreringen af data fra indformat (det format, der migreres fra) til udformat.

#### 6.1.1 Valg af udformat

En høj kvalitet i form af et avanceret udformat (funktionel bevaring) frem for et simpelt vil medføre væsentligt større omkostninger. Det skyldes, at man ved migreringen af data fra indformatet til udformatet skal fastlægge og anskaffe programmer til at håndtere, hvorledes alle data i indformatet migreres til en tilsvarende plads i udformatet. Desuden skal man ved selve migreringen kontrollere, at dette er sket, jf. nedenfor om fejltolerance. Eksempelvis vil en migrering fra ét tekstbehandlingsformat til et andet tekstbehandlingsformat medføre væsentligt større omkostninger end en migrering til et simpelt format i form af et grafisk bitmap format, da tekstbehandlingsformatet rummer væsentlig mere information end et grafisk format. Dette er et generelt forhold, for i praksis kan man være i den situation, at systemet, som migrerer data fra indformat til et avanceret udformat, er lavet væsentligt bedre end systemet som migrerer til et simpelt udformat. Valget af udformat er også afgørende for, hvor ofte migrering skal udføres (se herunder).

#### 6.1.2 Valg af fejltolerance

Angående fejltolerancen ved migreringen af data, vil en høj kvalitet i form af en lav fejltolerance medføre væsentligt større omkostninger end en høj fejltolerance. Det skyldes, at en lav fejltolerance typisk vil kræve, at man anvender ekstra midler til anskaffelse, drift og videreudvikling af systemet til migrering af de pågældende formater. Desuden vil man ved selve migreringen skulle anvende ekstra midler til fejlkontrol og især fejlretning. Som nævnt under valg af udformat vil man uanset valg af fejltolerance normalt have større omkostninger ved fejlhåndtering af et avanceret udformat end et simpelt. Det skyldes at der er flere muligheder for, at noget er gået galt, og det er dyrere at rette de enkelte fejl.

Valg af udformat og fejltolerance kan desuden kombineres. Afhængigt af formålet med bevaring og datas indhold, vil nogle ønske at vælge et avanceret udformat og en høj fejltolerance. Andre vil derimod ønske at vælge et simpelt udformat og en lav fejltolerance. Valg af et avanceret udformat behøver således ikke medføre valg af en lav fejltolerance, ligesom valg af et simpelt udformat ikke behøver medføre valg af en høj fejltolerance.

---

<sup>7</sup> Software Technology Support Center (STSC) Cost Analysis Group, U.S. Air Force, Software Development Cost Estimating Guidebook, 2010, [http://www.stsc.hill.af.mil/consulting/sw\\_estimation/SoftwareGuidebook2010.pdf](http://www.stsc.hill.af.mil/consulting/sw_estimation/SoftwareGuidebook2010.pdf)

## 6.2 *Hvornår skal man migrere?*

En vigtig faktor ved omkostninger ved migrering er, hvornår, herunder især hvor ofte, man skal migrere. Umiddelbart er svaret så sjældent som muligt med risiko for forældelse in mente. Både fordi hver migrering medfører en risiko for tab af information, når data migreres fra et format til et andet, og da hver migrering medfører omkostninger.

Der er forskellige taktikker for, hvornår man bedst og billigst skal migrere, herunder til hvilket udformat.

Én taktik kaldes for migrering til tidssvarende standardiseret format, og går ud på, at man så sjældent som muligt migrerer sine data til et tidssvarende standardiseret format. Argumentet bag taktikken er, at man ved at migrere til et tidssvarende standardiseret format mindsker antallet af migreringer og dermed risikoen for utilsigtede forandringer. Grunden er, at levetiden for et standardiseret format forventes at være væsentligt længere end for andre formater, da flere systemer vil have mulighed for at læse data i formatet og skrive dem i et andet. Endvidere burde det standardiserede format gøre det billigere at anskaffe, drifte og vedligeholde systemer til selve migreringen grundet større udbud.

Imod taler, at antallet, udbredelsen og systemunderstøttelsen af tidssvarende standardiserede formater skønnes at være beskedne. Man tvinges derfor til enten at vælge udformater af en lav kvalitet (simple formater), eller man kommer til at lave næsten lige så hyppige migreringer, som hvis man blot valgte den næste generation af indformatet som udformat.

En anden taktik kaldes for migrering til nyeste format, og går ud på, at man hele tiden migrerer sine data til det udformat, som p.t. er det nyeste. Argumentet bag taktikken er, at man dermed er i samme situation som resten af it-brugerne, der har et behov for at migrere sine data fra den forrige generation af formatet til den nyeste så korrekt og så billigt som muligt. Dermed kan man nyde godt af systemerne (programmerne) til den nyeste generation af formatet, som må antages at være de bedste til at læse den umiddelbart foregående generation af formatet.

Imod taler, at de hyppige migreringer er omkostningstunge og forøger risiko for utilsigtede forandringer. Endvidere er programmerne til den nyeste generation af formatet ikke altid de bedste til at fortolke forrige generation. Nogen gange må man vente til den næstfølgende generation for at få en bedre gengivelse. Desuden er interessen hos systemleverandører og brugere for at kunne skabe nye data i nye formater frem for at læse ældre data i ældre formater i de nye generationer af systemerne tilsyneladende så meget større, at der ikke gøres noget særligt for at lette migreringen. Således kniber det med at finde systemer som varetager masse migrering af den forrige generation af formatet til den nuværende.

En tredje taktik kaldes for migrering ved efterspørgsel, og går ud på, at hvis data er i et blot nogenlunde udbredt og dokumenteret format beholder man data i dette oprindelige format og migrerer først data til et andet format, når nogen efterspørger data. Argumentet bag taktikken er, at man skønner, at antallet og variationen i anvendelsen af dataformater bliver stadig mindre, og udbredelsen, åbenheden og dokumentationen stadig større. Derfor er sandsynligheden for, at man om en årrække vil kunne læse et tidligere nogenlunde udbredt og dokumenteret format så stor, at der ikke er grund til at foretage migreringer inden da. Dermed sparer man en lang række mellemliggende omkostningskrævende og farefyldte migreringer.

Imod taler, at nogen finder, at risikoen er for høj, altså sandsynligheden for, at der efter en årrække alligevel ikke er et system, som vil kunne fortolke formatet. Endvidere har man – afhængigt af udformat – ofte en fordel ved at migrere kort tid efter skabelsen af data, da mange formater ikke er afsondrede, men afhænger af ydre data, f.eks. skrifttyper i systemet, henvisninger til billeder eller andre data uden for formatet, som kan være ændret efter en årrække. Det er vel at mærke ydre afhængigheder, hvis indkapsling kræver systemer,



som man må anskaffe, drifte og videreudvikle. Nogle standardprogrammer, som f.eks. MS Word 2010, kan dog nu delvist indlejre skrifttyper.

### 6.2.1 Hvornår er et format forældet?

Tidspunktet for, hvornår et format er forældet, hænger sammen med den tidligere diskussion af, hvornår man skal migrere, og er primært et spørgsmål om omkostninger.

Næsten uanset deres format er data ikke meget anvendelige i sig selv; de kræver et system, et program, der kan fortolke deres format, så de kan gengives i en forståelig form. Formaters forældelse afhænger derfor af forældelsen af de programmer, som skal fortolke formatet.

Hvad angår bevaring af data, er det ikke tilstrækkeligt, at et program kan fortolke data i dets nuværende format (indformat), det skal også kunne skrive data i et passende, tidssvarende format (udformat), ellers er man endt i en blindgyde. Især tidligere har man hæftet sig ved, at der ofte var en fremviser – en viewer – til et format, og at det derfor ikke var forældet. Det er dog ikke et holdbart argument, da man så er blevet afhængig af nye generationer af fremviseren, og desuden ikke kan få data viderebehandlet i andre systemer/programmer.

Svaret leder altså blot til et nyt spørgsmål, hvornår er så de pågældende systemer, der kan læse data i dets nuværende format (indformat) og skrive det i et tidssvarende udformat, forældede?

Så længe der kommer nye generationer af systemer, der kan læse data i dets nuværende format (indformat) og skrive det i et tidssvarende udformat, er der ikke noget reelt forældelsesproblem. Det forudsætter dog, at man husker at afprøve det, herunder at gengivelsen er acceptabel. Når man på et tidspunkt kommer til en generation af systemet, hvor det pågældende dataformat ikke længere kan fortolkes, eller blot ikke kan skrives i et passende tidssvarende format, er man nødt til at anvende den *forrige* generation af systemet som kunne dette, og er dermed er man afhængig af dets levetid.

Efterhånden kan flere og flere formater fortolkes af systemer, som er flere generationer yngre, men der er selvfølgelig grænser. I bilag 11.1 er et eksempel på MS Excel 2007, hvor læsningen af en række ældre formater er blevet fjernet. Man er derfor nødt til at anvende den forrige generation af programmet for at kunne læse formaterne og gemme dem i et tidssvarende udformat. I øvrigt er der som ved de fleste nye generationer af programmer og dataformater en del funktionalitet og deraf afledt data, som ikke understøttes i den nyeste generation.

Levetiden for systemer/programmer ophører ikke den samme dag, som en ny generation af systemet ser dagens lys eller et konkurrerende system overtager markedet. Systemers/programmernes levetid er afhængig af de omkostninger, der er ved at anvende og vedligeholde dem. Så længe et system/program kan afvikles på tidssvarende maskinel og integreres med tidssvarende systemer, er omkostningerne ved at anvende og vedligeholde det overkommelige. Dermed er både systemet og de dataformater, det kan fortolke og skrive til, ikke forældede.

Et kendt eksempel på forældelse er BBC Domesday Project: BBC udgav i 1986 en omfattende moderne multimedieudgave af den berømte Domesday Book, der beskriver England i det 11. århundrede. BBC Domesday udgivelsen bestod af breve, kort, billeder, statistiske data, videoer mm. lagret på to interaktive laser discs, LaserVision Read Only Memory (LV-ROM). I 2002 blev man opmærksom på, at diskene var i fare for ikke at kunne læses pga. teknologisk forældelse af udstyr, og man var nødt til at anvende både migrering, emulering og gendigitalisering for at bevare data. Teknisk lod det sig gøre med stort besvær, og de store omkostninger var et tydeligt tegn på, at formaterne var blevet forældede:

*“The lesson of this digital preservation project is that if you have enough time, individual skill, dedication and imagination then almost anything is possible, provided that you don't leave it too late. If you start counting the cost this may seem an expensive project, but then the value of the record is high too - and that applies equally to the original Domesday Project. There is of course a great need to preserve other electronic records in a routine and predictable manner, and this rescue project is not a suitable model to be followed in such cases. The National Archives is working on ways to make this possible in future”<sup>8</sup>.*

Dette i øvrigt til trods for, at skaberne fra begyndelsen var klar over bevaringsrisikoen og i tide havde afleveret data og dokumentation til et arkiv, der dog ikke formåede at løfte opgaven:

*“The deputy editor of the Domesday Project, Mike Tibbets, has criticized the UK's National Data Archive to which the archive material was originally entrusted, arguing that the creators knew that the technology would be short lived but that the archivists had failed to preserve the material effectively.”<sup>9</sup>*

Er man altid afhængig af bestående systemer til at læse data i et givet format? I praksis ja, for selv med en udførlig dokumentation af formatet er det normalt en ganske krævende opgave selv at skulle udvikle eller få udviklet et system til at læse data i ét format og skrive det i et andet.

Undtagelsen er de meget enkle formater, hvor man med beskedne omkostninger selv vil kunne udvikle eller få udviklet systemer, der kan læse data i et format og skrive det i et andet. Eksempler på det er f.eks. TIFF, UTF-8, eller XHTML.

### **6.3 Hvor ofte skal man migrere?**

Ud fra den nuværende situation vil vores forsigtige skøn for, hvornår et format er forældet være 8 til 20 år efter dets indførelse på markedet.

De 20 år er baseret på den længste tidshorisont vi overhovedet tør skønne inden for digital bevaring. De 8 år er baseret på den tid, inden for hvilken vi skønner, at man endnu vil kunne afvikle et program, der kan læse data i dets indformat og skrive dem i et passende og tidssvarende udformat.

Begge yderkanter er forsigtige skøn, og det kan om nogle år vise sig, at levetiden skønnes endnu længere, om ikke andet, fordi man da kender den tid som nu er fremtid og igen skønner 20 år ud i fremtiden.

#### **6.3.1 Vigtige parametre for levetid**

Hvorledes man anslår levetiden for et givet format mellem yderrammerne på 8 og 20 år er særdeles vanskeligt, men vi vurderer, at det er formatets udbredelse, kompleksitet og dets dokumentation, som er afgørende. Levetiden er stigende med stor udbredelse, lille kompleksitet og god dokumentation. De tre parametre er indbyrdes afhængige, hvilket ikke gør skønnet lettere. Enkle og veldokumenterede formater er ofte udbredte, og enkle formater er ofte veldokumenterede.

Udbredelse i denne sammenhæng vedrører både antallet af brugere, men især antallet af forskellige systemer, som anvender formatet. It er et marked med betydelige netværkseffekter, og det gælder således om, at ens

---

<sup>8</sup> Darlington, J., Finney, A. & Pearce, A., Domesday Redux: The rescue of the BBC Domesday Project videodiscs, Ariadne Issue 36, 2003, <http://www.ariadne.ac.uk/issue36/tna/intro.html>

<sup>9</sup>Mike Tibbets, ACM Committee on Computers and Public Policy, Forum on Risks to the Public in Computers and Related Systems, Volume 25: Issue 44, Saturday 8 November 2008, <http://catless.ncl.ac.uk/Risks/25.44.html#subj7>

programmer fuldt ud kan læse konkurrentens format, men kun kan skrive i ens eget, ellers bliver man tvunget til at konkurrere på konkurrentens hjemmebane eller på lige grundlag.

Kompleksitet afhænger af mængden af typer af information i formatet, herunder den funktionalitet i systemerne, som afspejles i formatet. Formater med høj kompleksitet erstattes ofte hurtigere (end formater med lav) af nye generationer af formatet, da producenterne eller brugerne ønsker endnu mere funktionalitet. Som nævnt kan formater med en meget lav kompleksitet være uafhængige af bestående systemer, fordi det vil være muligt ud fra dokumentationen, hvis den er god, uden uoverstigelige omkostninger, at udvikle/få udviklet et system, som kan fortolke formatet.

Dokumentation vedrører beskrivelse af formatets struktur og anvendelse. God dokumentation er kendetegnet ved, at det giver andre end den oprindelige skaber af formatet en overkommelig mulighed for at udvikle systemer, der kan fortolke formatet. Til tider vil det også være nødvendigt med delvis dokumentation af systemet, for at forstå, hvorledes formatet skal fortolkes. For at dokumentation skal være god, skal den for det første være tilgængelig, dernæst dækkende for hele formatets struktur og anvendelse, samt forklarende, altså beregnet på, at andre udviklere end de oprindelige skal kunne forstå formatet.

## **6.4 Omkostninger til systemer til overvågning og migrering**

- løbende at overvåge, hvilke formater, der er forældede og hvis indhold skal migreres til andre formater
- selve migreringen af data fra et format til et andet, herunder kontrol af, at data ikke forandres utilsigtet, altså ud over hvad det valgte format måtte medføre

### **6.4.1 Omkostninger til løbende overvågning**

Der skal afholdes omkostninger til anskaffelse, drift og videreudvikling af system til identifikation og registrering af samtlige formater for samtlige data med angivelse af præcis version for hver enkelt dataenhed.

I praksis medfører det, at man ved indtagning af data i bevaringssystemet analyserer alle data, så man kan identificere og registrere deres formater, således at alle data i et givet format kan fremfindes, når de skal overføres til et andet format.

Opgaven kan foretages af afleverende part, såfremt modtagende part kan få denne til at udføre opgaven og stole på udførelsen, men i praksis vil de fleste bevaringsinstitutioner selv skulle varetage denne opgave.

Identificering bør i praksis efterfølges af validering og delvis karakterisering. Det skyldes, at alt for mange data ikke overholder deres format, og at mange formater er så indholdsrige, at man kan få brug for viden om deres karakteristika, altså hvilke dele af formatet der befinder sig data i.

#### **6.4.1.1 Anskaffelse**

Anskaffelse af et sådant system medfører p.t., at man udvikler det eller får det udviklet, omend der findes delvise løsninger i form af PRONOM, DROID og JHOVE.

Vi skønner, at omkostningerne til anskaffelse af kernen i et modulsystem, som via specifikke moduler for de enkelte formater kan foretage en rimelig identifikation, en delvis validering og en anelse karakterisering, vil være på 12-24 personmåneder.

Omkostningerne til udvikling af de enkelte moduler afhænger af formaternes kompleksitet og dokumentation, og skønnes at være hhv. eksponentielt stigende og aftagende. Vi skønner, at omkostningen per format vil være fra et par personuger for simple formater til adskillige for avancerede formater.

Går man ud over hvad vi uklart har kaldt for en rimelig identifikation, en delvis validering og en anelse karakterisering, skønner vi, at man får en voldsom eksponentiel stigning i omkostningerne. Eksempelvis er det stadig ikke lykkedes at lave en fuldstændig validering af PDF/A. P.t. må man bruge validatorer fra flere leverandører for at dække så mange områder som muligt.

Det vil ikke være muligt at undgå fejlidentifikation eller fejlformaterede data. I praksis må man forlade sig på, at programmerne, som skal migrere data til andre formater, er nogenlunde fejltolerante. Enkelte fejl vil ikke kunne undgås uden meget store omkostninger.

Er man i en situation, f.eks. i forbindelse med indtagning i bevaringssystemet, hvor man skal kunne sikre migrering af et næsten ubegrænset antal formater, vil omkostningen til identifikation af alle disse mange formater medføre, at man må nøjes med en generel grov identifikation med bedst identifikation af de mest anvendte formater.

#### **6.4.1.2 Drift**

Selve overvågningen sker ved at man manuelt gennemgår listen af anvendte formater og sammenholder den med udviklingen på markedet for de pågældende formater, for at vurdere, om nogle formater er ved at blive forældede. Der arbejdes også på at etablere et fælles internationalt format register, Unified Digital Format Registry (UDFR)<sup>10</sup>, som vil kunne effektivisere overvågningen. Overvågning af markedet medfører i praksis, at der for hvert format er ét eller flere systemer/programmer, som man må registrere og angive som nødvendige for at fortolke formatet. Disse systemer/programmers levetid må man så også vurdere, herunder om formatet er understøttet i nyeste generation af systemet/programmet.

Opgaven er meget manuel, og vi skønner, at omkostningen er proportional med formatets kompleksitet. På den baggrund anslås det, at overvågningen vil vare fra enkelte persondage til få personuger, og at den hyppigst vil skulle ske hvert andet år for et givet format.

#### **6.4.1.3 Vedligeholdelse**

Vedligeholdelse af systemet omfatter foruden almen vedligeholdelse, f.eks. i forbindelse med nyt styresystem, også udvikling af nye profiler for identifikation, validering og karakterisering for nye formater, som måtte blive anvendt i bevaringssystemet.

### **6.4.2 Omkostninger til migrering af data**

Migreringen af data fra ét format til et andet er selve kernen i migrering, og i forhold til omkostninger kan det opdeles i en række delopgaver.

#### **6.4.2.1 Anskaffelse**

Anskaffelse af et system til migrering indebærer, at man udvikler/får udviklet et system med moduler, som kan løse følgende opgaver:

- generelt modul, som ud fra den centrale registrering af data og deres format kan fremfinde de data i en informationspakke (en SIP eller AIP), hvis format er skønnet til at være forældet, og udpakke disse data.
- generelt modul, som kan holde styr på samtlige fremfundne informationspakker og data i de forældede formater og deres status igennem migreringsprocessen. Modulet skal for hver mængde data i et format kalde det specifikke modul, som er lavet for hvert format, registrere resultatet, og ved succes

---

<sup>10</sup> Unified Digital Format Registry (UDFR), <http://www.udfr.org/>

sende de migrerede data i deres nye format til ”ompakning” sammen med de uforandrede data fra pakken, således at der skabes nye pakker. Af hensyn til effektivitet skal modulet kunne parallelisere sine kald.

- specifikke moduler for hvert format, som sørger for, at data i formatet migreres med det system/program, som anses som bedst egnet til at migrere data fra indformatet til udformatet i den ønskede kvalitet. De pågældende programmer vil normalt være de samme, som man har registreret i forbindelse med overvågning af formatets forældelse. For at kunne automatisere migreringen vil man få behov for at modulet kan kontrollere dele af programmets opførsel, f.eks. så man ikke stoppes af forespørgsler fra programmet. Vælger man et avanceret udformat, vil man endvidere kunne risikere at få behov for yderligere at kunne styre programmet for at få alle de ønskede informationer med over i udformatet.

Omkostningerne ved at udvikle/få udviklet overstående system er betragtelige, og genbrug af andres løsninger er oplagt. Vi kender ikke til nogen færdige løsninger, men en række delløsninger, som f.eks. Apache Hadoop eller Berkeley Boinc kunne inddrages.

De generelle moduler er ikke mere krævende end at vi skønner, at de kan udvikles på 12-24 mandmåneder.

Omkostningerne for de specifikke moduler er ikke nødvendigvis proportionelle med antallet af formater, hvis en række formater anvender samme program til migrering. Dog er testen af, at modulet virker korrekt med et givet format, proportionel, og derfor kan omkostningen være næsten proportionel. Genbrug af andres løsninger er oplagt, men vi kender ikke til nogen. Til hvert format, primært de avancerede, hvor behovet findes, er der ofte hele eller delvise løsninger, som kan manipulere afviklingen af et program eller direkte tilgå formatet, som f.eks. Apache POI eller Microsoft Open XML Format SDK. Imidlertid vurderer vi, at omkostningen ved direkte at tilgå formatet for avancerede formater, som f.eks. ODF eller OOXML, overstiger hvad en enkelt bevaringsinstitution umiddelbart kan overkomme. Institutionerne må derfor afvente en udvikling i et større fællesskab, hvis kvaliteten skal overgå, hvad færdige programmer kan tilbyde.

Vi skønner, at omkostningen per format er eksponentiel med formatets kompleksitet og omvendt med fejltolerance.

Vi skønner, at et modul for et simpelt format med en lav fejltolerance vil tage nogle få personuger, medens et avanceret format med en lav fejltolerance vil tage adskillige.

#### **6.4.2.2 Drift**

Omkostningerne ved drift af systemet er primært i forbindelse med fejlhåndtering. Omfanget af fejlhåndtering afhænger som for så mange andre systemer af, hvor driftssikkert det er udviklet. I den henseende er omkostningerne til udvikling og til efterfølgende fejlhåndtering ofte omvendt proportionale, og det er ikke let at beregne den optimale fordeling.

Fejlhåndtering omfatter egentlige driftsnedbrud for de områder, hvor man ikke har udviklet systemet til at være driftssikkert nok. Desuden omfatter den konstatering af fejl i de enkelte moduler, når et format ikke kan migreres som forventet. Endelig omfatter fejlhåndteringen fejl, som systemet ikke ved det laver, og som kun kan detekteres ved efterfølgende stikprøver. Med andre ord en håndtering af fejl, som man ved, vil komme, fejl, som man formoder, vil komme, samt fejl, som man ikke forventer. Når fejlene er konstateret, skal man afgøre, om, og i givet fald hvorledes, de skal udbedres.

Afhængigt af valgt migreringskvalitet, primært udformatets kompleksitet og migreringens fejltolerance, skønner vi, at der per format per TB skal anvendes fra en persondag til få personuger til overvågningen. Fejlretning skønner vi tager op til 10 gange så lang tid som overvågning.

Selv om omkostningerne kan opgøres per format, er der alligevel besparelser ved at migrere flere formater samtidigt, f.eks. i forbindelse med ind- og udpakning, lagring og fejlhåndtering. Da formater ikke dør samme dag de erklæres for forældede, bør man samle flere forældede eller næsten forældede formater til samtidig migrering.

### **6.4.2.3 Vedligeholdelse**

Vedligeholdelse af systemet omfatter almen vedligeholdelse, f.eks. i forbindelse med nyt styresystem og udvikling af nye moduler til nye formater.

## 7 Omkostninger ved bitbevaring (Archival Storage)

Omkostninger ved fysisk bevaring omfatter omkostninger til at lagre data, særligt omkostninger til anskaffelse, drift og videreudvikling af systemer, der kan varetage:

- anskaffelse af tilstrækkeligt antal medier (lagerplads)
- i et tilstrækkeligt antal eksemplarer (kopier)
- fordelt på flere geografiske steder for at mindske risikoen for brand, tyveri, hærværk, terror og naturkatastrofer
- fordelt på flere typer og fabrikater af medier for at mindske risikoen for generelle produktionsfejl
- tilstrækkelig hyppig overvågning af indhold på medier for at sikre eller i det mindste sandsynliggøre, at de er uforandrede på kontroltidspunktet
- dokumenteret overførsel af data fra gamle medier til nye medier, herunder nye typer af medier, når bestående medier teknisk eller teknologisk er ved at være forældede.

## 8 Indhentning af omkostningsdata for Ingest

Der er udsendt et spørgeskema til en række offentlige myndigheder for at indhente oplysninger om deres reelle tids- og ressourceforbrug ved produktion af arkiveringsversioner af data fra it-systemer i forbindelse med aflevering til Statens Arkiver. De indhentede oplysninger skal bruges til at teste og justere omkostningsmodellen i forhold til den nyudviklede Ingest del. Hvis myndighederne har benyttet en ekstern leverandør til produktion af arkiveringsversionen, er de også blevet bedt om at indsende kopi af kontrakt/faktura på opgaven for at få et fuldstændigt overblik over omkostningerne.

Spørgeskemaet (se herunder) er udarbejdet med afsæt i PAIMAS standarden.

	Spørgsmål vedrørende aflevering af it-system til arkiv	Tid	Beløb	Bemærk.
A	Hvor mange timer blev der brugt på projektplanlægning? F.eks. deltagelse i møder, løbende kontakt til modtagende arkiv og leverandør, udarbejdelse af kontrakt, etc.			
B	Hvor mange timer blev der brugt på identificering og beskrivelse af afleveringsdata og dokumentation? F.eks. fremfindning af data, identificering af hvilke data der skulle afleveres og hvordan de skulle afleveres (evt. afleveringsmodel), fremfindning af teknisk og administrativ dokumentation, etc.			
C	Hvor mange timer/penge blev der brugt på produktion af arkiveringsversionen? Etablering og opsætning af hardware/software Udtræk af data fra myndighedens system Konvertering af data til arkivering (tabeller, typer, struktur) Dokumentkonvertering Maskintid Fejlretning (manuel, teknisk og administrativ) f.eks. håndtering af fejlbehæftede dokumenter Udarbejdelse af metadata Pakning af arkiveringsversionen			Vedlæg evt. leverandørkontrakt
D	Hvor mange timer blev der brugt på selve den fysiske aflevering af arkiveringsversion til modtagende arkiv?			
E	Hvor mange timer blev der brugt på at teste arkiveringsversionen frem til godkendelse?			

Spørgeskemaet blev udsendt til 34 myndigheder, hvoraf ca. halvdelen har svaret. Tabel 1 herunder gengiver omkostningsdata for udvalgte myndigheder ved aflevering til offentligt arkiv. De modtagne svar peger i mange retninger, og viser at myndighederne har haft svært ved at forstå spørgeskemaet og at opgøre ressourceforbruget. På baggrund af de indkomne svar kan det forsigtigt konkluderes ved større afleveringsprojekter (>160 timer), at projektledelse (A) koster ca. 13 % af det samlede afleveringsprojekt. Identifikation og beskrivelse af de digitale objekter og deres referencer (B) udgør ca. 16 %. Produktionen af arkiveringsversionen (normaliseringen) (C) antager ca. 66 %. Test af arkiveringsversionen (E) udgør ca. 5 %. De indkomne svar for, hvor lang tid der er brugt på selve den fysiske aflevering (D), indgår ikke, da svarene viste at spørgsmålet var blevet misforstået. Man bør endvidere bemærke, at det er de afholdte udgifter, som er angivet, da samtlige omkostninger ikke er kendt. Endvidere gør det sig gældende, at for de myndigheder, som har anvendt konsulenter, er en høj pris ikke nødvendigvis lig med en stor fortjeneste for konsulenten, da forholdet mellem pris og fortjeneste ikke er ens for alle konsulenter.



Tabel 1. Omkostninger for udvalgte myndigheder ved aflevering til offentligt arkiv.

Institution	Timeforbrug i alt	Beløb i alt	System	MB	Dokumentformat	Kontrakt dato	Projektplanlægning timer / kroner	Dataid og -beskrivelse - timer /	Produktion af arkiveringsversionen	Fysisk aflevering timer / kroner	Tes t	Producent af udstøttet	Afleveret til?	Pris pr. MB	Timer pr. MB	Timepris
A	49	17.100	E-J	10,0	TIF	2010	9	6	30	2	2	1	KSA	1.710,00	4,900	349
B	40	12.540	E-J	100,0	TIF	2010	9	6	22	1	2	1	KSA	125,40	0,400	314
C	681	487.918	ESDH	180.000,0	.xlsx, .xls, .if	2008	181	500	426.518			2	SA	2,71	0,004	716
D	20	36.000	Register	700,0	SQL				36.000			6	SA	51,43	0,029	1.800
E			ESDH	2.100,0	mails og dd	2010						2	SA			
F	160	207.628	ESDH	27.284,0	AcroExch.D	2008	40	40	207.628	70	10	2	SA	7,61	0,006	1.298
G	772	571.638	ESDH		ACE AVL, E	2007-08	60	60	571.638			3	SA			740
H	235	1.621.885	ESDH	86.603,0		2007	175	60	547	656.884		4	SA	18,73	0,003	6.902
I												5				
J	112	100.000	ESDH	44.000,0	alle	2006	31	74	50	100.000	7	1	KSA	2,27	0,003	893
K	182		ESDH	328,5	wordperfect	2006-10						1	KSA		0,554	
L	17,5	754.812	ESDH	189	se aftale m	2007	5	12,5	194.812	560.000		2			0,093	43.132
M		662.660	E-J og ESDH	134,0		2006			662.660			2	SA	4.945,22	0,000	
N	577	451.350	Register/ESDH	22.000,0	TIF	2010	20	5	531	451.350	20	1	KSA	20,52	0,026	782
O	218	158.950	Register	6,5	TIF	2009	10	5	187	158.950	15	1	KSA	24.453,85	33,538	729
P		2.159.999	ESDH	1.499.000,0				60.200	1.819.799					1,44		
Q	190		Registre	11,0	TIF	2009	5	20	160	5					17,273	
Middelv	255,3	557.114		124.164,4			49,5	72,9	227,8	15,0	9,3			3.132,04	4,323	1.452

## 9 Vejledning til CMDP model (regneark)

Som en del af denne rapport er der udarbejdet et regneark benævnt CMDP\_2.xls. Formålet med regnearket er, at man skal kunne anvende modellen, hvormed menes, at man kan vælge formater og mængder, man kan rette konstanter og tilmed formlerne.

Regnearket er lavet med det formål, at det skulle være så åbent og tilgængeligt som muligt, hvorfor oprettelse af egne makroer og funktioner er undgået. Imidlertid har det vist sig, at behovet for funktionalitet i form af opslag og beregninger er så stort, at mange formler bliver ganske komplekse, selv med brug af mellemregninger, som er lavet direkte i regnearket af hensyn til gennemsigtighed.

Regnearket er således åbent, men ikke så tilgængeligt eller gennemsigtigt som ønsket. Problemet er, at selve regnearksformatet ikke både kan opfylde de store krav til funktionalitet og krav til åbenhed og gennemsigtighed. I en næste udgave af regnearket vil der således blive behov for at implementere størstedelen af funktionaliteten i egne makroer og formler for at mindske kompleksiteten, hvilket vil medføre at regnearket til gengæld blive mindre åbent og formlerne knapt så gennemsikkelige.

### 9.1 Opbygning

Regnearket indeholder de vigtigste omkostningselementer fra de funktionsenheder (Functional Entity) i OAIS, som dette projekt har behandlet, dvs. Ingest (IN), Administration (ADM) og Preservation Planning (PP). Regnearket er opbygget med brug af flere ark (faneblade). Således bruges et ark for hver funktionsenhed.

Det første ark, "Overview", er blot en oversigt af OAIS funktionsenheder, herunder Ingest, Administration og Preservation Planning.

Det følgende ark, "Cost", sammenstiller omkostninger fra de enkelte ark for hver funktionsenhed. "Cost" angiver omkostningerne over en 20-årig periode baseret på angivelser af årlige og periodevise omkostninger, som de findes i de enkelte ark for Ingest (IN), Administration (ADM) og Preservation Planning (PP).

Arkene IN, ADM og PP er opbygget ud fra samme skabelon med angivelse af element og underelement. Hver omkostning har dens værdi angivet i persontid (typisk personuger) og fordeling på typer af arbejdskraft og hyppighed.

Til arkene Ingest (IN), Administration (ADM) og Preservation Planning (PP) hører andre mellemregninger i form af ark, benævnt PAIMAS og Migrate (regnearksformatet tillader ikke rangorden af ark, og denne orden er derfor ikke umiddelbar synlig.) Arket IN får nogle af sine data fra arket PAIMAS og fra arket Migrate. Arket ADM og PP får nogle af sine data fra arket Migrate. Regnearket indeholder et ark med data fra sager, arkets navn begynder med "Case".

### 9.2 Arket Migrate

I dette ark beregnes omkostninger ved at migrere fra ét format til et andet. Ideen er, at man vælger de formater, der skal migreres fra og til, hvilke mængder for hvert format, og hvorledes migreringssoftwaren fremskaffes (baseret på bestående programmer eller udvikling).

Ud fra de mange på forhånd fastsatte parametre, som man selv kan ændre på, fremkommer et resultat (opgjort i personuger) for både den første normalisering (Ingest) og de efterfølgende migreringer (fra bevaringsformat til bevaringsformat).

Arket består af en række blokke:

## **Format identification**

Her angives formatets type, navn og gruppe. Der er på forhånd angivet en række kendte typer og formater. Når et mere omfattende fælles formatregister er blevet udviklet, vil dets definitioner kunne anvendes. Indtil da må brugeren selv tilføje ny formater (ved at indsætte rækker).

## **User choices for size and format**

Her angives datamængden i Gigabytes (GB) samt udformatet.

Valg af udformat er for hvert indformat begrænset til ganske få formater (de er angivet i samme kolonneblok som indformaterne). Begrænsningen er lavet efter, hvilke udformater, som er fundet mest optimale ud fra et bevaringsmæssigt perspektiv. Kriteriet har været forventet levetid og lav kompleksitet. Alle udformater er også angivet som indformater.

Her angives også, hvorledes programmet til migrering tilvejebringes, dvs. om man anvender bestående program eller udvikler det.

Udgangspunktet er bestående program, da der for næsten alle indformater findes programmer, som kan læse dem og skrive dem i det valgte udformat. Disse bør vælges, da omkostningerne ved deres tilvejebringelse er langt mindre end ved udvikling. Som eksempel kan nævnes, at det vil være langt billigere at anskaffe f.eks. Word Perfect X5 til at læse formatet Wordperfect 7 og skrive i formatet ODT 1.0, end at udvikle program til det.

## **Format characteristics**

I denne blok angives for hvert format en række karakteristika, nærmere bestemt dets kompleksitet, dets dokumentation, og hvorledes det er implementeret. Alle værdier angives i forhold til formattypen uformateret tekst. De pågældende karakteristika er valgt, fordi vi anser de har betydning for omkostningerne ved bevaring.

Kompleksitet hænger nøje sammen med funktionalitet. Påstanden er, at komplekse formater hurtigere bliver forældede, da de bliver udskiftet med nye udgaver, netop fordi man ønsker mere eller anden funktionalitet. Desuden er komplekse formater i sig selv dyrere at håndtere end simple på mange områder, simpelthen på grund af det større udfaldsrum.

Det relative niveau for kompleksitet er baseret på den vurdering af formater, som blev foretaget i første del af projektet. Vurderingen var i høj grad opgjort ud fra den mængde af dokumentation, man skulle sætte sig ind i for at forstå formatet. Blandt de forholdsvis få typer af formater, som er angivet, har regneark og tekstbehandling den største kompleksitet.

Dokumentation angiver i hvilken grad et format har en dokumentation, som er tilgængelig, dækkende og anvendelig. I praksis er det kun de simple tekstformater, hvor dette er opfyldt 100 %, hvilket med UTF-8 kan være en tilsnigelse.

Under implementering er angivet, i hvilken grad der findes en implementering af formatet i åben kildekode, i hvilken grad formatet anvendes bredt, og i hvilken grad formatet p.t. anvendes. Angivelsen for åben kildekode er sjældent 100 %, hvilket skyldes, at det er sjældent, at alle dele af formatet er understøttet i åben kildekode. Med bred anvendelse menes, at formatet anvendes i mange forskellige brancher og lande. Det forbedrer muligheden for at finde programmet, der kan læse formatet.

Nuværende brug angiver, i hvilken grad formatet anvendes inden for dets område, omtrentligt hvilken andel af dets formattypes brug det har. Eksempelvis er Word Perfect 7 meget bredt anvendt i og med, at det findes alle mulige steder, men dets nuværende brug er minimal, hvorimod diverse nutidige udgaver af MS Word har over 90 % af det nuværende marked for tekstbehandling.

## **Format Life Time**

De beregnede levetider anvendes ikke direkte i modellen, men fungerer blot som ledetråde. Levetiden er beregnet ud fra formatets karakteristika med stor vægt (40 %) på formatets kompleksitet. Grunden til at vægtningen for dokumentation og for en implementering i åben kildekode er lav, er at det omkostningsmæssigt er mest fordelagtigt at basere sig på at migrere med bestående programmel, hvorfor udbredelse og nuværende anvendelse vægtes højere. En gennemlæsning af resultaterne vil få mange til at udbryde, at dette eller hint format forventes at få en meget kortere eller længere levetid. Det viser blot, at der er andre variable, som har indflydelse, variable, som vi endnu ikke har identificeret.

## **Software provision using existing software**

Her er omkostningen til at fremskaffe programmel til det enkelte format opgjort. Antagelsen er, at den optimale måde at migrere på, er ved at udvikle et generelt migreringssystem, hvortil der laves implementeringer for hvert format.

Omkostningen til hvert format er underopdelt i at fremskaffe og afprøve programmellet samt i at implementere det i migreringssystemet. Præcist hvilket konkret programmel man skal anskaffe til et givet format, er ikke angivet. Andre værktøjer, som Plato fra EU-projektet Planets, kan her være en hjælp til at vælge værktøj, men generelt er det antaget, at man anvender de programmer, som er skabt sammen med formatet, hvis de ellers kan skrive i det ønskede udformat.

Omkostningen er i høj grad baseret på formatets kompleksitet, og særligt på udformatets kompleksitet. Modellen tager i denne udgave ikke højde for, at en lang række formater ligner hinanden så meget, at omkostningerne er aftagende frem for ligefrem proportionale, f.eks. MS Word 2000 og MS Word 2002. Argumentet er, at hvis man migrerer fra noget komplekst til noget simpelt (f.eks. fra ODT til TIFF), er det udformatet, som er afgørende for, hvor megen kompleksitet man skal overføre, og dermed hvor mange omkostninger det medfører.

I modellen er udformatet vægtet med 75 %, og kompleksiteten medfører en stigning i omkostninger med en potens på 1,2 (eksponent på 1,2). Omkostningen er altså ikke ligefrem proportional med kompleksiteten, men stigende. Som eksempel til understøtning af denne antagelse kan nævnes, at erfaringerne fra behandlingen af de åbne dokumentformater ODT og OOXML viste, at omkostningerne til at afslutte, hvorvidt data var migreret korrekt fra lukkede formater til de åbne eller mellem de åbne, steg voldsomt med anvendelsen af de komplekse funktioner, som udformaterne gav mulighed for at lagre resultatet af.

Omkostningen til at implementere migreringsprogrammellet i migreringssystemet er sat til 50 % af omkostningerne ved at finde og afprøve programmellet. Antagelsen er, at programmellet skal kunne automatiseres, således at migreringssystemet uden risiko for afbrydelser kan kalde det pågældende programmel, når et af dets formater skal migreres. Som eksempel kan nævnes, at mange af de programmer, som er bedst til at læse de pågældende indformater og skrive i de ønskede udformater, ikke umiddelbart er egnede til automatisering (f.eks. i form af batchkørsel uden GUI). Egne erfaringer og fra ESDH-leverancer viser, at der kræves en betydelig indsats for at få ”indkapslet” programmer, så man får en automatisk behandling.

Levetiden af migreringsprogrammellet er angivet i denne blok, og det er en vigtig antagelse, at programmellet som alt andet programmel har en begrænset levetid. Formatet i sig selv lever ofte meget længere end det konkrete programmel man i sin tid valgte at migrere med. Den angivne levetid på 8 år synes kort, men vor begrænsede erfaring giver os ikke mange eksempler på programmel, som efter 8 år stadig er i fuld drift. Bindinger til styresystemer m.m. har hidtil medført en forholdsvis kort levetid, om end den synes at være stigende.

## **Software provision developing software**

Her er omkostningen til at fremskaffe programmel til det enkelte format opgjort. Antagelsen er her, at det ikke har været muligt at fremfinde et programmel, som har kunnet læse indformatet og skrive i udformatet. Basisomkostningen er flere gange højere og eksponenten er også højere, idet kompleksitet anses for endnu vigtigere, når man selv skal udvikle programmel til formatet.

Dokumentationen vægter meget mere end ved brug af bestående programmel, ligesom en bestående implementering i åben kildekode mindsker omkostningen.

Omkostningen ved implementeringen i migreringssystemet er lavere end ved bestående programmel.

### **Software provision for migration system**

I denne blok er omkostningen til migreringssystemet angivet. Det er det centrale migreringssystem, som antages at kalde andet programmel for hvert format. Migreringssystemet skal således sørge for at styre processen fra ind- til udformat, således at migreringen kan foregå så automatisk og hurtigt som muligt, herunder med færrest mulige afbrydelser. Systemet skal således kunne parallelisere migreringen, håndtere nedbrud for enkelte opgaver, og genstarte disse. Omkostningen er skønnet til 80 personuger, og det forventes, at systemet kan bestå i 8 år.

### **Processing**

Her beregnes den tid, der tager at migrere hvert format, både for maskiner og for de personer, som skal overvåge forløbet, lave stikprøvekontrol og rette fejl. Beregningerne er baseret på formatets kompleksitet med uformateret tekst som basis. For hvert enkel delproces reguleres beregningen, men med det udgangspunkt kan det ikke undgås, at formater som afviger meget fra tekst får en misvisende beregning (eksempelvis kan video være et komplekst format at migrere, men forholdsvist hurtigt at udføre stikprøvekontrol på). En mere retvisende beregning vil kræve flere parametre for hvert format, hvilket vi ikke har kunnet overkomme i dette projekt.

Migrering er baseret på formatets kompleksitet, og indformat og udformat vægter lige meget. Eksponenten er under én, idet det antages, at behandlingstiden ikke er ligefrem proportional med kompleksiteten, men aftagende. Eksempelvis antages, at selv om et regnearksformat er 14 gange mere komplekst end et uformateret tekstformat tager det kun 8 gange længere tid at behandle regnearksformatet.

Overvågningen er angivet til 5 % per maskintime. Stikprøvekontrollen er på 1 promille af datamængden. Det antages at kontrol koster 16 timer per GB uformateret tekst. Således vil en datamængde på 1.000 GB medføre en stikprøve på 1 GB, hvilket vil tage 16 persontimer, svarende til 1 min/MB udtaget stikprøve.

Fejlretningen er baseret på, at der for uformateret tekst er 1 promille fejl per GB, og at denne stiger mere end proportionalt med formatets kompleksitet, samt at det tager 5 minutter at rette en fejl for uformateret tekst. Det bemærkes, at samvarians mellem stikprøvestørrelse og fejlrate ikke er medtaget i denne udgave.

### **Format Interpretation Factor**

Her opgøres for hvert formats dets "format interpretation factor", som er direkte baseret på formatets karakteristika. Altafgørende er basisværdien på en personuge per 100 % af formatets kompleksitet. Hertil kommer små tillæg for dokumentation og implementering.

Format interpretation factor bruges bl.a. til at opgøre, hvor megen arbejdskraft arkivet skal anvende til at holde øje med om formatet er blevet teknologisk forældet.

### **Migration Factor**

I rækkerne efter alle indformaterne er de tilsvarende data for udformaterne angivet. Eftersom alle udformater også er indformater, har de samme karakteristika, og de er primært angivet for gennemskuelighedens skyld.

Udformater skal også migreres til nye udformater. Med andre ord skal f.eks. et format som JPEG2000 også på et tidspunkt migreres. Modellen er på dette punkt meget forenklet og meget forsigtig, idet alle formater migreres samtidigt efter 8 år. Det skal altså forstås således, at der hvert 8. år migreres til formater af tilsvarende kompleksitet (f.eks. migreres JPEG2000 til et udformat af tilsvarende kompleksitet efter 8 år, og dette nye udformat migreres igen efter yderligere 8 år til et nyt udformat af tilsvarende kompleksitet, osv.).

For hvert format opgøres omkostningen til udvikling af programmet til at migrere det pågældende format, og summen for samtlige formater kaldes for "Total Migration Factor".

### **9.3 Arket PAIMAS**

Arket PAIMAS er opdelt i rækker af blokke for hver hovedgruppe af opgaver i PAIMAS, mens de enkelte omkostningsparametre er angivet i søjler.

Hovedgrupperne er Project Management, Information to be preserved and users (Content), Digital Objects and Standards Applied to These Objects, Security Conditions, Legal and Contractual Aspects, Transfer samt Validation.

Ovennævnte opgaver udgør grundlaget for at opgøre omkostningen til at udarbejde en afleveringsaftale, -en omkostning som i OAIS er placeret i Administration under Negotiate Submission Agreement.

For hver hovedgruppe er der fastsat en andel af de samlede omkostninger. Disse andele er bestemt ud fra det typiske mønster i det spørgeskema, som projektet udsendte samt ud fra de erfaringer, projektmedlemmerne har. For hver hovedgruppe er der en underopdeling på de opgaver, som indgår i hovedgruppen.

Eksempelvis er hovedgruppen 3.12.5. Security Conditions tildelt en andel på 10 %. Disse 10 % er igen opdelt på de tre grupper P23 & P25, P24 og P26 med hhv. 45 %, 10 % og 45 %.

Omkostningsparametrene er angivet i søjler, f.eks. Producer/Archive Strength. For hver parameter angives værdien Low, Medium eller High, gældende for alle forhold. For hver opgave angives, i hvilken grad den pågældende opgave er påvirket af omkostningsparameteret, altså uanset om dets værdi måtte være Low, Medium eller High.

Eksempelvis kan Producer/Archive Strength generelt være "Low", hvilket betyder at Producer ikke har megen styrke over Archive. For nogle opgaver har dette parameter stor påvirkning af omkostningen, medens det for andre enten ingen indflydelse har, eller lav.

Værdien af påvirkningen er fastsat for hver søjle. Eksempelvis er den for Producer/Archive Strength for Low hhv. 4 %, 6 % og 9 % (L, M, H); for Medium hhv. 20 %, 30 % og 45 % og for High hhv. 100 %, 150 % og 225 %.

Afhængigt af hvilke parametre, der har betydning for hvilke opgaver, fremkommer derved et omkostningstilæg, som altså er grundlaget for at opgøre omkostningen til at udarbejde en afleveringsaftale.

Foruden ovennævnte opgaver, er der opgaverne Actions in the Transfer Phase (T) og Actions in the Validation Phase (V). Disse indgår ikke som grundlag for at opgøre omkostninger til udarbejdelse af en afleveringsaftale, men er udmøntningen af dele af denne aftale.

## 10 Konklusioner

Overordnet mener vi, at metoden til at fremfinde omkostninger fungerer. Modellen som helhed er dog endnu ikke detaljeret nok til at give retvisende resultater for alle typer af materialer. Al empiri stammer fra offentlige arkivalier, hvilket medfører, at modellen p.t. er bedst til at skønne omkostninger netop for den type materiale.

Generelt set viser vores arbejde med omkostninger ved digital bevaring, at bevaringsinstitutioner er stærkt afhængige af at kunne bruge standardiserede løsninger, da det bliver meget dyrt, hvis de selv skal udvikle alle løsninger.

Hvad angår projektets første fase, som var udvidelse af modellen med OAIIS funktionen Ingest, har projektet anvendt forholdsvis meget tid på at analysere PAIMAS standarden. Analysearbejdet resulterede i en detaljeret liste over aktiviteter og en gennemgang af de variable, der påvirker omkostningerne for de enkelte aktiviteter i en aflevering. Variablerne var f.eks. projektets og samlingernes størrelse, samlingernes ensartethed og kompleksitet, både mht. indhold, struktur og format, betydningen af forhold omkring f.eks. lovgivning og sikkerhed.

En vigtig konklusion i arbejdet med Ingest, som spørgeskemaundersøgelsen bekræfter, er, at normalisering af formater udgør den største omkostning i Ingest, nemlig omtrent 2/3 af de samlede omkostninger. At normalisering tillige indebærer omkostningsfølsomme valg såsom migreringshyppighed (hvornår skal det gøres?), migreringsniveau (til hvilket format?), aktør (hvem skal udføre opgaven?) og fejltolerance (hvilke og hvor mange fejl tillades?), understreger at netop denne omkostning kræver et helt særligt fokus både af hensyn til omkostningsmodellens præcision, men også i forhold til valg af migreringsstrategi, jf. kapitel 6.2, Hvornår skal man migrere?, side 16.

Producenterne kan medvirke til at nedbringe omkostningerne til normalisering, hvis de indkalkulerer den fremtidige normalisering i deres daglige brug. Særligt hvorledes man vil migrere fra brugsformat til bevaringsformat.

Herudover peger projektets analyser på, at styrkeforholdet mellem producent og bevaringsinstitution har stor indflydelse på omkostningerne samt på fordelingen af disse, hvorfor det udgør et essentielt parameter i modellen.

Omkostningsmodellens metode er i projektets anden fase blevet videreudviklet.

En vigtig konklusion er, at valget af det digitale objekt (formatet), dets kompleksitet og mængde som beregningsmæssigt udgangspunkt, gør modellen potentielt generisk og dermed i stand til at beregne omkostningerne for forskellige digitale samlinger.

For at opnå retvisende resultater for alle typer af digitalt materiale, er der dog behov for at udvide med flere parametre for hver objekttype, f.eks. antal objekter.

En anden konklusion vedrørende metoden er, at den er afhængig af en række skøn, såsom formatlevetid og dermed migreringsfrekvens. For at imødegå dette problem giver modellen mulighed for angivelse af andre værdier, end dem, CMDP foreslår.

Med et solidt metodisk udgangspunkt mener vi at modellens vigtigste udestående nu udgøres af dens manglende præcision. Vi mener at denne kan forøges ved udvidelse med flere parametre samt test af modellen vha. indhentede omkostningsdata.

Implementeringen af modellen i regneark har vist sig at være problematisk. Ønsket om gennemsigtighed og præcision kan ikke opfyldes samtidigt. I en kommende udgave af modellen med større præcision bliver det derfor nødvendigt at give køb på gennemsigtigheden og angive formlerne i kode.

Desuden mangler en egentlig brugervejledning til regnearket eller etablering af en brugergrænseflade, da modellen i sin nuværende form er meget vanskelig for udefrakommende at anvende.

Med henblik på videreudvikling af modellen har projektet holdt sig ajour med den internationale udvikling af omkostningsmodeller for digital bevaring. Muligheden for at dette fokus fører til formelle eller uformelle samarbejder i fremtiden er til stede, idet det vurderes, at både interessen for og nødvendigheden af at få større vished på dette område generelt anses for stor.

Alt i alt har CMDP projektet (projekt #1 og #2) forbrugt 24 mandmåneder (1,2 mio. kr.). Til sammenligning har LIFE projektet forbrugt 42 mandmåneder (4,6 mio. kr.), KRDS projektet 14 mandmåneder (1,0 mio. kr.) og NASA CET har siden 2002 brugt 216 mandmåneder (13,5 mio. kr.).

Såfremt CMDP skal fuldstændiggøres både metodisk og angående præcision samt udvides til at inkludere ikke blot alle OAIS funktionerne, men ligeledes indeholde mulighed for f.eks. beregning af forskellige bevaringsstrategier og forskellige samlinger, kræves flere ressourcer.



## 11 Litteraturliste

Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., & Wheatley, P. 2008. The LIFE2 Final Project Report. <http://eprints.ucl.ac.uk/11758/>

Beagrie, N., Chruszcz, J., & Lavoie, B. 2008. Keeping Research Data Safe. A Cost Model and Guidance for UK Universities, Copyright HEFCE 2008.  
<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>

Beagrie, N., Lavoie, B., Woollard, M., Keeping Research Data Safe 2, Final Report, Charles Beagrie Limited, 2010, <http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>

Björk, B.C., Economic evaluation of LIFE methodology, Research report. LIFE project, 2007, p. 5-6, 18-19.  
<http://eprints.ucl.ac.uk/7684/>

Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation, Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008. <http://brtf.sdsc.edu/>

Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Sustainable Economics for a Digital Plant: Ensuring Long-Term Access to Digital Information, Final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010, [http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)

Boehm, B., Abts, C., Winsor Brown, A., Chulani, S., Clark, B. K., Horowitz, E., Madachy, R., Reifer, D.J., Steece, B. 2000. Software cost estimation with COCOMO II. Englewood Cliffs, NJ:Prentice-Hall, ISBN 0-13-026692-2

CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval), Review of the State of the Art, 2007, [http://www.casparpreserves.eu/Members/cclrc/Deliverables/review-of-state-of-the-art-1/at\\_download/file.pdf](http://www.casparpreserves.eu/Members/cclrc/Deliverables/review-of-state-of-the-art-1/at_download/file.pdf)

Center for Research Libraries (CRL) & RLG OCLC Programs, Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist, Version 1.0., 2007 <http://www.dcc.ac.uk/tools/trustworthy-repositories/>

Charles Beagrie Limited and JISC, Keeping Research Data Safe Factsheet,  
[http://www.beagrie.com/KRDS\\_Factsheet\\_0910.pdf](http://www.beagrie.com/KRDS_Factsheet_0910.pdf)

Construx Estimate (værktøj til estimering af omkostninger forbundet med udvikling af programmer),  
<http://www.construx.com/Page.aspx?nid=68>

Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, Blue Book, 2002, (ISO14721:2003).  
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

Digital Curation Centre (DCC) & Digital Preservation Europe (DPE), Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), Version 1.0., 2007 <http://www.repositoryaudit.eu/>

Egger, A., Shortcomings of the Reference Model for an Open Archival Information System (OAIS). TCDL Bulletin Vol. 2(2). 2006. <http://www.ieee-tcdl.org/Bulletin/v2n2/egger/egger.html>

Foxite.com, Software Development Costs and Choice of Tools,

<http://www.foxite.com/articles/read.aspx?id=37&document=software-development-costs-and-choice-of-tools>

Grace, S., Knight, G. Montague, L., Investigating the Significant Properties of Electronic Content over Time (InSPECT), Final Report, 2009, <http://www.significantproperties.org.uk/>

Higgins, S. & Boyle, F., Response to CCSDS's comments on the OAIS Five-year review: recommendations for update, The Digital Curation Centre (DCC) and The Digital Preservation Coalition (DPC), 2006, <http://www.dpconline.org/events/previous-events/427-oais-5-year-review-follow-up>

Karjalainen, M., Large-scale migration to an open source office suite: An innovation adoption study in Finland, Academic Dissertation, Faculty of Information Sciences of the University of Tampere, 2010, <http://acta.uta.fi/pdf/978-951-44-8216-8.pdf>

Hole, B., Lin, L., McCann, P. & Wheatley, P., LIFE3: A Predictive Costing Tool for Digital Collections, In: Proceedings of iPRES 2010, 7th International Conference on Preservation of Digital Objects, Austria <http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/hole-64.pdf>

Marshall, C., How People Manage Information over a Lifetime, 2008, <http://www.csd.tamu.edu/~marshall/PIM%20Chapter-Marshall.pdf>

McLeod, R., Wheatley, P., & Ayris, P. 2006. Lifecycle information for e-literature: full report from the LIFE project. Research report. <http://eprints.ucl.ac.uk/1854/>

NASA Cost Estimation Toolkit (CET). <http://opensource.gsfc.nasa.gov/projects/CET/CET.php>

Nationaal Archief, Digital Preservation Testbed, Cost of Digital Preservation, 2005.

<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf>. Testbed computational model:

<http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=185&categorie=6>

Both links are broken.

OECD, International Standard Cost Model Manual, International SCM Network to reduce administrative burdens, 2004, <http://www.oecd.org/dataoecd/32/54/34227698.pdf>

Erhvervs- og Selskabsstyrelsen ([www.amvab.dk](http://www.amvab.dk)) vedligeholder forskellig dokumentation og en dansk udgave: Manual til Aktivitetsbaseret Måling af Virksomhedernes Administrative Byrder (AMVAB):

<http://www.amvab.dk/graphics/DK/Metode%20info/AMVAB-manual-okt2010.pdf>

Officeshots.org, Creates transparency in the Office market, <http://www.officeshots.org/pages/about>

Planets. 2007. Report on tool and service approach, pp. 1, 32-33. [http://www.planets-project.eu/docs/reports/Planets\\_PA4-D1\\_ReportOnToolAndServiceApproach-Final\\_Public.pdf](http://www.planets-project.eu/docs/reports/Planets_PA4-D1_ReportOnToolAndServiceApproach-Final_Public.pdf)

Rosenthal, D.S.H., Robertson T., Lipkis, T., Reich V., Morabito, S., Requirements for Digital Preservation Systems. A Bottom-Up Approach. D-Lib Magazine, Vol. 11 No. 11, ISSN 1082-9873, 2005,

<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>

Sanett, S. 2002. Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity. College & Research Libraries 63, 5, pp. 388-404.

Shenton, H., Life Cycle Collection Management, LIBER QUARTERLEY, 13, pp 294-272, ISSN 1435-5205, 2003. <http://liber.library.uu.nl/publish/articles/000033/article.pdf>

Software Technology Support Center (STSC) Cost Analysis Group, Software Development Cost Estimating Guidebook, [http://www.stsc.hill.af.mil/consulting/sw\\_estimation/SoftwareGuidebook2010.pdf](http://www.stsc.hill.af.mil/consulting/sw_estimation/SoftwareGuidebook2010.pdf)

Stephens, A., The application of life cycle costing in libraries. British Journal of Academic Librarianship 3, pp. 82-88, 1988.

Stephens, A., The application of life cycle costing in libraries: a case study based on acquisition and retention of library materials in the British Library. IFLA journal, 1994, 20(2) pp. 130-140

Stocholm, J. L., SaaS og dokumentformater, Version2, 2010, <http://www.version2.dk/artikel/16409-saas-og-dokumentformater>

Watson, J., The LIFE project research review. Mapping the landscape, riding a life cycle, 2005. <http://eprints.ucl.ac.uk/1856/>

Wheatley, P.: Costing the Digital Preservation Lifecycle More Effectively, iPRES2008 Conference, 2008 [http://www.bl.uk/ipres2008/presentations\\_day1/19\\_Wheatley.pdf](http://www.bl.uk/ipres2008/presentations_day1/19_Wheatley.pdf)

## 11 Ordliste

Danish	English	Definition, dansk	Definition, English
<b>Administrationsfunktionsheden</b>	<b>Administration</b>	I OAIS den funktionsenhed, som indeholder de services og funktioner, der er behov for i forbindelse med styringen af det OAIS kompatible arkivs andre funktionsenheder og som servicerer dets overordnede operationelle niveau. Administrationsfunktionerne udarbejder retningslinjer for afleveringsbestemmelser, kontrollere afleveringer for at sikre sig, at de overholder afleveringskravene og vedligeholder konfigurationsstyringen af hardware og software i test- og bevaringssammenhænge. Administrationsfunktionen tilvejebringer ligeledes systemudviklerfunktioner med henblik på at overvåge og forbedre de arkivtekniske forhold samt at styre, migrere, ajourføre og indberette om arkivets indhold. Funktionen er også ansvarlig for at etablere og vedligeholde arkivstandarder og -politikker samt at yde brugersupport.	This entity provides the services and functions for the overall operation of the archive system. Administration functions include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It also provides system engineering functions to monitor and improve archive operations, and to inventory, report on, and migrate/update the contents of the archive. It is also responsible for establishing and maintaining archive standards and policies, providing customer support, and activating stored requests.
<b>Afleveringsrevisionsfunktionen</b>	<b>Audit Submission</b>	I OAIS den funktion under Administration, som kontrollerer, at SIP eller AIP opfylder specifikationerne i afleveringsaftalen. Denne funktion modtager AIP/SIP reviews fra Bevaringsplanlægningsfunktionen. Revisionsprocessen skal kontrollere, at data opfylder kravene i arkivet. Den skal kontrollere, at der er tilstrækkelig repræsentationsinformation og bevaringsbeskrivende information for at sikre, at oplysninger om indholdet er forståelige og uafhængigt anvendelige til den relevante brugergruppe. Graden af formalitet af denne gennemgang vil variere afhængigt af de interne arkiv politikker. Høringsprocessen kan bestemme, at visse dele af SIP ikke er egnede til optagelse i arkivet og skal forelægges på ny. Høringsmetoder omfatter potentielt stikprøver, periodisk gennemgang og peer review.	The Audit Submission function will verify that submissions ( <i>SIP or AIP</i> ) meet the specifications of the Submission Agreement. This function receives <i>AIP/SIP reviews</i> from Preservation Planning and may also involve an outside committee (e.g., science and technical review). The audit process must verify that the quality of the data meets the requirements of the archive and the review committee. It must verify that there is adequate Representation Information and PDI to ensure that the Content Information is understandable and independently usable to the Designated Community. The formality of the review will vary depending on internal archive policies. The Audit process may determine that some portions of the SIP are not appropriate for inclusion in the archive and must be resubmitted or excluded. An <i>audit report</i> is provided to Ingest. After the audit process is completed, any <i>liens</i> are reported to the Producer, who will then resubmit the SIP to Ingest or <i>appeal</i> the decision to Administration. After the audit is completed, a <i>final ingest report</i> is prepared and provided to the Producer and to Negotiate Submission Agreement. Audit methods potentially include sampling, periodic review, and peer review.
<b>Afleveringsaftale</b>	<b>Submission Agreement</b>	I OAIS en aftale, der indgås mellem producent og arkiv i forbindelse med en aflevering. Aftalen definerer alle relevante forhold i forbindelse med et afleveringsprojekt, herunder hvilke data og anden dokumentation skal bevares, lov-, kontrakt- og sikkerhedsmæssige forhold, overførsel og godkendelsesprocedurer, samt budget og tidsplan.	An agreement (OAIS) concluded between the producer and the archive regarding a submission. The agreement defines all matters pertaining to a submission project, including which data and documentation to be preserved, legal, contractual and security aspects, procedures for transfer and admission, budget and time schedule.
<b>Afleveringspakke</b>	<b>Submission Information</b>	I OAIS de informationspakker som producenten afleverer til arkivet.	An Information Package that is delivered by the Producer to the OAIS for use in the con-

Danish	English	Definition, dansk	Definition, English
	Package (SIP)		struction of one or more AIPs.
Antal computere	Number of Computers (NoC)	I CMDP er NoC det antal computere, der benyttes til migreringen.	In CMDP NoC refers to the number of computers used in a migration processing.
Antal sider	Number of Pages (NoP)	I CMDP henviser NoP til det antal sider det er nødvendigt for en datalog til at læse og forstå for at kunne udvikle migreringssoftware.	In CMDP NoP refers to the number of pages it is necessary for a computer scientist to read and comprehend in order to be able to develop migration software for the format.
Arkivinformationsopdateringsfunktionen	Archival Information Update	I OAIS den funktion under Administration, der tilvejebringer en mekanisme til opdatering af indholdet af arkivet. Den modtager ændringsanmodninger, procedurer og værktøjer. Den opdaterer ved at sende anmodning til Access, ajourfører indholdet af de deraf følgende DIPs og genindsender dem som SIPs.	The Archival Information Update function provides a mechanism for updating the contents of the archive. It receives <i>change requests, procedures</i> and <i>tools</i> from Manage System Configuration. It provides updates by sending a <i>dissemination request</i> to Access, updating the contents of the resulting <i>DIPs</i> and resubmitting them as <i>SIPs</i> to Ingest.
Arkiveringspakke	Archival Information Package (AIP)	I OAIS de informationspakker, bestående af indholdsinformation og bevaringsbeskrivelsesinformation, som arkiveres i arkivet.	An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.
Basis omkostning software udvikling	Base cost software development	I CMDP et af de parametre, der anvendes til beregning af hvor meget tid det tager at udvikle migreringssoftware. Sat til 4 personuger for uformatet tekst	In CMDP one of the parameters used in the calculation of how much time it takes to develop migration software. Default is 4 person weeks (pw) for unformatted text
Beskrivende information	Descriptive Information	I OAIS den information som lagres i Data Management og som bruges til at finde, bestille og hente informationspakkerne. Kan bestå af både en unik identifikator samt titel og andre referencer.	The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers.
Bevaringsbeskrivende information	Preservation Description Information	I OAIS den der er nødvendig for tilfredsstillende bevaring af bevaringsværdige data. Den består af proveniensinformation, referenceinformation, integritetsinformation og kontekstinformation.	The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.
Bevaringsplanlægningsfunktionseheden	Preservation Planning	I OAIS den funktionsenhed, der overvåger arkivmiljøet og rådgiver om, hvordan det sikres, at digitale arkivalier forbliver tilgængelige over tid, selvom det oprindelige computermiljø forældes. Enheden evaluerer arkivets indhold og rådgiver om, hvornår det er nødvendigt at ty til f.eks. migrering (se migreringsstrategi) samt hvordan dette bør gøres. Enheden udarbejder også anbefalinger ift. arkivstandarder og -politikker og overvåger teknologiforandringer og ændringer i brugerne adfærd. Bevaringsplanlægningsfunktionen designer også informationspakker til specifikke datamodeller. Endelig fremstiller den detaljerede bevaringsplaner, software prototyper og testplaner med henblik på at muliggøre implementeringen af bevaringsinstitutionens bevaringsstrategi.	This entity provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. Preservation Planning also designs IP templates and provides design assistance and review to specialize these templates into SIPs and AIPs for specific submissions. Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of

Danish	English	Definition, dansk	Definition, English
			Administration migration goals.
<b>Bitbevaring</b>	<b>Bit preservation</b>	Aktiviteter, der sikrer, at bits forbliver intakte og læsbare over tid. Disse aktiviteter varetages i OAIS af funktionsenheden Archival Storage.	Bit preservation ensures that the bits remain intact and readable.
<b>Bitbevarings-funktions-enheden</b>	<b>Archival Storage</b>	I OAIS den funktionsenhed, som sørger for bevaring samt vedligeholdelse og fremfinding af materialer. Enhedens funktioner inkluderer modtagelse af AIPer fra indtagningsfunktionen, tilføjelse af disse til permanent bitlagring, styring af hierarchical storage management (HSM), udskiftning af lagringsmedier, udførelse af fejltjek, tilstedeværelse af katastrofeberedskab og servicering af Access.	This entity provides the services and functions for the storage, maintenance and retrieval of AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfil orders.
<b>Bruger</b>	<b>Consumer</b>	I OAIS de brugere eller systemer, som interagerer med arkivet.	The role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISs, as well as internal OAIS persons or systems.
<b>Brugergruppe</b>	<b>Designated Community</b>	I OAIS de brugergrupper, som arkivet har specificeret, som de der skal kunne bruge arkivets materialer.	An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.
<b>Brugerover-vågning</b>	<b>Monitor Designated Community</b>	I OAIS den funktion under Preservation Planning, der følger udviklingen i producenters og brugeres krav til arkivets services og teknologi, og rapportere ændringer til Develop Preservation Strategies and Standards.	The Monitor Designated Community function interacts with archive Consumers and Producers to track changes in their <i>service requirements</i> and available <i>product technologies</i> . Such requirements might include data formats, media choices, preferences for software packages, new computing platforms, and mechanisms for communicating with the archive.
<b>CMDP</b>	<b>Cost Model for Digital Preservation (CMDP)</b>	Omkostningsmodel til beregning og estimering af omkostninger for bevaring af digitale materialer over tid.	Generic cost model for calculating and estimating cost of long-term digital preservation.
<b>Data</b>	<b>Data</b>	I OAIS information, der er udtrykt på en formaliseret måde og som kan fortolkes. Eksempler på data er bitsekvenser, en tabel med numre, tegn på en side, en lydoptagelse af en person der taler, en prøve fra en månesten.	A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.
<b>Dataobjekt</b>	<b>Data object</b>	I OAIS enten et fysisk eller et digitalt objekt.	Either a Physical Object or a Digital Object.
<b>Datamodtagelsesfunktionen</b>	<b>Receive Data</b>	I OAIS den funktion under Archival Storage, som modtager anmodning om bitbevaring og AIPer fra Ingest og overfører pakkerne til arkivets det permanente lager.	The Receive Data function receives a storage request and an AIP from Ingest and moves the AIP to permanent storage within the archive. The transfer request may need to indicate the anticipated frequency of utilization of the data objects comprising the AIP in order to allow the appropriate storage devices or media to be selected for storing the AIP. This function will select the media type, prepare the devices or volumes, and perform the physical transfer to the Archival Storage volumes. Upon completion of the transfer, this function sends a storage confirmation message to Ingest, in-

Danish	English	Definition, dansk	Definition, English
			cluding the storage identification of the AIPs.
<b>Datastyrings-funktions-enheden</b>	<b>Data Man-agement</b>	I OAIS den funktionsenhed, der stiller services til rådighed med henblik på at udfylde, vedligeholde og tilgå både fremfindingsinformation, som identificerer og dokumenterer arkivalierne, og de administrative data, der bruges til at styre arkivet. Datastyringsfunktionerne inkluderer administration af arkivdatabasefunktionerne (vedligeholdelse af <i>sche-maer</i> og <i>view</i> definitioner samt den referentielle integritet), udførelse af databaseopdateringer (af fremfindingsinformation og administrative data), udførelse af databasefore-spørgsler med henblik på generering af resultatsæt og udarbejdelse af rapporter på baggrund af disse.	This entity provides the services and functions for populating, maintaining, and accessing both Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive. Data Management functions include administering the archive database functions (maintaining schema and view definitions, and referential integrity), performing database updates (loading new descriptive information or archive administrative data), performing queries on the data management data to generate result sets, and producing reports from these result sets.
<b>Digital bevaring</b>	<b>Digital Pres-ervation</b>	Bevaring af digitalt materiale. Materialet kan være skabt digitalt eller være resultat af en digitalisering af papirdokumenter, lyd eller levende billeder i analog form m.v.	Digital preservation designates the methods and systems which are needed to assure access to digital materials over time. Digital preservation may be divided into bit preservation, which must ensure that the bits remain intact and readable, and logical or functional preservation, which must ensure that the bits remain comprehensible. The concept digital preservation is also used in a broader context for the whole life cycle process from the production of the digital materials, acquisition by the OAIS archive, and long term preservation, until they are made available to the users.
<b>Dokumenta-tionsgrad</b>	<b>Degree of documenta-tion</b>	I CMDP afspejler det hvor godt dataformater er dokumenteret, og detaljeringsgrad af dokumentationen. Fx: En høj detaljeringsgrad forudsætter at producenten har dokumentation for alle data og alle formater, som f.eks. dokumenterer, at en bestemt fil er i et Word Perfect 5.1 format.	In CMDP this expresses how well the data formats are documented, what formats do data belong to, and how detailed are these formats documented. E.g.: A high level of documentation requires that the producer has documentation for all data and all formats, such as documenting that a certain file is in a Word Perfect 5.1 format and having documentation for that specific format.
<b>Etabler stan-darder og politikker-funktionen</b>	<b>Establish Standards and Policies</b>	I OAIS den funktion under Administration, som er ansvarlig for at etablere og vedligeholde standarder og politikker for hele arkivet, på baggrund af de overordnede retningslinjer (formål og budget) udstukket af Management. I relation til Ingest sender funktionen format og dokumentationsstandarder til Ingest samt de procedurer som skal følges i forbindelse med Ingest.	The Establish Standards and Policies function is responsible for establishing and maintaining the archive system standards and policies. It receives budget information and policies such as the OAIS charter, scope, resource utilization guidelines, and pricing policies from Management. It provides Management with periodic reports. It receives recommendations for archive system enhancement, and proposals for new archive data standards from Preservation Planning. It also receives performance information and archive holding inventories from Manage System Configuration. Based on these inputs, archive standards and policies are established and sent to other Administration functions and the other Functional Entities for implementation. The standards include format standards, documentation standards and the procedures to be followed during the Ingest process. It provides approved standards and migration goals to Preservation Planning. This function will also

Danish	English	Definition, dansk	Definition, English
			develop storage management policies (for the Archival Storage hierarchy), including migration policies to assure that archive storage formats do not become obsolete, and database administration policies. It will develop disaster recovery policies. It will also determine security policies for the contents of the archive, including those affecting Physical Access Control and the application of error control techniques throughout the archive.
<b>Fejlhåndtering</b>	<b>Error Handling (EH)</b>	I CMDP den tid det tager manuelt at rette fejlbehæftede filer som følge af en en migreringshandling.	In CMDP the time it takes to manually correct flawed files in a migration processing action.
<b>Forhandling af afleveringsaftale</b>	<b>Negotiate Submission Agreement</b>	I OAIS den funktion under Administration, der forhandler afleveringsaftalen mellem producent og arkiv på plads.	The Negotiate Submission Agreement function solicits desirable archival information for the OAIS and negotiates <i>Submission Agreements</i> with Producers. This function also negotiates a <i>data submission schedule</i> with the Producer. It maintains a calendar of expected Data Submission Sessions that will be needed to transfer one or more SIPs to the OAIS and the resource requirements to support their ingestion. This function receives <i>AIP/SIP templates</i> and <i>customization advise</i> from Preservation Planning and sends <i>SIP designs</i> and <i>SIPs</i> to the Audit Submission function as part of the submission approval process. The data submission formats and procedures must be clearly documented in the archive's data submission policies, and the deliverables must be identified by the Producer in the Submission Agreement.
<b>Format forståelsesfaktor</b>	<b>Format Interpretation factor</b>	I CMDP angiver faktoren hvor svært et format er at forstå, hvilket afhænger af kompleksiteten, dokumentationen og implementeringen. Komplexiteten er baseret på den tid som det tager målt i personuger at identificere og læse formatspecifikationen og anden relevant dokumentation. Der er et tillæg, hvis dokumentationen ikke er 100 % tilgængelig, dækkende og implementerbar, ligeledes er der tillæg hvis implementering ikke findes 100% i åben kildekode, er udbredt og i aktuel brug. Som udgangspunkt er det kompleksiteten, der udgør det vigtigste aspekt af forståelsesfaktoren.	In CMDP it denotes how difficult a format is to understand. It depends on the complexity, the documentation and the implementation. The complexity is based on the time it takes, measured in person-weeks, to identify and read the format specifications and any other relevant documentation. A addition is made if the documentation is not 100% available, covering and implementable, likewise additions are made if the format is not 100% implemented is open source, widespread and in current usage. The complexity level is by default the most important part of the format interpretation factor.
<b>Formategenskaber</b>	<b>Format characteristics</b>	For hvert format er angivet dets kompleksitet, dokumentation og implementering målt i forhold til formatet og uformateret tekst.	The characteristics of each format is given by its complexity, documentation and implementation measured relatively to unformatted text
<b>Formaters forventede levetid</b>	<b>Format Life Time (Expectancy)</b>	CMDP estimerer et formatters forventede levetid ud fra dets kompleksitet, dokumentation og udbredelse. I denne udgave af CMDP er der grundet usikkerhed sat en længste levetid for alle formater på 8 år.	In CMDP the average estimated lifetime of a format has been calculated, based on complexity, documentation and implementation. In this version of CMDP due to uncertainty the default max life time is set to 8 years.
<b>Funktionel bevaring</b>	<b>Functional preservation</b>	I digital bevaring er funktionel – eller logisk – bevaring en sikring af, at den bevarede fil kan fortolkes over tid.	Functional or logical preservation ensures that the bits remain comprehensible in the long term.
<b>Generér AIP</b>	<b>Generate AIP</b>	I OAIS den funktion under Ingest, der genere-	The <b>Generate AIP</b> function transforms one or



Danish	English	Definition, dansk	Definition, English
		rer AIPer, herunder omformer migrerer SIPer til AIPer. Omformningen Migreringen af pakker kan indebære, at formater og/eller strukturer migreres (normaliseres) til godkendte bevaringsformater og/eller strukturer, at repræsentationsdata ændres og/eller at informationen i pakkerne reorganiseres. Funktionen kan også bede Data Management om yderligere information, der er nødvendig for beskrive pakken fuldstændigt.	more <i>SIPs</i> into one or more <i>AIPs</i> that conform to the archive's <i>data formatting and documentation standards</i> . This may involve file format conversions, data representation conversions or reorganization of the content information in the <i>SIPs</i> . The Generate AIP function may issue <i>report requests</i> to Data Management to obtain <i>reports</i> of information needed by the Generate AIP function to produce the Descriptive Information that completes the AIP. This function sends <i>SIPs or AIPs for audit</i> to the Audit Submission function in Administration, and receives back an <i>audit report</i> .
<b>Generér beskrivende information</b>	<b>Generate Descriptive Information</b>	I OAIS den funktion under Ingest, der genererer beskrivende information, dvs. metadata og dokumentation om informationspakkerne. Den beskrivende information trækkes fra pakken selv eller tilføjes pakken fra Data Management.	The Generate Descriptive Information function extracts Descriptive Information from the <i>AIPs</i> and collects <i>Descriptive Information</i> from other sources to provide to Coordinate Updates, and ultimately Data Management. This includes metadata to support searching and retrieving <i>AIPs</i> (e.g., who, what, when, where, why), and could also include special browse products (thumbnails, images) to be used by Finding Aids.
<b>Generer rapport funktionen</b>	<b>Generate Report</b>	I OAIS den funktion under Data Management, der modtager bestillinger på rapporter fra Ingest, Access eller Administration, og som udfører forespørgsler og andre processer, som er nødvendige for at levere rapporten.	The Generate Report function receives a report request from Ingest, Access or Administration and executes any queries or other processes necessary to generate the report that it supplies to the requester. Typical reports might include summaries of archive holdings by category, or usage statistics for accesses to archive holdings. It may also receive a report request from Access and provides descriptive information for a specific AIP.
<b>Indflydelse på brugere</b>	<b>Influence on Consumption</b>	I CMDP den indflydelse OAIS arkivet har på hvilke formater brugerne bruger.	In CMDP the influence the OAIS archive has on what formats the users use. We assume, for example, that the entity Monitor Designated Community depends on how much influence the archive has on the production and use of formats: The more influence, the fewer costs.
<b>Indflydelse på produktion</b>	<b>Influence on Production</b>	I CMDP den indflydelse OAIS arkivet har på hvilke formater producenterne bruger.	In CMDP the influence the OAIS archive has on what formats the producers use. We assume, for example, that the entity Monitor Designated Community depends on how much influence the archive has on the production and use of formats: The more influence, the fewer costs.
<b>Indformat</b>	<b>Source format</b>	I CMDP det format, som skal migreres.	In CMDP the format to be migrated.
<b>Indholds-information</b>	<b>Content Information</b>	I OAIS den information, der skal bevares. Indholdsinformationen består af dataobjekt og repræsentationsinformation. Fx en tabel med tal der viser og er forståelig som temperaturer. Ikke dokumentation, der beskriver tabellens historie og oprindelse eller hvordan tallene relaterer til andre observationer mm.	The set of information that is the original target of preservation. It is an Information Object comprised of its Content Data Object and its Representation Information. An example of Content Information could be a single table of numbers representing, and understandable as, temperatures, but excluding the documentation that would explain its history and origin, how it relates to other observations, etc.

Danish	English	Definition, dansk	Definition, English
<b>Indholdskompleksitet</b>	<b>Content complexity</b>	I CMDP datas semantiske kompleksitet - hvor svære data er at forstå. Fx: Biomolekylær data har en større kompleksitet end en opgørelse af møbler.	I CMDP the semantic complexity of the data, how hard the data are to understand. E.g.: Understanding biomolecular data is of a higher complexity than understanding an inventory of furniture
<b>Indholdsvariation</b>	<b>Content variation</b>	I CMDP datas semantiske variation - hvor forskellige områderne er. Fx: Forståelse af en aflevering, der består af biomolekylære data, geografiske data og russisk poesi har en større kompleksitet end bare at forstå russisk poesi.	I CMDP the semantic variation of the data, how diverse the areas are. E.g.: Understanding a submission consisting of biomolecular data, geographical data and Russian poetry has a higher complexity than just understanding Russian poetry.
<b>Indtagning</b>	<b>Ingest</b>	I OAIS den funktionsenhed, som yder de services, som gør det muligt for et arkiv, at modtage informationspakker fra producenter og forberede dem til videre bevaring. Funktionerne inkluderer: Modtagelse af informationspakker (SIP), kvalitets sikring af informationspakkerne (SIP), generering af arkiveringsegnede informationspakker (AIP), udtrækning af beskrivende information til brug i arkivdatabaserne, koordinering af ajourføringer i forhold til Archival Storage og Data Management.	This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs, generating an Archival Information Package (AIP) which complies with the archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the archive database, and coordinating updates to Archival Storage and Data Management.
<b>Information</b>	<b>Information</b>	I OAIS alle typer viden, der kan udveksles. I en udveksling er information repræsenteret af data. Et eksempel på information er en bitsekvens (data) og den tilhørende beskrivelse af hvordan bitsekvensen skal fortolkes som numre, der udtrykker temperatur observationer målt i celsius grader (repræsentationsinformationen). Svarer til OAIS begrebet indholdsinformation.	Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret a string of bits as numbers representing temperature observations measured in degrees Celsius (the representation information).
<b>Informationspakke</b>	<b>Information package</b>	I OAIS består en informationspakke af indholdsinformation og bevaringsbeskrivelsesinformation, dvs. den information, der er nødvendig for at bevare indholdsinformationen. Informationspakkerne selv beskrives af pakkeinformation, der identificerer og afgrænser indholdsinformation og bevaringsbeskrivelsesinformationen.	The Content Information and associated Preservation Description Information which is needed to aid in the preservation of the Content Information. The Information Package has associated Packaging Information used to delimit and identify the Content Information and Preservation Description Information.
<b>Integritetsinformation</b>	<b>Fixity information</b>	I OAIS garantien for, at indholdet i en informationspakke ikke er blevet ændret på udokumenteret vis. Et eksempel på integritetsinformation er resultatet af Cyclical Redundancy Check (CRC) på en fil.	The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information object has not been altered in an undocumented manner. An example is a Cyclical Redundancy Check (CRC) code for a file.
<b>Kalenderuge</b>	<b>Calendar week</b>	I CMDP sat til 168 timer.	In CMDP the default is 168 hours.
<b>Kontekstinformation</b>	<b>Context Information</b>	I OAIS den information, der dokumenterer indholdsinformationens relationer til dets miljø. Fx hvorfor indholdsinformationen blev skabt og hvordan den relaterer til anden indholdsinformation.	The information that documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects.
<b>Koordinér opdateringer funktionen</b>	<b>Coordinate Updates</b>	I OAIS den funktion under Ingest, der koordinerer opdateringer af informationspakkeres metadata og dokumentation og som overfører informationspakkerne til Archival Storage og	The Coordinate Updates function is responsible for transferring the AIPs to Archival Storage and the <i>Descriptive Information</i> to Data Management. Transfer of the AIP includes a

Danish	English	Definition, dansk	Definition, English
		sender metadata og anden dokumentation til Data Management.	<i>storage request</i> and may represent an electronic, physical, or a virtual (i.e., data stays in place) transfer. After the transfer is completed and verified, Archival Storage returns a <i>storage confirmation</i> indicating (or verifying) the storage identification information for the AIP. The Coordinate Updates function also incorporates the storage identification information into the Descriptive Information for the AIP and transfers it to the Data Management entity along with a <i>database update request</i> . In return, Data Management provides a <i>database update response</i> indicating the status of the update. Data Management updates may take place without a corresponding Archival Storage transfer when the SIP contains Descriptive Information for an AIP already in Archival Storage.
<b>Kvalitetssikring</b>	<b>Quality Assurance</b>	I OAIS den funktion under Ingest, der kontrollerer at alle bits er intakte og fuldstændige efter overførsel af SIP'er til arkivet. Kontrollen sker typisk vha. checksums algoritmer. Kvalitetssikringen omfatter ikke kontrol af pakkens indholdsmæssige fuldstændighed.	The Quality Assurance function validates ( <i>QA results</i> ) the successful transfer of the <i>SIP</i> to the staging area. For digital submissions, these mechanisms might include Cyclic Redundancy Checks (CRCs) or checksums associated with each data file, or the use of system log files to record and identify any file transfer or media read/write errors.
<b>Ledelse</b>	<b>Management</b>	I OAIS den rolle, som fastsætter de overordnede OAIS politikker som en del af en større strategi.	The role played by those who set overall OAIS policy as one component in a broader policy domain.
<b>Løn, høj</b>	<b>Salary, high</b>	I CMDP et af de tre lønniveauer. Sat til 900 euro / personuge.	In CMDP one of the three salary levels. The default value is 900 E/PW.
<b>Løn, lav</b>	<b>Salary, low</b>	I CMDP et af de tre lønniveauer. Sat til 300 euro / personuge.	In CMDP one of the three salary levels. The default value is 300 E/PW.
<b>Løn, middel</b>	<b>Salary, medium</b>	I CMDP et af de tre lønniveauer. Sat til 600 euro / personuge.	In CMDP one of the three salary levels. The default value is 600 E/PW.
<b>Maskintid</b>	<b>Machine processing time</b>	I CMDP er maskintiden baseret på kompleksiteten af indformat og udformat i forhold til uformateret tekst. Standardværdien for uformateret tekst er 1 min/GB.	I CMDP the machine time is based on the complexity of the source and destination formats relative to unformatted text. Default for unformatted text is 1 min/GB.
<b>MB/s</b>	<b>MB/s</b>	Megabytes pr. sekund.	Megabytes per second.
<b>Migreringsfaktor</b>	<b>Migration factor</b>	I CMDP angiver migreringsfaktoren omkostningen i personuger til at tilvejebringe programmet til migrering af et format	In CMDP the migration factor is the cost in person weeks of providing software for migration of a format.
<b>Migreringsfrekvens</b>	<b>Migration frequency</b>	I CMDP er migreringsfrekvensen baseret på formatets levetid, men af hensyn til enkelhed i modellen og forsigtighed sat til hvert 8. år for alle formater baseret på levetiden af migreringssoftwaren.	In CMDP the frequency of migration is based on average estimated lifetime of formats, which we for simplicity and to be conservative have set to be 8 years, based on the lifetime of the migrations software.
<b>Modtag afleveringsfunktionen</b>	<b>Receive Submission</b>	I OAIS den funktion under Ingest, der modtager afleveringer og placerer dem på et midlertidigt lager, hvor overførslen kvalitetssikres. Funktionen sender også en kvittering på modtagelsen tilbage til producenten. Hvis afleveringen er mangelfuld eller fejlbehæftet, anmoder funktionen producenten om at genoverføre afleveringen.	The Receive Submission function provides the appropriate storage capability or devices to receive a <i>SIP</i> from the Producer (or from Administration). Digital SIPs may be delivered via electronic transfer (e.g., FTP), loaded from media submitted to the archive, or simply mounted (e.g., CD-ROM) on the archive file system for access. Non-digital SIPs would likely be delivered by conventional shipping procedures. The Receive Submission function may represent a legal transfer of custody for

Danish	English	Definition, dansk	Definition, English
			the Content Information in the SIP, and may require that special access controls be placed on the contents. This function provides a <i>confirmation of receipt</i> of a SIP to the Producer, which may include a <i>request to resubmit</i> a SIP in the case of errors resulting from the SIP submission.
<b>Modtag databaseopdateringer funktionen</b>	<b>Receive Database Updates</b>	I OAIS den funktion under Data Management, der opdaterer metadata og anden dokumentation lagret i Data Management.	The Receive Database Updates function adds, modifies or deletes information in the Data Management persistent storage. The main sources of updates are Ingest, which provides Descriptive Information for the new AIPs, and Administration, which provides system updates and review updates. Ingest transactions consist of Descriptive Information which identifies new AIPs stored in the archive. System updates include all system-related information (operational statistics, Consumer information, and request status). Review updates are generated by periodic reviewing and updating of information values (e.g., contact names, and addresses). The Receive Database Updates function provides regular reports to Administration summarizing the status of updates to the database, and also sends a database update response to Ingest.
<b>OAIS</b>	<b>OAIS</b>	Et Open Archival Information System (OAIS) er et arkiv, der har accepteret ansvaret for at bevare information og gøre denne tilgængelig for en specificeret brugergruppe ved at leve op til ISO standarden OAIS (14721:2003).	An archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.
<b>Omkostningskritisk aktivitet</b>	<b>Cost critical activity</b>	I CMDP er en omkostningskritisk aktivitet en opgave, som anslås at tage mere end én personuge om året til at udføre.	I CMDP a cost critical activity is a task which is estimated to take more than one person week (pw) per year to complete.
<b>Omkostningsparameter</b>	<b>Cost parameter</b>	I CMDP et parameter som har indflydelse på de omkostningskritiske aktiviteter og som udgør en del af variableerne i regnearket.	In CMDP the parameters that influence the cost critical activities and that constitute part of the variables in the spreadsheet formula.
<b>Overvågning</b>	<b>Supervision</b>	I CMDP angiver overvågning hvor meget manuelt arbejde, det kræver (beregnet i personuger) at overvåge processeringen af en migrering. Modellen anslår udgifterne til 5 % af maskintiden.	I CMDP supervision designates how much manual work it takes (calculated in person weeks) to supervise the processing of a migration. The model estimates the cost of man power monitoring this process to be 5% of the machine processing time.
<b>PAIMAS</b>	<b>PAIMAS</b>	PAIMAS er en ISO standard, hvis formål er at identificere, definere og strukturere relationer og interaktioner mellem en producent og et arkiv. Standarden definerer en metode til styring af aktiviteter mellem producenten og arkivet. Disse aktiviteter omfatter første fase af indtagningsprocessen (Ingest + Administration).	The purpose of the Producer-Archive Interface Methodology Abstract Standard (PAIMAS) is to identify, define and provide structure to the relationships and interactions between a Producer and an Archive. The standard defines the methodology to follow to manage the activities and the interfaces from the initial contact between the Producer and

Danish	English	Definition, dansk	Definition, English
			the Archive until the objects of information are received and validated by the Archive. These activities cover the first stage of the Ingest Process as defined in the Reference Model for an Open Archival Information System (OAIS).
<b>Periodisk</b>	<b>Periodic</b>	I CMDP omkostninger som kun forekommer med et givent interval	In CMDP a cost that only occurs with a certain time span
<b>Personuge</b>	<b>Person week (PW)</b>	I CMDP svarer en personuge til 32 timer.	In CMDP one person week is set to 32 hours.
<b>Producent</b>	<b>Producer</b>	I OAIS de personer eller systemer, der leverer den information til arkivet, som skal langtidsbevares.	The role played by those persons, or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems.
<b>Producent / Arkiv styrkeforhold</b>	<b>Producer / Archive strength</b>	I CMDP angiver det det formelle (juridiske) forhold mellem producent og arkiv. F.eks. den indflydelse arkivet på producentens valg af bevaringsformater	In CMDP this indicates the formal (legal) relationship between Producer and Archive E.g. the influence A has on P in the choice of preservation formats.
<b>Producentens gennemsigtighed</b>	<b>Transparency of producer</b>	I CMDP udtryk for at en organisations digitale forvaltning skal være dokumenteret på en forståelig måde og være tilgængelig for alle medarbejdere og relevante berørte parter.	In CMDP it designates that the processes and activities of an organization's record keeping program shall be documented in an understandable manner and be available to all personnel and appropriate interested parties. E.g. : Existence of a record keeping program, documentation of processes
<b>Proveniensinformation</b>	<b>Provenance Information</b>	I OAIS den information, der dokumenterer indholdsinformationens historie: Beskriver indholdsinformationens oprindelse eller kilde, ændringshistorik, ejerskabshistorik. Proveniensinformation kan fx være hvem, der har optaget data og hvordan indholdsinformation har været lagret, håndteret og migreret.	The information that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. Examples of Provenance Information are the principal investigator who recorded the data, and the information concerning its storage, handling, and migration.
<b>Referenceinformation</b>	<b>Reference Information</b>	I OAIS den information, der identificerer og evt. beskriver en eller flere mekanismer, der bruges til at tildele globalunikke persistente identifikatorer til indholdsinformation. Herunder identifikatorer, der gør det muligt for eksterne systemer at referere entydigt til indholdsinformation. Referenceinformation kan fx være et ISBN nummer.	The information that identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides identifiers that allow outside systems to refer, unambiguously, to a particular Content Information. An example of Reference Information is an ISBN.
<b>Repræsentationsinformation</b>	<b>Representation Information</b>	I OAIS den information, der forbinder et dataobjekt til mere meningsfuld information – og som er nødvendig for at brugerne kan forstå dataobjekterne. Repræsentationsinformation giver mening (semantik) og format (struktur) til dataobjektets digitale bits, således at det sikres, at et dataobjekt kan forstås af mennesker eller maskiner på sigt. Den strukturelle information om et dataobjekt oversætter objektets bits til noget meningsfuldt, f.eks. værdier som bogstaver, tal, pixels, tabeller, osv. Den semantiske information er f.eks. information om, hvilket sprog teksten er skrevet i, hvad den indeholdte videnskabelige terminologi betyder. Et eksempel er, at en bitsekvens tolkes af den strukturelle information som	The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol.

Danish	English	Definition, dansk	Definition, English
		værende en række af tal: 2, 1, 3, 8, 14, 16, 18, 19, 13, 7, 4, 3. Dem semantiske information fortæller os, at der er tale om grader i celsius, og at hvert tal i den angivne rækkefølge er et givet lands månedlige gennemsnitstemperatur.	
<b>Software tilvejebringelsesfaktor</b>	<b>Software Provision factor</b>	I CMDP angiver denne faktor omkostningen i personuger for at tilvejebringe migrationssoftware for en valgt kombination af indformat og udformat. Tilvejebringelsen sker ved at fremskaffe bestående software eller udvikle den, og dernæst indføre den i migrerings-systemet. Omkostningen afhænger af kompleksiteten og dokumentationen for formaterne. Levetiden for formatsoftware er sat til 8 år.	In CMDP the Software Provision factor is the cost (in person weeks) of providing migration software for a chosen combination of source and destination format. Provision is made by using existing software or developing software and following implementing it in the migration system. The cost is dependent on the complexity and documentation of the source and destination formats. The format software life time is set to 8 years.
<b>System/format kompleksitet</b>	<b>System/ Format complexity</b>	I CMDP datas/systemers syntaktiske kompleksitet. Fx: Multidimensionale formater for geodata og biokemiske strukturer er langt mere komplekse end formater til 2-dimensionel grafik.	In CMDP the syntactical complexity of the data, how complex are the systems and formats. E.g.: Multidimension formats for geodata, biochemical structures are much more complex than formats for 2-dimensional graphics
<b>System/format variation</b>	<b>System/ Format variation</b>	I CMDP den syntaktiske variation i systemer og formater, der skal indtages. Fx: mange forskellige filformater eller datastrukturer, såsom forskellige formater til lyd, inkluderer også variation i formatkompleksitet.	In CMDP the syntactical variation in the systems and formats to be submitted. E.g.: many different file formats or data structures, such as various formats for sound, also includes variation in complexity of formats
<b>Teknologi-overvågning</b>	<b>Monitor Technology</b>	I OAIS den funktion under Preservation Planning, der løbende overvåger de formater, strukturer og systemer som arkivet benytter, og rapportere ændringer til Develop Preservation Strategies and Standards.	The Monitor Technology function is responsible for tracking emerging digital technologies, information standards and computing platforms (i.e., hardware and software) to identify technologies which could cause obsolescence in the archive's computing environment and prevent access to some of the archives current holdings.
<b>Tid pr. side</b>	<b>Time per Page (TpP)</b>	I CMDP henviser det til den tid, det tager en datalog til at læse og forstå en side i et formats dokumentation for at kunne udvikle migreringssoftware. Parameteren er sat til 20 minutter per side, og er reguleret af formatkompleksitet samt basis tid, og er dermed 20 minutter gange 0 %, 25 % eller 50 %.	In CMDP this refers to the time it takes a computer scientist to read and comprehend a page in a format's documentation in order to be able to develop migration software for the format. The parameter is set to 20 minutes per page and is regulated by the format's complexity, thus 20 minutes times 0%, 25% or 50%.
<b>Tilgængeliggørelses-funktions-enheden</b>	<b>Access</b>	I OAIS den funktionsenhed, som omfatter de services og funktioner, som gør arkivbeholdningen synlig og tilgængelig for brugerne. Funktionen kan således oplyse brugerne om et arkivalies eksistens, beskrivelse, placering og tilgængelighed. Tilgængeliggørelsesfunktionen gør det ydermere muligt for brugerne at anmode om og konsultere arkivalier. Funktionen skal tillige sørge for en sikkerhedsopmærkning af beskyttede arkivalier.	This entity provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allowing Consumers to request and receive information products. Access functions include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers.

Danish	English	Definition, dansk	Definition, English
<b>Total format forståelsesfaktor</b>	<b>Total Format Interpretation factor</b>	I CMDP er den totale forståelsesfaktor summen af forståelsesfaktorer for hvert format som haves i en bevaringsinstitution.	In CMDP the Total Format Interpretation factor is the sum of an institution's format interpretation factors, each of which pertains to a specific format.
<b>Trin</b>	<b>Stages</b>	I CMDP svarer hvert trin til en funktionel enhed i OAIS, f.eks. Access.	In CMDP each stage corresponds to a functional entity in OAIS, ie. Ingest, Access, etc.
<b>Udformat</b>	<b>Destination format</b>	I CMDP det format, der migreres til.	In CMDP the destination format of a migration.
<b>Udvikling af bevaringsstrategier og standarder funktionen</b>	<b>Develop Preservation Strategies and Standards</b>	I OAIS den funktion under Preservation Planning, der på baggrund af rapporter fra Monitor Designated Community og Monitor Teknologi, løbende udvikler og anbefaler strategier og standarder for arkivet med henblik på at undgå forældelse.	The Develop Preservation Strategies and Standards function is responsible for developing and recommending strategies and standards to enable the archive to better anticipate future changes in the Designated Community service requirements or technology trends that would require migration of some current archive holdings or new submissions.
<b>Udvikling af informationspakke design og migreringsplan funktionen</b>	<b>Develop Packaging Designs and Migration Plans</b>	I OAIS den funktion under Preservation Planning, som udvikler nye designs for informationspakker, migreringsplaner og migreringsprogrammer med henblik på at implementere politikker og retningslinjer fra Administration.	The Develop Packaging Designs and Migration Plans function develops new IP designs and detailed migration plans and prototypes, to implement Administration policies and directives.
<b>Understøttende funktionsenhed</b>	<b>Common Services</b>	I OAIS den funktionsenhed som inkluderer: Operativsystemtjenester – er de kerneydelser som er nødvendige for at drive og administrere applikationsplatformen. Netværkstjenester som tilvejebringer mekanismer til understøtning af distribuerede applikationer, der kræver adgang til data og applikationsinteroperabilitet i heterogene, netværksbaserede miljøer. Sikkerhedstjenester - mekanismer til at beskytte følsomme oplysninger i informationssystemet.	Operating system services provide the core services needed to operate and administer the application platform, and provide an interface between application software and the platform. Network services provide the capabilities and mechanisms to support distributed applications requiring data access and applications interoperability in heterogeneous, networked environments. Security services capabilities and mechanisms to protect sensitive information and treatments in the information system. The appropriate level of protection is determined based upon the value of the information to the application end-users and the perception of threats to it.
<b>År</b>	<b>Year</b>	I CMDP refererer det til omkostninger for et konkret år, bestående af årlige omkostninger og periodiske.	In CMDP this refers to the cost for a specific year, consisting of annual and periodic cost.

## 12 Bilag

### *12.1 File format support in Microsoft Excel 2010*

Historically Excel has supported many different data formats. We have determined that a number of these older formats are seldom, if ever used. We are removing support for some file types to allow us to devote more of our efforts towards the file formats that are being used. These formats are being deprecated in 2 ways. For the set of file formats with the lowest usage, we will be discontinuing support for opening and saving of these formats. For the second set that has some minimal usage, we will support loading the files in Excel 2007 to allow you to save them in a newer format.

The following formats cannot be opened or saved in Excel 2007:

WK1 (1-2-3), WK4 (1-2-3), WJ3 (1-2-3 Japanese) (.wj3), WKS (1-2-3), WK3,(1-2-3), WK1,FMT(1-2-3), WJ2 (1-2-3 Japanese) (.wj2), WJ3, FJ3 (1-2-3 Japanese) (.wj3), DBF 2 (dBASE II), WQ1 (Quattro Pro/DOS), WK3,FM3(1-2-3), Microsoft Excel Chart (.xlc), WK1,ALL(1-2-3), WJ1 (1-2-3 Japanese) (.wj1), WKS (Works Japanese) (.wks).

The following formats may be opened, but not saved to in Excel 2007:

Microsoft Excel 2.1 Worksheet, Microsoft Excel 2.1 Macro, Microsoft Excel 3.0 Worksheet, Microsoft Excel 3.0 Macro, Microsoft Excel 4.0 Worksheet, Microsoft Excel 4.0 Macro, Microsoft Excel 97- Excel 2003 & 5.0/95 Workbook, Microsoft Excel 4.0 Workbook, DBF 3 (dBASE III), DBF 4 (dBASE IV).

Microsoft Excel 2010 - The official blog of the Microsoft Excel product team. Deprecated features for Excel 2007 by David Gainer, 24 Aug 2006 2:18 PM<sup>11</sup>

---

<sup>11</sup> <http://blogs.msdn.com/b/excel/archive/2006/08/24/718786.aspx>



## 12.2 Omkostning i Statens Arkiver vedr. Ingest

	2008		2007		2006		2005		2004		2003		2002		2001										
	Produktion	Timeforbrug	pr. enhed	Produktion	pr. enhed	Produktion	pr. enhed	Produktion	pr. enhed	Produktion	pr. enhed	Produktion	pr. enhed	Produktion	pr. enhed										
<b>A. Godkendelse og kassation</b>																									
Edb-registre	16	506,5	31,7	59	862,0	14,6	73	352,0	4,8	84	359,0	4,3	65	269,4	4,1	31	205,3	6,6	105	420,0	4,0	120	466,0	3,9	
E-journaler	13	132,3	10,2	14	120,0	8,6	13	213,0	16,4	59	280,5	4,8	24	328,8	13,7	51	460,3	9,0	40	674,3	16,9	25	434,0	17,4	
ESDH, type1	32	665,8	20,8	24	621,0	25,9	22	576,0	26,2	22	425,5	19,3	22	641,8	29,2	19	978,8	51,5	8	261,4	32,7	9	399,0	44,3	
ESDH, type2	10	356,0	35,6	10	702,0	70,2	18	689,0	38,3	2	139,5	69,8													
<b>B. Afleveringsforberedelse</b>																									
Edb-registre	148	1.577,3	10,7	138	1.804,0	13,1	88	1.510,0	17,2	109	1.135,0	10,4	224	1.649,5	7,4	202	1.673,0	8,3	122	1.862,0	15,3	118	1.448,0	12,3	
E-journaler	81	819,5	10,1	38	983,0	25,9	59	883,5	15,0	48	533,0	11,1	44	489,0	11,1	33	636,4	19,3	34	688,0	20,2	103	692,0	6,7	
E-arkiver/ESDH-systemer	22	756,5	34,4	19	359,0	18,9	9	320,5	35,6	7	213,0	30,4	7	176,0	25,1	3	119,5	39,8	3	57,0	19,0	0	22,0		
<b>C. Testning</b>																									
Edb-registre	203	2.548,5	12,6	219	2.956,0	13,5	66	2.142,5	32,5	118	2.005,0	17,0	225	3.050,0	13,6	110	1.459,5	13,3	134	1.621,0	12,1	56	1.717,0	30,7	
E-journaler	33	369,0	11,2	69	690,0	10,0	43	395,0	9,2	26	537,5	20,7	30	463,0	15,4	35	358,5	10,2	102	719,0		38	820,0	21,6	
E-arkiver/ESDH-systemer	4	269,0	67,3	11	653,0	59,4	2	234,5	117,3	5	358,0	71,6	0	143,5			5,0					0	0,0		
<b>2001-2008</b>																									
<b>A. Godkendelse og kassation</b>																									
Edb-registre	553	3.440,2	6,2																						
E-journaler	239	2.643,2	11,1																						
ESDH	198	6.455,7	32,6																						
<b>Ialt</b>	<b>990</b>	<b>12.539,0</b>	<b>16,6</b>																						
<b>B. Afleveringsforberedelse</b>																									
Edb-registre	1.149	12.658,8	11,0																						
E-journaler	440	5.724,4	13,0																						
E-arkiver/ESDH-systemer	70	2.023,5	28,9																						
<b>Ialt</b>	<b>1.659</b>	<b>20.406,7</b>	<b>17,6</b>																						
<b>C. Testning</b>																									
Edb-registre	1.131	17.499,5	15,5																						
E-journaler	376	4.352,0	11,6																						
E-arkiver/ESDH-systemer	22	1.663,0	75,6																						
<b>Ialt</b>	<b>1.529</b>	<b>23.514,5</b>	<b>34,2</b>																						