

Digital bevaring i 100 år

Digitale materialer skal udvælges, bevares og tilgængeliggøres for eftertiden. Det er udfordrende fordi det kræver mere end blot at sikre materialerne eksistens og beskrive dem, dataformaterne skal også kunne læses i fremtiden.

Af Birgit Nordsmark Henriksen, Det Kongelige Bibliotek bnh@kb.dk

Har du prøvet at stå i den situation, at du skulle bruge et digitalt foto eller dokument, som du havde i din samling, men da du fandt frem til det, så var det alligevel uden for din rækkevidde?

Det har jeg selv gjort for nylig. En serie ældre fotos, som var skannet i 1995, viste sig stadig at ligge i det zip-arkiv, som samme år var blevet udlæst og fordelt dels på 1.44 MByte floppydiske dels på en 135 MByte 3,5" SyQuest removable harddisk cartridge. Fordi de dengang begrænsede og dyrt-indkøbte harddiske skulle bruges til nogle nye projekter. 15 år senere blev de gamle løse diske fundet i en papkasse, og der var problemer med at få udpakket data korrekt, nutidens PC'er kunne ikke tilsluttes det gamle SyQuest-device, nogle af filerne kunne ikke længere åbnes, fordi de var blevet korrupte med årene, og for flere af billedernes vedkommende var informationerne om, hvem de afbillede, gået tabt.

"Digital bevaring" er alle de ting, der skal sikre, at denne situation ikke opstår, og at digitalt materiale fortsat er læsbart og forståeligt ikke blot om 15, men også om rigtig mange år – jeg plejer at sige mere end 100 år, for at alle bliver klar over, at det skal kunne anvendes i systemer, som vi slet ikke kender og af brugere, der kun har den viden til rådighed om materialet, som vi har overleveret.

Opgaverne er mange

Som man kan ane, omfatter digital bevaring en lang række opgaver: Man skal først have en politik for, hvilket materiale der skal bevares, og hvilket man blot vil sikre på vanlig vis i de eksisterende systemer.

Dernæst skal man sikre det udvalgte materiales blotte eksistens: At det digitale materiale kommer fornuftigt ind i institutionens samlinger, og at det hele tiden flyttes fra udgåede medier til tidsvarende medier (mediemigration), og at alle bits til stadighed er til rådighed og i den rigtige sekvens, så data ikke korrumpere (aktiv bitbevaring). Man skal også sikre sig, at nogle i institutionen har til opgave at sikre, at data findes og er tilgængelige som ønsket samt værner om det, når budgetterne skal beskæres (samlingsansvarlige).

Det er dog ikke tilstrækkeligt blot at sikre materialet eksistens. Det skal også sikres, at det til stadighed kan forstås. Det kalder man logisk eller funktionel bevaring. Man skal sikre, at data er vel beskrevet – såvel for den indholdsmæssige del som for den tekniske del (beskrivende metadata, tekniske og bevarings-metadata). Man skal sikre at de anvendte dataformater fortsat kan læses af tidssvarende læseprogrammer enten ved om nødvendigt

at ændre dataformatet (data migrering) eller ved at ændre omgivelserne til de gamle data (forskellige former for emulering). Man skal følge den teknologiske udvikling, samt overvåge det materiale man har omfattet af digital bevaring, så man ved, hvornår og hvordan man skal gennemføre de forskellige bevaringsaktiviteter (bevaringsplanlægning), og endelig skal man også eksekvere sin bevaringsplan.

Nogle af de nævnte opgaver kan med fordel løses af nogle få institutioner i fællesskab, f.eks. teknologiovervågning samt tilvejebringelse af værktøjer og platforme til at teste og gennemføre bevaringsprocesserne, mens andre, som f.eks. at fastlægge hvilket materiale der skal omfattes af digital bevaring og hvordan, må ske i den enkelte institution. Her hører de løbende aktiviteter med at sikre de nødvendige ressourcer og samlingsrevisioner også hjemme. Samt udvælgelsen af materialet.

"I Det Kongelige Biblioteks afdeling for Digital Bevaring har vi arbejdet med bevaring af materiale fra internettet samt tekster og stillbilleder, og listen er i 2009 udvidet til også at omfatte interaktive digitale objekter som f.eks. spil, videnskabelige datasæt og videnskabelig korrespondance."

Den nødvendige selektion

For langt fra alt digitalt materiale i en institution behøver at blive omfattet af digital bevaring. På Det Kongelige Bibliotek har vi foreløbig besluttet kun at lade digital bevaring omfatte samlinger der a) enten er født digitale og dermed kun findes som



Pixeleret udgave af Apple Macintosh med floppydrev, 1986.

sådanne, b) er fremkommet ved substitutionsdigitalisering, hvor den digitale kopi erstatter en fysisk original, som er ved at gå til grunde eller c) består af digitaliseret materiale, som har været særlig kostbar at fremstille. Årsagen til denne afgrænsning skal findes i økonomien – eller snarere de manglende ekstra midler til formålet. Digital bevaring koster nemlig som alt andet penge, og fordi det er aktiviteter og dermed udgifter, som skal afholdes årligt, så længe materialet ønskes bevaret, kan det blive dyrt.

Behovet for at kunne danne sig et overblik over den samlede udgift opstår, når man skal beslutte, hvorvidt en given digital samling skal langtidsbevares. Statens Arkiver og Det Kongelige Bibliotek har derfor med støtte af Kulturministeriet i 2008 igangsat et projekt, hvor man har set på, hvordan en engelsk omkostningsmodel, kaldet LIFE, kan tilpasses til danske forhold og erfaringer. Målet er at tilvejebringe en model for beregning af omkostninger forbundet med bevaring af digitalt materiale på statens arkiver, biblioteker og museer. Modellen skal kunne bruges i forbindelse med strategisk beslutningstagning og budgetlægning. Den omfatter endnu ikke alle faser i den digitale livscyklus, men vi håber i 2010 både at kunne udvide modellen og at gøre såvel modellen som det tilhørende værktøj til beregning tilgængeligt for alle.

Statsbiblioteket og Det Kongelige Bibliotek har begge længe modtaget digitalt materiale som et led i pligtafleveringen og har derfor de sidste 5-10 år haft medarbejdere til at arbejde med disse problemstillinger. Dette arbejde har i høj grad også omfattet involvering i internationale projekter og organisationer både for at løse så mange opgaver som muligt i et internationalt samarbejde med institutioner med identiske opgaver, men også for at kunne trække den nødvendige viden hjem til de danske bevaringsinstitutioner.

I Det Kongelige Biblioteks Afdeling for Digital Bevaring har vi således arbejdet med bevaring af materiale fra internettet

samt tekster og stillbilleder, og listen er i 2009 udvidet til også at omfatte interaktive digitale objekter som f.eks. spil, videnskabelige datasæt og videnskabelig korrespondance. Særligt for de to sidste områder gælder, at vi arbejder med at identificere, hvor de digitale bevaringsaktiviteter skal indarbejdes.

Digital bevaring er i vores øjne ikke blot en opgave, som ligger til sidst i en proces, men er noget der skal indarbejdes i hele det digitale materiales livscyklus, hvis materialet til stadighed skal være tilgængeligt og genbrugeligt.

Fælles bit magasin

De løsninger, som institutioner vælger til opbevarelse af sine data, er ofte forskellige, fordi data, formål og anvendelse tilsvarende er forskellige. I det nederste lag – der hvor vi skal sikre materialets korrekte sekvens af 0'er og 1-taller, er opgaven derimod helt identisk. Derfor er de tre største danske kulturbevaringsinstitutioner, støttet af Kulturministeriet, i færd med at tilrettelægge en arkitektur til et "fælles bit magasin", der grundlæggende bygger på, at institutionerne kan hjælpe hinanden i opgaveløsningen ved at lade deres egne systemer indgå i en samlet løsning. Gennem ny software, som vi lægger oven på de forskellige fysiske installationer, vil vi gøre det muligt for en institution at gemme og bevare sine bits i det antal kopier, med den tilgangshastighed, med det valg af lagringsmedie, med den geografiske spredning og den indbyrdes bit integritetskontrol som man ønsker og er villig til at betale for. Projektet blev i oktober 2009 præsenteret for en stor gruppe af institutioner under Kulturministeriet og forventes at fortsætte i de kommende år med en målsætning om pilotdrift i 2011.

I 2010 er det ligeledes planen at de store danske bevaringsinstitutioner i fællesskab vil lancere et nyt websted – digitalbevaring.dk – hvor mange af de problemstillinger, som har nævnt i denne artikel, behandles. Målgruppen vil være danske arkiver, biblioteker og museer, men alle yderligere interesserede forventes at kunne få nytte af initiativet. 