

Udredning A: Status for indsamling og bevaring af digitalt materiale

Af

Birte Christensen-Dalsgaard

Statsbiblioteket

Revideret 26/6 2002, korrektion 10/2 2003

Indholdsfortegnelse:

1	Indledning _____	1
2	Materiale – før, nu og i fremtiden _____	3
2.1	Terminologi _____	3
2.2	Eksempler _____	5
2.2.1	Online _____	6
2.2.2	Fast form _____	9
3	Roller og rammer _____	10
3.1	Terminologi _____	11
3.2	OAIS _____	11
3.2.1	OAIS: Informationsmodellen _____	12
3.2.2	OAIS: Den funktionelle model _____	13
3.3	Aktører - livscyklus for materiale _____	15
3.3.1	Punkter til afklaring _____	17
4	Strategi og handlingsinitiativer _____	19
4.1	Status Danmark _____	19
4.1.1	Status – biblioteker _____	19
4.1.2	Status – arkiver _____	21
4.1.3	Status – private institutioner _____	22
4.2	Nationale handlings- og strategiinitiativer. _____	22
4.2.1	De nordiske lande _____	22
4.2.2	England _____	23
4.2.3	Holland _____	24
4.2.4	USA _____	25
4.2.5	Australien _____	27
4.3	Internationalt relevante projekter og initiativer _____	27
4.3.1	EU _____	28
4.4	Spørgsmål _____	29
5	Metadata _____	30
5.1.1	Repræsentationsinformation: _____	30
5.1.2	Bevaringsbeskrivelsesinformation _____	32
5.2	Opsummering af spørgsmål _____	32
6	Indsamling _____	33
6.1	Terminologi _____	33

6.1.1	Indsamlingsstrategi	34
6.1.2	Indsamlingsmetoder	34
6.2	Afvejning af indsamlingsmetode i forhold til materiale type	34
6.3	Eksisterende erfaring	36
6.3.1	Selektiv indsamling	36
6.3.2	Eventbaseret indsamling	37
6.3.3	Tværsnit	37
6.3.4	Aflevering	37
6.3.5	Høstning	37
6.4	Opsummering af spørgsmål	38
7	<i>Ingest</i>	40
7.1	Terminologi	40
7.2	Autentifikation og kvalitetskontrol	40
8	<i>Bevaring</i>	41
8.1	Materialetype, formater og bevaring	41
8.2	Terminologi	42
8.3	Bevaring af informationsobjektet	43
8.3.1	Migrering	43
8.3.2	Konvertering/standardformat	43
8.3.3	Emulering	44
8.3.4	Museal strategi: Teknisk Museum	44
8.3.5	Konvertering til analog form	44
8.3.6	Sammenligning	45
8.4	Bevaring af dataobjekt	45
8.4.1	Lagermedier	46
8.4.2	Decentral opbevaring	47
8.5	Sikkerhedsforanstaltninger	47
8.5.1	Autenticitet	47
8.5.2	Audit	48
8.6	Opsummering af spørgsmål	48
9	<i>Adgang/Tilgængeliggørelse</i>	49
9.1	Oversigt over strategier i forskellige lande	49
9.2	Tekniske-juridiske overvejelser i forbindelse med adgang	50
9.3	Opsummering/Spørgsmål	50
10	<i>Kompetenceudvikling</i>	51
11	<i>Økonomi/Forretningsmodel</i>	52
12	<i>Det videre arbejde</i>	54
13	<i>Litteraturliste</i>	58

1 Indledning

The vision of creating digital libraries that will be able to preserve our heritage currently rests on technological quicksand. There is as yet no viable long-term strategy to ensure that digital information will be readable in the future. Not only are digital documents vulnerable to loss via media decay and obsolescence, but they become equally inaccessible and unreadable if the software needed to interpret them—or the hardware on which that software runs—is lost or becomes obsolete.

Rothenberg, (R 99) i: Avoiding Technological Quicksand

.....men, som kunne være konklusionen i denne udredning, der arbejdes hårdt på at finde tekniske og organisatoriske løsninger.

Målet med denne udredning er

- at give en introduktion til området og introducere en terminologi relevant i forbindelse med digital bevaring,
- at illustrere, at rigtig mange af problemstillingerne er velbeskrevne,
- at demonstrere, at selv om der ikke findes en entydig, rigtig løsning, er der flere forslag til forskellige dele af bevaringsprocessen og endelig
- at pointere, at på trods af ovenstående, er der store udfordringer – specielt hvis man ikke selv kan vælge sin bevaringsstrategi og må håndtere alle typer formater. Hvilken, som er rigtig vil afhænge af situationen, materialet, organisationen bag, økonomi osv.

Digital bevaring og formidling er forbundne:

- Digitalt materiale bevares med det formål at kunne formidles.
- Analogt materiale digitaliseres af hensyn til formidling – hvorefter den digitaliserede version skal bevares.

Det er en forudsætning for etableringen af digitale biblioteker, arkiver og museer, at en passende mængde relevant materiale forefindes på elektronisk form. Et eksempel indenfor digitale biblioteker er JSTOR, hvor en lang række biblioteker digitaliserer gamle tidsskrifter og udbyder dem online. Der er dog her ikke tale om digitalisering som bevaringsstrategi og kravene til langtidsopbevaring er forskellige - specielt sikkerheden. Der tales om forskellige tidsskalaer i forbindelse med bevaring. Fokus i denne rapport er langtidsbevaring af digitalt materiale, hvor den digitale kopi er selve målet for opbevaring – dvs. der er tale om et unika.

I den analoge verden hænger det at stille materiale til rådighed ofte sammen det at eje et eksemplar. Denne binding er brudt i den digitale verden, hvor det handler om at formidle adgang til materiale. I eksemplet med elektroniske tidsskrifter, får biblioteker mulighed for at give adgang til en langt større korpus end tidligere – uden at eje et eneste eksemplar. Andre materialetyper er på vej – i Danmark kan man leje adgang til noder, online musikdistribution osv. osv.

Denne distributionsform rejser imidlertid en række spørgsmål, når bevaringsstrategier skal fastlægges, såsom hvem der har ansvaret for bevaring, hvordan det sker, og rettighedsproblematikken.

Udviklingen fortsætter og i dag findes meget materiale kun på digital form.

I flere artikler og rapporter understreges vigtigheden af at indtænke digital bevaring i livscyklens for informationer. Der er primært to grunde til, at dette er vigtigt:

- pga. den korte levetid af software og læseudstyr skal sikring af informationer, nødvendige for bevarelse, påbegyndes stort set med det samme
- er der ikke klare retningslinier for hvem, der har ansvar for hvad, vil information gå tabt pga. af misforstået ansvarsfordeling.

Underforstået i sidste pind er, at en bredere skare af institutioner kan være medansvarlige for bevaring. I dag påhviler ansvaret for bevaring af kulturarv en veldefineret række kulturinstitutioner, med veldefinerede ansvarsområder. I afsnittet om aktører vender vi tilbage til dette spørgsmål.

Bevaring af kulturarven kræver ikke bare deltagelse af mange aktører, det kræver også en bred sammensat kompetence hos de institutioner, som er ansvarlige for langtidsbevaring. Typisk har bevaringsopgaven ligget hos konservatorer, og det er en gængs opfattelse, at bevaring af den digitale kulturarv påhviler it-folk. Dette billede er ved at ændre sig, og der er en forståelse for, at bevaring kræver en bred vifte af kompetencer – herunder konservatorer, biblioteksfolk, arkivarer og it-folk.

Spørgsmålene omkring digitalt materialer og bevaring er rejst af mange i mange forskellige sammenhænge. Og mængden af litteratur, som adresserer et eller flere af de relevante spørgsmål, er rimelig overvældende. Mens denne rapport blev lavet, er udkommet 3 strategiske dokumenter:

Council on Library and Information Resources Washington, DC and Library of Congress: *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving*, April 2002 (<http://www.clir.org/pubs/abstract/pub106abst.html>)

OCLC/RLG. *Trusted Digital Repositories: Attributes and Responsibilities*. An RLG-OCLC Report, Maj 2002 (<http://www.rlg.org/longterm/repositories.pdf>)

Digital Preservation Coalition (DPC): *The Handbook*, Juni 2002, (<http://www.dpconline.org/graphics/handbook/index.html>)

Fremstilling her kan ikke være fuldt uddybende, men har til hensigt at give et billede af, hvad der er gjort, og hvilke initiativer der er undervejs. Formålet er dels at give en oversigt over, hvad der foregår, og hvilke spørgsmål man arbejder med internationalt omkring indsamling, arkivering og bevaring af digitale dokumenter, dels at definere relevante begreber og endelig at formulere konkrete spørgsmål, som er nødvendige inden en handlingsplan kan formuleres.

Jeg vil slutte denne indledning med at takke Birgit Henriksen, KB, for mange inspirerende diskussioner om digital bevaring – både før og under dette arbejde. Jeg vil også sige tak til Birgit Hansen, Statens Arkiver, for en oplysende eftermiddag og stor imødekommenhed.

2 Materiale – før, nu og i fremtiden

”New media” er materiale, som skal fortolkes af en maskine, før signalet giver mening. I PRO01 gives en let læst, underholdende introduktion til disse materialetyper. Et af de første eksempler på denne type materiale, var de hulkort, som blev brugt i vævene. Hulkortene i sig selv siger ikke noget, men fortolket bliver det til instruktioner, som får vævene til at producere specielle mønstre. Et senere eksempel, som nok er det første, som er relevant for denne gruppes arbejde, er de første valser. For at høre indholdet af valsen kræves et instrument, som konvertere overfladen til lydbølger. Fælles for alle ”new media” er, at en bevaring ikke kun er et spørgsmål om at bevare den fysiske enhed, men også om at bevare enten maskinen eller en nøje beskrivelse af, hvordan maskinen fungerer – herunder, hvordan maskinen aflæser signalet på den faste enhed.

Med fremkomsten af det digitale objekt er problemstillingerne omkring bevaring ændret radikalt. Komplexiteten omkring afspilleudstyr er steget væsentligt, hvorimod problemstillinger som forringelse af kvaliteten af kopier og slitage er forsvundet, idet enhver digital kopi er identisk med originalen, ikke nødvendigvis den fysiske form, men bitindholdet og dermed oplevelsen.

Et problem, specielt omkring webmateriale, er en afgrænsning af, hvad en materialeenhed er. I ’de gode gamle dage’ var en bog en bog, en plade en plade og et regnskab et regnskab, og det var rimelig veldefineret, hvem der var ansvarlig, og hvad en enhed var. Der var i forbindelse med udgivelsen en vis udvælgelse, som resulterede i såvel en kvalitetssikring som en begrænsning i antallet, som blev udgivet. Desuden var kunstnere en identificerbar størrelse, som selv under kunstnernavn var kendt af den ansvarlige udgiver. I dag kan alle komme til udtryk – som dem selv eller under pseudonymer, som kun er kendt af dem selv, og sider linker til hinanden på kryds og tværs, så det kan være svært at definere, hvad en enhed er. Og så er der jo ikke længere kun tale om dokumentlignede information – internettet bruges til et kommunikation, transaktioner og lignende.

En enhed inden for arkiververdenen er veldefineret og relaterer sig til en arkivfond. Arkivskaberen er den myndighed mv., som har skabt/tilvebragt en bestemt arkivfond.

Inden vi vender tilbage til denne diversitet af information og kommunikation, vil det være på sin plads at introducere en fælles terminologi omkring digitalt materiale.

2.1 Terminologi

I det følgende bruges følgende overordnede kategorier for digitalt materiale:

- Konverteret digitalt
- Født digitalt
 - o Fast
 - o Online

Hvor

- **Konverteret digital**, som navnet indikerer, er materiale, som er fremkommet via en digitaliseringsproces. Dette er typisk trykte

dokumenter, billeder, lyd eller videoer. På mange måder er det den nemmeste kategori at håndtere begrebsmæssigt. Vi er selv herre over format og kvalitet og kan derfor vælge dem, som giver den højeste chance for bevaring.

- **Født digital –fast** er materiale, som distribueres på fast form, typisk cd'er, dvd'er, cd-rom'er med information eller med spil. Tidligere distributionsformer omfatter disketter, hulkort, hulstrimler og magnetbånd af forskellige størrelse og tykkelse. Ofte vil der være tale om publiceret materiale, som er omfattet af pligtafleveringsloven. Afhængig af format vil denne gruppe høre til alt fra det meget simple til det uhyre komplekse.
- **Født digital – online** er materiale, som ikke er distribueret på fast form. Det er ikke sikkert, at materialet som sådan distribueres; det kan være at der kun produceres et svar i forhold til tidligere aktioner. Visse typer er nemme at håndtere, visse er ekstremt komplicerede.

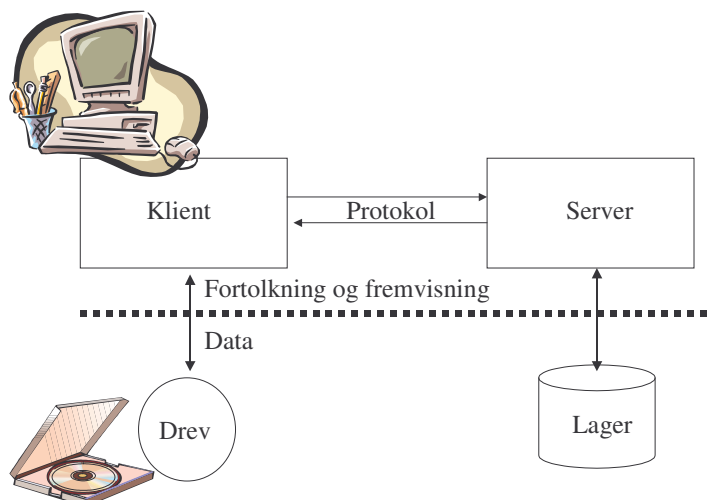
I det efterfølgende vil vi bruge termerne klient, drev, server og lager:

- **Klient** – er en maskine samt det software, som bruges til at vise data
- **Server** – er en maskine samt software, som leverer data til en klient
- **Drev** – sidder på en klient og bruges til at gemme data – er i dag 3 ½ ” diskettedrev, cd-rom-drev eller dvd-drev.
- **Lager** – sidder på en server og bruges til at gemme data – er typisk store harddiske, men kan også være tape-diskarrays.

En klient kan kommunikere med en server via en protokol:

- **Protokol:** En formel beskrivelse af besked (message) format samt regler, som de to computere skal følge for at udveksle disse beskeder. To meget almindelige protokoller er http, som bruges i forbindelse med web, samt ftp, som bruges til at overføre file. En anden, som er velkendt i biblioteksverdenen, er Z39.50.

Disse begreber er illustreret på tegningen nedenfor:



En vigtig kategori af født digitalt online materiale er det, vi normalt forbinder med WWW. Vi vil her bruge følgende terminologi i forhold til webmateriale.

- **URL-objekt:** En enkelt fil, det kan være tekst, grafik, lyd eller andet
- **URL-side:** En samling af URL-objekter – det, som fremkommer på en skærm
- **Site:** Bruges om en samling URL-sider – ofte inden for samme domæne
- **Domæne:** Et område med samme domænenavn
- **Hjemmeside:** Den første side, som fremstår som indgangen til en eller flere sider. Den side, som vises, når man på siden klikker på ”hjemmeside” – ofte logoet. Er en speciel side.

Desuden bruges i forbindelse med arkivmateriale:

- **Arkivfond:** den totale sum af arkivalier skabt og/eller samlet af en bestemt myndighed/organisation/institution eller person i forbindelse med udøvelsen af vedkommendes opgaver eller funktion.
- **Arkivskaberen** er den myndighed mv., som har skabt/tilvejebragt et bestemt arkivfond.

2.2 Eksempler

Digitalt materiale er et meget bredt begreb, idet digital information gennemsyrrer alle aspekter af vores samfund: fra hjemmebanken til det internationale finansmarked, fra hjemmets brug af internetbutikker til firmaers globale transaktioner, fra private e-mails til dialog med offentlige institutioner via internettet, fra chat til koncerners intranet med fokus på fænomener som vidensdeling, fra download af mp3-filer til koncerter spillet af kunstnere fra forskellige studier, vores adgang til kommunikation, til virtuelle verdener, adgang til nyheder, til at gense udsendelser, til underholdning osv. osv.

I den efterfølgende beskrivelse vil vi skelne mellem fire typer materiale:

Offentlig Information (én til mange distribution): typisk websider, men kan også være artikler i en database, noder i et licensbeskyttende, offentligt system, lyd- og videofiler.

Lukket Information (én til få): informationerne i sagsbehandlingssystemer, patenter, farmaceutisk information, data, statisk information

Kommunikation (én til mange, én til én og mange til mange): Mest almindelige form er e-mails, men omfatter også chat, videokonference, IP telefoni, virtuelle verdener, MUDS osv.

Services: (én til system) Eksempler er skatteindberetningssystemer, bibliotekssystemer, krak.dk, dsb.dk og flyselskabers selvbooking systemer, virtuelle eksperimenter ...

For alle typer gælder, at de kan være såvel synkrone som asynkrone.

2.2.1 Online

2.2.1.1 Offentlig Information

Eksemplerne nedenfor er taget fra internetverdenen.

Diversiteten inden den offentlige del af internetverdenen strækker sig dels over forskellige formater, forskellige strategier for præsentation og dels over mange forskellige mål med informationerne og forventninger over brug. Denne diversitet stiller forskellige krav i forbindelse med indsamling og bevaring. Lad os se på nogle eksempler:

Forskellige afviklingsomgivelser:

Kulturministeriet (www.kum.dk): Denne site benytter sig af såkaldt active server pages (ASP – kort), som giver mulighed for en dynamisk vedligeholdelse og dynamik på siderne. ASP er et specifikt Microsoftprodukt, og en afvikling kræver samme version af serversoftware.

Danmarks Elektroniske Forskningsbibliotek (www.deff.dk) Denne side er genereret vha. PHP, et sprog som minder om ASP – men kører under UNIX.

Denne type sites kan høstes, men der er ingen garanti for, at man får hele funktionaliteten med. Hvis de afleveres, vil det enten være nødvendigt at konvertere dem, eller også vil der forestå en kompliceret emuleringsopgave.

Personificerede sites:

My- sites: Visse sites tilbyder en personificeret indgang. En variant af dette er de sites, hvor man har en brugeradgang for alle, men er man medlem, kan man få adgang til flere services. Et eksempel på det sidste er FDM's hjemmeside www.fdm.dk.

Funktionaliteten kan afhænge af, hvem man er, og hvor man sidder. Et eksempel herpå er adgang til elektroniske tidsskrifter, hvor adgangen i øjeblikket styres via IP-numre.

Ovenstående er eksempler på, hvordan man kan benytte forskellige metoder til at udvikle og vedligeholde sin webside, og hører til det, man nu kalder 2. generations websites. Udviklingen er gået fra det, man ofte kalder første generations websider, hvor alt er håndholdt og det meste kodet i HTML over anden generations websider, som vedligeholdes via "content management systemer" og er på vej til 3. generations sites, hvor begrebet webservices bliver dominerende. (JISC 01 a,b). Vi vil se flere sites hvis hovedformål er at sammensætte en præsentation af information og tilbyde kommunikation over for brugeren. Det digitale bibliotek er et eksempel på en sådan udvikling – som flere og flere informationer bliver digitale, går vi fra at eje en ting over til at leje den (licens). Bibliotekets rolle går derfor fra at vedligeholde og beskrive materiale over til at tegne licenser og stille materiale til rådighed – samt sikre en fornuftig præsentation af dette. Et eksempel er Dansk Artikelbase, som er tilgængelig fra mange bibliotekshjemmesider. Vi vender tilbage til services senere.

Disse eksempler rejser de strategiske spørgsmål:

- Hvad skal bevares – præsentationen eller indholdet?

- Hvor mange præsentationer af samme personificerbare site skal gemmes?

Eksempler på forskellige materialtyper:

Danmarks Radio (www.dr.dk), hvor man kan genhøre og gense flere udsendelser. Disse er streamede og kræver, at man har realmedia eller realvideo/audio.

Kulturnet Danmark – soldaten i baghaven (<http://www.soldatenibaghaven.dk/>): Denne fortælling kræver Flash 5 og Quicktime for at blive afviklet.

For mange materialetyper gælder, at det ikke med dagens teknologi er muligt at høste dem. Eneste mulighed for at indsamle disse er derfor aflevering.

Hvad med de nye kunstværker? Besser skriver i *Archiving the Avant Garde: Documenting and preserving variable digital art* (www.bampfa.berkeley.edu/ciao/avant_garde.html):

Works of variable media art, such as performance, installation, conceptual, and digital art, represent some of the most compelling and significant artistic creation of our time. These works are key to understanding contemporary art practice and scholarship, but because of their ephemeral, technical, multimedia, or otherwise variable natures, they also present significant obstacles to accurate documentation, access, and preservation. The works were in many cases created to challenge traditional methods of art description and preservation, but now, lacking such description, they often comprise the more obscure aspects of institutional collections, virtually inaccessible to present day researchers.

Men sikkert er det, at dette arbejde er med i den gruppe af materiale, som er med til at skubbe grænsen for brugen af det digitale medium.

At gemme sådanne værker er komplekst, men vil af mange grunde være utrolig interessant for eftertidens forståelse af dette fænomens opståen.

Offentlig information er ikke nødvendigvis ensbetydende med fri, dvs. ikke rettighedsbelagt, information. Der er flere eksempler på materiale, som er tilgængeligt, men som eksempelvis ikke kan gemmes på pc'ens harddisk eller printes.

Skal vi vente med at igangsætte arkiveringsinitiativer, til vi kan håndtere denne form for materiale, vil vi miste rigtig meget, som vi nu er i stand til at gemme. Det er derfor vigtigt ikke at lade disse uhyre komplekse, meget interessante materialetyper skygge for videre initiativer. Udviklingen går så stærkt, at uanset hvad vi gør, vil vi stå over for materiale, som ikke kan håndteres. I en nylig artikel med titlen "change the metaphor: the archive as an ecosystem" (MC02) argumenteres for, at der ikke findes én løsning, som kan håndtere arkivering og langtidsbevaring, men at bevaringsstrategien konstant skal opdateres.

2.2.1.2 Lukket Information

Et oplagt eksempel her er indholdet af journalsystemer. Statens arkiver har håndteret spørgsmålet om journalsystemer, hvor man har fået gennemført, at såfremt informationen er bevaringsværdig, skal journalsystemet godkendes, forstået på den måde, at SA har defineret en metode til aflevering, og denne skal opfyldes af journalsystemet.

Andre eksempler er kirkebøger og medicinalindustriens afleveringer.

Et andet meget omtalt eksempel er det såkaldte ”dybe web”, som estimeres til at størrelsesorden større end det offentlige web.

Meget materiale kan findes frem – og kan være licensbelagt. Et glimrende eksempel er AmazingMusicWorld (www.SheetMusicNow.com), som sælger noder.

2.2.1.3 Kommunikation

Kommunikation er den aktivitet, som genererer mest aktivitet på Internettet. Der er mange typer:

- Chat
- E-mail
- MUD
- Samarbejde
- Video-konference
- Audio-konference

I hvor høj grad det er muligt at fange diskussionen, afhænger af protokollen.

I mange lande bruges meget energi på at arkivere specielt e-mails fra offentlige personer. Det mest kendte eksempel er NARA's håndtering af e-mails efter Bill Clinton.

2.2.1.4 Services

Services her dækker over sites, som tilbyder en service, det være sig

- Hjemmebank
- Bibliotekssystemer
- Skat
- KRAK
- DSB
- Online bestilling af flyrejser
- Virtuel reference vagt – biblioteksvagten eller
- Amazone.com – recommender system

Ovenstående er og vil formentlig forblive selvstændige services. Man kan imidlertid godt forestille sig, at der vil komme en generation af services, som er beregnet til at blive indlejret på websider. Afhængig af forretningsmodel kan deres oprindelse være mere eller mindre oplagt. Et simpelt eksempel på dette er Google søgebox'e på en række sider, og CNN ticker øverst på f.eks. Netscapes hjemmeside.

2.2.2 Fast form

Selv om denne materialetype i dag er omfattet af pligtafleveringsloven og derfor indsamlingsmæssigt er simplere end websider, udgør denne materialetype en potentiel stor udfordring. Der findes mange forskellige faste medier, som bruges i forbindelse med forskellige computere (Picoline, Commodore 64, IBM 286, MAC, osv.), musikanlæg, videoanlæg, digitale kameraer osv. herunder 5 ¼" disketter, 3 ½" disketter, CD-ROM, minidisk, ZIP-drev, Compact Flash osv. osv.

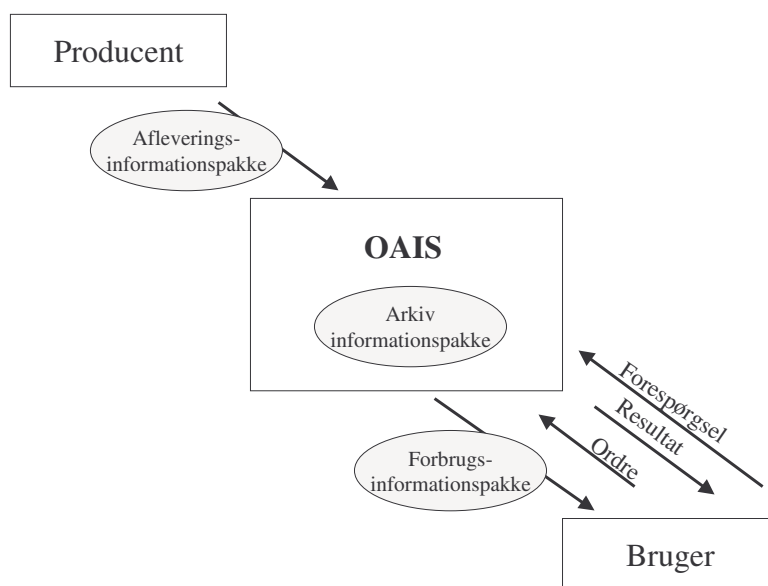
Ikke nok med, at man skal finde et drev, hvor en given diskette, cd-rom eller andet kan læses, man skal også have det rette udstyr med det rette styresystem og de rette drivere til drevene.

Selv færdiglavede, eksempelvis cd'er udgør en stor udfordring. Disse er velbeskrevne, men der er flere forskellige: cd-i, CDRW, dvd, cd-rom osv. beskrevet i det, som hedder hhv. Red book, Green book etc.

3 Roller og rammer

Materiale i et arkiv har haft et livsforløb, som er startet med, at det er skabt og derefter enten via aflevering eller via aktiv hentning blevet indlemmet i arkivet.

Arbejdet omkring indsamling og forberedelse til arkivering forventes at være forskellig afhængig af, om der er tale om billeder, fjernsyn, digital kunst online aviser, data i forbindelse med folketællinger, indhold af sagsbehandlingssystemer osv., ligesom det forventes, at der vil være forskel på adgang. I OAIS-modellen, som beskrives efterfølgende, har man denne tegning af det overordnede flow:



Figur 2-4 fra OAIS referencemodel (OAIS 01)

Det, som er indeholdt i OAIS kassen, selve det at sikre langtidsbevaring, forventes imidlertid at være det samme – på tværs af traditionelle faggrænser og på tværs af landegrænser. Det er derfor meget vigtig at identificere det, som er fælles, og sikre en harmonisering på et niveau, som gør, at man kan arbejde sammen. Samarbejdet kan have mange former, fra fælles udvikling af komponenter til deling af arbejdsopgaver. Samtidig er det også vigtig at definere de specielle nationale opgaver og gøre sig klart, hvad man ønsker selv at stå for.

Biblioteksverdenen er et eksempel til efterfølgning, men også til advarsel. Her udviklede Library of Congress et format (MARC) til beskrivelse af den bibliografiske information. Dette betød, at leverandører kan udvikle bibliotekssystemer, som kan bruges på biblioteker rundt om i verden – i teorien. De enkelte lande har af forskellige grunde set sig nødsaget til at modificere MARC, således at vi i dag har mange forskellige versioner, i Danmark DANMARC2. Udviklingen nu er, at lande begynder at slutte op om en fælles standard, MARC21. Som det blev demonstreret ved en udbudsforretning fra tre

store biblioteker, kan man vist godt sige, at brugen af DANMARC2 er en teknisk handelshindring.

Dette er ikke et indlæg i en begyndende debat om MARC-standard, men en illustration af, hvad man kan opnå/gå glip af, hvis man kan identificere og beskrive fælles processer. OAIS-modellen, som er beskrevet nedenfor, kan ses som et første skridt i retning af harmonisering.

3.1 Terminologi

I dette afsnit defineres en række termer, som lige nu kan virke ret abstrakte, men som introduceres mere uddybende i de efterfølgende afsnit.

Arkiv: en organisation, som har til hensigt at bevare information og sikre adgang og brug over for udvalgte brugergrupper.

Arkiv Informationssystem: samling af hardware, software samt menneskelige ressourcer, som er ansvarlig for indsamling, bevaring og efterfølgende distribution af den arkiverede (underforstået digitale) information.

Værk: afgrænser mængde information, som betragtes som en afsluttet og selvstændig enhed. Abstraktion – den intellektuelle skabelse. For at værket kan publiceres, skal det have en manifestation.

Udgivet værk: en manifestation, som afhængig af værkets natur kan være på papir, film eller i digital form

Informationsobjekt: De nødvendige komponenter og beskrivelser, som muliggør genskabelsen af et værk. Et informationsobjekt består af to komponenter: dataobjektet og repræsentationsinformation:

Dataobjekt: et fysisk eller digitalt objekt.

Digitalt objekt: en samling bit.

Repræsentationsinformation: information, som gør det muligt at afbilde dataobjektet på en måde, som kan opfattes af mennesker.

Bevaringsinformation: information, som giver oplysning om informationsobjektet og som er nødvendige af hensyn til genfinding, sikkerhed og forståelse af informationsobjektet.

Arkivinformationspakke: Det, som gemmes, dvs. en samling af dataobjekt såvel som relevant repræsentationsinformation som bevaringsinformation.

3.2 OAIS

OAIS står for Open Archivals Information System, hvor et *arkiv informationssystem* defineres som en mængde hardware, software samt menneskelige ressourcer, som er ansvarlige for indsamling, bevaring og efterfølgende distribution af den arkiverede information.

OAIS-referencemodellen (OAIS 01) giver en beskrivelse af den overordnede problemstilling. Ideen med modellen er at give en forståelse af, hvilke processer som skal adresseres, og hvilke informationer, som er nødvendige, og ikke at give en opskrift på, hvordan langtidsbevaring håndteres. Ideen i modellen er at adskille

hovedinformation fra sekundær information, at adskille hardware fra software og at indtænke de planlægningsmæssige og administrative aspekter i processen.

Formålet med at gennemgå modellen her er at introducere såvel et skelet som en terminologi, som den videre diskussion kan hænges op på.

Consultative Committee for Space Data Systems (CCSDS) påbegyndte udviklingen af OAIS-modellen miljø i deres forsøg på at standardisere langtidsbevaring af videnskabelige data. Modellens fordel er, at den fokuserer på det generiske, det som alle, uanset materialets oprindelse, er nødt til at adressere.

Parallelt var NEDLIB-projektet efter i et år at have prøvet at forstå forskellighederne på tværs af nationalarkiverne i Europa, gået over til at prøve at identificere det, som var fælles. I dette arbejde identificerede de koncepter og flow, som viste sig at svarer til OAIS. Der opstod derfor hurtigt et samarbejde mellem CCSDS og NEDLIB-projektet, og erfaringerne fra NEDLIB blev inkorporeret i processen, således at det forslag, som nu er ved at blive standardiseret, er meget generelt anvendeligt for mange informationstyper.

Det system til digital bevaring, som såvel KB, Holland som BL England samarbejder med IBM omkring, bygger på denne model. Modellen bruges også af NARA i forbindelse med deres projekt om e-mails.

Referencemodellen giver en beskrivelse af bevaring fra to synsvinkler: Hvilke informationer, som er nødvendige, Informationsmodellen, og hvilke processer/funktioner, som er nødvendige, den funktionelle model.

3.2.1 OAIS: Informationsmodellen

OAIS arbejder med tre forskellige informationspakker:

Afleveringsinformationspakke (SIP), som er den information - om nogen - som udgiveren afleverer sammen med dataobjektet/-objekterne.

Forbrugsinformationspakke (DIP), som er det brugeren præsenteres for, såfremt han/hun har fundet en bestemt information, som ønskes forevist. DIP kan være alt fra et simpelt dokument til en database med søgeinterface og indhold.

Arkivinformationspakke (AIP) er det, som bevares. Dvs. denne indeholder såvel dataobjektet som information til at fortolke det og information om objektets historie, både før det blev arkiveret, og mens det har befundet sig i arkivet.

Antag, at vi ønsker at bevare en videnskabelig artikel på elektronisk form. Det giver imidlertid ikke mening at sige, at man kun bevarer artiklen, værket, man er også nødt til at bevare den i et format – Word, TeX, PDF, SGML, XML eller lignende, som egner til artikler. Gemmer vi den i PDF, er vi også nødt til at gemme den nødvendige information om, at dette objekt kan fortolkes af eksempelvis en acrobat reader version 2.0.

I OAIS modellen vil der gælde følgende i forhold til eksemplet ovenfor:

Filen, som indeholder bittene, vil blive kaldt et dataobjekt. Information om programmet, som bruges i forbindelse med fremvisningen, kaldes for "repræsentations information", de to tilsammen er informationsobjektet og resultatet, det vi kan se, er værket.

Repræsentationsinformationen kan være simpel, hvis eksempelvis åbne standarder er brugt – f.eks. HTML. I sin yderste konsekvens kan den være meget kompleks og indeholde beskrivelse af styresystemer, nødvendige drivere, oversætterprogrammer samt kode til afvikling.

Informationsobjektet kan være logisk sammenhæng med andet materiale, det kan være nødvendige billedfiler, bagvedliggende databaser eller, som typisk ved websider, andre informationsobjekter. Alle disse informationer, som placerer informationen i en sammenhæng, kaldes bevaringsbeskrivelsesinformationen (PDI: Preservation Description Information).

I OAIS-modellen arbejdes med fire kategorier af bevaringsbeskrivelsesinformation:

- **Reference information:** intern såvel som eksternt identifikation af materialet (f.eks. ISBN, URN)
- **Sammenhæng (Context):** Placerer materialet i sammenhæng, det kan være i forhold til andre dokumenter, men også i forhold til hardware og software.
- **Proveniens:** Dokumenterer materialet historie (oprindelse, IP, hvad er gjort for at bevare det osv.)
- **Fiksering (Fixity):** Information, som sikrer autenticiteten af materialet (eksempelvis checksum, digital signatur)

Herudover foreslår NEDLIB at tilføje endnu en kategori:

- **Forandringer (change history)**

En Arkiv Informations Pakke (AIP i det følgende) er det, som gemmes, og som sikrer, at man i fremtiden kan se informationsobjektet. Den består af:

- Indholdsobjektet, dvs. dataobjektet samt repræsentationsinformationen
- Bevaringsbeskrivelsesinformationen (PDI).

AIP'er kan bestå af mange fysiske filer med de forskellige informationer. En beskrivelse af, hvor disse befinder sig, gives i en "pakkedliste" – eller pakkeinformationer i OAIS-modellens terminologi.

3.2.2 OAIS: Den funktionelle model

Den funktionelle model opererer med seks hovedfunktioner:

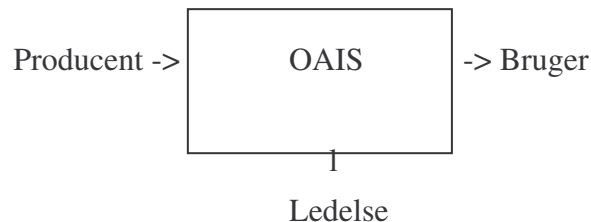
- Ingest
- Arkivering/Lagring
- Data management
- Adgang
- Administration
- Bevaringsplanlægning

Grænsen for modellen er:

Producenter som afleverer information, som skal bevares

Brugere som i fremtiden måtte ønske at komme til og se informationen

Ledelse som er ansvarlig for den overordnede politik mht. arkivering



Vekselvirkningen mellem Producent og OAIIS består af aflevering og pre-ingest aktiviteter:

- Definere selektionskriterier for materiale til digital bevaring
- Forhandle rettigheder
- Aftaler om format
- osv.

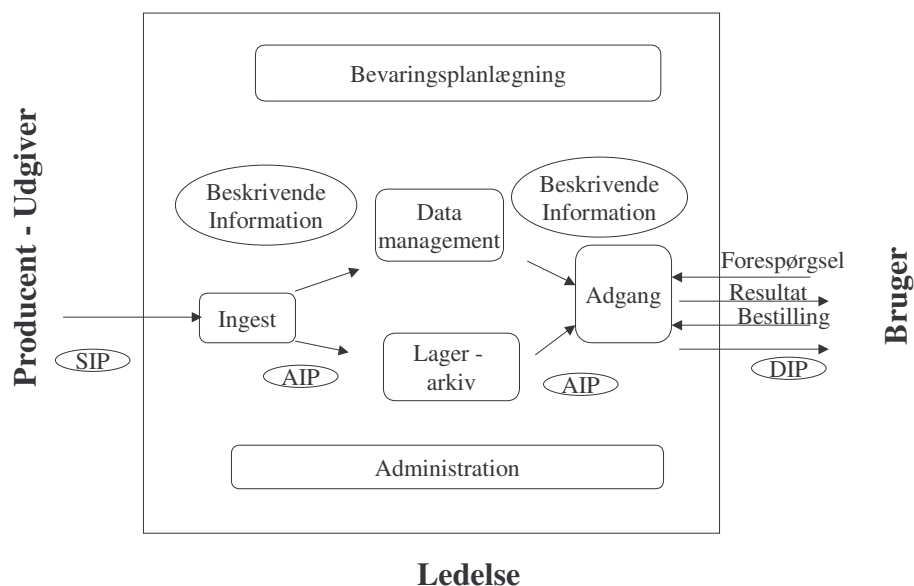
Disse er beskrevet i CCDSD white book (OAIIS 02)

Vekselvirkningen mellem ledelse og OAIIS består i

- Økonomi og prioriteringer
- Prispolitik, såfremt der er betaling i forbindelse med brugen af arkivet
- o.lign.

Vekselvirkningen mellem bruger og OAIIS er primært et spørgsmål om at stille værktøjer til rådighed, som gør, at brugeren kan identificere det, hun/han er interesseret i – og sikre, at disse kan leveres på en måde, så brugeren kan fortolke informationen.

Den funktionelle model tegnes ofte som vist nedenfor:



På tegningen repræsenterer de afrunde firkanter funktioner og ovalerne informationerne, som blev beskrevet i foregående afsnit. De forskellige funktioner er kort beskrevet:

Ingest: Sikrer, at alle informationer er til stede, at rettigheder er afklaret, at dataobjektet ligger på en i forhold til bevaringsstrategien optimal måde, at kvaliteten er i orden osv.

Datamangement: Sikring af adgang til såvel beskrivende data som administrative data, som er nødvendige for at drive arkivet.

Lager/Arkiv: Opbevaring og vedligeholdelse af arkivpakker.

Administration: Administrative procedurer, som er nødvendige for den overordnede vedligeholdelse af arkivet.

Adgang: Service over for brugere.

3.3 Aktører - livscyklus for materiale

Aktører har i det efterfølgende to betydninger:

Aktører er de, som på et eller andet tidspunkt har en rolle i forbindelse med materiale; det kan være i forbindelse med at det skabes, at det distribueres, det forbruges, eller det bevares. De fleste af disse aktører har en interesse i, at materialet bevares, men få har det som deres primære opgave

Aktører bruges også om de, som enten har lavet en forretning ud af at opbygge et arkiv, eller de, som har det som en kulturbevarende opgave. Denne gruppe aktører tæller datacentre bredt, arkiver, medicinalindustri, informationsdatabaser ved universiteter, nationalbiblioteker o.lign.

Der synes at være en erkendelse af, at digital bevaring afviger fra sædvanlig bevaringsopgaver på en række områder. En væsentlig forskel er nødvendigheden af at planlægge bevaring stort set på samme tidspunkt, som materialet skabes. Som indledning til afsnittet om digitalisering (dpconline):

.....the crucial role of the creator in helping to ensure long-term access to the digital resources they create..... Clearly, most individual creators cannot be expected to take on long-term commitment to preserving the digital content they create ([DLM Forum 1997](#)). Every digital resource has a life cycle and different stakeholders and interests within this. However, it is both achievable and highly desirable that a dialogue is established between them when issues of long-term preservation are involved.

Det er derfor nødvendigt såvel

- at analysere livscyklus for det digitale materiale

som det er nødvendigt

- at identificere de relevante aktører og deres rolle

Livscyklen er selvfølgelig afhængig af, hvilken type materiale man undersøger, om det er videnskabelige produktioner, data, medicinsk information, sagsjournaler, kirkebøger, online aviser osv. osv. Afhængig af materialetype vil forskellige aktørerne være relateret til materialets livscyklus. Aktørerne omfatter bl.a.:

- Udgivere
- Forfattere
- Videnskabelige foreninger
- Universiteter
- Biblioteker
- Arkiver
- Distributionscentre
- Information Service Provider

Både JISC/NPO-studiet (HSJB 97) og ICSTI rapporten (HC 99) har undersøgt livscykel aktørernes holdning til bevaring af publiceret materiale. De fandt, at

- forfattere og brugere traditionelt ikke har været specielt involveret, men at dette er ved at ændre sig, idet forfattere ofte bliver involveret i at skabe metadata i forbindelse med deres arbejde
- faglige foreninger. Visse er begyndt at se en mission omkring sikring af deres medlemmers produktion – gælder specielt for forskning.
- udgivere er involveret, men typisk fordi deres økonomi hænger sammen med et levende arkiv. Det betyder, at deres strategi går ud på at migrere arkivinformerer – i alt fald så længe der er økonomi i det

- Biblioteker og arkiver har været de drivende kræfter i arbejdet. Disse er imidlertid ofte hæmmet af, at der trods politisk opbakning ofte ikke følger et budget med.

I et forsøg på at deles om byrden med at bevare informationer på webben initierede Nationalbiblioteket i Australien en projekt (HSC 02) ”Safekeeping”, hvor ideen var at opfordre og støtte materialeejere til selv at tage ansvaret for bevaring og sikre adgang nu og i fremtiden. Jeg anser initiativet som interessant, men det er vigtigt at det følges op af kontrol (f.eks. et system bygget over LOCKSS, (Lots of Copies Keeps Stuff Safe omtalt i afsnit 8.4).

Vender vi os mod aktører, som en involveret i indsamling af information identificerer ICSTI rapporten fire organisatoriske modeller:

- Datacentre
- Institutionelle arkiver
- Tredjeparts depoter
- Pligtaflevering

For disse gælder forskellige modeller for informationsflow, for materialehåndtering (skabelse, management, bevaring og adgang), ansvar og ejerskab til data og den økonomiske model bagved.

I fremtiden vil disse de netbårne informationer række ind over alle disse kategorier – og nye vil opstå. Det betyder, at vi meget snart kommer i en situation, hvor de traditionelle rammer kan blive en forhindring for en effektiv indsamlings- og bevaringsstrategi.

Et eksempel på nødvendigheden af nytænkning er samspelet mellem Alexa og archive.org, hvor Alexa er et kommercielt firma, som har som forretningsstrategi at gemme sider for firmaer, og hvor archive.org er et non for profit firma, som bl.a. står for waybackmachine-siten (web.archive.org) og dels sammen med Alexa leverer materiale til Library og Congress. I Danmark kan man drage en parallel til Statens Information, som tilbyder en kommerciel tjeneste over for virksomheder, som ønsker af forsikringsgrunde at bevare deres site og et evt. samarbejde med nationalbibliotekerne om langtidsbevaring af online materiale..

Bl.a. i forbindelse med analysen af situationen ved Library of Congress (omtales i afsnittet om USA) blev det fremført, at ingen enkeltinstitution hverken kunne eller burde alene løfte opgaven omkring digital bevaring. Samme udsagn er kommet fra andre, som har beskæftiget sig med denne opgave. Den fulde opgave er ganske simpelt for stor. Tænk på det arbejde, som ligger i en vedligeholdelse af et softwaredepot?

3.3.1 Punkter til afklaring

Det bør overvejes, om man gennem samarbejde og koordination kan effektivisere en fremtidig indsamling og bevaring. Eksempelvis kunne man forestille sig, at en institution får ansvaret for e-mails – uanset hvilken type organisation eller privatperson, som donerer/ønsker at donere sit e-mail arkiv – og får til ansvar at sikre, at de vigtigste af disse fremover samles.

Det er her vigtigt at afgøre, om konsortier og deling af arbejde styres af teknikken bag eller dem, som der træffes aftaler med.

Kunne man forestille sig, at en organisation var ansvarlig for inddrivelse af materiale, en anden for den videre bevaring?

Et spørgsmål, som blev adresseret i UK, var: Hvem er aktørerne – og hvad forventer de, og hvad har de planlagt mht. langtidsbevaring?

Hvem bevarer hvad? Eksempel DR – kan man få en browse-kopi? Eller hvilken kvalitet skal afleveres – og hvad må den bruges til?

Der bør laves en undersøgelse af, hvem som i Danmark har bevaring som en primær eller sekundær rolle – og hvad deres rolle er.

4 Strategi og handlingsinitiativer

4.1 Status Danmark

4.1.1 Status – biblioteker

Den meste digitalisering, som sker i Danmark, er af hensyn til formidling og er ikke en del af en bevaringsstrategi. Dette betyder, at kravet til langtidsbevaring ikke vil være på samme niveau, som hvis der havde været tale om unika. KB anvender såvel cd-rom som alm. tape som lagermedie, SB anvender et SUN tapesystem.

Digitalt materiale kommer på to måder: via pligtafleveringen som fast materiale og via pligtaflevering.dk som afsluttede, stationære værker. Krav i forbindelse med pligtaflevering er, at såvel dataobjekt som information om standardprogrammer som evt. proprietære programmer afleveres(SIP). I øjeblikket sikres, at beskrivelsen er komplet – ellers foretages ikke noget i forbindelse med denne aflevering. Online materiale hentes til værkarkivet. Nødvendige plug-inns afgøres via mimetype af filerne.

Denne situation er ikke holdbar i længden – både mht. til bevaringsstrategi og mht. indsamlingsstrategi. Derfor blev i efteråret 2000 nedsat en arbejdsgruppe med deltagelse fra Biblioteksstyrelsen, Det Kongelige Bibliotek og Statsbiblioteket med henblik på en afklaring af problemstillingerne i tilknytning til bevaring af den virtuelle kulturarv.

Formålet var at opstille forslag til konkrete tiltag inden for:

- Metoder til indsamling og bevaring
- Registreringsproblematikken
- Forslag til lovtiltag på området for den virtuelle kulturarv
- I hvilket omfang indsatsen på området for bevaring af den virtuelle kulturarv kan erstatte pligtaflevering af faste medier.

Arbejdsgruppens arbejde resulterede i en rapporten (20 sider), som opsummerede problemstillinger omkring indsamling, bevaring og adgang primo 2001.

Parallelt udarbejdede KB en bevaringsrapport: ”*Langtidsbevaring af digitale data for Det Kongelige Bibliotek. Forslag til handlingsplan for fastlæggelse af metoder*”, som dels identificerede KB’s digitale materiale og opgjorde tilvækst, dels identificerede udestående opgaver i forbindelse med implementering af en bevaringsstrategi. Rapporten indeholder endelig dels en oversigt over tiltag i andre lande og deres budget samt et forslag til en strategiproces.

I det efterfølgende arbejde blev bl.a. identificeret følgende hovedområder:

1. Sikring af den løbende produktion af den nationale digitale kulturarv
 - 1.1 Sikring af fuldstændighed i indsamling af netpublikationer efter den gældende lov for pligtaflevering af netpublikationer.

1.2 Web-arkivering. Sikring af den fremtidige adgang til den digitale kulturarv, der produceres løbende på nettet
2. Langtidsopbevaring af digitale ressourcer

Der blev foreslået en proces, som startede med en konference og blev efterfulgt af pilotforsøg. Derfor har et konsortium bestående af CFI, KB og SB med støtte fra DEF dels arrangeret en international konference (juni, 2001) med deltagelse af nøglepersoner inden for webarkivering og igangsat et pilotforsøg omkring webarkivering (august 2001 – august 2002).

Konferencens formål var at påbegynde formuleringen af en strategi for, hvordan den Digitale kulturarv sikres for fremtiden med fokus på strategier og metoder til indhøstning og bevaring af information, som er tilgængelig på internettet.

Proceeding kan findes via DEF's hjemmeside www.deff.dk eller via www.netarkivet.dk.

Arrangørerne havde håbet at få en debat i gang mellem det Australske Pandora projekt og det svenske Kulturarw3 projekt, idet disse to rent anvender hhv. en selektiv og en tværsnitsmetode. Inden mødet var de imidlertid blevet enige om, at det i fremtiden ville være nødvendigt at lave en kombination af de to.

Ud over en beskrivelse af disse initiativer afdækkede konferencen flere problemstillinger, bl.a. problematikken omkring konsistens af indhøstet materiale, uoverensstemmelse mellem det, som pligtafleveringsloven foreskriver gemt, og det, som rent faktisk bruges, og muligheder for overhovedet at høste relevant materiale.

Flere indlæg vil blive omtalt i afsnittet om indsamlingsstrategi.

Efterfølgende har DEF støttet et pilotprojekt omkring webarkivering, netarkivet.dk, som gennem forsøg analyserer en eventbaseret indsamling samt dansk deltagelse i det nordiske Nordisk Web Arkiv -projekt, som ser på fælles adgang til forskellige arkiver.

Pilotforsøget ser dels på de forskningsmæssige problemstillinger i forbindelse med at definere relevant materiale i forbindelse med en event, dels på de aftalemæssige forhold, og endelig undersøges de tekniske problemstillinger. Forsøget skal slutte af med en økonomisk vurdering.

Konkret skal følgende problemstillinger undersøges og belyses:

- Hvad skal arkiveres?
- Hvordan skal indsamlingen tilrettelægges?
- Hvordan lagres det indsamlede?
- Hvordan kan der gives adgang til det indsamlede?

Undersøgelsen blev gennemført i et koordineret, tostrengt forløb, hvor den ene streng rettede sig mod arkivering og efterbearbejdning, mens den anden streng rettede sig mod belysning af de forskningsmæssige behov for at have adgang til [netarkivmaterialer](http://netarkivmaterialer.dk).

Status er, at materiale er hentet og er ved at blive analyseret. Den vigtigste rapport nu er delrapport 2, som kan findes på www.netarkivet.dk, som redegør for

indsamlingsfasen – både metode til udvælgelse af sites og metoder til høstning af disse. Slutrapporten vil foreligge ultimo juli.

Der kan nævnes et par delkonklusioner:

Det viste sig endog meget vanskeligt og meget tidskrævende at lave frivillige aftaler i forbindelse med forsøget.

Den tekniske kvalitet af sites var meget variabel og til tider meget mangelfuld.

Det er nødvendigt med softwareudvikling i forhold til eksisterende programmer.

4.1.2 Status – arkiver

Statens Arkiver (SA) består af: Rigsarkivet, landsarkiverne i København, Odense, Viborg og Aabenraa, Erhvevsarkivet i Århus og Dansk Data Arkiv i Odense.

SA har som den eneste offentlige institution etableret en procedure mht. digital bevaring.

IT-afdelingen blev etableret i 1997 som en bevidst satsning med det formål at matche kravene til det digitale Danmark.

I en artikel af Jan Danielsen og Ulla Mortensen (DM 00) gennemgås fordele og ulemper af fire bevaringsstrategier:

- Den museale strategi
- Emuleringsstrategi
- Migreringsstrategi (eller konverteringsstrategi)
- Non Custodial Archiving.

Strategierne omtales senere, men her vil jeg nøjes med at sige, at ud fra en samlet betragtning af arkivmæssige, økonomiske og tekniske parametre, har man inden for SA valgt den migreringsstrategi, som vi senere vil kalde konverteringsstrategien. Som omtalt nedenfor har SA formået at komme ind tidligt i livscyklens for materialet, hvilket sammenholdt med bevaringsstrategien betyder, at man kun vil modtage digital information i veldefinerede formater.

SA har etableret en praksis, hvor de kommer ind tidligt i livscyklens for materiale. Konkret betyder det, at SA er med til at stille krav til journalsystemer i forbindelse med udbud inden for den offentlige forvaltning. I øjeblikket inden for staten, men lignende initiativer er på vej over for amter og kommuner.

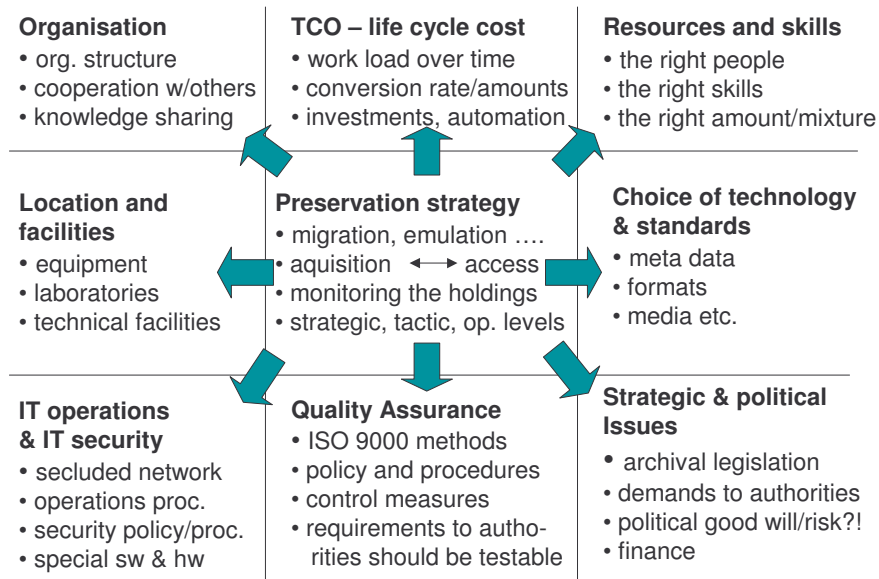
Pga. den eksplosive udvikling inden for IT indsamles det digitale materiale allerede 5 år efter dannelsen og ikke som sædvanlig efter 30 år.

SA har udviklet et system, som gør det nemt at tjekke, at modtaget materiale svarer til forskrifterne. Tjekket sker manuelt, men systemet gør, at det er indlysende, hvor evt. problemer opstår. Denne praksis er mulig pga. det relativt lille antal sager, som modtages. Til gengæld kan en enkelt sag godt være meget omfattende.

SA har arbejdet meget med journalplaner og emneord.

De efterfølgende overhead er udarbejdet af Birgit Hansen og viser SA's helhedssyn på digital bevaring.

Preservation – a holistic view



4.1.3 Status – private institutioner

Statens Information tilbyder en service over for private, hvor de høster deres site og brænder en cd-rom.

Aviser har opbygget arkiver med avisartikler. Fokus her er indhold, hvorfor de kan sammenlignes med vilkårlige informationsdatabaser mht. migrering og sikkerhed. Der tages backup, og der er taget forholdsregler for sikkerhedskopier på et andet sted.

Der er her klart tale om et levende arkiv, som vedligeholdes, fordi det genererer en indtægt.

Online aviser bevares ikke systematisk.

4.2 Nationale handlings- og strategiinitiativer.

Målet med dette afsnit er at identificere arbejde, som er gjort i forbindelse med formulering af handlingsplaner eller strategier og i forbindelse med etablering af en testbed eller et produktionssystem.

4.2.1 De nordiske lande

Traditionelt er der et tæt samarbejde mellem de nordiske lande, både inden for arkivsektoren og inden for bibliotekssektoren.

To af de nordiske lande var blandt de første, som rent faktisk begyndte webarkivering. Sverige har klart den største erfaring idet de der siden 1996 har gennemført regelmæssige høstninger af det svenske område.

Finland har været med i en række EU-projekter, som har været med til at placere dem strategisk inden for specielt digital bevaring. Nationalbiblioteket har en

alliance med CSC, det finske computercenter, som står for det meste af den tekniske udvikling i forbindelse med denne type aktiviteter. Bl.a. er det CSC, som har stået for udviklingen af NEDLIB høsteren, som bruges flere steder, bl.a. Island og Norge.

4.2.2 England

I 1995 stod British Library Research and Development Department for en konference "Warwick-konferencen", hvor man begyndte en diskussion af, hvad der var nødvendigt at afklare i forhold til digital bevaring (BL-41). Der blev identificeret en række områder, som burde belyses yderligere:

- En analyse af relevansen af anbefalingerne i en rapport fra RLG, som var på vej
- Identifikation af de vigtigste datatyper og formater samt problemstillinger, som påvirker langtidsbevaring af forskellige materialekategorier
- En undersøgelse af forfattere (originators), udgiveres og rettighedshaveres holdning til ansvarsfordeling for digital bevaring
- En undersøgelse af bevaringsbehov blandt universiteter og dem, som støtter forskning.
- Undersøgelse af arbejde gjort mod udvikling af guidelines for digital bevaring
- Rapport om metoder til at samle og indsamle materiale, om form og indhold af institutionelle elektroniske arkiver og om relevansen af den nuværende arkivpraksis over for digital bevaring

Efterfølgende besluttede JISC efter opfordring fra Management Committee of the National Preservation Office at støtte undersøgelseerne. Resultatet blev en række rapporter, som er summeret i "The digital Culture: Maximising the Nations's Investment", som blev udgivet i 1999 (F 99). Arbejdet resumerer resultater fra studierne, som belyser ovenstående spørgsmål:

- Beagrie, N. and Greenstein, D. (1998): *A strategic policy framework for creating and preserving digital collections*. British Library Research and Innovation Report 107. London: The British Library.
- Bennett, J.C. (1997) *A framework of data types and formats, and issues affecting the long term preservation of digital material*. British Library Research and Innovation Report 50. London: The British Library.
- The Data Archive, University of Essex (1998) *An investigation into the digital preservation needs of universities and research funders: the future of unpublished research materials*. British Library Research and Innovation Report 109. London: The British Library.
- Haynes, D., Streatfield, D., Jowett, T., and Blake, M. (1997) *Responsibility for digital archiving and long term access to digital data*. British Library Research and Innovation Report 67. London: The British Library.

- Hendley, T. (1998) *Comparison of methods and costs of digital preservation*. British Library Research and Innovation Report 106. London: The British Library.
- Matthews, G., Poulter, A., and Blagg, E. (1997) *Preservation of digital materials*
- Ross, S. and Gow, A. (1998) *Post-hoc rescue of digital materials. Digital archaeology? Rescuing neglected or damaged digital resources*. British Library Research and Innovation Report 108. London: The British Library.

BL har defineret en politik omkring pligtaflevering. Denne kan findes på:

<http://www.bl.uk/about/policies/workreplegdep.html>

BL har, sammen med Holland, lavet en kontrakt med IBM om levering af et system til håndtering af digitale dokumenter, som følger OAIS-modellen. Størrelsen af denne kontrakt kendes ikke, men det har været væsentlig dyrere end Holland.

Alle relevante parter inden for digital bevaring er gået sammen i "Digital Preservation Coalition", som tager initiativer til konferencer, og som nu står for at vedligeholde en håndbog med praktiske råd og oplysninger i forbindelse med bevaring.

4.2.3 Holland

Det Kongelige Bibliotek i Holland var projektleder på NEDLIB projektet og har gennem dette projekt opbygget stor ekspertise inden for systemer, som skal muliggøre digital bevaring.

Deres krav til et arkivsystem kan findes på nettet. Resultatet af deres udbud var et IBM system, som nu er ved at blive implementeret.

Holland har procedurer for, hvordan digitalt materiale lægges ind i et sådant system. F.eks. installeres programmer på en virtuel maskine, og dette diskbillede samles sammen med data.

Problemet omkring sikring af, at informationsobjektet kan fortolkes i fremtiden, er endnu ikke løst. De vil meget gerne invitere til samarbejde om dette (og andre) problemer.

Nationalarkiverne har netop etableret en testbed. Fra deres meddelelse om etableringen af denne hedder det:

DIGITAL PRESERVATION TESTBED

It has been said that the last decennium will be the worst documented era of the 20th century. Information will be recorded exponentially in digital form, without there being good methods to preserve this information for the long term. So, it is questionable whether future generations will be able to reconstruct what happened in our time.

The current Cabinet aims to carry out 25% of its transactions between government and citizens through digital means by

2002. Because of this, there is currently a great deal of work going on to develop strategies, methods, techniques and tools to handle the digital produce of the government in a responsible way.

The most important problem with the preservation of authentic digital records is technological obsolescence. Technological change is increasing exponentially. This brings up many questions, such as what to do with files that were made with old hard and software, which cannot be used anymore? Unless action is taken now, there is no guarantee that the current computing environment (and thus also records) will be readable by future computing environments with standard technology of that time.

To research solutions and strategies for this situation, the Ministry of the Interior and Kingdom Relations, and the Ministry of Education, Culture and Sciences (the National Archives) decided to establish a "Testbed" to gain the essential knowledge and experience.

<http://www.digitaleduurzaamheid.nl/>

4.2.4 USA

I USA blev det klart, at selv om Library of Congress (LC) havde langt den største kulturarv, var de på ingen måde førende mht. de nye muligheder, som den digitale alder giver. Det resulterede i nedsættelsen af en arbejdsgruppe, som skulle undersøge status på en række områder og give råd til LC. Resultatet af denne undersøgelse er udgivet i en bog: LC21 "A digital strategy for the Library of Congress", hvor specielt kapitlerne 3 "Building digital collections" og 4: "Preserving a digital heritage" er relevante. Generelt gav rapporten anledning til et initiativ National Digital Information Infrastructure and Preservation Program (NDIIPP) og en bevilling på \$5 mio. til en afklarings- og forslag til implementerings fase med lovning på yderligere 20 mio til en implementeringsfase. Endelig har Congressen stillet 75 mio i udsigt i en \$ til \$ matchning for private initiativer.

Rapporten identificerer 3 roller for LC i forbindelse med langtidsbevaring:

- Som skaber, aktivindsamler og havende det museale ansvar
- Som hovedansvarlig for en fail-safe mekanisme
- Som partner ved bevaring af distribuerede digitale samlinger.

Det bemærkes, at efterhånden som LC identificerer områder, hvor de har den primære rolle, bør andre institutioner justere deres bevaringsstrategi.

Desuden påpeges, at der er stor forskel på, som tilfældet er nu, selv at vælge formater og metadata standarder i forbindelse med egen digitalisering – og på, som det kræves i forbindelse med en fail-safe rolle, at have specialister, som kan håndtere og forstå vilkårlige formater. Som eksempel nævnes JSTOR.

Det konstateres, at byrden om at bevare de digitale samlinger er enorm og skal deles med andre institutioner med arkiveringsforpligtigelse..

Det konstateres videre, at LC ikke har den ledende rolle inden for dette område, og det anbefales, at der opbygges ekspertise gennem nyansættelser og personaleudvikling, som giver LC denne rolle. Desuden anbefales det, at LC arbejder sammen med Congress om at omformulere de IP begrænsninger, som der er på kopiering og migrering af data.

Rapportens anbefalinger er forsøgt efterfulgt via en række arbejdsgrupper, som adresserer forskellige aspekter. Bl.a. er initieret en tilbundsgående undersøgelse af state of the art i fem områder: Elektroniske tidsskrifter, E-bøger, Web, Lyd og video. Resultatet af dette arbejde er netop udkommet (LC02)

National arkivet, NARA har de sidste år brugt en del ressourcer på strategi for og forsøg med bevaring af e-mails. De har nu udviklet et system, hvor e-mails mærkes op i et XML format. Denne proces er meget effektiv.

To organisationer spiller en meget stor rolle i identifikation og løsning af problemstillinger i forbindelse med digital bevaring, nemlig RLG og CLIR.

RLG i samarbejde med OCLC netop publiceret [RLG 02] en rapport: *Attributes of a trusted Digital Repository*

Hvor følgende anbefalinger gives:

1. Udvikl en proces til certificering af digitale depoter
2. Forsk i og udvikl værktøjer til at identificere signifikante attributes i digital materiale, som skal bevares
3. Forsk i og udvikl modeller for kooperative depot netværker og services
4. Udvikl et system til entydig, blivende identifikation af digitale objekter som understøtter langtidbevaring
5. Undersøg og udbred information om den komplekse sammenhæng mellem digital bevaring og IP (Intellectual Property) rettigheder
6. Beslut, hvilken teknisk strategi som sikrer bedst adgang nu og i fremtiden
7. Definer et minimumsniveau for metadata, som er nødvendige for langtids management, og udvikl værktøjer til automatisk at generere og/eller udtrække så megen minimumsinformation som muligt.

CLIR står for en række meget relevante publikationer, som analyserer, rapporterer om forsøg og giver vejledning i emner omkring bl.a. digital bevaring.

Nedenfor er et relevant udvalg publiceret siden januar 1999. Disse samt den fulde liste kan findes på <http://www.clir.org/pubs/reports/reports.html>. Listen skal primært ses som en illustration af, hvilke typer initiativer, som støttes/igangsættes.

Pub106 *[Building a National Strategy for Preservation: Issues in Digital Media Archiving](#)* April 2002. \$20

pub100 *[Building and Sustaining Digital Collections: Models for Libraries and Museums](#)* August 2001. \$15

pub97 *[A Collaborative Approach to Collection Storage: The Five-College Library Depository](#)* by Willis E. Bridegam. June 2001. \$15

pub93 *Risk Management of Digital Information: A File Format Investigation* by Gregory W. Lawrence, William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney. June 2000. \$20

pub92 *Authenticity in a Digital Environment*
May 2000. \$20

pub89 *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment* by Anne J. Gilliland-Swetland. February 2000. \$15

pub87 *The Making of America II Testbed Project: A Digital Library Service Model* by Bernard J. Hurley, John Price-Wilkin, Merrilee Proffitt, Howard Besser. December 1999. \$15

pub77 *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* by Jeff Rothenberg. January 1999. \$20

4.2.5 Australien

Australien har været banebrydende omkring webarkivering. De startede et projekt PANDORA, som nu er gået i produktion. Strategien i Australien er at udvælge en række strategiske sites og følge dem med jævnlige høstninger – frekvensen afhænger af den enkelte site.

Man havde opstillet en vision:

By the year 2001 the information and cultural heritage sectors will be implementing strategies which will enable Australians to have appropriate, timely and equitable continuing access to significant Australian digital objects

som er gennemført.

Som beskrevet i udredningen om de retlige forhold, har Australien valgt en model, hvor de laver frivillige aftaler med ejerne af de forskellige sites.

Et glimrende overblik over aktiviteterne er givet i:

<http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html>

Australien betragter bevaring af fast materiale som et forskningsfelt, se:

A Draft Research Agenda for the Preservation of Physical Format Digital Publications, <http://www.nla.gov.au/policy/rsagenda.html>

Arkivet i Victorie har netop fået en bevilling på 8,2 mio. \$ (AUS) til videreudvikling af deres system VERS til håndtering af digitalt arkivmateriale.

Der er ikke samarbejde mellem biblioteket og arkivet, men der er tæt kontakt.

4.3 Internationalt relevante projekter og initiativer

Der er blandt alle, jeg har snakket med, en erkendelse af, at problematikken omkring digital bevaring er kompleks og i øjeblikket meget svær at prissætte, men også, at mange af problemerne kan løses internationalt. Samarbejdet kan ske på mange måder – fra fora, hvor man kan udveksle erfaringer, til mere struktureret samarbejde, hvor man går sammen om at få løst en opgave, som BL og KB har

gjort – til en tredje model, hvor man foretager en fælles softwareudvikling via et open source samarbejde. Præcist hvem, hvordan og hvor mange skal etableres.

En opgave, som er oplagt international (og dyr) er opbygningen af et system, som kan stille en service til rådighed, hvor man kan få fortolket et vilkårligt format. En sådan service kan opbygges på mange måder – at man deler forskellige formater mellem sig, at man opbygger en virtuel maskinpark eller på anden vis sikrer en fornuftig afvikling af en fil. Vi vender mere tilbage til denne problemstilling i afsnittet om bevaring.

4.3.1 EU

Digital bevaring er et indsatsområde under EU. Det spanske formandskab har netop udsendt et strategipapir, som er resultatet af en række møder med repræsentanter for alle EU-landene. Arbejdet har udmøntet sig i en resolution (udkast rundsendt 30/5). I denne står:

FORESLÅR følgende mål og vejledende foranstaltninger med henblik på yderligere analyse:

- stimulering af udviklingen af politikker for bevaring af den digitale kultur og arv samt deres tilgængelighed gennem
 - o oprettelse af samarbejdsrammer og –mekanismer mellem [...] medlemsstaterne med henblik på udveksling af erfaringer vedrørende politikker, programmer og dertil knyttede lovgivningsmæssige spørgsmål og på udvikling af fælles metoder;
 - o kollektiv og individuel støtte til de relevante lagringsorganisationer (f.eks. arkiver, biblioteker og museer) i forbindelse med deres ansvar for at indsamle digitalt indhold og sikre **dets** tilgængelighed fremover;
 - o undersøgelse af, hvilke organisatoriske infrastrukturer og tekniske standarder der er behov for til at støtte stabile og kompatible net af pålidelige lagre.
- øget oplysning og bevidstgørelse gennem net, der støtter deling af erfaringer og fremskridt, vedtagelse af passende standarder og måling og formidling af god praksis;
- undersøgelse af passende investeringer og analyse af omkostninger og virkninger for nuværende og fremtidig finansiering samt mulighederne for synergi mellem offentlig og privat investering;
- forbedring af færdighedsbasen ved at skabe mekanismer til udveksling af viden og færdigheder og til fortsat kortlægning af nye færdighedskrav og uddannelsesbehov;
- fremme af forskning i problemer og løsninger gennem udvikling af forskningsprogrammer, teknologitests og større eksperimentelle applikationer.

OPFORDRER KOMMISSIONEN OG MEDLEMSSTATERNE til i relevant omfang og inden for deres respektive kompetenceområder og med fuld respekt for subsidiaritetsprincippet at træffe eller udvikle disse foranstaltninger eller undersøge, om de kan gennemføres;

På det konkrete plan har kommissionen via specielt IST-programmet støttet og støtter fortsat en række projekter, som skal samle og udbrede viden om digital bevarelse, bl.a. ErpaNet (www.erpanet.org), og dels en arbejdsgruppe med

repræsentanter fra EU og USA, som skal identificere forskningsområder relevant for 6. rammeprogram.

ErpaNet holder en række workshops – den næste afholdes i november i København.

Der er etableret en række fora, hvor digital bevaring er med. Inden for arkiver afholder DLM årlige møder (se bl.a. www.dlmforum2002.org), hvor erfaringer kan udveksles. Inden for biblioteker har DELOS tages initiativer til en række workshops og står for den Europæiske ECDL-konference (www.ecdl2002.org).

Endelig skal nævnes en rapport sponsoreret af International Council for Scientific and Technical Information, Information Policy Committee og CENDI (HC 99). Den gennemgår situationen på en række arkiver rundt om i verden.

4.4 Spørgsmål

Der er allerede lavet mange undersøgelser om forskellige aspekter af digital bevaring. Det meste af det tekniske og strategiske arbejde kan umiddelbart overføres til den danske situation. Det skal overvejes, om man i en implementeringsfase af et arkiv til publiceret materiale skal forsøge at etablere et strategisk samarbejde med andre. Det bør overvejes, i hvor høj grad andres løsninger kan overføres direkte. Konkret bør man se på samarbejde med Holland og England samt USA og Australien. Det bør overvejes, i hvor høj grad man vil tilstræbe brugen af open source software.

Langtidsbevaring er klart internationalt, og Danmark skal sikre en repræsentation i relevante projekter og initiativer. Det skal overvejes, hvordan Danmark ønsker at udnytte 6. rammeprogram.

Det er også vigtigt at gøre sig klart, hvad man ønsker skal ligge nationalt, og hvad der kan løses internationalt.

Skal snapshot nødvendigvis ske i Danmark – kunne vi ikke lave et fælles nordisk eller europæisk arkiv? Hvad med den videnskabelige produktion – den ligger nu typisk i udenlandske databaser – er det ønskværdigt?

Som i andre lande bør det overvejes, om Danmark skal lave en gruppe, som kan tage initiativer på tværs af institutionsgrænser. Det skal overvejes, hvilke beføjelser og hvilken økonomi man ønsker at give en sådan gruppe.

5 Metadata

Metadata er på mange måder ryggraden i et informationsbevaringssystem.

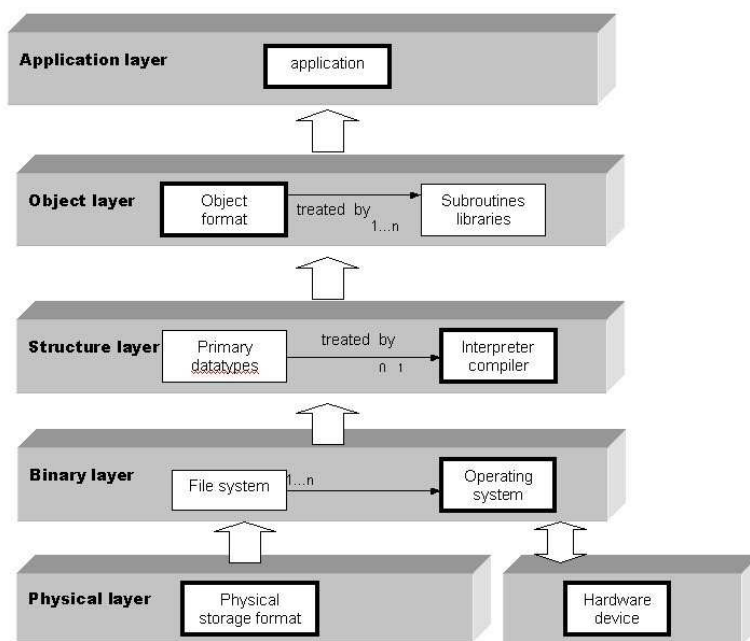
Metadata er strukturerede data om data. Traditionelt tænkes metadata som information, som er relevant i forbindelse med formidling. I forbindelse med digital bevaring vejer imidlertid andre prioriteter. En illustration af omfanget af metadata, som man forestiller sig er nødvendig i forbindelse med bevaring, gives af OAIS modellen, hvor to forskellige kategorier af metadata er identificeret:

- Repræsentationsinformation og
- Bevaringsbeskrivelsesinformation, som har fire/fem undergrupper (her er også den af NEDLIB foreslåede ”ændring” medtaget, (se diskussion 3.2.1):
 - Reference
 - Sammenhæng
 - Proveniens
 - Fixity
 - Ændringer

Ser man på mængden af data, og begynder man at overveje omkostninger ved at generere disse manuelt, forstår man, hvorfor automatisk generering af metadata er i fokus i flere lande.

5.1.1 Repræsentationsinformation:

Forskning omkring automatisk analyse af materiale – og dermed at mange af repræsentationsinformationerne kan genereres automatisk. At der vil være rigtig meget at vinde ved at automatisere genereringen af metadata kan ses af følgende struktur for repræsentationsinformation formuleret af NEDLIB projektet:



Lag	Beskrivelse
Fysisk lag	Et digitalt dokument opbevares på et fysisk medium, det kan være en CD-ROM, en floppy eller det kan være på en harddisk. Device drivers og chips sammenbygget med det fysiske lagermedium håndterer levering af bitstrømmen til det binære lag.
Binære lag	Bitstrømme samles i navngivne blokke uafhængig af det fysiske medium. Operativsystemet, som styrer filsystemet, giver funktionaliteten på dette niveau.
Strukturlag	Bit'ene samles til simple datastrukturer, som kan manipuleres af programmer. Den nødvendige funktionalitet stilles til rådighed af compilere og/eller fortolkere.
Objekt lag	Data struktureres til objekter, som meningsfuldt kan fortolkes af applikationer. Objektformater kan være åbne eller lukket/proprietært
Applikationslag	Programmer, som konverterer objekter til meningsfulde oplevelser

Figure 6. The layered information model from <http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm>

Information for ovenstående lag skal i princippet kendes for alt online materiale. I praksis er det i forbindelse med brug af standard- eller meget udbredt software selvfølgelig tilstrækkeligt at gemme informationen om navnet på softwaren. Parallelt opbygges så et softwaredepot, hvor ovenstående angives.

Som er berørt før, vil det give mening at pålægge et sådant softwaredepot centralt og knytte det sammen med en opgave om at sikre, at dette software kan afvikles på en fremtidig platform.

5.1.2 Bevaringsbeskrivelsesinformation

Som allerede omtalt har denne fire/fem kategorier, som skal defineres ved input/ingest.

OCLC/RLG har netop udgivet en rapport (januar 2002) med en oversigt over bevaringsmetadata. (http://www.oclc.org/research/pmwg/presmeta_wp.pdf) samt en rapport (april 2002) med anbefalinger (http://www.oclc.org/research/pmwg/pres_desc_info.pdf)

I rapporter opstilles grafiske modeller for, hvordan informationen struktureres samt gives anbefalinger til, hvordan disse udfyldes.

Skal det overhovedet være muligt økonomisk at etablere digitale arkiver er det vigtigt, at så mange af disse informationer genereres automatisk, og at man ser på metoder, som underbygger en høj grad af automatik i forbindelse med genfindning. Det gælder om i så høj grad som muligt at hente eksisterende informationer, såfremt der allerede er knyttet metadata til dataobjektet. Eksempler her er:

I forbindelse med langtidsbevaring planlægger SB at rippe cd'erne for at lægge musikken ind i et tapesystem. Meget af musikken ligger beskrevet i nationaldiskografien, hvorfor en proces skal tage højde for en udnyttelse af denne information.

Et andet SB-eksempel er opbygningen af et digital arkiv med pligtafleveringspligtigt radio/tv. Her vil man prøve at udnytte de metadata, som sendes med signalet i forbindelse med eksempelvis kabeltjenester.

Et tredje eksempel er pligtaflevering.dk, hvor materialet allerede er beskrevet i forbindelse med aflevering. I stedet for at gentage dette arbejde, udnyttes anmeldeblanketten. Denne suppleres med fritekstsøgning.

I forbindelse med aflevering og høstning kan automatisk laves en række kvalitetstest.

5.2 Opsummering af spørgsmål

Metadata er essentielle for langtidsbevaring for såvel det, at genfinde materiale som for at sikre den korrekte håndtering af materialet som led i bevaringsstrategien. Generering af relevante data er imidlertid også omkostningstung, hvorfor det er vigtigt at analysere denne proces igennem for at identificere muligheder for automatisering af processen.

I hvor høj grad kan metadata genereres automatisk eller ud fra eksisterende informationer?

Hvilke data er minimumskravet?

Eksisterende beskrivelse for metadata bør gennemgås og det bør overvejes, hvilke informationer er essentielle, hvilke, som kan genereres automatisk eller ved kilden, og hvordan kvalitetskontrol af data gennemføres

6 Indsamling

The digital world is one where objects are dispersed across the Web and where they are continually subject to change. Increasingly, digital objects are themselves dynamic information handling systems capable of interacting with the user. This changes the user experience as the result of user-interaction and choice, and the context and way in which the object is used become an important element in what may be preserved. New methods and concepts are needed to understand how and what to archive so that the integrity of the digital resource and the context of its creation and use are preserved.

Målet med indsamlingen er at skabe en samling af dataobjekter som hurtigst muligt suppleres med repræsentationsinformation og hvor relevant, den nødvendige software. Disse kan være af relativ simpel type som billeder, tekstorienterede artikler, lydfiler osv.; de kan være mere komplekse, som eksempelvis sager fra statslige institutioner, eller der kan være store datamængder. Det kan være afsluttede, afgrænsede publikationer, som de, der er omhandlet af den nuværende pligtafleveringslov, eller det kan være materiale, som er skabt til konstant at ændre form og fremtoning.

Der skal være enighed om, hvad en enhed er. I forbindelse med websider – er det den enkelte side, er det alle sider inden for et givet domæne eller er det siden plus alle links et step ud eller noget helt tredje? Der er mange muligheder og ingen sandhed – det skal defineres.

Grunden til det kan være vigtigt at afklare dette spørgsmål er, at det digitale objekt typisk vil have sin beskrivelse, og det vil typisk være til den enhed, som der senere skabes direkte adgang til.

Diskussionen her kan deles i to

- en som omhandler strategier for udvælgelse af materiale og
- en som omhandler metoder til implementation

Som ved det meste, er de to spørgsmål ikke fuldstændig afkoblede, men man kan godt f.eks. diskutere strategi for udvælgelseskriterier uden at tage stilling til, om materialet høstes, eller om det afleveres.

En forudsætning for, at det overhovedet giver mening at se på langtidsbevaring er, at der er noget at bevare. En meget vigtig komponent er derfor det første punkt ovenfor, en formulering af selektionskriterier for, hvad der skal bevares. Skal vi bevare så meget som muligt? Hvilke selektionskriterier skal vælges? Hvem er aktørerne i de forskellige dele af processen? Hvilken procedure anvendes i forbindelse med indsamling af materialet – aflevering eller hentning/høstning?

6.1 Terminologi

Vi skelner i det følgende mellem indsamlingsstrategi, som handler om kriterier for at vælge materiale ud, og indsamlingsmetoder, dvs. hvordan det valgte materiale kommer fra udgiveren til det digitale arkiv.

6.1.1 Indsamlingsstrategi

- Tværsnit: Ideen ved tværsnitsarkivering er, at man med veldefinere mellemrum gemmer en passende mængde af det materiale, som er offentlig tilgængeligt. Eksempler er Internetarkivet i USA og strategien med at indsamle alle radioudsendelser i en bestemt uge her i Danmark.
- Selektive: Værdien af enkelte objekter for fremtiden vurderes og de, som vurderes som værdifulde, bevares. Eksempler er Statens Arkiver i Danmark og Pandora i Australien.
- Begivenhedsorienteret: En begivenhed vurderes som havende betydning og man søger at gemme materiale, så eftertiden kan følge såvel reaktioner som indhold af begivenheden. Eksempler er September 11 arkivet, præsidentvalget i USA og kommunevalgspilotforsøget her i Danmark.
- (Pligt)aflevering – Regler definerer kategorier af materiale, som skal afleveres. Udgivere beslutter, om et givet materiale skal afleveres iht. Loven. Forskellen på denne kategori og den selektive er, at her defineres typer af materiale og ikke enkeltstående dokumenter. Et eksempel er den nuværende pligtaflevering af statiske værker: andre, som ikke er omfattet af pligtafleveringsloven, er online aviser.

6.1.2 Indsamlingsmetoder

- Aktiv indsamling: Materialet hentes manuelt eller automatisk enten vha. en høster eller på anden vis.
 - o Manuel indsamling: Enkeltfiler hentes og lægges ind i et arkivsystem. Et eksempel er pligtaflevering.dk, hvor enkeltværker anmeldes og efterfølgende hentes.
 - o Automatisk indsamling: Materiale høstes automatisk ud fra kriterier såsom et domæne, en række manuelt eller automatisk generede URL'er og lægges automatisk ind i et arkiv. Eksempler her er webarkiveringsprojekterne Kulturarw3 i Sverige og Internet Archive i USA
- Passiv indsamling: Ved passiv indsamling afleveres eller doneres materialet til arkivet.
 - o Aflevering: Udgiveren har ansvaret for at aflevere materialet på en på forhånd aftalt måde, evt. sammen med en beskrivelse. Et eksempel er afleveringen, som den sker til Statens arkiver.
 - o Donation: En samling af materiale doneres. Skal ofte efterfølgende struktureres.

6.2 Afvejning af indsamlingsmetode i forhold til materiale type

Forskellige strategier og materialetyper egner sig til forskellige indsamlingsstrategier. Pga. kompleksiteten ser vi, at forskellige institutioner og lande vælger en teknologisk løsning. Som det kom frem på konferencen i juni i

København, er dette mere et udtryk for manglende økonomi, mere end det er et resultat af tekniske overvejelser.

Nedenfor er givet en oversigt over, hvad forskellige lande og institutioner gør

	Eventbaseret	Selektiv	Tværsnit
Aktiv (automatisk og manuel hentning)	Strategi i USA Mulighed i forbindelse med pligtaflevering.dk	Strategi for Australien	Kulturarw3 Nordiske lande Archive.org
Pasiv (Aflevering og donering)		Strategi hos Statens Arkiver. Generelt strategi hos arkiver.	Sker i dag med lokalfjernsyn

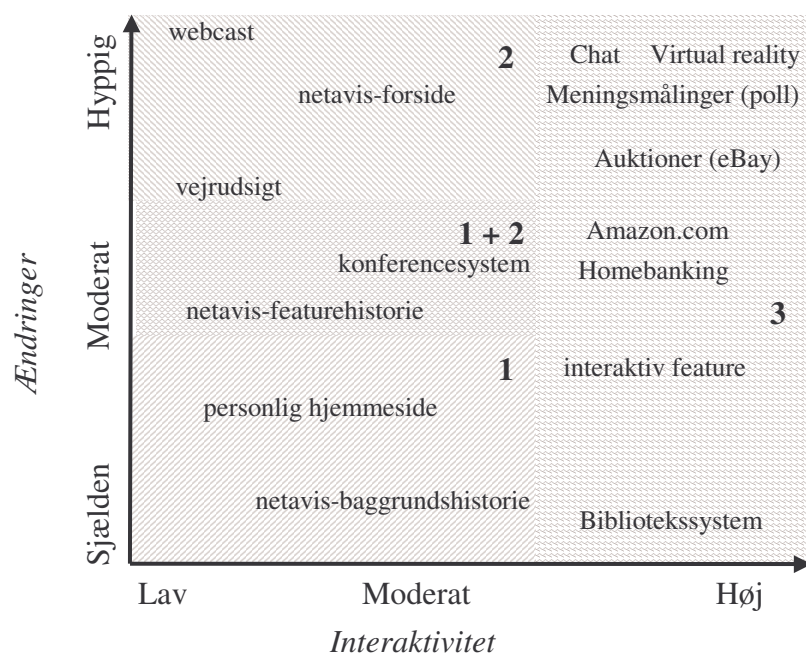
I forbindelse med den selektive strategi sker der en ret detaljeret beskrivelse af de udvalgte emner (arkivfond, websted osv.). I forbindelse med den eventbaserede høstning i USA forventer man på LC kun at lave en beskrivelse for hver event. Genfinding inden for begivenheden vil ske på anden vis.

I forbindelse med den amerikanske strategi forventes, at eventen som helhed beskrives i bibliotekssystemet, men ikke de enkelte komponenter

I forbindelse med tværsnitshøstninger er en beskrivelse af de enkelte sites umulig. Det er derfor nødvendigt her at bruge alternative strategier så som fritekstsøgning.

Forskellige informationstyper egner sig til forskellige indsamlingsstrategier og – metoder.

To vigtige parameter i forbindelse med indsamlingsstrategi for webmateriale, opdateringsfrekvens og grad af interaktivitet (ofte = kompleksitet). Afhængig af værdien af disse bør vælges forskellige strategier mht. høstning, som illustreret på figuren nedenfor:



En sammenhæng mellem arkiveringsstrategi og –metode og parametre bag materialet. På tegningen står 1 for tværsnit og 2 for selektiv strategi. I område 1 og 2 kan traditionelle høstemetoder generelt anvendes. I område 3 kan selektiv eller afleveringsstrategier bruges – men strategier kan ikke implementeres med en traditionel høstemetode.

6.3 Eksisterende erfaring

6.3.1 Selektiv indsamling

Australien benytter selektiv indsamling. Fra deres strategipapir får vi deres kategorier: <http://pandora.nla.gov.au/selectionguidelines.html>:

- Annual Reports
- Commonwealth Government Publications
- Digitised Materials
- Educational Material
- Exhibitions
- Ethnic Community Sites
- Indexed Publications and Articles
- Juvenile Publications
- Literary Works
- Newspapers
- Organisational and Personal Sites
- Other Protocols, eg. FTP, Gopher
- Pacific Publications
- Religious Sites
- State Government Publications
- Maps

Man har ca. 1200 sites, som høstes selektivt. Som omtalt i kapitel III har man her frivillig aftale og aftaleformularen vedlagt i bilag 3. Man indhenter kun materiale, hvor man kan få godkendelse til også at give adgang.

Hver site er registreret i biblioteksdatabase

6.3.2 Eventbaseret indsamling

LC i USA benytter en eventbaseret indsamling. De planlægger at indsamle websider i forbindelse med præsidentvalgene. De har også gennemført en September 11 indsamling. Denne blev iværksat samme dag vha. bibliotekernetværket – den ansvarlige for projektet skrev ud til alle fagreferenter, forskere og andre, som var en del af LC's sædvanlige netværk, og bad dem om dels at indrapportere relevante sites. Dette lykkedes over al forventning.

I Danmark gennemføres et pilotprojekt, som så på aktiviteterne omkring kommune- og amtsvalget. 2000 sites blev brugt som start URL'er til de to høstmaskiner.

6.3.3 Tværsnit

Alle de nordiske lande er i dag involveret i webarkivering i større eller mindre skala – med undtagelse af Sverige, som er midlertidigt stoppet.

Sverige startede webarkivering i 1996 og har siden foretaget 1-2 tværsnitshøstninger om året. Sidste høstning var i 2001.

Der er etableret et samarbejde mellem nationalbiblioteker i de fem nordiske lande om at udvikle en fælles adgang til webarkiverne. Dette projekt, Nordisk Webarkiv, er støttet af Nordunet2.

En andet problem i forbindelse med tværsnitshøstning er, at den synlige del af Weber er meget mindre end den samlede mængde materiale. Man taler om det dybe web – det, som ikke umiddelbart er tilgængeligt. Dette kan være Intranet, det kan være information i databaser, det kan være information skjult bar brugernavn og password. Langt den største informationsmængde er det dybe web (LV 00)

6.3.4 Aflevering

Arkiverne benytter afleveringsmetoden. De kan være med til at stille krav til leverandører af sagsbehandlingssystemer, så de sikrer, at systemet kan udtrække relevante informationer og låse disse, når en sag er færdig. Systemet fungerer til alles tilfredshed.

I webarkiveringsprojektet er aflevering også undersøgt. Her fik man en site fra Bornholm. Ved ankomsten viste det sig, at den var udviklet i ASP. Denne kunne umiddelbart installeres og kører – men at langtidsbevare den i denne form vil være besværligt.

6.3.5 Høstning

Som er blevet bemærket i forbindelse med afviklingen af pligtaflevering.dk og bekræftet af alle andre, som kører høstninger, findes der meget på internettet, som

ikke automatisk kan høstes. Bl.a. vil brugen af javaskrift forhindre dagens høstere i at fungere, ligeledes volder streamet audio og video problemer. Kræver en side input (f.eks. tekst i et felt) for at komme videre, stopper systemet. Der kan dog kodes ud af denne situation. Aschenbrenner (A 01) fra Østrig har undersøgt denne problematik og er kommet med forslag til, hvordan man kan lære systemet at gætte, hvad der er et relevant input.

Nedenstående oversigt blev udarbejdet i forbindelse med netarkivet.dk (delrapport 2) og viser, hvilke materialetyper der vil kunne indsamles med hver af de 3 indsamlingsprogrammer:

<i>Materialetype</i>	<i>Optimal høster 2002</i>
Ukomplerede/lidt komplicerede web-sider	Ja
Komplicerede websider	Nej
Streamet video	Nej
Streamet audio	Nej
Chat	Delvist, ja hvis det er HTML baseret
Net-konferencer	Kun dem styret af traditionel http-authentication
Mailinglister	Nej
Nyhedsgrupper	Nej
WAP-sider	Nej
SMS-beskeder	Nej

6.4 Opsummering af spørgsmål

Der skal laves en strategi for, hvad skal indsamles, og hvordan denne indsamling effektueres.

Afvejningen mellem aktiv indsamling og passiv indsamling.

Er der visse typer materiale, som hører til den selektive gruppe, f.eks. on-line aviser og magasiner?

Har vi begivenheder, som bør følges regelmæssigt, f.eks. valg (såvel folketings- som kommune- og amtsvalg), kongelige begivenheder ...?

Er der mere, som fortjener regelmæssig opfølgning, eksempelvis integrationsdebatten?

Skal der laves tværsnitshøstninger – og i givet fald hvilke domæner og hvor tit?

I hvor høj grad kan man få værkskaberne til at generere den beskrivende information?

Man kan godt forestille sig, at noget materiale vil egne sig til kortere opbevaring og andet til langtidsbevaring. Der bør formuleres en strategi for, hvad man ønsker at bevare – og hvor lange man forventer at kunne komme til materialet.

Vi har identificeret to basalt forskellige materialetyper: født digital og digitaliseret. Ofte vil den digitaliserede kopi være fremstillet af hensyn til formidling og ikke af hensyn til bevaring. Der ligger dog en stor investering i denne digitaliseringsproces. Såfremt der ikke er midler nok til bevaring af begge skal formuleres en prioritering mellem unika og dyre kopier.

7 Ingest

Ser vi på OAIS modellen, er ingest den process, hvor materiale modtages, kontrolleres for såvel korrekthed som kompletthed, og hvor evt. ekstra informationer, som er nødvendige af hensyn til bevaring, tilføjes.

7.1 Terminologi

Autentifikation: En proces, hvor man får bekræftet, at en bruger, service udbyder, ressource eller information er bona fide. I sikkerhedssammenhæng bruges termen ofte og hentyder til den proces, hvorved en modtageren af en digital besked er sikker på identiteten af afsender, og at meddelelsen ikke er ændret.

Autoritetskontrol: Processen at verificere og autorisere valget af entydigt identifikation såsom personnavn og emne og sikre, at dette identifikation er entydigt brugt. (autoritetsposter) Inden for sikkerhed – at afgøre om en identitet har ret til at foretage en handling, eksempelvis få adgang til en ressource.

7.2 Autentifikation og kvalitetskontrol

Som en del af modtagelsen er det vigtigt at sikre, at

- objektet er det, som det foreskriver at være,
- det er komplet,
- det ikke er virusinficeret,
- metadata er komplette – hvis ikke, supplere disse
- der er reference til såvel oprindelse som ny placering
- osv.

8 Bevaring

En bevaringsstrategi falder naturligt i tre dele:

- en, som sikrer dataobjektet og
- en, som sikrer, at dataobjektet kan fortolkes og vises
- en, som sikrer, at data er identisk med de oprindelige, og i tilfælde af ændringer, at disse er veldokumenterede

Et fundamentalt spørgsmål er, hvad man ønsker at bevare – det fulde system eller ønsker man at bevare informationen? Lad mig prøve at illustrere spørgsmålet med et par eksempler.

Ønsker man at bruge hulkort for at få svar på datamateriale, som blev optaget i 70'erne?

Ønsker man at se online information fra 80'erne på en sort/hvid skærm med en 480*640 opløsning?

Svaret afhænger selvfølgelig af, hvilket formål man har med at opsøge dataene, illustreret ved et andet eksempel – ønsker man at finde indholdet af en avisartikel (via Polinfo), eller ønsker man at se hele siden, hvor artiklen forekommer.

På samme måde kunne man forestille sig, at man ønskede at forstå forskningsmetoder dengang i 70'erne, hvor en 300 baud TTY terminal forbundet med computercentret var vild luksus.

Bevaringsstrategien vil klart afhænge af formålet. Er det vigtigt at bevare oplevelsen, må man sikre migration af såvel data som programmer til at vise disse på en måde, som er ”trofast” over for originalen. Er man på den anden side primært interesseret i oplevelsen, vil det eksempelvis være nok at specificere, at dette billede er gemt i med G4 kompression, hvorimod den aktuelle viewer ikke er vigtig. Og vigtigere endnu, man behøver ikke migrere den relevante software, man behøver blot i fremtiden at kunne skrive software, som kan fortolke den G4-komprimerede bitstrøm.

8.1 *Materialetype, formater og bevaring*

Et gennemgående tema i litteraturen omkring bevaring er, at man skal forsøge at bruge åbne standarder. Skal man om 50 år fortolke et dokument, er det vigtigt, at man ved, hvordan bittene skal fortolkes. Har man brugt et lukket format som eksempelvis Word, er eneste chance for at fortolke bitstrømmen, at man har bevaret et eksemplar af denne software, og at man kan etablere en platform, hvor man afvikle det. Brugtes et åbent format, kunne man skrive ny software, som kunne fortolke bitstrømmen. At et format er åbent, behøver ikke betyde, at det en åben standard. Eksempelvis Adobes postscript og pdf formater er åbne.

Indsamles eller afleveres materiale i lukkede formater, bør det overvejes, om man kan lave en overspilning, som bringer det på en standardiseret form. Arkiverne i Danmark kræver aflevering i en række veldefinerede og veldokumenterede formater, hvilket klart øger chancen for, at man kan fortolke disse i fremtiden men det giver også en chance for, at omkostningerne ved bevaring er væsentlig lavere, end hvis vilkårlige formater blev accepteret.

Traditionelt har biblioteker accepteret alle formater. Det bør overvejes, om alle formater skal bevares, eller om man kun vil indsamle visse (nok væsentlig flere end arkiverne), som stadig giver en rimelig mulighed for at kunne indsamle et repræsentativt udsnit af det publicerede materiale.

I 1997 publiceredes resultatet af en undersøgelse: "A Framework of data types and formats, and issues affecting the long term preservation of digital material" (B 97). I denne gennemgås materialetyper og filformater for hver gives en basisværdi, en kompleksitetsværdi samt en risikoanalyse. Også medityper gennemgås. Der bør laves en opdateret version af denne analyse, og denne bør benyttes til at lave en prioriteringsliste over foretrukne filformater. Det bør overvejes, om man ønsker at indsamle materiale, hvor chancen for at vise det om ti år er meget lille.

At dette forekommer, kan illustreres af et eksempel. I 1991 udgav DR og UNI-C et multimedieprodukt "Siulleq", som samlede materiale om Grønland. Udgivelsen bestod af en cd-rom (til MAC), en laservision plade samt en bog med strekkoder. For at afvikle dette produkt kræves:

- En MAC, som kører
- En laservision afspiller
- Et fjernsyn

Ved demonstrationer krævedes en stationcar for at medbringe det relevante udstyr. Laserdisk afspilleren blev styret af MAC'en. Materialet er interessant, for det var banebrydende og et af de første eksempler på multimedia produktion.

Men hvad er chancen for at bevare dette produkt?

Og kunne vi – skulle det så i fremtiden afvikles på en MAC med 11" s/h skærm samt et 50 Hz 28" fjernsyn?

8.2 Terminologi

Overspilning (Refreshing) – en periodisk overspilning af data til samme type lagermedium uden ændring af bitstrømmen.

Modificeret overspilning: Overspilning til et lignede medium uden at ændre bitstrømmen – eksempelvis fra 5 ¼" floppy til 3 ½" floppy eller fra et tapedevice til et andet.

Nedenfor vil en række metoder til bevaring af fortolkningen blive omtalt. Disse er:

Emulering: at skabe computeromgivelser, som giver mulighed for at afvikle et computerprogram udviklet til en platform på en anden. Eksempler på emulering er Softwindows, som giver mulighed for at emulere et Intelmiljø på en ikke-intel platform.

Migrering: Materialet konverteres med mellemrum til nye version af den relevante software. Dokumenter i Word er et eksempel.

Konvertering: Speciel form for migrering, nemlig den hvor materialet konverteres til et standardformat.

Museal strategi: Man opretter et computermuseum, som indeholder alle typer hardware, styresystemer og softwarepakker.

8.3 Bevaring af informationsobjektet

Informationsobjektet er det, vi ser, hører eller på anden vis oplever. Dette kan bevares på forskellig vis, enten som de oprindelige data samt relevante programmer eller ved, at det er transformeret til et standardformat.

Nedenfor bruges et meget simpelt eksempel – et tekstdokument. Dette er gjort for overskuelighedens skyld. Man skal være opmærksom på, at vi ofte har at gøre med meget komplekse digitale objekter, hvor udsagnene ikke er så oplagte, som de måske kan forekomme nedenfor.

En metode, som ofte omtales i forbindelse med bevaring, er digital arkæologi. Jeg har ikke medtaget denne strategi, da jeg anser den for at være resultatet af manglende strategi. Men det er da godt at vide, at der findes firmaer, som er ved at lave en forretning ud af at genskabe folks tabte data.

8.3.1 Migrering

Ideen i migrering er, at man forsøger at holde materialet så tæt på sin oprindelige form som muligt, men at man løbende laver en formatkonvertering, når nye versioner kommer på markedet. Eksempelvis konverteres et Word 2.0 format løbende, således at det stadig kan læses med Word, men nu Word 2000.

Fordelen er, at det er relativt billigt at komme i gang her og nu. Ulempen, at denne metode kan være omkostningsfuld på længere sigt. Endelig er det langt fra alle materialetyper, som egner sig til migrering – og hvis en sådan gennemtvinges, kan det forårsage tabt funktionalitet.

Ønsker man at gennemføre migrering, kræver det, at man har defineret en politik for, hvilke formater som understøttes på et givet tidspunkt.

8.3.2 Konvertering/standardformat

Standardisering går ud på, at man enten kræver, at ting afleveres i bestemte formater, som eksempelvis er tilfældet hos Statens arkiver, eller at man konverterer data til et standardformat. Eksempelvis kan et Word-dokument konverteres til pdf, HTML eller XML – og gemmes som sådan.

Fordelen ved metoden er, at det er meget simplere og nok billigere at sikre langtidsbevaring. Man skal dog være opmærksom på, at migreringsmetoden sikkert vil kræve, at der overspilles til nye standardformater, efterhånden som disse udvikles.

Ulempen er, at det er omkostningsfuldt her og nu, og at metoden har et relativt begrænset område af medietyper. Det skal dog ikke undervurderes, at mængden af materiale, som faldet inden for disse medietyper, er meget stort.

8.3.3 Emulering

Emulering består i at sikre, at det oprindelige signal kan fortolkes ved at emulere gamle systemer på nye. Emulering kan etableres på (mindst) to fundamentalt forskellige måder:

Gennem udvikling af software (se R 99) eller ved at udvikle en såkaldt virtuel maskine, som foreslået af Lorie fra IBM (se f.eks. Lorie 01).

Fordelen ved denne metode er, at den giver mulighed for at genskabe det oprindelige look and feel. En anden fordel er, at man gemmer data og de oprindelige programmer, dvs. der bliver ikke tale om at overspille og forholde sig til formatkonvertering. Ulempen er, at det kan blive dyrt at etablere emulatorer for samtlige systemer. Ydermere er denne metode stadig på forskningsstadiet, om end der allerede er etableret en lang række emulatorer til f.eks. Commodor 64, Amiga og andre.

Et andet problem er, at der skal sikres, at software er legalt, at det ikke udløber og andre lignede potentielle problemer.

Der er ingen tvivl om, at etableringen af disse emulatorer vil være en ekstremt kostbar ting, på den anden side er det en potentiel service, som kun behøves få steder i verden. Der er altså tale om et oplagt område for internationalt samarbejde.

En variant, som fungerer over for standardformater, er at etablere et arkiv med software, som er skrevet i åben, standardiseret kode, og som kan læse data af givne mime-typer og med veldefinerede strukturer.

8.3.4 Museal strategi: Teknisk Museum

En tredje strategi vil være at gemme eksemplarer af computere såvel som de forskellige drev til forskellige lagringsmedier. Alternativt kunne man gemme dokumentation for såvel hardware som software (i det omfang det overhovedet eksisterer). En anden variation blev nævnt ovenfor, nemlig virtuelle maskiner. Ideen i disse er at kunne programmere chips til at opføre sig som tidligere chips.

Et andet problem, som jeg vil beskrive her, selv om det også er relevant for eksempelvis emulgeringsstrategien, er problemet med drev. Et potentielt problem for fastform materiale er muligheden for at læse dette. Hvor mange kan stadig læse de engang meget udbredte 5 ¼ ” disketter? Det kan stadig lade sig gøre. Hvad chancen så er for, at disketten ikke er ødelagt er så en helt anden sag – som hører under sikring af data.

Igen er der forslag om at gemme specifikationerne, således at man i fremtiden kan læse disse på nykonstruerede drev.

8.3.5 Konvertering til analog form

Det lyder måske lidt barokt i vore dage, men der tales faktisk om konvertering af specielt tekst og billeder til analog form. Grunden er, at bevaringsstrategi for disse er velkendt, og at disse medier og at har en længere forventet levetid end det digitale medium. Sos fra NARA sagde i sin præsentation på CLIR-mødet i marts,

at de foretrak mikrofilm frem for en digital version. Warner Brothers gemmer film og ikke en digitaliseret version.

Der er flere ulemper ved denne form. Adgang er den vigtigste, men det skal også erkendes, at der ofte tabes funktionalitet ved at konvertere det digitale signal til analog form.

8.3.6 Sammenligning

I DM 2000 gives følgende sammenligning af de tre metoder (i artiklen er også indført en fjerde, non custodial, der dog, som også påpeges i artiklen, ikke er en bevaringsstrategi):

Strategi	Arkivmæssigt	Teknisk	Økonomisk
Museal	God	Umulig	Dyr
Emulering	God	Svær	Dyr
Konvertering	Rimelig	Rimelig	Rimelig

E forskellige metoder har fordele og ulemper – og som det siges flere steder, man skal ikke forvente, at man kan vælge en universel metode. Den vil afhænge af omstændighed og muligheder.

8.4 Bevaring af dataobjekt

En bevaringsstrategi går ud på, at dataobjektet bevares intakt, og der sker en løbende udvikling af den applikation, som skal fortolke bitstrømmen. I denne strategi er det derfor essentielt at have en strategi for, hvordan man sikrer disse dataobjekter.

Dataobjektet er filen med 0'er og 1'taller, som enten ligger på et fast medium eller ligger online i et eller andet filsystem.

Målet med bevaring er at sikre, at denne bitstrøm kan genskabes, dvs. at de givet det rette operativsystem og drev kan læses.

Der findes forskellige strategier for dataobjekter, som findes online, herunder

- overspilning evt. modificeret overspilning hvert femte år
- test af kvalitet – overspilning (og dataopretning) hvis kvaliteten forringes (bruges ofte i forbindelse med tapearkiver)
- Statistisk bestemt sampling – og efterfølgende kvalitetsbedømmelse

En ting, som også hører med her, er hvordan man sikrer sig mod ved f.eks. en katastrofe at miste informationen. I Danmark har man, som i de fleste andre lande, en centralistisk politik, dvs. at bittene gemmes i specielle institutioner, som har en meget høj grad af sikkerhed omkring bevaring. En anden mulighed, en decentral lagring, kan ske for specielle typer materiale, eksempelvis forskningspublikationer. Et meget interessant projekt er LOCKSS. Ideen i LOCKSS (Lots Of Copies Keep Stuff Safe) er, som navnet siger, at man kan sikre langtidsbevaring (I alt fald på dekadenniveau) ved at sikre, at der forefindes et passende antal kopier af materialet. LOCKSS fokuserer på noline artikler og er et

samarbejde mellem 40+ universitetsbiblioteker og 30+ udgivere. Ideen er, at systemet løbende tjekker kvaliteten af en kopi. Er denne ikke god nok, sørger systemet for at overskrive den med en korrekt kopi.

Generelt anbefales, at man har etableret en katastrofebeskrivelse, som er godkendt af direktionen, og som har etableret relevante procedurer.

8.4.1 Lagermedier

...Reading and understanding information in digital form requires equipment and software, which is changing constantly and may not be available within a decade of its introduction. Who today has a punched card reader, a Dectape drive, or a working copy of FORTRAN II? Even newer technology such as 9-track tape is rapidly becoming obsolete. We cannot save the machines if there are no spare parts available, and we cannot save the software if no one is left who knows how to use it.

Rapid changes in the means of recording information, in the formats for storage, and in the technologies for use threaten to render the life of information in the digital age as, to borrow a phrase from Hobbes, "nasty, brutish and short." Some information no doubt deserves such a fate, but numerous examples illustrate the danger of losing valuable cultural memories that may appear in digital form. Consider, for example, the now famous, but often misrepresented, case of the 1960 Census.

As it compiled the decennial census in the early sixties, the Census Bureau retained records for its own use in what it regarded as "permanent" storage. In 1976, the National Archives identified seven series of aggregated data from the 1960 Census files as having long-term historical value. A large portion of the selected records, however, resided on tapes that the Bureau could read only with a UNIVAC type II-A tape drive. By the mid-seventies, that particular tape drive was long obsolete, and the Census Bureau faced a significant engineering challenge in preserving the data from the UNIVAC type II-A tapes. By 1979, the Bureau had successfully copied onto industry-standard tapes nearly all the data judged then to have long-term value.

Though largely successful in the end, the data rescue effort was a signal event that helped move the Committee on the Records of Government six years later to proclaim that "the United States is in danger of losing its memory."

[http://www.rlg.org/ArchTF/tfadi.intro.htm#fragility:](http://www.rlg.org/ArchTF/tfadi.intro.htm#fragility)

Den forventede levetid for forskellige lagermedier er gennemgået bl.a. i Data Storage Technology Assessment 2000 (nml 00) Her er givet følgende overslag:

Tape: 20-50+++ år

CD-R: 50+++ år

CD-ROM: 5-100 år

I ovenstående giver antallet af plusser fabrikanters forventning til antal ekstra dekader udover der 50. Som med meget, er der store variationer i kvalitet og dermed til forventet livstid.

Det er også vigtigt at sikre optimale betingelser for arkivområdet. Afhængig af arkivmediet er forskellige anbefalede temperaturer og luftfugtighedsprocenter.

Der synes dog enighed om, at det største problem ikke er levetiden af lagermediet, men af læseudstyr. Det nævnes endvidere, at det er alm. Praksis at overspille det dataobjektet hvert femte år.

Dette får indflydelse på omkostningerne ved at bevare det digitale objekt. Skal man under alle omstændigheder overspille hvert femte år, uanset om mediet kan holde hundrede år, er det vigtigt, at overspilningsprocessen er så automatisk som mulig. Jeg anser dette for at være et argument for, at tape ofte anbefales i forbindelse med digital bevaring – på trods af de væsentlig bedre levetider for CD mediet.

8.4.2 Decentral opbevaring

Materialet kan efter aftale opbevares uden for de traditionelle institutioner. Sker dette, skal der aftales:

- Certificering
- Kvalitetskontrol
- Adgang for brugere

Denne metode kan komme på tale over for eksempelvis udgivere med en forsvarlig bevaringsstrategi. Da dette vil være dyrt, forventer jeg ikke, at mange vil udnytte denne mulighed. I USA er JSTOR et eksempel på denne bevaringsstrategi.

SA kalder denne strategi for Non custodial Archiving og argumenterer imod den – i alt fald hvis der ikke er kontrol med de decentrale opbevaringsinstitutioner.

8.5 Sikkerhedsforanstaltninger

Tager man en 200 år gammel bog, og er dens indbinding intakt, er der en ret stor sandsynlighed for, at den er den originale. I alle tilfælde vil et forsøg på forfalskning ved at skifte et par sider ud, være ekstremt dyrt og yderst kompliceret. Dette er ikke tilfældet med elektroniske materiale, og i flere kredse ytres ønske om eksempelvis versionskontrol. Et andet problem er, at man faktisk uden skyld kan komme til at ændre i signalet. Dette gør, at man i forbindelse med langtidsbevaring bør overveje, hvordan man vil sikre integriteten og autenticiteten af signalet, og hvordan man vil sikre, at evt. planlagte ændringer dokumenteres.

8.5.1 Autenticitet

I dag vil man fokusere på følgende måder:

- Digital signatur til sikring af, at bittene ikke er ændrede
- Digital vandmærkning – for at sikre materiale og ophav

- Digital tidsstempling

Sammen med versionskontrolsystemer.

8.5.2 Audit

Skal arkivet være troværdigt, er det vigtigt at evt. ændringer dokumenteres.

8.6 Opsummering af spørgsmål

Der skal etableres strategier og efterfølgende handlingsplaner for, hvordan vi sikrer langtidsbevaring, når først digitalt materiale er indsamlet. Det forventes, at strategien vil afhænge af såvel type som oprindelse af materialet. Herunder skal det overvejes, for hvilket materiale det er vigtigt at bevare den fulde funktionalitet, for hvilket look and feel er vigtig, og for hvilket det er indholdet, som er vigtigt.

Anbefalinger i forbindelse med lagermedium og overspilning bør fastlægges.

Er flere involveret i bevaringsforpligtigelserne, skal der fastlægges procedurer for kontrol, ligesom der skal fastlægges procedurer for, hvordan en situation som at tredje part nedlægger sin forretning, håndteres.

Der skal undersøges, hvordan man sikrer, at der løbende følges med i de tekniske muligheder.

9 Adgang/Tilgængeliggørelse

Det giver kun mening at indsamle materiale, som man har en rimelig forventning til, at man kan give adgang til i fremtiden. Det lyder som en tautologi, men er det ikke nødvendigvis.

At kunne bruge materiale engang i fremtiden har i den analoge verden krævet, at materialet er beskrevet på en måde, så man genfinde det – på samme måde, som bøger beskrives, så man kan genfinde det, som handler om noget bestemt, er skrevet af en bestemt osv. uden at skulle have samtlige bøger i hånden. Med digitalt materiale behøver det ikke at forholde sig sådan.

Digitalt materiale adskiller sig fra analogt ved selv at have al information søgbart. Vi skal ikke her gå ind på en længere diskussion af mulighederne for automatisk klassifikation, men blot konstatere, at digitalt materiale er fundamentalt forskelligt fra analogt materiale, og at alle oplysninger i det er tilgængelige.

Til gengæld, som allerede diskuteret i sidste kapitel, stiller visningen af det digitale materiale ofte store krav til udstyr – og det er langt fra sikkert, at vi kan give adgang såfremt det rette arbejde ikke er gjort i indsamlingsøjeblikket.

Det meste materiale på webben findes vha. søgemaskiner som eksempelvis Google. Det er også velkendt, at disse medfører meget støj, hvilket er et af hovedargumenterne for at udvikle emneportaler. At disse er nyttige kan ses af PADI, som inden for det område, som er behandlet i denne rapport, indeholder mange relevante henvisninger

Ovenstående diskussion afspejler et meget fundamentalt spørgsmål: Hvordan vil fremtiden finde informationerne, og hvordan vil de forvente at få dem præsenteret. I en vilkårlig tid har vi en tendens til at tro, at fremtidens muligheder ligner dem, vi har nu. Efter at være vant til at aflevere mine hulkort ved en disk og vente 24 timer på at få en stak papir tilbage virkede det som en revolution, da jeg i 1976 kom til Cambridge og fik adgang til en 300 baud TTY terminal, som stod i et nabokontor. Kunne man forestille sig mere? Ikke dengang. Men i dag giver det anledning til at trække på smilebåndet.

Hvordan vil vi i dag forvente at få adgang til de informationer? Via en 300 baud TTY terminal via et i vore øjne arkaiske styre- og filsystem? Nok ikke – vi vil forvente, at informationen kan søges med set med i vore øjne moderne metoder, og at vi kan få vist dem på en skærm og ikke via 9-nåls print på endeløse baner papir.

Det er derfor vigtigt at gøre sig klart, at selv om arkivobjekterne bevares på original form, skal adgangen følge med tiden – og at denne ind i mellem foretager kvantespring i sin adgangsmetafor.

9.1 *Oversigt over strategier i forskellige lande*

Strategierne i andre lande spreder sig fra den australske, hvor der kun indsamles fra sites, hvor man via en frivillig aftale har sikret sig, at materialet må offentliggøres, til den svenske model, hvor der indsamledes til dødarkiv. En mellemvej er ved at blive testet i USA, hvor man i de første høstninger gjorde webmasteren opmærksom på, at man høstede, og havde han/hun indsigelser,

skulle de henvende sig, og nu vil prøve den omvendte strategi, at de skal henvende sig, før det offentliggøres. I USA havde man den politik, at alt blev høstet, men information gemt bag robot.txt blev ikke offentliggjort.

9.2 Tekniske-juridiske overvejelser i forbindelse med adgang

I Danmark har vi traditionelt bundet adgang sammen med fysisk tilstedeværelse. Dette har god mening, når der er tale om analoge objekter, hvor der jo gives adgang til selve det museale objekt. Dette giver imidlertid ikke mening for digitale objekter, hvor brugeren sidder med en kopi. I forbindelse med pligtaflevering.dk kan brugeren fra sin arbejdsplads se, om et givet dokument ligger i basen, men vil han/hun se det, skal vedkomne begive sig til en af to specielle maskiner, som befinder sig på hhv. KB og SB. I dagens samfund, hvor ting kan beskyttes på mange måder, bør der løses op for denne binding – ligesom der nu i forbindelse med det DEF-finansierede LDAP-projekt løses op for, at man skal sidde på et bibliotek eller på et universitet for at få adgang til elektroniske tidsskrifter – man skal blot være registreret på et bibliotek – og ens privilegier afgør, hvilke tidsskrifter man har adgang til.

9.3 Opsummering/Spørgsmål

Et springende spørgsmål er altid, om man skal indsamle materiale, som ikke må stilles tilgængeligt over for publikum. Man kan indsamle og håbe på, at loven ændres. Anser man det for overvejende sandsynligt, at lovgivningen ændres, bør der indsamles.

Hvordan skal materiale stilles tilgængeligt? Som det var på dannelsesstidspunktet, eller som man kan på det tidspunkt, hvor en forsker ønsker at komme til materialet?

Traditionelt er der en forbindelse mellem fysisk placering og adgang – denne er meningsløs for digitalt materiale. Det bør overvejes, hvilke alternative muligheder der er for at sikre, at kun den gruppe/de, som retmæssigt har adgang, får adgang f.eks. fra deres arbejdsplads.

10 Kompetenceudvikling

Det er nævnt flere gange i denne rapport, at håndteringen af digitale arkiver kræver flere forskellige typer ekspertise. Skal dette fungere i praksis, er det nødvendigt, at disse har en fælles forståelse af problemer og et fælles sprog. Dette har betydning for såvel basisuddannelsen og for etablering af relevante efteruddannelsesforløb.

Rektor for biblioteksskolen nævnte en kulturarvsuddannelse, hvor studerende fra biblioteksskole, fra IT-højskole og universitet kunne tage en overbygningsuddannelse, som gives af de tre institutioner.

Det er imidlertid også vigtigt, at der foretages et generelt kompetenceløft blandt de potentielle aktører, som i dag sidder som bibliotekarere, it-folk, konservatorer, arkivarer o.lign.

Som led i gruppens arbejdet bør det undersøges, hvordan den nødvendige ekspertise kan opbygges.

11 Økonomi/Forretningsmodel

Som ved enhver aktivitet er der en økonomi forbundet med såvel at etablere en indsamling såvel som løbende at vedligeholde det digitale arkiv. Der findes mig bekendt ikke nogen anerkendt algoritme for beregning af omkostninger.

Som vejledning kan ses på økonomien i andre lande. Mig bekendt der ingen tal for hele cyklen med såvel materialeopbygning som bevaring.

Eksempler på bevaring:

Holland havde et budget på knap 15 mio. kr. over en fireårig periode til at etablere et asset management system. Herefter er der bevilget ca. 8 mio. kr. pr. år til driften. Dette budget dækker over etablering af arkivinformationspakker samt lager. Budgettet indeholder ikke udvikling af emuleringssoftware eller hardware emulatorer.

Congressen i USA har bevilget i alt 100 mio. \$ til igangsætning af program omkring digital bevaring. Pengene er fasedelt. Fase 1 var på 5 mio. til LC og skulle bruges til dels at formulere strategier og dels komme med en deltaljeret plan for, hvordan digital bevaring skal foretages. Sker dette på tilfredsstillende vis, vil LC modtage 20 mio. \$ til igangsætning af disse initiativer. Samtidig er 75 mio. \$ reserveret til en \$ til \$ matchning af non-government initiativer.

De to modeller afspejler forskellene i tilgang til løsning af opgaven omkring digital bevaring. I Holland er fokus på kulturarvsinstitutionerne, hvorimod man i USA ser man opgaven løst gennem en række aktører – under ledelse af LC. I en samtale med deres leder af bevaring, Mark Roosy, blev det klart, at en sådan strategi kræver en form for certificeringssystem, som skal give en garanti for, at institutionerne kan leve op til deres opgave.

I Australien har Arkivet i Victoria netop fået bevilget xx mii. \$ svarende til xxx DKr.

Eksempler på budget til webarkivering – dette er kun til indsamling – ikke langtidsbevaring!:

Pandora: 3-5 mio. Nationalbiblioteket har ikke fået øget deres bevilling, dvs. beløbet er fremkommet via omlægning af opgaver

Kulturarw3: 3 mio.

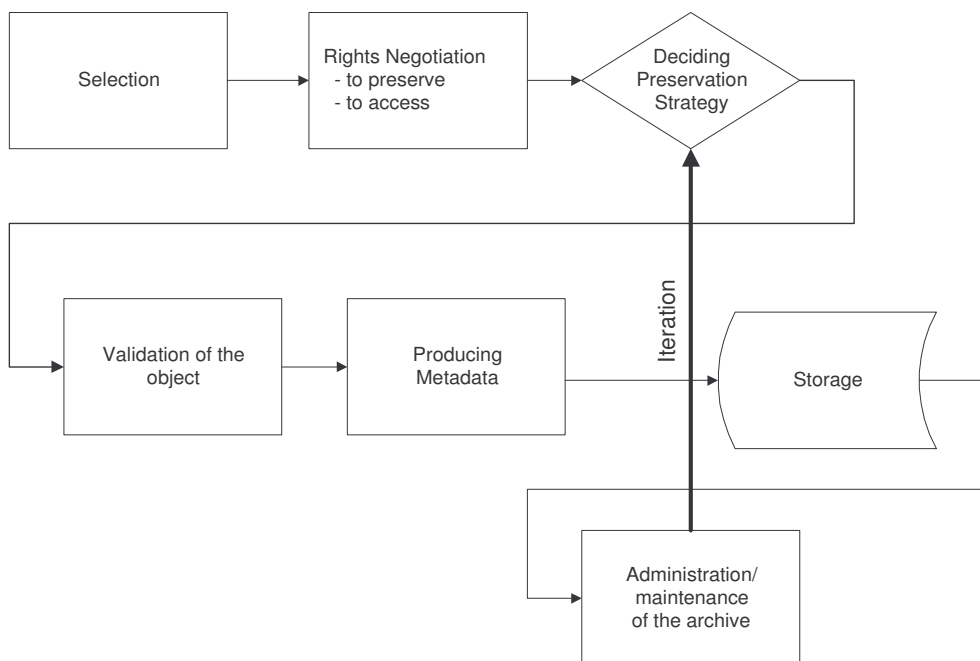
Spørgsmål: Hvilken form for økonomi vil være realistisk i forbindelse med denne opgave? Sættelse såvel som størrelse. Ultimativt formålet med gruppens arbejde.

Stort set alle artikler, jeg har set, starter med at sige, at prisen for bevaring er en ukendt størrelse. I "Comparison of Methods and Cost of Digital Preservation" (H 98) opstilles en matrix, som kan danne baggrund for en undersøgelse i dag.

Det skal siges, at der ikke er enighed om de konkrete konklusioner i rapporten. Modellen kan dog godt overføres til danske forhold.

Cedar projektet (<http://www.leeds.ac.uk/cedars/colman/costElementsOfDP.doc>) har også analyseret økonomiske modeller uden dog at sætte tal på. Deres model er baseret på OAIS modellen, som er brugt i denne rapport. Efter at have

gennemgået forskellige arkivmodeller sluttet af med følgende generiske model for elementer til omkostninger:



Den iterative loop er indført for at indikere, at bevarelsesstrategi vil ændre sig over tid.

I det videre arbejde skal der sættes tal på de dele af processen, hvor det giver mening. I arbejdet bør overvejes, hvilke aktører som med fordel kan inddrages i arbejdet. Forskellige organisatoriske modeller skal overvejes.

Det, som koster, er de menneskelige ressourcer. Det skal nøje overvejes, hvilke processer som kan automatiseres, og såfremt en sådan automatisering resulterer i en dårligere service, hvor alvorlig denne serviceforringelse er.

I opstilling af forskellige økonomiske scenarier bør der opstilles prioriteret liste over materialekategorier, som skal bevares. Elementer i denne er bl.a.:

- Født digital/digitaliseret
- Repræsentativ brug – kvalitetsvurderet
- Materialekategorier

12 Det videre arbejde

Undervejs i udredningen er stillet en række spørgsmål:

Afsnit 3: - Roller og rammer

Det bør overvejes, om man gennem samarbejde og koordination kan effektivisere en fremtidig indsamling og bevaring. Eksempelvis kunne man forestille sig, at en institution får ansvaret for e-mails – uanset hvilken type organisation eller privatperson som donerer/ønsker at donere sit e-mail arkiv – og får til ansvar at sikre, at de vigtigste af disse fremover samles.

Det er her vigtigt at afgøre, om konsortier og deling af arbejde styres af teknikken bag eller dem, som der træffes aftaler med?

Kunne man forestille sig, at en organisation var ansvarlig for inddrivelse af materiale, en anden for den videre bevaring?

Et spørgsmål, som blev adresseret i UK var: Hvem er aktørerne – og hvad forventer de, og hvad har de planlagt mht. langtidsbevaring?

Hvem bevarer hvad? Eksempel DR – kan man få en browse-kopi? Eller hvilken kvalitet skal afleveres – og hvad må den bruges til?

Der bør laves en undersøgelse af, hvem som i Danmark har bevaring som en primær eller sekundær rolle – og hvad deres rolle er.

Afsnit 4: Strategi og handlingsinitiativer

Der er allerede lavet mange undersøgelser om forskellige aspekter af digital bevaring. Det meste af det tekniske og strategiske arbejde kan umiddelbart overføres til den danske situation. Det skal overvejes, om man i en implementeringsfase af et arkiv til publiceret materiale skal forsøge at etablere et strategisk samarbejde med andre. Det bør overvejes, i hvor høj grad andres løsninger kan overføres direkte. Konkret bør man se på samarbejde med Holland og England samt USA og Australien. Det bør overvejes, i hvor høj grad man vil tilstræbe brugen af open source software.

Langtidsbevaring er klart international, og Danmark skal sikre en repræsentation i relevante projekter og initiativer. Det skal overvejes, hvordan Danmark ønsker at udnytte 6. rammeprogram.

Det er også vigtig at gøre sig klart, hvad man ønsker skal ligge nationalt, og hvad der kan løses internationalt.

Skal snapshot nødvendigvis ske i Danmark – kunne vi ikke lave et fælles nordisk eller europæisk arkiv? Hvad med den videnskabelige produktion – den ligger nu typisk i udenlandske databaser – er det ønskværdigt?

Som i andre lande bør det overvejes, om Danmark skal lave en gruppe, som kan tage initiativer på tværs af institutionsgrænser. Det skal overvejes, hvilke beføjelser og økonomi man ønsker at give en sådan gruppe.

Afsnit 5 – Metadata

Metadata er essentielt for langtidsbevaring for såvel det at genfinde materiale som for at sikre den korrekte håndtering af materialet som led i bevaringsstrategien. Generering af relevante data er imidlertid også omkostningstung, hvorfor det er vigtigt at analysere denne proces igennem for at identificere muligheder for automatisering af processen.

I hvor høj grad kan metadata genereres automatisk eller ud fra eksisterende informationer?

Hvilke data er minimumskravet?

Eksisterende beskrivelse for metadata bør gennemgås, og det bør overvejes, hvilke informationer er essentielle, hvilke som kan generes automatisk eller ved kilden, og hvordan kvalitetskontrol af data gennemføres.

Afsnit 6 – Indsamlingsstrategier

Der skal laves en strategi for, hvad skal indsamles og hvordan denne indsamling effektueres.

Afvejningen mellem aktiv indsamling og passiv indsamling.

Er der visse typer materiale, som hører til den selektive gruppe, f.eks. online aviser og magasiner?

Har vi begivenheder, som bør følges regelmæssigt, f.eks. valg (såvel folketing som kommune og amtsvalg), kongelige begivenheder ...?

Er der mere, som fortjener regelmæssig opfølgning, eksempelvis integrationsdebatten?

Skal der laves tværsnitshøstninger – og i givet fald hvilke domæner og hvor tit?

I hvor høj grad kan man få værkskaberne til at generere den beskrivende information?

Man kan godt forestille sig, at noget materiale vil egne sig til kortere opbevaring og andet til langtidsbevaring. Der bør formuleres en strategi for, hvad man ønsker at bevare – og hvor længe man forventer at kunne komme til materialet.

Vi har identificeret to basalt forskellige materialetyper: født digital og digitaliseret. Ofte vil den digitaliserede kopi være fremstillet af hensyn til formidling og ikke af hensyn til bevaring. Der ligger dog en stor investering i denne digitaliseringsproces. Såfremt der ikke er midler nok til bevaring af begge, skal formuleres en prioritering mellem unika og dyre kopier.

Afsnit 8 – Lagring

Der skal etableres strategier og efterfølgende handlingsplaner for, hvordan vi sikrer langtidsbevaring, når først digitalt materiale er indsamlet. Det forventes, at strategien vil afhænge af såvel type som

oprindelse af materialet. Herunder skal det overvejes, for hvilket materiale det er vigtigt at bevare den fulde funktionalitet, for hvilket look and feel er vigtig, og for hvilket det en indholdet, som er vigtigt.

Anbefalinger i forbindelse med lagermedium og overspilning bør fastlægges.

Er flere involveret i bevaringsforpligtigelserne, skal der fastlægges procedurer for kontrol ligesom der skal fastlægges procedurer for, hvordan en situation som at tredjepart nedlægger sin forretning, håndteres.

Der skal undersøges, hvordan man sikrer, at der løbende følges med i de tekniske muligheder.

Afsnit 9 – Adgang

Et springende spørgsmål er altid, om man skal indsamle materiale, som ikke må stilles tilgængeligt over for publikum. Man kan indsamle og håbe på, at loven ændres. Anser man det for overvejende sandsynligt, at lovgivningen ændres, bør der indsamles.

Hvordan skal materiale stilles tilgængeligt? Som det var på dannelsesstidspunktet, eller som man kan på det tidspunkt, hvor en forsker ønsker at komme til materialet?

Traditionelt er der en forbindelse mellem fysisk placering og adgang – denne er meningsløs for digitalt materiale. Det bør overvejes, hvilke alternative muligheder der er for at sikre, at kun den gruppe/de, som retmæssigt har adgang, får adgang f.eks. fra deres arbejdsplads.

Afsnit 10 – Kompetence

Som led i gruppens arbejdet bør det undersøges, hvordan den nødvendige ekspertise kan opbygges.

Afsnit 11 – Økonomi

I det videre arbejde skal der sættes tal på de dele af processen, hvor det giver mening. I arbejdet bør overvejes, hvilke aktører, som med fordel kan inddrages i arbejdet. Forskellige organisatoriske modeller skal overvejes.

Det, som koster, er de menneskelige ressourcer. Det skal nøje overvejes, hvilke processer som kan automatiseres, og såfremt en sådan automatisering resulterer i en dårligere service, hvor alvorlig denne serviceforringelse er.

I opstilling af forskellige økonomiske scenarier bør der opstilles prioriteret liste over materialekategorier, som skal bevares.

Elementer i denne er bl.a.:

- Født digital/digitaliseret
- Repræsentativ brug – kvalitetsvurderet

- Materialekategorier

Spørgsmålene ovenfor kan samles i grupper:

1: Som gennemført i såvel USA, Australien og England, bør der etableret en arbejdsgruppe, som undersøger samspillet mellem de potentielle aktører. Gruppen skal dels afdække muligheder for aflevering som en integreret del af fremstillingsprocessen, dels undersøge aktørers holdning til langtidsbevaring (og finansiering af denne) samt undersøge aktørers villighed til selv at påtage sig bevaringsrollen (under opsyn af den museale institution på det relevante område). Forskellige organisatoriske modeller bør opstilles og undersøges. Endelig bør det undersøges, i hvor høj grad biblioteker fortsat skal acceptere alle formater af materiale til langtidsbevaring

2: Der skal laves en strategi for, hvad der indsamles med henblik på bevaring – og metoder bør vælges i forhold til materialetype. Strategien bør operere med forskellige faser for at sikre, at man ikke venter på at kunne løse det mest besværlige problem.

3: Økonomien for forskellige komponenter omkring indsamling og bevaring bør gennemgås, herunder:

- Niveau for metadata
- Mulighed for at indtænke digital bevaring i skabelsesproceduren for materiale
- Krav i forbindelse med adgang skal undersøges, og man skal overveje, i hvor høj grad denne kan tilfredsstilles via et maskingeneret interface. Det skal undersøges, i hvor høj grad de resterende bevaringsmetadata kan genereres (semi)automatisk.
- Omkostning og konsekvenser af brug af forskellige lagermedier, cd-rom, tape osv. bør klarlægges.

4: Der skal opstilles kriterier for en prioritering mellem materialekategorier, som skal bevares. Elementer i denne er bl.a.:

- Født digital/digitaliseret
- Repræsentativ brug – kvalitetsvurderet
- Materialekategorier

5: Der bør allerede nu etableres en organisatorisk gruppe med deltagelse fra de museale institutioner (Statens Arkiver + nationalbiblioteker) samt andre interessenter med de mål at få

- koordineret initiativer og
- lavet et fælles vidensløft på området

13 Litteraturliste

- A 01 Andreas Aschenbrenner Speciale dec. 2001 *Long term Preservation of Digital material – building an archive to preserve digital cultural heritage from the Internet*,
<http://www.ifs.tuwien.ac.at/~aola/publications.html>
- B 97 Bennet: *A Framework of data types and formats, and issues affecting the long term preservation of digital material*
- BL-41 *Policy and strategy issues for the UK: report of a meeting on the CPA/RLG report, December 1996*. British Library Research and Innovation Report 41. London: The British Library.
- DLM 99 Proceedings of the DLM forum 1999:
http://europa.eu.int/ISPO/dlm/dlm99/Proceed99-down_en.htm
- DM 00 Jan Danielsen og Ulla Mortensen, *Strategier for elektroniske arkivaliers overlevelse*, Arkiv 5, 78
- Dpconline *The Handbook*. <http://www.dpconline.org>
- EDCL 01 http://www.bnf.fr/pages/infopro/dli_ECDL2001.htm
- F 99 Mary Feeney (editor) *A synthesis of JISC/NPO studies on the preservation of digital materials*.
<http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/intro.html>
- HC 99 Hodge, G and Carroll, B. C., 1999; Information International Associates Inc: *Digital Electronic Archiving: the State of the Art and State of Practice*:
http://www.icsti.org/99ga/digarch99_TOCP.pdf
Resume og hovedrapport findes på:
http://www.icsti.org/99ga/digarch99_ExecP.pdf and
http://www.icsti.org/99ga/digarch99_MainP.pdf,
- HSC 02 Hilera Berthon, Susan Thomas and Colin Webb: *Safekeeping: a cooperative approach to building a digital preservation resource*
D-lib magazine Jan. 2002
<http://www.dlib.org/dlib/januar02/berthon/01berthon.html>
- HSJB 97 David Haynes, D. Streatfield, T. Jowett and M.Blake 1997: *Responsibilities for digital archiving and long term access to digital data*, British Library Research and Innovation Report 67.
<http://www.ukoln.ac.uk/services/papers/jisc-npo67/digital-preservation.pdf>
- JISC 01a Content management systems
<http://www.jisc.ac.uk/techwatch/reports/>
- JISC 01b Jave Portals: <http://www.jisc.ac.uk/techwatch/reports/>
- LC 02 Building a National Strategy for Preservation: Issues in Digital Media Archiving

- <http://www.clir.org/pubs/abstract/pub106abst.html>
- LC21 Committee on an Information Technology Strategy for the Library of Congress, Computer Science and Telecommunications Board, National Research Council July 2000 *LC21: A Digital Strategy for the Library of Congress*, National Academy Press
<Http://books.nap.edu/books/0309071445/html>
- Lorie 01 <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>
- MC 02 The archive as an ecological system
- MP 01 Margereth Philips *Ensuring Long-Term Access to Online Publications* <http://www.press.umich.edu/jep/04-04/phillips.html>
- Nml 00 K. Sadashige, *Data Storage Technology Assessment 2000*. Kan findes via www.nta.org
- OAIS 01] CCSDS 650.0-R-1, ISO/DIS 14721.2, July 2001 *Reference Model for an Open Archival Information System Issue 2*
http://www.ccsds.org/red_books.html
- OAIS 02 CCSDS – 651.0-W-1 *Producer-Archive Interface Methodology Abstract Standard* http://www.ccsds.org/white_books.html
- PADI *Preserving Access to Digital Information* Portal til relevant materiale om digital bevaring
<http://www.nla.gov.au/padi/topics/3.html>
- PRO 01 Farley, J. (original, 1999). An Introduction to Archival Materials; new media (PRO Preservation Guide series). :published in the series of **Introduction to Archival Materials** Crown ...
www.users.globalnet.co.uk/~farley/jon/conservation/deadmedia/newmedia.pdf
- R 01 Rauber 2001 *Part of Our Culture is Born Digital – On Efforts to Preserve it for Future Generations*: Internet-Zeitschrift für Kulturwissenschaften Vol 10, Juni 01:
<http://www.inst.at/trans/10Nr/rauber10.htm>
- R 99 Rothenberg *Avoiding the technological Quicksand* CLIR report
<http://www.clir.org/pubs/reports/rothenberg/>
- RLG 02 Neil Beagrie, Marianne Doerr, Margaret Hedstrom, Anne Kenney, Catherine Lupovici, Kelly Russel, Colin Webb, Robin Dale, Meg Bellinger, maj 2002. *Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources*
<http://www.rlg.org/longterm/repositories.pdf>
- WebDK 01 <http://www.deflink.dk/arkiv/dokumenter2.asp?id=695>
- WG 96 D Waters and J Garrett 1996. *Preserving digital Information: Final report and recommendations*. <http://www.rlg.org/ArchTF>
- LV 00 P.Lyman og H.R.Varian *How much Information?*
<http://www.sims.berkeley.edu/how-much-info/>