

Sikring af adgang til digitalt materiale og OAIS referencemodel

Price, Adrian; Christensen-Dalsgaard, Birte; Sandfær, Mogens; Henriksen, Birgit; Lau, Birgitte; Kværndrup, Hanne Marie

Published in:
DF Revy

Publication date:
2004

Document Version
Også kaldet Forlagets PDF

Citation for published version (APA):

Price, A., Christensen-Dalsgaard, B., Sandfær, M., Henriksen, B., Lau, B., & Kværndrup, H. M. (2004). Sikring af adgang til digitalt materiale og OAIS referencemodel. *DF Revy*, 27(2), 16-23. <http://www.dfrey.dk/>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact rucforsk@ruc.dk providing details, and we will remove access to the work immediately and investigate your claim.

Sikring af adgang til digitalt materiale og OAIS referencemodel

Af Adrian Price, Birte Christensen-Dalsgaard, Mogens Sandfær, Birgit Henriksen, Birgitte Lau og Hanne Marie Kværndrup
Fotos Birte Christensen-Dalsgaard

Adrian Price

Bibliotekskonsulent ved Danmarks
Veterinær- og Jordbrugsbibliotek
ap@kvl.dk

Birgit Henriksen

Digitaliseringschef
ved Det Kongelige Bibliotek
bnh@kb.dk

Birte Christensen-Dalsgaard

IT-udviklingschef ved Statsbiblioteket
bcd@statsbiblioteket.dk

Birgitte Lau

Chefkonsulent
ved Roskilde Universitetsbibliotek
blau@ruc.dk

Mogens Sandfær

Udviklings- og IT-chef
ved Danmarks Tekniske Videncenter
ms@dvt.dk

Hanne Marie Kværndrup

Bibliotekskonsulent
ved Biblioteksstyrelsen
hmk@bs.dk

Lagtingsbevaring af og adgang til digitalt materiale er for alvor på vej ind på bibliotekernes dagsordener. Bevaring af digitale objekter stiller helt anderledes krav end bevaring af papirmateriale. Der er derfor en stor opgave foran bibliotekerne, før de kan deklare sig selv i stand til at garantere fuld adgang for brugerne. Der er mange spændende aktiviteter i gang, hvor der arbejdes på den metodemæssige, tekniske og organisatoriske infrastruktur, som er nødvendig.

Indledning

I oktober gik turen for seks DEF-folk (Adrian Price, DVJB, Birgit Henriksen, KB, Birte Christensen-Dalsgaard, SB, Birgitte Lau, RUB, Mogens Sandfær, DTV, Hanne Marie Kværndrup DEF/BS) til den amerikanske østkyst for at studere Digital Preservation og E-publishing.

Studieturen udsprang af FC-kollegiets rapport om digital arkivering og institutional repositories samt af DEF's programområde for E-publicering, hvor bl.a. institutional repositories er et indsatsområde. Gennem de sidste år har arbejdet i DEF samt på Det Kongelige Bibliotek (KB) og Statsbiblioteket (SB) vist, at opgaverne med webarkivering og digital bevaring er meget væsentlige emner at få sat på dagsordenen i den danske bibliotekssektor.

Det er ikke længe siden, at vi indførte de første digitale bibliotekssystemer i

Danmark. Det er slet ikke længe siden, tekstbehandling begyndte at vinde indpas for alvor i informationsproduktionen, og det er kun "i forgårs", at Internettet og webteknologien for alvor tog fart. Allerede nu er vi bagefter med tekniske standarder, softwareudvikling, workflow-systemer, organisering og planlægning af, hvorledes vi skal håndtere de digitale objekter og informationsprodukter, som hver eneste dag "fødes".

Umiddelbart kan det synes som en simpel opgave at bevare digitalt materiale! Vi har jo en forholdsvis ny pligtafleveringslov, der også omfatter digitale værker, og KB og SB har som national opgave at bevare disse produkter! Ja vel – men hvad med de dynamiske værker, det vil sige websites, e-tidskrifter med videre, og al øvrig vidensproduktion fra universiteter, uddannelsesinstitutioner – alt det der ikke kan kaldes værker – det kan vel bare gemmes i digitale depoter eller biblioteker? Men så nemt er det ikke, vi har travlt, for hver dag forsvinder information, som ikke genskabes.

Studieturen gik først til Boston, hvor vi besøgte MIT og specielt så nærmere på systemet DSpace [DSPACE], som er et software for institutional repositories. Næste stop var Ithaca, NY, nærmere betegnet Cornell University, hvor e-publiceringsprojektet Euklid [EUKLID] og softwareudvikling af Fedora [FEDORA] blev studeret. Cornell University var vært

ved en uges workshop om Digital Preservation, og vi blev grundigt indført i OAIS modellen samt i mange højst relevante emner omkring emnet digital bevaring. DEF har bidraget økonomisk til turen, og hensigten er selvfølgelig, at den viden, vi har bragt med os tilbage, skal indgå i det videre arbejde i DEF og derved komme bibliotekerne til gode. Vi vil efterfølgende forsøge kort at præsentere resultatet af vores studietur.

Open Archival Information System (OAIS) referencemodel og Trusted Digital Repositories (TDR)

Bibliotekerne har i de sidste cirka ti år koncentreret sig om adgang til digitale materialer og ikke så meget om bevaring af dette materiale. I "papirets tid" var bevaring for fremtidige generationer ellers opfattet som en af bibliotekernes kerneopgaver. Biblioteker er nu mange steder begyndt også at sætte fokus på bevaring af digitalt materiale.

Årsagen til denne tøven med at kaste sig over bevaringsproblematikken vedrørende digitale materialer er også begrundet i, at de skal behandles på en anden måde end det materiale, som bibliotekerne hidtil har været vant til at behandle. Der skulle først udvikles og indføres metoder til, langtidbevaring af dette "e-materiale".

Disse metoder knytter sig til de organisatoriske forhold, de informationsmæssige og de tekniske: Institutioner skal være klar

til kontinuerligt at udføre bevaringsopgaverne. E-materialer skal beskrives med metadata, en forudsætning for en sikker bevaring, og de nødvendige tekniske processer, som knytter sig til det enkelte e-materiale, skal være på plads. Disse metoder er nu ved at være så udviklede, at de kan give bibliotekerne et solidt grundlag for, hvordan opgaven skal takles.

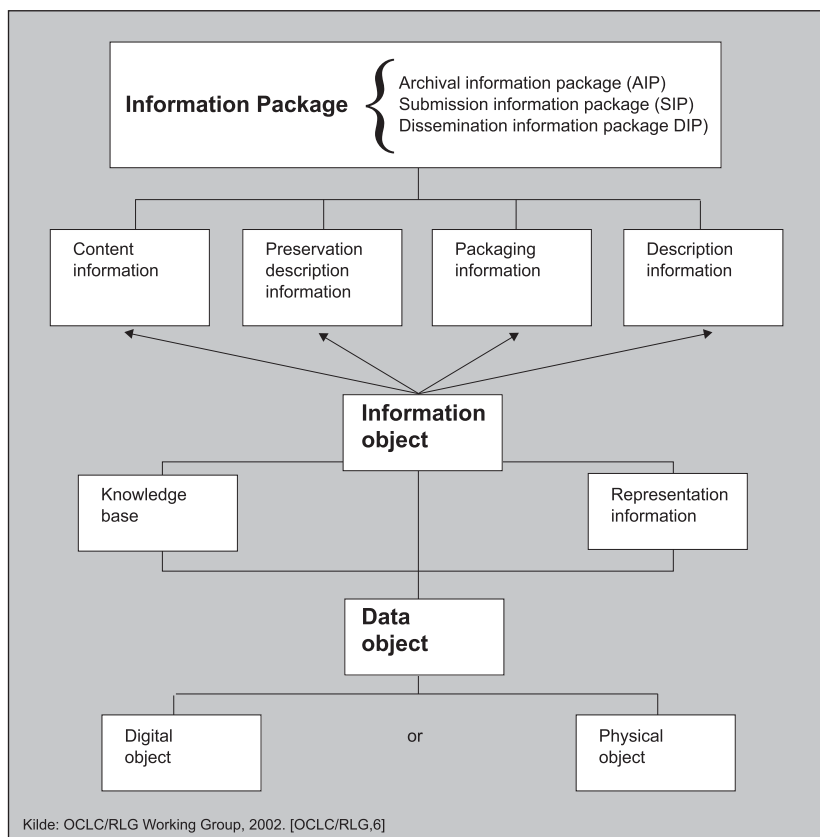
OAIS referencemodel

I januar 2002 udkom *Reference Model for an Open Archival Information System* (OAIS), en anbefaling udarbejdet af Consultative Committee for Space Data Systems (CCSDS) [OAIS].

OAIS referencemodel fastlægger det organisatoriske og/eller systemmæssige grundlag, som er nødvendigt for langtidsbevaring af "informationsbærende objekter". En "referencemodel" bruges til at klarlægge terminologi og standarder og til at fastlægge et hensigtsmæssigt work- og informationsflow inden for et givent område.

Grundlæggende tager OAIS-modellen ikke stilling til, om objekterne er digitale eller fysiske, men når modellen udfoldes, er det de digitale materialer, man koncentrerer sig om. Modellen er udviklet til at blive brugt i design af såvel specifikke systemer til håndtering og bevaring af digitale objekter som organisationer med arkivforpligtelser. Modellen i sig selv forudsætter ingen specifik metodik, IT platform, sprog, grænseflade, teknologi, database eller andet element, som måtte indgå i en bestemt implementering eller anvendelse af modellen: "The OAIS Reference Model is designed as a conceptual framework in which to discuss and compare archives....it attempts to address all the major activities of an information preserving archive in order to define a consistent and useful set of terms and concepts." [OAIS, 1-3]. Modellen kan anvendes som ramme for udvikling af systemer (arkiver, software), som vil sikre langtidsbevaring af og -adgang til digitale materialer.

Fig. 1 viser de forskellige elementer, som indgår i OAIS informationsmodel. Det er ikke muligt her at gennemgå modellen i detaljer, kun en summarisk gennemgang.



Figur 1: OAIS informationsmodel

Modellen er bygget op omkring et "dataobjekt" (som kan være fysisk eller digitalt), hvor dataobjektet fortolkes ved at kombinere en nærmere defineret brugergruppes nødvendige viden ("knowledge base" i modellen) med information, som er nødvendig for en korrekt gengivelse, forståelse og fortolkning af et dataobjekts indhold. [OCLC/RLG, 6-7] Et "informationsobjekt" er et dataobjekt kombineret med objektets repræsentationsinformation, som er den information, som er nødvendig for at gengive objektet og gøre det forståeligt. Der er fire typer informationsobjekter: indholdsinformation (den information som arkivet skal bevare), bevaringsinformation (information om objektet som skal bruges til bevaring), indkapslingsinformation (packaging information, information om objektet og dets metadata) og deskriptiv information (som bruges til genfindning af objektet i arkivet). I modellen er disse fire typer henholdsvis "content information", "preservation description information", "packaging information" og "descriptive information".

Disse fire informationsobjekter indgår i en "informationspakke", som er en af tre typer: en "submission information package" (SIP), en "archive information package" (AIP) og en "dissemination information package" (DIP). Hvordan disse "pakker" behandles af arkivet, gennemgås i næste afsnit.

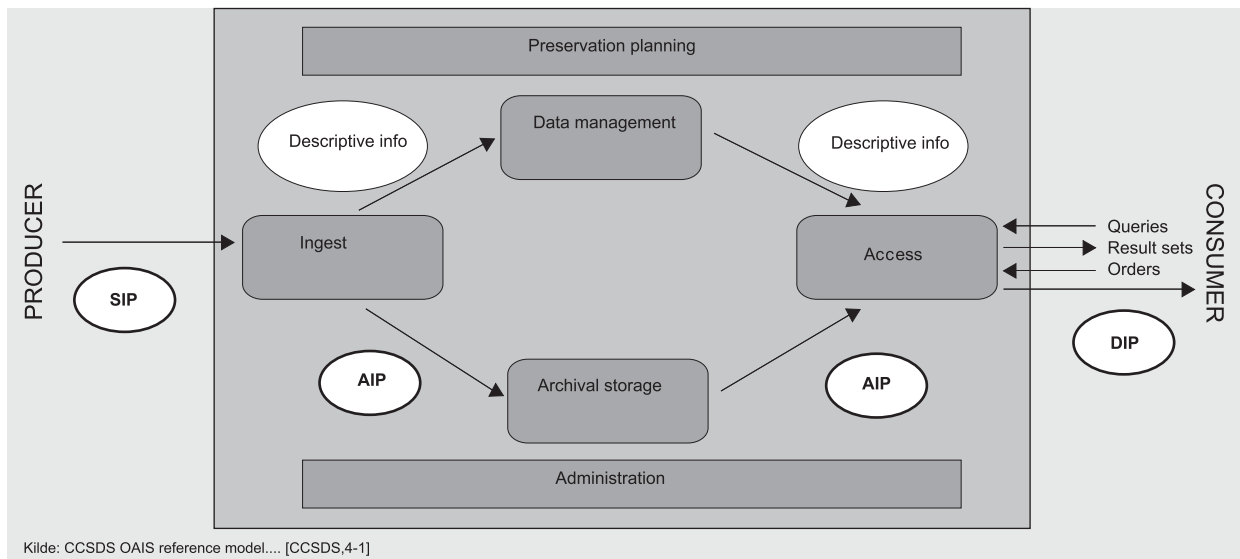
OAIS funktionel model

OAIS referencemodellen fastlægger også et arkivs funktionelle komponenter, som skal håndtere information i et "OAIS arkiv".

Fig. 2 (se næste side) er på et højt niveau, og kun et udvalg af informationsstrømmen er vist. Figuren viser OAIS-modellens seks funktionelle enheder (ingest, data management, archival storage, preservation planning, access og administration) og informationspakkernes (SIP, AIP, DIP) placering i informationsstrømmen. Deskriptiv information hentes hovedsageligt fra pakkerne og sørger for genfindning af objekterne.

Fra venstre mod højre viser figuren, hvordan en informationspakke (SIP-pakke) indlemmes i arkivet, deskriptive metadata hentes fra pakken til databasen, så objektet kan genfindes af brugerne ved søgning (DIP-pakken), og en "AIP pakke" dannes til objektets langtidsbevaring. Hele operation overvåges af administrationsenheden.

Informationsstrømmen i hver af disse funktionelle enheder kan brydes yderligere ned. [CCSDS, 4-5 – 4-16]



Figur 2: OAIS funktionelle enheder

TDR

Hvis OAIS-modellen fastlægger den informations- og funktionsmæssige infrastruktur, som hører til digital bevaring i et arkiv, så er der behov for fastlæggelse af arkivets politiske og organisatoriske mål, som vil sikre langtidsbevaring og adgang til e-materialer. I forlængelse af en rapport publiceret i 1996 *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* [CPA/RLG], som beskrev behovet for langtidsbevaring af digital materiale samt arkivinstitutternes opgaver og forpligtelser, blev der nedsat en Research Library Group (RLG) og OCLC arbejdsgruppe, som fik til opgave at beskrive de funktioner og ansvarsområder som digitale arkiver ("repositories"), skulle være i besiddelse af [RLG/OCLC].

Arkiver, som skal sikre langtidsbevaring af digitalt materiale, skal stræbe efter at få betegnelsen "trusted digital repository (TDR)", det vil sige arkiver, som er i besiddelse af den tilstrækkelige og nødvendige infrastruktur og kompetence til at sikre adgang til digital information i fremtiden. Der indgår i arbejdsgruppens rapport en hensigtserklæring om certificering af institutioner som "trusted digital repository" [RLG/OCLC, 33-35]. Institutioner skal anvende metoder og procedurer, som overholder kvantitative og kvalitative mål til sikring af adgang til digital information, og disse metoder og procedurer skal dokumenteres.

TDR er ligesom OAIS-modellen defineret i forhold til det, som betegnes en "designated community", det vil sige en gruppe

brugere, som kan forventes at skulle kunne anvende og forstå det materiale, arkivet har ansvar for. Det er denne definerede brugergruppe, som arkivet skal etablere adgangs- og brugsbetingelser for. Hertil kommer, at brugerne også skal kunne stole på integriteten af det materiale, arkivet giver adgang til.

Som fremlagt i RLG-OCLC rapporten, er det hensigten, at en TDR's arbejdsprocesser og procedurer skal være "OAIS compliant", det vil sige være bygget op med en informations- og funktionsmæssig infrastruktur, som er i overensstemmelse med OAIS-modellen (og som nu er en ISO standard).

Bevaringsmetadata

Som det fremgår af fig. 1, er forståelse og anvendelse af et "informationsobjekt" et spørgsmål om en brugers videnbase (den viden, brugeren er i besiddelse af, og som skal anvendes til at bruge og forstå et bestemt informationsobjekt) og et spørgsmål om et sæt data, som er blevet tilknyttet objektet. Dette sæt data skal bruges til både at sikre forståelse og anvendelse hos brugeren og til at sikre, at objektet forbliver forståeligt og anvendeligt fremover.

Metadata, som er blevet tilknyttet et informationsobjekt, kan opdeles i forskellige typer afhængig af, hvad det skal bruges til, for eksempel genfindingsmetadata (metadata for "resource discovery" også kaldet deskriptive metadata) og administrative metadata. Metadata, beregnet til at sikre adgang over tid, kaldes bevaringsmetadata. (Der er endnu ikke konsensus for opdeling af metadata i typer, og bevarings-

metadata kan opfattes som en særlig form for administrative metadata eller som et selvstændigt sæt metadata).

Bevaringsmetadata bruges til for eksempel at fastlægge, hvilket filformat et dataobjekt er i, og hvilken software, hardware og operativsystem der er nødvendigt for at anvende objektet. Er et objekt et sammensat objekt? Hvis ja, hvad består den enkelte del af (tekst, billeder, lyd...)? Hører der dokumentation til objektet? Og så videre.

Der er blevet udarbejdet en række metadataformater for digitale objekter [CEDARS] [JISC] [NLNZ] [AU] [OCLC/RLG], og nogle af disse formater har valgt at følge OAIS modellen mere eksplicit end andre, for eksempel CEDARS og OCLC/RLG Working Group's forslag. Ligesom på andre områder er det både kritisk og nødvendigt at fastlægge et passende niveau for tildeling af metadata med hensyn til formål og anvendelse.

Med OAIS-modellen og en række metadataformater, som knytter sig til modellen, er der et modent grundlag at basere nationale tiltag på vedrørende langtidsbevaring af digitale objekter. Herudover er der en hel del internationale aktiviteter på området, som sikrer videreudvikling af modellen og formaterne.

Pligtaflevering af netpublikationer – set i lyset af OAIS modellen

I Danmark har pligtafleveringsloven siden 1998 omfattet statiske netpublikationer. Publikationer anmeldes via www.pligtaflevering.dk, hvorefter de hentes – eller hvis dette ikke kan gen-



Cornell University – Ithaca, NY, var vært ved en uges workshop om Digital Preservation

nemføres, afleveres efter aftale. Nedenfor er aktiviteterne beskrevet i relation til det begrebsapparat, som er formuleret i OAIS modellen for at give en illustration af denne.

Som beskrevet ovenfor er der to synsvinkler – den funktionelle og den informationsorienterede. Den funktionelle beskrivelse opererer med funktioner som administration, indsamling (ingest), arkivering, informationsstyring (data management), bevaring og adgang. Disse funktioner beskrives først. Derefter vender vi os mod struktur og indhold af de informationspakker, som håndteres.

Den helt centrale funktion er administrationsfunktionen, der samler organiseringen af opgaven med at arkivere netpublikationerne. Da ledelse ligger uden for modellen betyder det i et webarkiveringsscenario, at de grundlæggende forudsætninger som den lovgivningsmæssige beslutning om, at dele af nettet skal arkiveres, samt tilvejebringelse af resurser i form af økonomi, personale, platform (hardware og software) samt gode netværksforbindelser ikke er en del af det, der er omfattet af modellen.

Administrationsfunktionen omfatter derimod opgaverne med at etablere og sikre overholdelse af standarder, at gennemføre den grundlæggende opsætning af arkiveringssystemet, etablering af aftaler om aflevering af netpublikationer eller adgang til sådanne, der eventuelt er beskyttet af brugerid og password, effektivering af den fastsatte politik for hvad der skal indsamles (statiske netpublikationer), og hvordan det skal gøres (indsamling efter anmeldelse), og hvordan det skal bevares, kvalitetssikring (får vi de publikationer, vi skal have) og endelig også brugeradministration.

Selve indsamlingen (ingest) omfatter såvel den semi-automatiserede indhøstning, der sker, når programmer indhenter de statiske netpublikationer til pligtafleveringssystemet, samt indlemmelse af netpublikationer, der er afleveret ifølge en forudgående aftale.

Før selve indsamlingstidspunktet (pre-ingest) sker der i dag en anmeldelse af publikation til arkivering via www.pligtaflevering.dk, og undertiden træffes der beslutning om, at det anmeldte ikke kan indsamles efter loven, eller at

det anmeldte ikke egner sig til at indsamle via høstning, og derfor bør afleveres på en anden måde, der så aftales.

Når en netpublikation er modtaget, lagres den i det fysiske arkiv. Selve den arkivmæssige opbevaring omfatter både den initiale lagring, det vil sige opbevaring af den modtagne bit-strøm og lagring af alle de efterfølgende repræsentationer af denne bit-strøm, som fremkommer, når vi har flyttet en bit-strøm fra et dataformat til et andet. Løbende tager vi kopi af hele arkivet både i form af backup og ved at spejle hele arkivet fra Det Kongelige Bibliotek til Statsbiblioteket.

En lang række oplysninger gemmes uden for selve netpublikationen. Disse vedligeholdes gennem informationsstyringsfunktionen (Data Management). Det er for eksempel oplysninger, der knytter sig til hele indsamlingen i modsætning til indsamlingen af hvert enkelt lille dataobjekt, men også opsætning af indsamlingsværktøjer, bibliografiske metadata samt andre data, der for eksempel stammer fra anmeldelsesblanketten, teknisk-administrative metadata som proveniens, tidsstempler, adgangsrestriktioner, dataformat og auten-

citet, og endelig etableres et indeks til fritekstsøgning.

Planlægning af bevaringsaktiviteterne (Preservation planning) omfatter etablering af en bevaringsstrategi for alle de indsamlede netpublikationer, det vil sige både en plan for, hvorledes for eksempel HTML-, PDF- og worddokumenterne med deres indlejrede billeder i for eksempel GIF eller JPEG eller de indsamlede lyd- eller flash filer skal formatkonverteres, og en plan for, på hvilke medier de må ligge, så vi er sikre på, at vi for eksempel stadig har båndstationer, der kan læse vores bånd.

Ud over udarbejdelsen af strategier omfatter aktiviteterne også helt konkret at overvåge de lagringsmedier og dataformater, der forekommer i arkivet, samt for de sidstes vedkommende løbende at sikre, at der kan gives adgang til filerne, og at der er adgang til de viewere og plug-ins, der skal til, for at dette kan ske på tilfredsstillende måde. Endelig omfatter bevaringsplanlægning også en løbende kontakt til arkivets forskellige brugergrupper.

Endelig skal der jo kunne gives adgang (access) til indsamlede og bevarede publikationer. Denne funktion omfatter både en regulering af adgangen til de arkiverede netpublikationer, der kun må gives fra Det Kongelige Biblioteks og Statsbibliotekets læsesal, samt redskaber til dels at få overblik over, hvad der ligger i arkivet via søgegrænsefladen på www.pligtaflevering.dk, dels at få adgang til (se/høre og eventuelt udskrive) selve den bevarede netpublikation.

Vender vi os mod informationspakkerne, opererer vi med tre typer:

- Det, som afleveres og skabes i forbindelse med aflevering (SIP – submission information package)
- Det, som opbevares (AIP – Archival information package)
- Det, som brugeren får (DIP – dissemination information package).

Ser vi på den nuværende håndtering af netdokumenter har vi:

En SIP i det nuværende webarkiveringsarkiv består af navnet på den URL, der arkiveres, samt den http-respons, vores programmer modtager (HTTP-header

information med blandt andet mime type, samt sidens egentlige indhold), når vi beder en webserver om at få den pågældende URL.

En AIP i samme arkiv indeholder ud over SIP'en også oplysninger, der har med indsamlingen at gøre (høsttidspunkt, autentificeret...), deskriptive metadata indtastet af anmelderen, for eksempel udgiver, titel og forfatter, oplysning om hvorvidt publikationen stammer fra en offentlig institution eller ej, en entydig identifikation (værknummer) samt oplysninger om URL'ens liv efterfølgende i arkivet.

En DIP i et eksisterende adgangssystem, som det for eksempel kan ses på www.pligtaflevering.dk, indeholder et uddrag af AIP'en (få udvalgte bibliografiske metadata, oplysninger om formater og tilgængelige repræsentationer) samt i dette tilfælde oplysninger om den URL, hvor publikationen oprindeligt er indsamlet fra, og den URL, hvor dokumentet nu kan findes i arkivet.

Institutional Repositories og OAIS

I disse år arbejdes der intensivt overalt i den akademiske verden for at etablere digitale institutionsarkiver, de såkaldte "institutional repositories". I flere lande, for eksempel UK, kører ambitiøse og landsdækkende programmer, som har som mål at opbygge et netværk af digitale arkiver på samtlige universiteter.

Under vores besøg på MIT, hvor vi mødte udviklingsstaben bag det mest udbredte software på området, DSpace, blev vi blandt andet orienteret om samarbejdsprojektet CMI mellem MIT og Cambridge, som med et budget på 18 mio. kr. skal indføre et insitutional repository på Cambridge. Samtidig er en række træningsseminarer for andre UK universiteter under afvikling, så man formentlig kan forvente en bølge af implementeringer i de kommende måneder og år.

På MIT er situationen den, at forskerne har næsten uindskrænkede rettigheder til deres forskning, og at universitetet arbejder med meget store eksterne forskningsprogrammer, som sætter rammer for forskningspubliceringen. Så selv om DSpace på MIT selvbevidst erklærer, at deres formål er "*to create and establish an electronic system that captures, preserves and communicates*

the intellectual output of MIT's faculty and researchers", fik vi indtrykket af, at det gik noget langsomt med at overbevise forskerstaben om de fordele, de kan have af et Institutional Repository. Deres råd til andre, der skulle begynde processen, var at tilbyde så meget assistance som muligt, for eksempel til registrering og datahåndtering, i en lang indkørfase.

Det er ingen tilfældighed, at bevarings- og kommunikationsaspektet understreges i MIT's mission statement. For selv om mange i forskningsbibliotekssektoren har den bagtanke med disse arkiver, at det kunne være fint at etablere nogle rammer for publicering, som kunne udfordre kommercielle tidsskrifter med deres ublu prissætninger, er der ingen tvivl om, at dette formål er en langsigtet strategi, som arkiverne forhåbentlig vil bidrage til. På kort sigt er de to vigtigste aktuelle argumenter for oprettelsen af arkiverne, at dokumenterne bliver lettere tilgængelige og mere synlige, og at de vil blive opbevaret sikkert. Forskeren skal kunne stole på, at dokumentet opbevares i en autentificeret udgave og vil være tilgængeligt på længere sigt.

Det kan lyde som en selvfølgelighed, at de tilgængelige softwarepakker til institutional repositories også er klædt på til opbevaringen, men det er ikke uden videre tilfældet, i hvert fald ikke, hvis man tager OAIS modellen som udgangspunkt. Flere af de softwarepakker, som er fremme i øjeblikket, har ikke tilstrækkelige faciliteter til sikring af dokumenternes fremtidige tilgængelighed. Det gælder begge de udviklingsprojekter, som gruppen fik lejlighed til at mødes med i forbindelse med studierejsen.

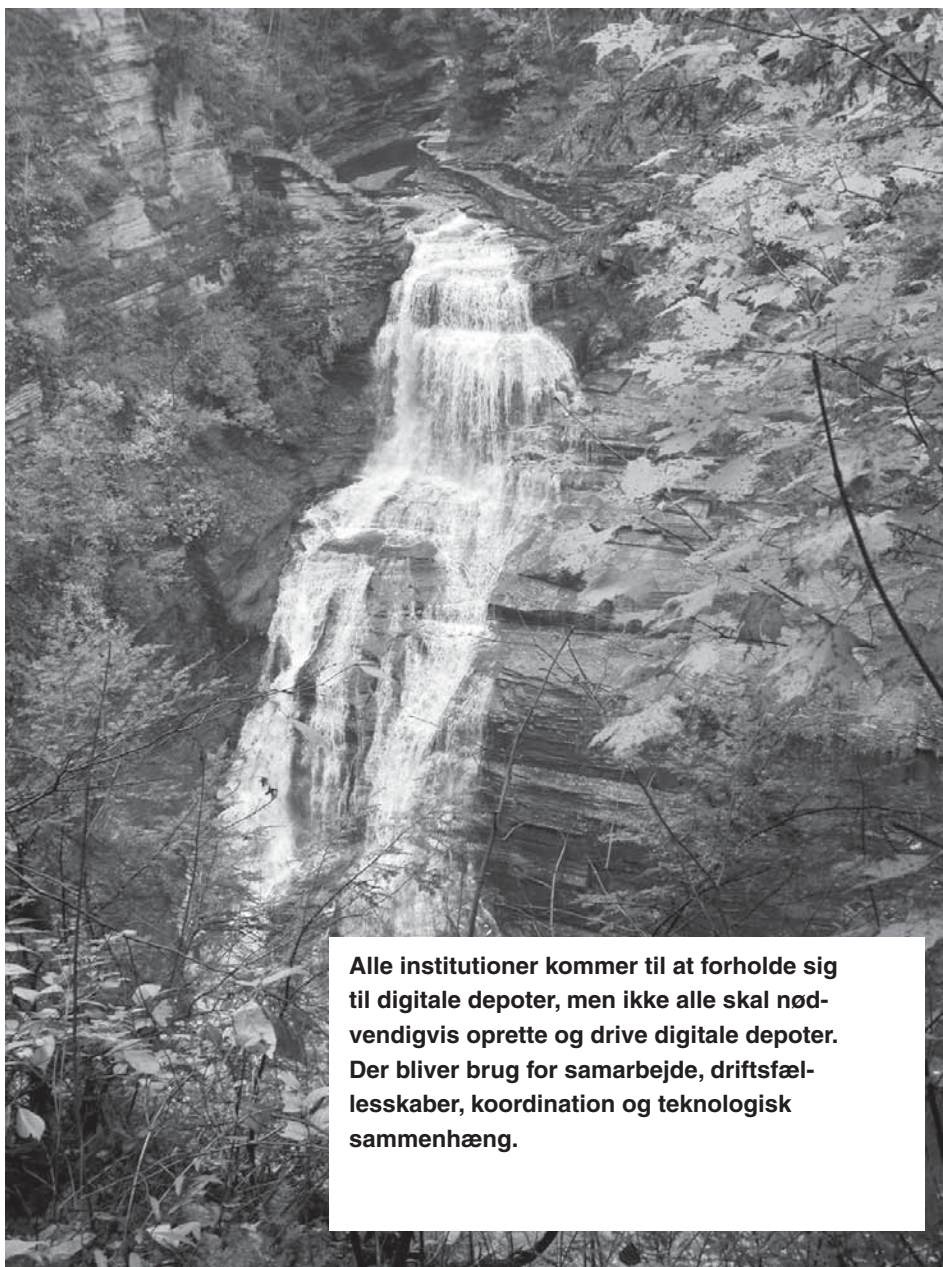
Ud over DSpace, som er omtalt ovenfor, var der tale om projektet Fedora, som er udviklet på Cornell, og som i sin opbygning forekommer mere fremtidsrettet end DSpace. Hvor DSpace har karakter af en færdig softwarepakke, som man kan downloade og installere uden yderligere dikkedarer, har Fedoras software en stærkere grundlæggende opbygning, både hvad angår databasestruktur og faciliteter til strukturering af det enkelte digitale objekt. Til gengæld har Fedora ikke fokuseret på de dele af systemet, som skulle supplere den overbevisende systemkerne med en brugervenlig og salgbar indpak-

ning – for eksempel et brugerinterface og nogle web-baserede administrationsrutiner. Vælger man Fedora, markerer man, at man systemmæssigt har muskler til selv at sætte systemkernen ind i sin egen ramme, mens DSpace er til at installere for næsten enhver institution, forudsat at man kan leve med dens defaultformater og metadata.

Hvis man tager OAIS modellen som det, den egentlig er tænkt som – en slags blueprint for ”det ansvarlige arkiv” – så har hverken Fedora eller DSpace endnu udviklet de faciliteter, der skal til, for at styre samlingerne ud fra et bevaringsaspekt. Hele det område af OAIS modellen, som vedrører arkivets samspil med ”preservation planning” og ”preservation administration”, indgår endnu ikke i systemerne, og selve strukturen i opbygningen af såvel ”submission packages” (SIP) som arkivets dokumentenheder, ”archive packages” (AIP), er ikke fremtidssikrede, hvis man forestiller sig, at man løbende skal kunne overvåge behov for bevaringsmæssige indgreb for at sikre visse af arkivets objekter.

Men selv om det tilgængelige open source software altså endnu ikke lever op til OAIS modellen, kan der være gode argumenter for at vælge det som udgangspunkt for at starte en nødvendig proces, som med sikkerhed vil være langvarig og lærerig. Selve afklaringen og erfaringsopbygningen omkring denne type arkiver kan ikke vente på, at det perfekte software er udviklet og til rådighed. Man må snarere se for sig en proces, hvor vi fra dansk side byder ind med bidrag til de udviklingsmiljøer, som er opstået omkring for eksempel DSpace og DIVA, som Statsbiblioteket arbejder med. Set i dette lys, er udviklingsmiljøet omkring softwaret muligvis en vigtigere evalueringsparameter, end det øjeblikkelige systemindhold. Det var ikke mindst af hensyn til denne evaluering vigtigt for studiegruppen at få mulighed for et reelt møde med udviklerne bag DSpace og Fedora.

I mellemtiden – det vil sige mens der arbejdes på, at softwaret for institutional repositories nærmer sig OAIS standarder for bevaring – er det vigtigt at sikre sig bedst muligt mod de risici for databtab, som vitterlig er til stede. Det kan og bør dels ske ved, at samspillet mellem institutionsniveauet



Alle institutioner kommer til at forholde sig til digitale depoter, men ikke alle skal nødvendigvis oprette og drive digitale depoter. Der bliver brug for samarbejde, driftsfællesskaber, koordination og teknologisk sammenhæng.

og det nationale niveau for digital bevaring snarest bliver implementeret, dels ved at der sker en systematisk kompetenceopbygning hos dem, der skal implementere systemerne lokalt. Hvis vi ikke sætter massivt ind på disse områder, i forbindelse med at vi åbner vores institutionsarkiver, risikerer vi i værste fald at tabe troværdighed i forhold til de forskningsgrupper, hvis arkiveringsbehov vi tilbyder at dække.

Nationale aktiviteter

I Danmark er mange biblioteker involveret i opbygning af digitale samlinger. Disse kan opstå som resultat af en bevaringsstrategi for det analoge materiale, det vil sige gennem digitalisering, som resultat af en formidlingsstrategi, igen digitalisering eller ved at lave samlingsopbygning af ”født digitalt” materiale. Det er klart, at bevaringsindsatsen for det digitale materiale må og skal afhænge af,

om det er eneste kopi, eller om der eksisterer en analog bevaringskopi. Ligger det museale ansvar på den digitale version, bør bevaring sikres gennem en aktiv bevaringsproces. En forudsætning er selvfølgelig, at materialet er blevet indsamlet.

Tager vi det sidste, indsamling, først, eksisterer der flere initiativer, som kan ses som arbejdende hen mod dette mål. Opbygningen af de tidligere omtalte institutionelle arkiver (eller hvad nu institutional repositories kaldes) er et eksempel, hvor man forener formidling med opbygning af et digitalt arkiv med forskningsmateriale. Et andet eksempel er opbygningen af et digitalt arkiv hos DAB med publikationer fra staten, et sidste er opbygning af preprintservere og understøttelse af elektroniske tidsskrifter. Alle disse er eksempler, hvor bibliotekerne tidligere håndterede de trykte

versioner, men nu ved håndtering af digitale versioner opnår såvel et mere brugervenligt produkt som en umiddelbar rationaliseringsgevinst. Prisen er imidlertid, at materialet nu er i en form, som på anderledes vis kræver overvågning – både mht. bevaring af filerne og til fortolkning af disse.

Hvad, hvordan og hvor meget, som skal bevares, er i nogen grad et politisk spørgsmål. Dette blev behandlet i kulturarvsrapporten, hvor alle aspekter af den digitale kulturarv blev behandlet. I forbindelse med webarkivering blev det konstateret, at der var et behov for at udvide mandatet, hvis man i fremtiden skulle kunne danne sig et indtryk af, hvordan informations- og kommunikationsmønstre var i starten af det 21'ende århundrede.

I webarkiveringsprojektet [netarkivet.dk] er målet at undersøge mulighederne for at etablere en samling af publiceret materiale – stationært, dynamisk og interaktivt, bestående af tekst, billeder, musik, video og animationer og dækkende både informationsformidling og kommunikation. Der skal etableres grundlag for en vurdering af, hvad der kan lade sig gøre og til hvilken pris. Der er her tale om store mængder data og om en stor diversitet i materialet. Bevaringsudfordringerne for dette materiale er derfor meget store.

Vender vi os mod bevaring kan det først konstateres, at der er tale om to typer bevaring:

- Bitbevaring, som går ud på at bevare filerne intakte (det vil sige alle bit'ne).
- Logisk bevaring, som går ud på at sikre, at filerne kan omsættes til noget, som et menneske kan opfatte.

Tidligere har fokus været på det første punkt, men med de hastige ændringer i soft- og hardware er der ved at komme fokus på den anden pind. Et nyligt eksempel på en bevaringsindsats er det engelske Domesday-projekt [DOMESDAY]. Generelt, som omtalt i NDIPP rapportens bilag II [NDIPP, 40], har der siden 1960'erne eksisteret omkring to hundrede medier, hvoraf kun meget få har eksisteret i mere end ti år.

Det er klart, at bitbevaring er en forudsætning for al bevaring. Er bittene forsvundet kan man have nok så mange emulatorer eller konverteringsalgoritmer. Det er derfor naturligt, at fokus har ligget og stadig ligger

her. Der er også stadig tale om, at man kan fortolke de fleste filformater. Har man en gammel Wordfil på en gammel 5 ¼ tommer disk, er chancen for at finde relevant drev lille, og chancen for, at denne endnu kan læses, endnu mindre. Men har man i tide sørget for at kopiere den over til en 3 ½ tommer diskette eller på sin institutions lager, ja så er chancerne betydeligt højere.

Ovenstående er et eksempel på mediamigrering – og en sådan er en meget vigtig komponent i en handlingsplan. Et eksempel på massemediakonvertering er Statsbibliotekets konvertering fra CD-ROM til tape af musik udgivet i Danmark. 20.000 cd'er er rippet og lagt ind i det digitale bogtårn som wav-filer. Disse er også konverteret til mwa-filer af hensyn til formidling (Phonofile). Wav-filerne er arkivfiler, hvortil knytter sig det museale ansvar. Det er derfor vigtigt at sikre, at filerne sikres gennem mindst to kopier, og at det løbende sikres, at man kan fortolke wav-filerne.

I dag findes forskellige strategier mht. logisk bevaring:

- Emulering, hvor man enten i software eller i hardware sikrer en emulering af de oprindelige omgivelser.
- Formatkonvertering, hvor man sikrer, at man til stadighed har materialet på en form, som kan fortolkes med den til enhver tid gængse fortolkere.
- Museal strategi, hvor man gemmer hardware og software, så man til enhver tid kan genskabe de rette omgivelser (også ofte kaldt teknisk museums strategien).

Institutioner med ansvar for digital bevaring anser den museale strategi for noget nær umulig, hvorfor interessen drejer sig om de to første. Der er mange argumenter for og imod, som det vil være for vidt at komme ind på her. Her vil vi blot konstatere, at begge anses for realistiske alternativer. Har man et rimeligt homogent materiale, synes formatkonvertering i dag at være et godt bud på en strategi, og den vælges ofte, hvis mulighederne er til stede.

Ovenfor er nævnt mange beslutninger i relation til digital bevaring. Dette leder os hen mod et sidste punkt, som vi stiftede bekendtskab med på vores kursus, nemlig udfærdigelse af en handlingsplan.

Handlingsplan

Digital bevaring handler selvsagt om evnen til at sikre digitalt materiale over en lang tidshorison. I modsætning til eksempelvis bøger kræver dette, at institutioner er indstillet på løbende at overvåge og agere i forhold til ændringer. Digitalt materiale kan ikke stilles på en hylde og gemmes, men kræver konstant overvågning – både med hensyn til at sikre filerne og sikre adgang og korrekt fortolkning af disse.

En institutions mulighed for at langtidssikre vil afhænge af en række parametre, såsom hvor centralt det er for dens forretningsområde, økonomisk sikkerhed, lovgivningsmæssig forpligtelse og så videre. For at sikre åbenhed omkring disse spørgsmål bør bevaringsinstitutioner offentliggøre en handlingsplan, på samme måde som flere biblioteker offentliggør deres servicemål. Aktiviteten kan også sammenlignes med den, som ligger bag processen bag ISO9000 certificering, hvor en del af certificeringen netop er en klar dokumentation af de nødvendige processer.

På kurset blev gennemgået en model for, hvordan en sådan handlingsplan kunne opbygges. Denne blev angrebet fra tre vinkler:

- Organisatorisk infrastruktur
- Teknisk infrastruktur
- Finansielle aspekter

Den organisatoriske infrastruktur har tre komponenter, som har hver sit strateginiveau:

- Framework, som på et meget overordnet, politisk plan afstikker rammerne for og indholdet af bevaringsaktiviteterne. Det er den del af handlingsplanen, som forventes at have den længste levetid. Indholdet skal på overordnet plan positionere institutionen i forhold til omgivelserne med hensyn til relevante aspekter af langtidssikring. Det betyder, at områder som mandat, andre aktører, interessenter, indsamlingsstrategi og potentielle brugsgrupper omtales.
- Politikker og procedurer, som giver en funktionel tilgang til aktiviteterne og arbejder mere specifikt med de relevante processer. Denne del er baseret på et godkendt framework og omfatter de overordnede retningslinier til politikker og procedurer, som efterføl-

gende kan omsættes til handling. Ting, som behandles, er roller og ansvar, bevaringsstrategi, teknisk infrastruktur samt en række aspekter ved det digitale materiale. Blandt disse er afleveringsprocedure, selektions- og kassationspolitik og kvalitetskontrol.

- Implementation, som fokuserer på implementeringsplaner her og nu.

Beskrivelse af den tekniske infrastruktur vil fortælle, hvordan man konkret løser de relevante opgaver, det vil sige, hvordan man teknisk behandler materialet, hvordan man sikrer det fysiske rum, hvilket udstyr, man har, hvordan man fysisk sikrer dette, hvilket personale man har og så videre. Beskrivelsen her falder i fem områder: Materiale, personale, udstyr, IT-omgivelser og det fysiske rum.

Afsnittet om de finansielle aspekter skal synliggøre omkostningerne – både de direkte og de afledte. Et klart mål er at sikre, at organisationen er bevidst om niveauet for de påbegyndte aktiviteter. Det er ikke nok at kunne påbegynde opbygning af digitale arkiver, der skal også være økonomi til den løbende drift.

Som beskrevet i TDR afhænger en del af tilliden af, i hvilket omfang disse ting er dokumenteret som en integreret del af arkivet. Alle institutioner, som påtænker at etablere/allerede har etableret et digitalt arkiv, bør forholde sig til disse punkter. Planen behøver ikke at definere aktiviteten i forhold til alle punkter, derimod kan man godt forestille sig en række strategiske samarbejder med klare rollefordelinger.

Afslutning

Desværre findes der endnu ikke et dækkende udtryk på dansk for opbygning af et digitalt arkiv, men vi tror, vi kan være enige om følgende:

- At opbygge et institutional repository betyder, at man ved hjælp af teknik, organisation, information samt workflow skaber de nødvendige rammer, for at digitale informationer og objekter kan deponeres, synliggøres og formidles, samt sikre fremtidig adgang til disse.

Resultatet af vores studietur er, som beskrevet, en opdateret viden om de mange spørgsmål, som DEF og bibliotekerne på nationalt niveau skal have svar på. Kort

fortalt handler det om at bevare adgang til vores videnproduktion i den digitale tidsalder. Mere konkret handler det om at besvare bl.a. følgende spørgsmål:

- Hvem er ansvarlig?
- Hvem føler sig ansvarlig?
- Hvad skal bevares?
- Hvorledes og hvem gennemfører selektion?
- Hvem betaler?
- Hvorledes håndteres kvalitetssikring?
- Hvorledes håndteres sikkerhed?
- Hvordan sikres migration?
- Hvem har adgang til hvad?

På nationalt niveau har Kulturministeriet udarbejdet Kulturarvsrapporten som tidligere nævnt vedrørende den nationale forpligtigelse. Men det er ikke nok.

Den ”grå” videnproduktion, som finder sted på universiteter, på forskningsinstitutioner, i sektorforskningen samt i uddannelsesinstitutioner, er lige så væsentlig for den fremtidige forskning. Det betyder, at alle institutioner kommer til at forholde sig til digitale depoter, men ikke alle skal nødvendigvis oprette og drive digitale depoter. Der bliver brug for samarbejde, driftsfællesskaber, koordination og teknologisk sammenhæng. Der er behov for at udnytte kompetencer, tekniske løsninger og udviklingsressourcer på tværs af institutioner og geografi. Derfor bliver der brug for et overordnet ”framework”, som beskriver alle disse aspekter.

DEF har stillet sig den opgave at medvirke til, at digital bevaring af videnproduktion på de nævnte institutioner kommer på dagsordenen. DEF kan gennem afklaring af de nødvendige spørgsmål og aktiviteter medvirke til at skabe dialog og sætte fokus på at sikre adgang til dansk videnproduktion gennem etablering af Institutional Repositories.

Et pilotprojekt er allerede i gang, nemlig test af forskellige softwareprodukter for opbygning af institutional repositories. DSpace og DIVA afprøves på Roskilde Universitetsbibliotek og Statsbiblioteket.

Resultatet er svar på nogle af de mange spørgsmål nemlig:

- Hvilke systemkoncepter er bedst egnet?
- Hvilken rolle har bibliotekerne i e-publicering i samspil med egen moderinstitution?

- Hvilke informationsmaterialer er relevante at bevare digitalt?
- Hvorledes sikres kvalitet i videnproduktionen?
- Hvilke samarbejdsmodeller er mulige, teknisk og organisatorisk?

Alle disse spørgsmål og flere til vil vi forhåbentlig have svar på i løbet af det kommende år.

DEF's ønske kan udtrykkes således, som vore australske kolleger har udtrykt det i *Digital preservation – a many layered thing* – The National Library of Australia, 2002:

- Ikke alle skal gøre alting, men der er nogen, der skal gøre noget!
- Der skal være nogle, som er villige til at gå forrest på næsten alle områder
- Og der skal være nogen, der må være villige til at tage ansvaret, selvom ansvaret kan være ikke at gøre noget!

Referencer

[AU]: *Preservation Metadata for Digital Collections*, <http://www.nla.gov.au/preserve/pmeta.html>, 1999

[CEDARS]: *Metadata for digital preservation*, <http://www.leeds.ac.uk/cedars/colman/metadata/metadataaspec.html>, 2000

[CPA/RLG]: *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*: <http://www.rlg.or/ArchTF/index.html>, Commission on Preservation and Access and RLG, 1996

[DOMESDAY]: <http://www.atsf.co.uk/dottext/domesday.html>

[DSpace]: <http://www.dspace.org/>

[FEDORA]: <http://www.fedora.info/>

[JISC]: *Feasibility and requirements study on preservation of e-prints*, http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf, 2003

[NDIPP]: (http://www.digitalpreservation.gov/report/ndiipp_appendix.pdf)

[NLNZ]: *Metadata Standards Framework – Metadata Implementation Scheme*, http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf, 2003

[OAIS]: *Reference model for an Open Archival Information System (OAIS)*: [http://www.ccsds.org/ConsultativeCommitteeforSpaceDataSystems\(CCSDS\),2002](http://www.ccsds.org/ConsultativeCommitteeforSpaceDataSystems(CCSDS),2002)

[OCLC/RLG]: *Preserving metadata and the OAIS information model: a metadata framework to support the preservation of digital objects*: <http://www.oclc.org/research/pmwg/>, OCLC/RLG Working Group on Preservation Metadata, 2002

[RLG/OCLC]: *Trusted digital repositories: attributes and responsibilities*: <http://www.rlg.org/>, 2002