

**Norsk Arkivforum**

**18**



# Norsk Arkivforum

18

Frank Meyer (red.)

**Digitalisering av kulturarvmateriale**

Arkivarforeningen

Oslo 2004

ISBN: 82-996884-1-8  
ISSN: 0800-3106 (hft)

Norsk Arkivforum er en skriftserie  
utgitt av Arkivarforeningen,  
c/o Statsarkivet i Oslo

Trykk: Edgar Høgfeldt AS, Kristiansand, 2004



## **Innhold**

<i>Frank Meyer</i>	Forord	7
--------------------	--------	---

### **I. Hvordan ser det ut i dag i den norske digital-jungelen?**

<i>Snorre D. Øverbø</i>	Digitalisering utenfor Arkivverket	10
-------------------------	------------------------------------	----

<i>Yngve Nedrebø</i>	Digitalisering innenfor Arkivverket	16
----------------------	-------------------------------------	----

### **II. Digitalisering – Gjør vi de riktige ting, gjør vi ting riktig?**

<i>Hilde Sandvik</i>	Gjør vi de riktige ting? Et brukerperspektiv på digitalisering av arkivmateriale	26
----------------------	--	----

<i>Jon Holmen</i> <i>Christian-Emil Ore</i>	Gjør vi ting riktig? Kvalitetssikring av bilder i digitalisert form	30
--	---	----

<i>Gunnar Thorvaldsen</i>	Gjør vi ting riktig? Kvalitetssikring av digitalisert data	41
---------------------------	--	----

### **III. Debatt: Har Norge en digitaliseringspolitikk?**

<i>John Herstad</i>	Arkivverkets digitaliseringspolitikk	54
---------------------	--------------------------------------	----

<i>Kjersti Løkken</i>	Nasjonalbibliotekets digitaliseringsprinsipper og -politikk	60
-----------------------	---	----



*Frank Meyer*

## **Førord**

Denne boken inneholder de skriftlige versjonene av foredragene som ble holdt på Arkivarforeningens seminar Digitalisering av kulturarvmateriale i slutten av oktober 2001. Seminaret ble planlagt og organisert av under- tegnede i samarbeid med styret i Arkivarforeningen og generøst støttet av Forskerforbundet.

Det spesielle utgangspunktet for seminaret var forslagene fra arbeidsgruppen for *IT og andre sektorovergripende spørsmål innenfor bibliotek, arkiv og museum*, utnevnt av Kirke-, utdannings- og forskningsdepartementet (KUF) og Kulturdepartementet (KD), som Arkivarforeningen ønsket å drøfte. På seminaret ble følgende spørsmål diskutert: Hvordan ser arbeidsfeltet vårt (ABM-sektoren) ut i dag, og hvordan ønsker vi at det skal se ut i fremtiden? Hvilke ønsker og behov har våre samarbeidspartnere, hva kan vi lære av dem, hva har vi å gi? Hvilke sektorovergripende virkemidler kan vi ta i bruk? Hvem kan vi alliere oss med? Hvor kan vi se oss om etter nye medlemmer?

Det er umulig å gjengi i denne seminarrapporten de mangfoldige perspektivene som kom fram i diskusjonene. Det ble imidlertid klart at digitalisering av og til blir betraktet som en vidundermedisin i forhold til arkivinstitusjonenes plass-, konserverings- og formidlingsproblemer. Men som medisiner ellers er kanskje heller ikke denne helt uten bivirkninger.

Det denne rapporten kan dokumentere er de forberedte innleggene som ble holdt på seminaret. Disse var (og er) delt inn i tre bolker.

I en første bolk tar Snorre D. Øverbø og Yngve Nedrebø for seg å tegne hovedtrekkene i et kart over den norske digital-jungelen, og da henholdsvis utenfor og innenfor Arkivverket.

Den andre bolken inneholder tre innlegg som svarer på spørsmålene om vi gjør de riktige ting og om vi gjør ting riktig. Jon Holmen og Christian-Emil Ore, Hilde Sandvik og Gunnar Thorvaldsen gir oss ulike perspektiver, ett fra et bruker- og to fra produsentstandpunkt.

Avslutningsvis forklarer John Herstad og Kjersti Løkken henholdsvis Riksarkivarens og Nasjonalbibliotekets digitaliseringspolitikk.

Når bidragene utgis først nå, er dette helt og holdent redaktørens ansvar. Det at publikasjonen er forsinket betyr imidlertid ikke at innholdet har mistet sin relevans eller aktualitet. Dette er ikke minst tydelig, når det gjelder behovet for å samordne digitalisering av kulturarvmateriale i Norge. Da det ble vedtatt at Statens senter for arkiv, bibliotek og museum skulle opprettes og være operativ fra og med 1. januar 2003, var det mange som nærte forhåpninger om at organet skulle utøve denne viktige nasjonalt koordinerende funksjonen. Forhåpningene ble styrket av at institusjonen formulerte kriterier for tildeling av prosjektmidler våren 2003 som gikk ut på at en ville prioritere prosjekter med høy overføringsverdi til andre prosjekter og tuftet på samarbeid med andre institusjonene innen ABM-sektoren. Oversikten over innvilgede søknader fra ABM-utvikling, som ligger ute på organets hjemmeside, viser at det er betydelig avstand mellom liv og lære. Prosjektmidlene ble bokstavelig spredt i øst og vest og fordelt på en rekke mindre prosjekter, der verken samarbeidet med andre ABM-aktører eller overføringsverdien i forhold til andre prosjekter er synlig. Så viktig disse midlene enn er for å få til viktige bevarings- og formidlingstiltak lokalt, så lite bidrar en slik bevilgningspolitikk til å løse den store utfordringen som hele ABM-sektoren har å gjøre med, nemlig å få til mer kostnadseffektivitet ved samordning.

På vegne av Arkivarforeningen takker jeg bidragsyterne for deres tålmodighet.

**Del I.**

**Hvordan ser det ut  
i den norske digital-jungelen?**

## Digitalisering utenfor Arkivverket

Takk for at jeg fikk komme hit for å snakke om digitalisering. Det er et interessant område, som kommer til å sette sitt preg på vårt arbeide som arkivarer i årene som kommer. Overskriften på denne delen av seminaret er formulert som et spørsmål: "Hvordan ser det ut i dag i den norske digitaljungelen?" Det er et interessant formulering, og umiddelbart tenkte jeg å ville svare: - Det ser ikke ut der! Slik kan det i hvert fall fortone seg, hvis man skal gå ut å få oversikt over landskapet. Kanskje særlig i den delen av reservatet som er utenfor Arkivverkets høye gjerder. Floraen er viltvoksende, og like hurtigvoksende som i tropene. Det er alle slags vekster der, både store og små, og jungelen ser ut til å bli tettere og tettere. Det er jo bra på en måte, men det burde vel bli ryddet litt av og til. Foreløpig ser det ikke ut til at avskoging og klimaendringer vil gjøre seg gjeldende i nær framtid.

Som førstemann ute bør jeg vel forsøke å resonnerer litt omkring definisjoner, begrepsbruk og avgrensinger. Så la oss spørre oss selv: Hva er digitalisering? Ordet digitalisering er ikke lett å finne i de mest vanlige oppslagsverkene. Ordet "digital" er derimot forklart flere steder. Vi kan plukke ut den som står i "Mitt dataleksikon": "**digital**: informasjon basert på en kombinasjon av det binære tallsystemet, 0 og 1." I andre oppslagsverk settes digital gjerne i kontrast til analog, som *ikke* er basert på det binære tallsystemet, men "som er basert på fysiske, kontinuerlig varierende størrelser".

En tredje definisjon av begrepet "digitalisering" har jeg funnet i Caplex på Internett, og definisjonen lyder som følgende:

"**digitalisering** (av eng. *digit*, siffer), koding av informasjon som tall. Bilder, tekst, osv. kan kodes som tall i et binært tallsystem, slik at informasjonen kan behandles av datamaskiner, for senere å bli omdannet til den opprinnelige formen igjen. "

Ut fra dette må en kunne si at å "digitalisere" er å konvertere/overføre informasjon fra analogt medium til digitalt binært tallsystem. Konverteringen/overføringen kan skje rent maskinelt/teknisk eller at menneskehender er inne bildet, i større eller mindre omfang.

Vi arkivarer digitaliserer *verneverdige* analoge arkivdokument. Det vil si at

vi konverterer informasjonen som er lagt på analoge arkivdokument og over til arkivdokument som har sin informasjon basert på en kombinasjon av det binære tallsystemet.

Men det er ikke tilstrekkelig at dokumentet er verneverdig for at det skal bli digitalisert. Det store gross av arkivsektorens arkivdokumenter vil aldri bli digitalisert, i hvert fall ikke i vår levetid. I tillegg til verneverdi må det ligge andre motiv bak, det vil si at det ut fra forskjellige hensyn er hensiktsmessig å digitalisere dokumentet. Slike hensyn kan være konservering, tilgjengeliggjøring, formidling med mer.

På bakgrunn av det som hittil er sagt kan vi så reformulere spørsmålet: *Hva slags verneverdige arkivdokument blir digitalisert utenfor Arkivverket?*

Historisk sett har digitalisering for arkivarer dreid seg mye om persondata, og slik vil det vel bli fremover også. Det har gått mye på å registrere/overføre opplysninger fra folketellinger, kirkebøker med mer til digitalt format, som har blitt håndtert i databaser. Dette er informasjon som gjennom sin skjematiske karakter, passer særlig godt inn i databaser. Utover på 1990-tallet har en digitalisert flere kildegrupper, og først og fremst gjelder det historiske fotos.

Enten vi liker det eller ikke: Det er mange utenfor "arkivbransjen" som forvalter verneverdige arkiv. Utenfor Arkivverket er dette museer, biblioteker, privatpersoner, bedrifter, i tillegg til arkivinstitusjoner. Noen av disse, temmelig mange faktisk, digitaliserer arkivdokument eller registre til slike. Landskapet er svært uoversiktlig - og for å holde oss til naturfagterminologien -, og både topografi og vegetasjon er til dels uoversiktlig.

Vi har større museer (for eksempel Norsk Folkemuseum) som digitaliserer fotos. Det samme gjør Nasjonalbiblioteket i Mo i Rana. De sistnevnte har spennende digitaliseringsprosjekter på gang i forhold til bilde, lyd og trykte medier. Også Tromsø Museum har tidligere i høst lagt ut deler av sin fotosamling knyttet til "Fagenhet for nyere historie". Omtrent 50 000 bilder er tilgjengelig i databasen på Internett.

Som dere vet består Arkiv-Norge utenfor Arkivverket av en svært heterogen samling av institusjoner, foreninger og privatpersoner som arbeider med historiske arkiver. Det finnes relativt store institusjoner med 10 til 20 ansatte og med et sterkt faglig miljø, og det finnes lokalarkiv og samlinger som utelukkende baserer seg på frivillig arbeidskraft, på det vi kaller "amatører". Jeg har naturligvis ingen totaloversikt over hva alle disse arbeider med på

området ”digitalisering”. Men jeg skal forsøke å gi dere noen hovedtrekk. Og så tar jeg alle mulige forbehold om at det jeg sier kan være feil på den måten at jeg da kan unnlate å omtale prosjekter som ikke jeg er klar over.

Jeg syns for øvrig det er beklagelig at det ikke fins en instans som har til oppgave å registrere det som skjer av slikt arbeid. For det er mye viktig som skjer, og det er stedvis ikke små ressurser som er lagt inn i digitaliseringsprosjekter. Jeg avventer et koordinerende organ (både nasjonalt og regionalt) som alle som digitaliserer kan forholde seg til, det gjelder både i faglig som i økonomiske henseender.

## **Arkivkataloger**

De aller fleste byarkiv, fylkesarkiv, interkommunale arkiv og lokalhistoriske arkiv har arkivkatalogene sine i et elektronisk format. Mange bruker ASTA (gjerne de større institusjonene), andre bruker mer hjemmesneakra system. Noen har dem inne som tekstdokument på en pc. En del større byarkiver/fylkesarkiver har lagt hele eller deler av sitt katalogverk ut på Internett.

Det store bildet på arkivområdet utenfor Arkivverket er naturligvis at analoge formater – i særdeleshet papir - er totalt dominerende. Situasjonen er vel stort sett lik den en finner innenfor Arkivverket. Det er kun et helt marginalt innslag av digitaliserte arkivdokument. Der en har kommet litt på vei – hvis en kan si det sånn – er digitalisering av fotografi. Her er det flere arkivinstitusjoner som gjør en god jobb. Vi kan for eksempel nevne Oslo byarkiv og Fylkesarkivet i Sogn og Fjordane. Men den store dominerende delen av norske arkivbestander består som kjent av papir. Dette hurtigvoksende berget av papir har man knapt begynt på. Spørsmålet er vel også om man skal begynne på det.

Her er vi ved en problemstilling – kjære kolleger – som jeg gjerne vil utfordre dere litt på: Skal digitalisering kun sees som et formidlingstiltak? Kan digitalisering bli en integrert del av arkivsektorens bevaringsarbeid? Kan utvalgte deler av arkivberget digitaliseres, og deretter makuleres? Hva er egentlig poenget med digitalisering hvis det ikke også kommer som en erstatning for noe, men bare blir et tillegg til alt det andre vi skal rekke over i vår travle hverdag. Dette er kontroversielle spørsmål, som jeg tror vi som arkivarer vil måtte forholde oss til i årene som kommer.

Noen steder har man så smått begynt å digitalisere arkivdokument ved å skanne/avfotografere tekstdokument, det vil si arkivdokument med skriftlig innhold. Det kan være referater fra møtebøker, korrespondanse,



saksdokument med mer. IKA Kongsberg har begynt med et prosjekt med innskanning av arkivdokument fra kommunale arkiv. For de interkommunale arkivinstusjonene i Norge, vil digitalisering av kommunale arkivsaker få økt aktualitet i nær framtid. Det arbeides nå intenst med planer mange steder i landet om å etablere felles arkivdepot for de eldre kommunale arkivsakene. Som dere vet oppbevares arkivene nå hos kommunene. Men i kjølvannet av arkivlovens krav til arkivlokale og lignende, har dette med felles arkivdepot blitt mer aktuelt. Jeg vil tro at flere IKA'er nå ser for seg løsninger, der en vil ta i bruk digitaliseringsverktøy for å takle de utfordringene en slik arkivhåndtering innebærer, for eksempel med henblikk på tilbakelån til kommunene med mer, formidling til publikum og avlevering av elektroniske arkiv.

Vi i Fylkesarkivet i Sogn og Fjordane har drevet med noen forsøk i det små med digitalisering av utvalgte arkivdokument. Disse har vi koplet til ASTA-katalogen og lagt ut på Internett. Ellers er vi i gang med det vi kaller "Møtebok-register". Der registrerer vi alle saker i kommunenes formannskapsprotokoller i Sogn og Fjordane. Vil legger så dataene i databasen og kopler dem til digitaliserte, det vil si i vår sammenheng avfotograferte elektroniske dokument av den aktuelle siden i møteboka. Vi tar sikte på å digitalisere brorparten av formannskaps- og kommunestyreprotokollene i Sogn og Fjordane, fra 1837 til omtrent 1945. Det brukes et særskilt digitalt kamera til dette, og Statsarkivet i Bergen har gjort dette arbeid for oss hittil. Dette er et prosjekt som skal utvikles til et fullskalaprojekt i 2002 og 2003.

Lokalhistoriske arkiv, museums lag og historielag arbeider flere steder med å digitalisere lokalhistoriske kilder. Det være seg historiske fotos, utlegging av bygdebøker på nettet, lokalhistoriske artikler med mer. Mange steder foregår det en imponerende formidlingsvirksomhet, som ofte helt eller delvis er båret fram av frivillig arbeid. Flere steder er kommunen inne i bildet både som finansieringskilde og faginstans.

Kort sagt: Det skjer ganske mye spennende rundt omkring i Norge.

Problemene er imidlertid mange. Jeg vil bare kort nevne følgende: Digitalisering av arkivdokument er et særlig ressurskrevende arbeid. Det koster mye i form av utstyr, personalressurser og faglig utvikling. Vi har per i dag ikke tekniske løsninger som tilfredsstillt våre krav til hurtighet og kvalitet, i hvert fall ikke i kombinasjon. Det er et langt stykke fram før man kan si at digitalisering kan bli en fullblods bevaringsmetode for arkivsektoren.

Hvis vi fortsetter å skue fremover er det helt klart at vi til nå kun har sett noen ytterst vage konturer av hvordan den digitale jungelen vil bli utenfor Arkivverket. Vegetasjonen endrer seg raskt, og nye arter vil bli introdusert. Jeg vil anta at dette med digitalisering av arkivdokument vil eksplodere de neste 10 år. Grunnene til det er blant annet:

1. Brukerne våre ønsker det. Og det er jo for dem vi eksisterer. Det må vi ikke glemme, selv om det kanskje er lett av og til.
2. Teknologi og pris gjør dette faglig og økonomisk realiserbart for flere enn i det gjør i dag.
3. Bredbånd. Gjennom høyhastighetslinjer kan selv til dels store dokumenter passere uten for store problem.

Utfordringene er mange for dem som arbeider med digitalisering utenfor Arkivverket. Jeg tror det fortsatt vil variere sterkt fra institusjon til institusjon hvor mye en vil og kan satse på digitalisering. Fortsatt er det slik at det vil være store forskjeller innen arkivområdet, og naturligvis størst forskjeller utenfor Arkivverket. Teknologien som ligger der, vil ennå i mange år fram i tid kreve kunnskap ut over det gjennomsnittlige, ja mer enn det. Programvare og utstyr vil fortsatt koste relativt mye. Videre kreves det ikke minst en sterk faglig orientering mot området "Digitalisering". Arbeidet vi gjør på området må stå på sikker faglig grunn. Vi har bare så vidt begynt på den veien.

Tilgjengelige ressurser og faglige prioriteringer vil legge sterke føringer på hvor sterk digitaliserings-innsatsen blir for den enkelte. Jeg tror at vi kan få til en god utvikling, og jevne ut forskjellene innen "Arkiv-Norge". Gjennom visse stimuleringsordninger kan en forsere utviklingen innen arkivsektoren. Et fond eller en romslig årlig pott med penger kan støtte pionerprosjekter, utviklingsprosjekter innen digitalisering, også utenfor Arkivverket. Det er bare igjennom praktisk arbeid med digitalisering av arkivdokumenter, at en får grunnlaget for å sette standarder, faglige og teknologiske. Det er også viktig at det etableres et nasjonalt ressurs- og kompetansesenter for digitalisering i vid forstand, som kan virke støttende for planlagte og pågående digitaliseringsprosjekter.

Etter hvert bør vel målsettingen være at det ikke lenger er tale om *prosjekter* innen digitalisering, men at digitalisering blir en del av det løpende arbeide

som arkivarer til daglig utfører. Som en helt integrert del av institusjonens bevarings- og formidlingsarbeid.

I årene fremover må vi ta stilling til om vi ønsker å ha den digitale delen av "Arkiv-Norge" som en jungel. Eller skal vi prøve å lage en park ut av det hele? Har den sin verdi? Fins det ressurser til det? Er det i det hele tatt mulig å sende gartnere inn i jungelen, og er det ønskelig?

Mitt viktigste budskap er at vi som profesjonelle arkivarer bør ha en offensiv og positiv holdning til utfordringen som alltid vil ligge der: Hvordan vi på nye og bedre måter kan bevare og formidle kulturarvmateriale. I møte mellom arkivaren og våre nye teknologiske verktøy kan mye spennende oppstå.

## **Digitalisering av kulturarvmateriale innenfor Arkivverket**

Jeg vil begynne med å trekke fram noen vel kjente fakta om Arkivverket og Arkivverkets mening og mål. Det vil jeg ha med som et naturlig bakteppe for det jeg skal berette om, digitalisering i Arkivverket, og min intensjon vil være å forsøke å vise sammenhengen mellom det som har vært gjort og gjøres på feltet digitalisering, og den virksomheten etaten er satt til å drive med.

Jeg har min erfaringsbakgrunn fra statsarkivene i Hamar og Bergen, og fra arbeid med Digitalarkivet. Det vil uten tvil bli speilet i min framstilling.

\*\*\*

”Arkivverket skal sørge for at arkiver som har vesentlig kulturell eller forskningsmessig verdi, eller som dekker administrative, rettslige eller velferdsmessige dokumentasjonsbehov, blir bevart og gjort tilgjengelig for bruk.” (Etatens virksomhetsidé, formulert i strategiplanen for perioden 2001-2005.)

Gjennom flere tiår har Arkivverket delt inn sin virksomhet etter tre hovedinnsatsområder:

- Stimulering til bevaring av arkiver.
- Bevaring av arkiver.
- Formidling av arkivinformatjon.

Den som vil gjøre seg kjent med Arkivverkets mål eller politikk bør begynne med å lese etatens strategiplan. I planen som løper nå er det gitt strategimål som går på forholdet til digitalisering av arkivmateriale – et program for å gi publikum elektronisk tilgang til kataloginformasjon og til et bredt utvalg av etterspurte arkiver, og med konkret anvisning av hvilke tiltak som skal gjennomføres for at disse målene skal nås innen utgangen av 2005.

Leser vi årsmeldingene og årsstatistikken ser vi at vi år om annet bruker knapt 20 årsverk på konservering (“bevaring av arkiver”), og nesten 50 på å betjene publikum, på 8 lesesaler, og til å besvare forespørsler fra et bredt

sammensatt publikum, om kopier, om avskrifter, om attester og til å svare på konkrete spørsmål om rettigheter, om faglige behov, eller om hobby. Totalt har etaten rundt 200 årsverk til disposisjon.

Vi ser videre at vi år om annet har knapt 40 000 registrerte lesesalsgjester, og at vi får mellom 30 000 og 35 000 journalførte spørsmål. Det betyr at vi i snitt bruker 0,7 årsverk for hver 1000 gjester/henvendelser. Videre kan vi lese ut av statistikkene at mellom 70 og 75% av etatens kontaktflate mot publikum ligger i ytre etat, i statsarkivene. Går vi ned under statistikken og ser på hvor publikums interesse ligger, vil vi oppdage at bruken er konsentrert om et relativt lite antall kilder, og at de samme kildene har vært i brukernes fokus tiår etter tiår. Jeg har ofte hevdet at 97% av arkivbruken i Statsarkivet i Bergen er rettet mot bare rundt 3% av den arkivmengden vi forvalter. Det er selvsagt spissformulert, men likevel ganske nær realitetene.

Den konsentrerte bruken har sin pris. De kildeseriene som er mest populære har vært utsatt for en betydelig slitasje, og det har vært nødvendig å forsøke å skjerme originalene effektivt. Det følger av vårt formulerte mål om å bevare arkivene. Men den konsentrerte bruken har også sine klare fordeler, det gjør det overkommelig å sette inn tiltak som effektivt kan flytte bruken fra originalene og over til andre informasjonsbærere.

Tradisjonelt har dette vært forsøkt ivaretatt gjennom mikrofilming, som har vært drevet i godt over 50 år, først av "mormonerne" og fra 1990-tallet i etatens egen regi. Mikrofilmsamlingen vår omfatter nå rundt 12 000 filmruller, og vokser med 6-700 hvert år. Det er et erklært mål å øke ressursbruken på dette feltet. På hver filmrull er det fra noen hundre til opp mot 2000 eksponeringer, hver eksponering vil normalt omfatte to sider i originalen. Det betyr at vi har fått filmet atskillige millioner sider. Siden 1995 har etaten drevet et prosjekt for å få mikrofilmet alle etatens tinglysingsarkiver. Prosjektet vil ta i overkant av 10 år, og omfatte nesten 10 millioner sider. Det har hele tiden vært en forutsetning at mikrofilmene med tinglysingsmaterialet bare er det første skrittet, og at en digitalisering av mikrofilmene er det neste.

\*\*\*

Siden begynnelsen av 1970-tallet har etaten vært med på digitaliseringsprosjekter. Det første prosjektet gjaldt 1801-tellingen, registrert ved Historisk Institutt i Bergen. Jeg har latt meg fortelle at 1801-tellingen i sin tid var Riksarkivets mest brukte kilde. Den digitale versjonen har vært tilgjengelig for publikum siden slutten av 1970-tallet, først produsert og formidlet på papir og på mikrokort, etter hvert også på nettet. Etterspørselen

etter det digitaliserte materialet har vært enorm. Originalene har fått stå i fred.

Statsarkivet i Bergen kom med i arbeidet med digitalisering av kilder i 1975. Det var jubileumsåret for norsk utvandring til Amerika, og etterspørselen etter informasjon fra emigrantprotokollene var enorm. Til noen av emigrantprotokollene for Bergen fantes det håndskrevne registre, men stort sett var man henvist til å bla og leite i originalene. Og siden den som spurte sjelden hadde presis informasjon, måtte søkene ofte gjøres gjennom mange årganger. Materialet var i ferd med å bli bladd i filler. Mikrofilm ville gjøre situasjonen bedre for originalene, men ikke i særlig grad for brukerne, som på samme måte som før ville være henvist til ørkesløs leiting.

1801-prosjektet hadde vist at det var mulig å taste inn store informasjonsmengder for en relativt liten kostnad, og at interessen for å delta i slike prosjekter var til stede. Det ble lagt planer for en registrering av rundt 100 000 poster fra emigrantprotokollene for Bergen fra perioden 1874-1924. Prosjektet var beregnet til 1,2 årsverk, og med økonomisk støtte fra blant annet Norsk Kulturråd, Nordmanns-Forbundet og Norsk Slektshistorisk Forening, ble registreringen satt i gang i januar 1975. Kalkylene holdt, for rundt 40 000 kroner og en viss egeninnsats, var materialet ferdigregistrert våren 1976. Så fulgte en lang og tung prosess med korrekturlesing og oppretting av materialet, og først i 1979 var man kommet så langt at emigrantmaterialet i form av papirutskrifter og mikrokort kunne sendes ut til institusjoner og bibliotek i inn- og utland. Filene gav også grunnlag for en rekke hovedoppgaver. Originalene, som midt på 1970-tallet var i ferd med å bli slitt ned, har siden fått stå temmelig ubrukte i magasinene, og få eller ingen har spurt etter mikrofilmer av dem.

Selv om emigrantprosjektet til tider hadde vært følt som en belastning for de involverte, hadde det også vist at det både var muligheter for og interesse for satsing på digitalisering av kilder. I 1980 ble det inngått en samarbeidsavtale mellom Historisk Institutt i Bergen, Registreringssentralen for Historiske Data ved Universitetet i Tromsø og Statsarkivet i Bergen om registrering av kirkebøkene for Bergen 1816-1880. Vi mente at dette var en sterk allianse: mellom forskermiljøer som tradisjonelt hadde vært tunge arkivbrukere, og som selv etterspurte kilder i digital form, og Arkivverket, med sine sentrale kilde-serier, med ønsket om å formidle kildeinformasjon til et bredt sammensatt publikum, og med behovet for å skjerme originalene.

Det var i utgangspunktet historikermiljøet i Bergen vi fant det naturlig å samarbeide med, men vi har i løpet av årene som har gått hatt gleden av å

levere digitaliserte kilder til jurister, teologer, etnologer, filologer og medisiner. Og vi har funnet samarbeidspartnere i flere andre land.

Selve arbeidsprosessen med digitaliseringen var i den første fasen komplisert, med skriving på skrivemaskin med strenge krav til skrifttyper og margbredder og med OCR-lesing av de maskinskrevne arkene – etterfulgt av en tilsynelatende uendelig korrekturfase, der den optiske leserens mange nykker skulle rettes opp. Etter hvert som PC-ene gjorde sitt inntog midt på 1980-tallet ble prosessene sterkt forenklet, og det ble mulig å ta styringen over registreringsprosjektene selv.

Digitalarkivet (DA) ble organisert i januar 1998 som et prosjekt underlagt Riksarkivaren. Igjen var det de gamle samarbeidspartnerne Historisk Institutt ved Universitetet i Bergen og Statsarkivet i Bergen som stod sentralt. Digitalarkivet var i utgangspunktet et prosjekt i tid avgrenset til to års satsing, seinere utvidet med ytterligere to år, fram til utgangen av 2002. Strategiplanen slår fast at DA innen utgangen av 2002 skal være etablert på permanent basis. Målet for DA er å presentere digitaliserte kilder fra Arkivverket for et bredt publikum.

Vi forventet at publikumsinteressen skulle være stor, fordi både 1801-tellingen og emigrantmaterialet fra Bergen hadde ligget søkbare gjennom Internett siden 1995, og hadde hatt stor trafikk. Vi hadde like etter oppstarten rundt 400 pålogginger hver dag, og rundt 15 000 daglige søk. Etter at vi i juni 1998 fikk legge ut materiale fra 1865- og 1900-tellingene økte trafikken betydelig, og etter hvert som tilbudene er blitt bedre kjent, og tallet på tilgjengelige kilder har økt, har også interessen økt. Vi vil i 2001 registrere i overkant av 1,3 millioner pålogginger, og rundt 100 000 søk per dag i snitt. Kostnaden med å holde tilbudet oppe 24 timer i døgnet, 7 dager i uken, året rundt, vil være godt under et halvt årsverk! Sammenligner vi dette med de tradisjonelle tilbudene, med 0,7 årsverk på hver 1 000 henvendelse, bør vi kanskje undersøke om vi ikke bør arbeide for å flytte henvendelser over til internettbaserte tilbud.

Sommeren 1995 fikk vi ved Statsarkivet i Bergen laget en undersøkelse av våre lesesalgjester og deres holdninger til oss og til sin rolle. En av de viktigste "oppdagelsene" var at lesesalgjestene i svært høy grad var fra lokalmiljøet. 80% hadde postadresse Bergen, 90% kom fra kommunene i Hordaland. Her gir statistikkene fra Digitalarkivet oss ganske andre tall. Mellom 55 og 65% av brukerne er norske, men 20-25% er amerikanske, og daglig har vi brukere fra 15-25 ulike nasjoner. Utleggingen av materiale på Internett gir oss ikke bare nye og større brukergrupper, men åpner materialet for mange som tidligere har vært avskåret fra å bruke det.

Arkivmaterialet vil ofte være vanskelig tilgjengelig for nye brukere. Den gotiske skriften vil alene representere en betydelig terskel for mange. Gjennom inntasting gjør vi denne terskelen mye lavere, og det bidrar til å gjøre informasjonen tilgjengelig for flere enn tidligere.

Digitalarkivet har to registreringsenheter, en i Stavanger og en på Voss, som arbeider med å taste inn informasjon fra kilder. De to enhetene, som samlet disponerer i overkant av 7 årsverk, ble overtatt fra Nasjonalbiblioteket i juni 1998. De gjorde seg først ferdige med arbeidet med 1865-tellingen, før de tok på seg fullføringen av emigrantdatabasene, og siden høsten 2001 har de arbeidet med inntasting av 1910-tellingen, som skal være klar for utlegging på nettet når den åpnes for bruk ved årsskiftet 2010/2011.

Riksarkivaren har gjennom avtaler med flere andre aktører (for eksempel Registreringsentral for historiske data, Tromsø, Emigrantsenteret i Stavanger, Vesterheim Genealogical Center i Madison, Wisconsin og Fylkesarkivet i Sogn og Fjordane), som har arbeidet med registrering av historiske kilder, sikret Digitalarkivet tilgang til en rekke databaser. Riksarkivaren har gjennom en avtale med DIS-Norge også fått åpnet for et samarbeid som også i fortsettelsen vil sikre god tilgang på digitaliserte kilder for utlegging i Digitalarkivet.

Digitaliserte kilder er selvsagt ikke bare inntastet materiale i form av databaser eller i form av avskrifter. Ved hjelp av skanning kan vi oppnå langt billigere "datafangst", og ved skanning av mikrofilmet materiale, kan vi raskt og ganske billig digitalisere store og viktige serier. Arkivverkets strategiplan slår fast at man i 2001 skal få laget en utredning om overføring fra mikrofilm til elektronisk medium. Utredningen, forfattet av Lars Nygaard og Svein Warberg, foreligger. Den gir status for mikrofilmingen, for de tekniske mulighetene, og den har kartlagt de økonomiske sidene. Og den gir grunnlag for å slå fast at det nå er teknisk mulig å gjennomføre skanning av mikrofilm med tilfredsstillende resultat og til overkommelige kostnader.

\*\*\*

Hvordan har utleggingen av materiale på Internett påvirket brukermønsteret ved Statsarkivet i Bergen? Vi så naturligvis for oss at det var tre muligheter:

- Vi kunne få redusert trafikk på lesesalen og i "brevbunken" fordi mange av brukerne selv kunne finne informasjonen de etterspurte på Internett.



- Vi kunne få økt trafikk, fordi langt flere mennesker enn tidligere ble oppmerksomme på oss og den informasjonen vi forvaltet.
- Eller kanskje ville det ikke ha noen målbar effekt overhodet.

Ved Statsarkivet i Bergen hadde vi hatt en svært sterk vekst i tallet på henvendelser fra 1990 til 1997. Tallet på lesesalsgjester var gått opp fra 3500 i året i 1990 til 5500 i 1997, og tallet på henvendelser om kopier, avskrifter, attester eller med spørsmål til oss var økt tilsvarende. Det eneste vi så reduksjon på, var ressurstilgangen. Lesesalen, med plass til 18-20 gjester, var blitt så rent for liten, med daglig gjennomsnittsbesøk på rundt 30 personer. Og tallet på arkivvolumer framfunnet til lesesalen steg mot 13 000 i året. En ytterligere vekst på toppen av dette kunne ha gitt oss alvorlige problemer.

Det er derfor med stor tilfredsstillelse vi kan vise tallene og kurvene for lesesalsbesøk og utleverte arkivstykker (se vedlegg). Umiddelbart etter etableringen av Digitalarkivet så vi nedgang, etter 8 år med sammenhengende vekst. Og da 1998 var omme kunne vi slå fast at tallet på lesesalsgjester hadde falt fra 5582 i 1997, til 4867 i 1998. Utviklingen har fortsatt, med reduksjon til 4184 lesesalsgjester i 1999, til 3701 i 2000, og til 3585 i 2001. Enda sterkere har tilbakegangen i antall utleverte arkivstykker vært, fra 12 671 i toppåret 1996, til 7528 i 1998, og til 3802 i 2000. Når det gjelder tallet på utleverte volumer er det ikke bare Digitalarkivet som kan forklare utviklingen. Nye og bedre mikrokort, flere kilder i kopiert form, og flere kildeutskrifter på lesesalen står for en del av reduksjonen.

Vår erfaring så langt har altså vært at vi har fått flyttet trafikken bort fra lesesalen, og i enda sterkere grad bort fra originalene, og over på digitaliserte medier, først og fremst gjennom Internett. I 2000 hadde 3,5 millioner søk i de databasene vi hadde registrert og lagt ut i Digitalarkivet basert på kilder i Statsarkivet i Bergen. Vi mener at vår egen satsing på digitalisering gjennom snart 30 år har vært fornuftig, i tråd med de visjoner og mål etaten har satt opp, og at det i et driftsøkonomisk perspektiv har vært særdeles rentable investeringer, og der investeringene vil gi økende avkastning i årene som kommer.

\*\*\*

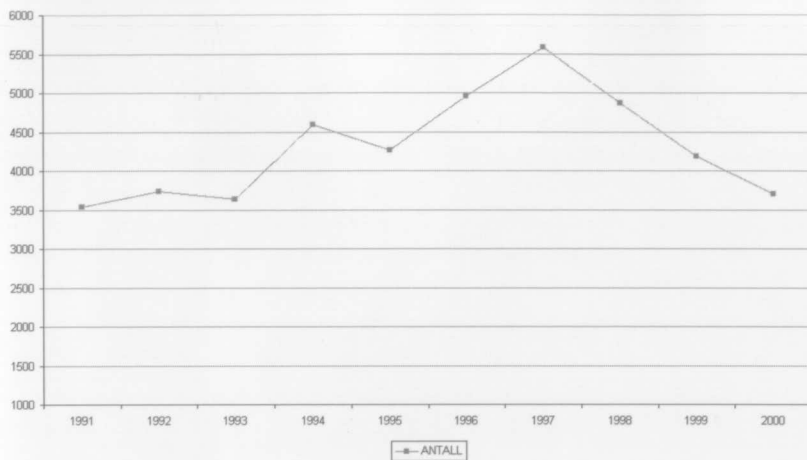
Skal jeg forsøke å trekke trådene sammen får det bli slik: Arkivverket har gjennom snart 30 år vært engasjert i ulike digitaliseringsprosjekter. I utgangspunktet skjedde dette i et samarbeid med forskermiljøer, men fra Arkivverkets side var det også et ledd i en strategi for å flytte bruk bort fra

slitasjeutsatte originaler. Etableringen av Digitalarkivet har vist at vi kan nå langt større brukergrupper enn de vi tradisjonelt har hatt. Vi kan gi tilbud til store brukergrupper som tidligere har hatt liten eller ingen mulighet til å gjøre bruk av materiale som ligger hos oss.

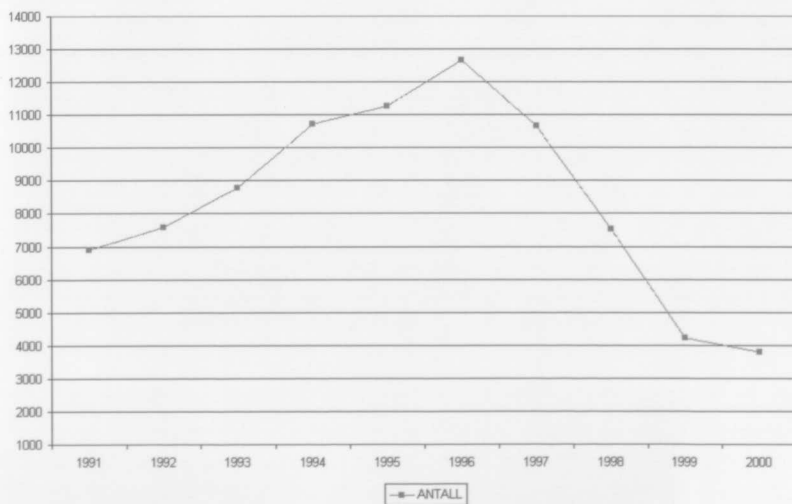
Jeg var invitert til å rapportere fra den norske digital-jungelen. Jungelen er kjent for sin store artsrikdom, sin tette underskog og sin eksplosive vekstkraft. Måtte det da avslutningsvis være meg tillatt, med en smule uærbødighet, å innrømme at jeg i min tid i Arkivverket oftere kan ha følt at jeg har befunnet meg i en oase i en ørken, enn i en jungel.

## Vedlegg:

Lesesalsbesøk 1991-2000



Utleverte arkivstykker 1991-2000





**Del II.**

**Gjør vi de riktige ting –  
gjør vi ting riktig?**

## **Gjør vi de riktige ting? Et brukerperspektiv på digitalisering av arkivmateriale**

Digitalisering av arkivmateriale er et svært aktuelt spørsmål for historiestudiet på universitetene. Særlig har den pågående studiereformen av universitet og høyskoler (Kvalitetsreformen) aktualisert behovet for digitalisering. Reformen, som ble satt i verk i 2003, tar som kjent sikte på å legge studiene bedre til rette, slik at studietiden kan kortes ned, studieintensiteten økes og at studentene dermed får en mulighet til å gjennomføre studiene på normert tid. Samtidig lanseres reformen som en kvalitetsreform: kvaliteten av studiet skal ikke bli dårligere, selv om studietiden skal ned.

Hvordan vil dette slå ut for historiefaget? Det nye masterstudiet vil være normert til to år, likt dagens hovedfag, men kravet til gjennomføring og studieintensitet vil bli høyere – om ikke på annen måte, så ved at studiefinansieringen vil oppmuntre til det. Jeg tror de fleste vil være enige om at dagens hovedfag er en svært god utdanning. Trening i kritisk analyse sammen med konstruktiv forskning som ender i et selvstendig arbeid, er verdifull utdanning i et kunnskapssamfunn. Ikke minst i dagens skole er lærere med denne kompetansen nødvendig for å gi de prosjektbaserte, problembaserte arbeidsmetodene i skolen innhold og retning. Utfordringen for instituttene, men også for Arkivverket, er derfor å beholde hovedoppgavens karakter av å være forskning, men gjøre det mulig å gjennomføre på to år.

Jeg tror samarbeid med Arkivverket om digitalisering av kilder kan være ett fornuftig bidrag til dette. Ikke fordi vi skal slutte å sende studentene til arkivene, men fordi en ved digitalisering av kilder kan legge kildene bedre til rette for studentene. Vi ser det allerede, for eksempel i hovedoppgavene som skrives om tidlig nytid historie: Ved hjelp av digitaliserte folketellinger skaffer studenten seg raskt oversikt over befolkningen i området som skal studeres, og går så vel forberedt løs på de øvrige kildene.

Også på de lavere trinnene i historiestudiet ser vi hvordan digitaliserte kilder fint lar seg passe inn i nye studieopplegg med varierte arbeidsformer, der studentene allerede fra første semester arbeider både med primærkilder og ulike historiske teorier.

For å oppsummere så langt: Når Arkivarforeningen stiller spørsmålet om hvilken målgruppe en skal digitalisere for, så er mitt svar at det er studenter og forskere som stadig blir en viktigere målgruppe for digitalisering.

## Hvordan skal en digitalisere?

Selv har jeg vært begeistret for fritekstsøk i store tekstmengder. Jeg mener det er fascinerende at gamle middelalderdokument og tingbøkene fra tidlig nytid kan skummes igjennom av elektronikken på jakt etter ord og uttrykk. For den øvede bruker er dette et utrolig flott hjelpemiddel, som jeg tror vi ennå ikke helt har forstått hvilke muligheter det kan gi. Og i forhold til studiereformens vekt på fleksibilitet er det helt riktig verktøy. Det er bevist at det er mulig å sitte i Saragossa og skrive hovedoppgave basert på de norske middelalderdokumentene som ligger på nettet!

Det er like fascinerende hvordan strukturering av data i de digitale databasene på nettet gjør det mulig at en raskt kan skaffe seg oversikt. Ved oppslag i Digitalarkivet og Registreringssentralen for historiske data får en svar på spørsmål en tidligere ville brukt uker på!

Etter mitt skjønn er det nødvendig å holde fram med å legge til rette for både fritekstsøk i ustrukturerte data og mer strukturerte data. Det er god investering, og bruksmulighetene er uante. Jeg tror også det er helt riktig å bruke nettet som et felles arbeidssted, åpent for mange typer brukere og tilrettelagt for mange ulike interesser.

Utover dette kan en tenke seg flere måter å tenke digitalisering av kilder:

I tillegg til databasene og omtale av kildene kan en også legge ut "flate filer", det vil si avskriftene uten søkestruktur, bare rene tekstfiler samt bilder av kildene. Dette letter kildekritikken, som er så nødvendig del av kildestudiet, og avmaktfølelsen vi får fordi vi ikke helt vet hvordan datane er blitt puttet inn i "den svarte boksen", blir trolig mindre.

Dessuten mener jeg at vi må se fram i tid. Dette er et felt der nye muligheter stadig åpner seg. Verken fritekstsøk eller Internett er særlig gamle! Ser en framover, er det først og fremst muligheten for raskt å overføre bilder på nettet som er det nye. Og her burde en undersøke mulighetene til å skanne eller fotografere viktige eldre arkiv.

Arkivkatalogene som nå er tilgjengelige på Arkivverkets hjemmesider kan også bygges ut med bilder eller databaser av kildene. Slik kan en få en systematisk illustrasjon av hvilke kildetyper arkivet har.

## Hva skal en digitalisere?

Arkivverket har mange brukergrupper som dels kan ha ulike interesser. Mens slektsforskeren først og fremst er interessert i navn, er historikeren kanskje mest interessert i gruppen. Slektsforskeren kan være svært fornøyd med en kirkebokbase der det er lett å søke på navn, mens historikeren som vil studere spedbarnsdødelighet i kirkebokbasene ønsker seg en kode-funksjon der de ulike innførlene om alder, blir gjort om til år. De ulike interessene er først og fremst en *utfordring* for Arkivverket, ikke en uløselig konflikt.

Hva skal en digitalisere? Mitt opplagte svar på dette spørsmålet er: Hold stø kurs! Det gjelder å utvide allerede eksisterende baser, fortsette med å digitalisere tingbøker, kirkebøker og annet for å hindre slitasje på originaldokumentene. Men dessuten kan en prøve noe nytt, for eksempel å digitalisere bilder. Og som første prosjekt ville jeg foreslått å undersøke om det var mulig å legge ut på Internett noen "mormoner-filmer" av de gamle tingbøkene. Både filmene og leseapparatene finnes bare på lesesalen i Riksarkivet, slik at flere brukere kunne få tilgang til disse verdifulle kildene ved en publisering på Internett. Det jeg er usikker på er om kvaliteten på filmene er god nok til digitalisering, og om basens struktur ville blitt for stor og tungvinn.

Tenk om for eksempel Statsarkivet i Tromsø la ut en filkatalog som inneholdt alle tingbøker som er filma helt fram til 1800, med muligheten til at en kunne bla seg fram side for side:

### **Statsarkivet i Tromsø**

Sorenskriveren i Senja

Tingbøker

- 1730 - Astafjord 1.1, f. 1-2b (bilder av de to sidene
- Sand 2.1. f. 2b-3 (bilder av de to sidene) osv.
- 1731 osv.

For øvrig mener jeg at vi historikere må bli flinkere til å synliggjøre at vi har bruk for digitalisering av kilder i søknader om forskningsmidler til Forskningsrådet og andre instanser. Digitalisering av kilder åpner for at forskningsprosjekt kan ha egne web-sider der kildene presenteres. Men for systematikkens og kildekritikkens del bør kildene selvsagt ligge under arkivet der de hører hjemme. Fra forskerens web-side går det lenker til arkivet.



Jeg vil skissere et framtidsscenario. Scenarioet begynner med at jeg skriver om en sak fra kommerseskollegiet. Fra Norges Forskningsråd har jeg søkt om data-assistanse på kr. 30 000,-. Midlene brukes til å lønne en ansatt i arkivet i noen uker. Denne har ansvar for skanning eller fotografering og elektronisk tilrettelegging. Det digitaliserte materialet legges under Riksarkivets filkatalog, slik at en systematisk kan klikke seg dit. Jeg har en direkte henvisning fra min hjemmeside til disse kildene. Jeg kan senere bruke dette i undervisning og nasjonalt og internasjonalt forskningssamarbeid der en har bruk for å kikke hverandre i kortene.

Samtidig ser jeg innvendingen: Det kan lett bli enorme digitale kildemengder som er lite i bruk. Her må Arkivverket og historikere sammen vurdere hva som er kostnadseffektivt å satse på.

## **Gjør vi ting riktig? Erfaringer fra Dokumentasjonsprosjektet/ Museumsprosjektet**

I invitasjonen til foredraget i Arkivarforeningen ble vi spurt om å snakke under tittelen "Digitalisering – gjør vi det riktig?". I denne artikkelen vil vi beskrive hvordan vi har arbeidet med å bygge opp samlinger av digitale bilder av ulike slag. "Vi" er i denne forbindelse kjernegruppen i "Dokumentasjonsprosjekt" (1991-1997) som har fortsatt i universitetsmuseenes "Museumsprosjekt" (1998-2005). Begge prosjektene hadde og har som mål både å konvertere tekst og bilder til elektronisk form samt å lage databaseløsninger for både spesialistene og for folk flest.

### **Arbeidet med fotografisamlingene ved museene**

Arbeidet med fotobevaring kom for alvor i gang i Norge på midten av 1970-tallet. I 1976 kom det en innstilling om registrering og bevaring av fotografier (Norsk kulturråds fotoutvalg). Gjennom 1980-tallet ble det lagt ned mye arbeid fra Humanistisk datasenter i Bergen og Sekretariatet for fotoregistrering (SFFR, opprettet i 1977) ved Riksarkivet for å utvikle en (edb)-registreringsstandard for fotografier. Dette konkretiserte seg i forslaget til feltkatalog for norske kunst og kulturhistoriske museer (NKKM, Jon Birger Østby 1992). På 1980-tallet kom også museumssystemet Regimus, senere WinRegimus. Dette har moduler for fotoregistrering og hadde på begynnelsen av 1990-tallet en Apple/MacIntosh-versjon som kunne lagre katalogbilder. Norsk Folkemuseum utviklet sitt eget magasinsystem som i annen halvpart av 1990-tallet ble videreutviklet gjennom PRIMUS-prosjektet til et svært godt museumssystem med en fullverdig fotomodul.

I 1993 begynte en også i Dokumentasjonsprosjektet å arbeide med fotografisamlingene. Vi begynte med en ren registrering av de arkeologiske fotografisamlingene ved Tromsø museum og Oldsaksamlingen i Oslo. Da vi digitaliserte nynorskarkivet på denne tiden, ble sedlene lagret som ren svarthvitt uten gråtoner og sterkt komprimert (4 KBytes per seddel). Det var meningsløst å lage digitale fotografier av en så dårlig kvalitet. Vi bestemte oss derfor bare å skrive informasjonen som stod på fotolistene/kartotek-kortene. Dette svarte til hva som var praksis ved andre museer, vel med unntak av Norsk Folkemuseum som allerede da hadde begynt med større mengder digitale foto både selv og i regi av Nasjonalbibliotekavdelingen i

Rana (Bauta-prosjektet som ble til GalleriNor). Referansedataene ble skrevet inn i DBase og Access. Disse databasene var en erstatning for de manuelle arkivkortene, men ga ingen muligheter til å se bildene på skjermen. For den vanlige bruker føltes nok dette som et savn, men fotografene ved museene var generelt sett skeptiske til kvaliteten på digitale bilder. De mente vel at det ikke var verdt kostnadene å lage store samlinger med dårlige digitale bilder. Noe som må sies å være en fornuftig holdning.

Det neste foto-delprosjektet ved Tromsø Museum var å lage en database over fotografisamlingene ved Avdeling for nyere kulturhistorie (NyK) og Samisk etnografisk avdeling (SE). Dette delprosjektet ble startet i 1995. Fagpersonalet ved avdelingene vil gjerne ha digitale bilder sammen med referanseinformasjonen slik at en kunne få et inntrykk av selve bildet. Den teknologiske utviklingen var da kommet så langt at det var mulig å lagre større mengder digitale bilder av det en kaller katalogkvalitet (i motsetning til arkivkvalitet). Digitale bilder med arkivkvalitet skal ha så god gjengivelse at de er mulig å bruke som erstatning for det originale bildet, mens katalogkvaliteten er bedre enn ren gjenkjenning, men kan ikke brukes til reproduksjoner av annet enn i beste fall mindre formater. En arkivkvalitet krever at dias i glassrammer åpnes og overflaten renses for støv. Det hadde vi ikke ressurser til og fagpersonene mente det ikke var realistisk, jf ovenfor. Delprosjektet baserte seg på katalogkvalitet, men med høy oppløsning og store bildefiler. Fotografiene i samlingene var for det meste svart-hvitt-negativer og dias som ble skannet i en diasskanner ved museet. Men i tillegg ble det skannet inn (kontakt)kopier som var klebet opp på arkivkortene. Selve innskrivningen av referanseinformasjonen ble utført av personer på arbeidsmarkedstiltak ved en registreringssentral i Fauske. Frem til 1997 da Dokumentasjonsprosjektet ble avsluttet, ble det skannet inn og registrert rundt 40 000 fotografier. Referanseinformasjonen ble skrevet inn i en egenprodusert Access-database. Vi brukte altså ikke daværende NKKMs feltkatalog eller fotodatabaser som WinRegimus. I ettertid viste det seg uproblematisk å flytte dataene over i en database som var i samsvar med feltkatalogen.

Det er imidlertid andre forhold rundt registreringen som kan være verdt å bemerke. Da delprosjektet ble utformet, hadde fagavdelingene liten kunnskap og erfaring med å etablere databaser på bakgrunn av eksisterende samlinger. Det var nok derfor en viss grad av "nå starter vi frisk på ny med en ny samling"-innstilling. En ønsket å bruke noe av den gamle informasjonen og supplere den med ny. Blant annet ble registratorene bedt om å beskrive stikkordsmessig hva de så på bildet. I tillegg ble motivet

forsøkt kodet med temakoder tatt fra den norske varianten av "Outline of cultural material", en hierarkisk kodeliste mye brukt av antropologer verden over. Outline-kodingen skal i prinsippet gjøre det mulig med språk-uavhengige søk. På den måten kan forskere verden over søke i bildematerialet. Det viste seg at kodelisten fungerte dårlig for detaljert motivbeskrivelse av bilder fra fiske og sjøfart. Som et eksperiment prøvde en også med dobbelt registrering av bildene. To registratorer beskrev uavhengig bildet både med outline-koding og stikkord. En slik dobbeltregistrering med påfølgende maskinell kontroll av uoverensstemmelser hadde fungert svært godt på annet materiale. Men det viste seg temmelig fort at motivbeskrivelse er en svært subjektiv oppgave. To mennesker vil sjelden gi den samme beskrivelse av et og samme bilde. Det er kanskje ikke et overraskende resultat, men det betyr at enhver motivbeskrivelse i en bildedatabase i bunn og grunn vil være noe tilfeldig.

Som nevnt over, ønsket fagavdelingene å bruke deler av den gamle informasjonen og supplere den med ny. Det var liten forståelse for å ta vare på all den gamle referanseinformasjonen på kartotekkortene. I den nye databasen er det derfor ikke mulig å si om informasjonen i feltene er original fra kortene eller stammer fra registreringsprosessen. For den daglige alminnelige bruk av fotodatabasen er dette kanskje mindre viktig. Men det setter i praksis startdatoen for den digitale fotosamlingen til 1995. Det vil vanskeliggjøre faghistorisk arbeid på samlingen med utgangspunkt i referanseinformasjonen. Situasjonen er ikke så gal som den kunne ha vært siden vi hadde utstyrt registreringssystemet med et ekstra felt for utfyllende informasjon der registratorene skrev inn den teksten på kortene som de ikke fant plass til ellers. Dette feltet har i ettertid vist seg svært nyttig for semi-automatisk uttrekk av informasjon om motiv og fotograf.

Vi har av og til fått kritikk fordi vi ikke brukte et registreringskjema som var basert på NKKMs og senere NMUs feltkatalog. Ideelt sett bør en holde seg mest mulig til standarder, og vår nyeste fotodatabaser er kompatible med den nåværende feltkatalogen. Det hadde imidlertid ikke gjort noe fra eller til med hensyn på det å ta vare på den gamle originale informasjonen på kortene. Her er det nok mer snakk om en bevisstgjøring av personalet som arbeider med samlingene og ikke minst ressurser. Gamle fotolister og kartotekkort oppfattes nok som svært sekundært i forhold til selve fotografiene. Men vi vil på det sterkeste anbefale at originale lister og referanseinformasjon tas med inn i den nye databasen så langt råd er. Med dagens billige datautstyr vil vi foreslå at en som et minimum skanner inn og lagrer elektroniske faksimiler av originalforeleggene. Disse faksimilene kan lagres i databasen i tilknytning til bildene. Dette kan også lette registrerings-

arbeidet i det innskrivningen kan gjøres utenfor avdelingen ved at faksimilene hentes opp på nettet eller sendes med på CD.

Da Dokumentasjonsprosjektet ble avsluttet ved utgangen av 1997, ble det et opphold i arbeidet med bildedatabasen ved Tromsø Museum. Arbeidet ble tatt opp igjen i Museumsprosjektet i perioden 1998-2000. Det ble laget et nytt og forbedret databasesystem som vi vil gi en beskrivelse senere i artikkelen. Hovedforbedringene er at systemet nå er fullstendig nettbasert. Bildene skannes i Tromsø. I prinsippet kunne de så sendes over nettet til den felles databasen i Oslo. I praksis har det vist seg at nettet er for dårlig til at dette er hensiktsmessig. I praksis har derfor bildene blitt lagret på CD-er og så sendt med posten til Oslo. Her legges bildene på CD-ene inn i databasen. Selv innskrivningen av referanseinformasjonen og motivbeskrivelsen ble gjort ved skrivestuer i Karasjok og Kautokeino. Innskriverne henter opp bildene som er lagret i Oslo på skjermen og skriver inn informasjonen i registreringsskjemaet. Korrekturlesingen gjøres i Tromsø ved hjelp av det samme databasegrensesnittet. Dette er et av utallige eksempler på at Internett fjerner avstander. Diassamlingen fra utgravningen på Bryggen i Bergen registreres på samme måten i et annet delprosjekt i Museumsprosjektet.

## **Fotodatabasen**

Vårt arbeid med fotografier var i utgangspunktet ikke fotobevaring. Delprosjektet i Tromsø ble startet for å gi en bedre oversikt og tilgang til fotosamlingen. I 1995 ønsket avdelingene, som nevnt, bare et lite katalogbilde for å lette gjenfinningen. At vi i 2002 har et svært stort antall digitale fotografier av relativt høy kvalitet, skyldes den tekniske utviklingen. Det er i Tromsø ikke gjort noen tiltak for å restaurere eller forbedre fotografiene i vår regi. Fotodatabasen fra 1995 ble derfor laget som en ren registreringsdatabase som var det vanlige på den tiden. I 1998-2000 ble den forbedret til også å kunne håndtere store digitale bilder. I Bergen hadde delprosjektet et annet utgangspunkt. Digitaliseringen av diasene fra Bryggen-utgravningen er en redningsoperasjon for dias som falmer og vil være svært forringet om noen år. Men fotodatabasen i 2000 var fremdeles en database for registrering av informasjon om vanlige fotografier med muligheter for lagring av ulike digitale versjoner av fotografiet.

Datamodellen som ligger til grunn for databasen forutsetter at det er fysiske bilder i en verden utenfor datamaskinen som det skal lagres informasjon om. Digitalt skapte bilder som aldri har hatt en eksistens utenfor datamaskiner var ikke tatt med. Dette er også i dag den mest vanlige måten å modellere

kulturhistorisk orienterte bildedatabaser. Men det er åpenbart at vi nå står foran eller faktisk er midt inne i en massiv omlegging fra kameraer med vanlig film til kameraer med digital lagring av bildene. I april 2002 hadde salget av digitale kameraer gått forbi salget av kameraer med tradisjonell analog film (Aftenposten 11/4-2001). Det er ikke urimelig å anta at om et par år vil digitale bilder være regelen mens ordinær analog film vil være unntaket.

Høsten 2001 besluttet vi derfor å lage en ny datamodell og nytt databasesystem for bildebasene våre. En forenklet fremstilling av modellen for databasen er vist i figuren (se figur på side 36). I denne nye databasen skiller vi ikke mellom bilder som har digital eller analog opprinnelse. Et vanlig digitalt stillbilde kan sees som et spesialtilfelle av levende bilder (digital video). Tilsvarende kan lydopptak tenkes som video uten bilde. Vi har derfor introdusert et databasebegrep vi kaller "medieobjekt" som favner alle opptakstyper av "virkeligheten" med unntak av tekst. Den øverste stiplede rammen i figuren omfatter de delene av databasen som har med selve medieobjektet (bildet) å gjøre. Den nederste stiplede rammen omfatter den delen av databasen som holder klassifikasjonen av bildet og motivbeskrivelsen. De tre andre boksene angir den felles persondatabasen og den felles stedsdatabasen. I figuren er det også satt inn en boks for en generell gjenstand eller museumsobjektdatabase for å indikere at et bilde kan knyttes direkte til en gjenstandspost som for eksempel en illustrasjon.

Medieobjektdelen er den delen av databasen som skiller seg mest fra en vanlig fotodatabase der en vanligvis bare har med referanseinformasjon om et analogt fotografi og en eller flere digitale replika av det. Denne delen kan muligens virke litt abstrakt og trenger vel derfor en ekstra forklaring.

Termen medieobjekt er ment å dekke ikke-tekstlige ting slik som analoge og digitale fotografier, faksimiler, levende bilder og lydopptak, altså alle slags ikke-tekstlige avbildninger. En tegning eller et maleri sett som bilde og ikke som gjenstand vil kunne gå inn her. Hvor langt vi skal strekke begrepet er for tiden noe uavklart, men det er ikke av betydning for å håndtere fotografier. Et negativ i en fotosamling vil være et medieobjekt. Til medieobjektet vil en derfor knytte oppbevaringssted, tilstand, eier og lignende. Et digitalt bilde på en CD eller på en disk i en datamaskin vil også være et medieobjekt. Det samsvarer vel med termen 'eksemplar' i Åndsverkloven. I databasen vil det kunne være knyttet en fil med den digitale versjonen av medieobjektet. Men siden databasen også skal kunne beskrive ikke-digitaliserte (foto-)samlinger, trenger det ikke være en slik fil.

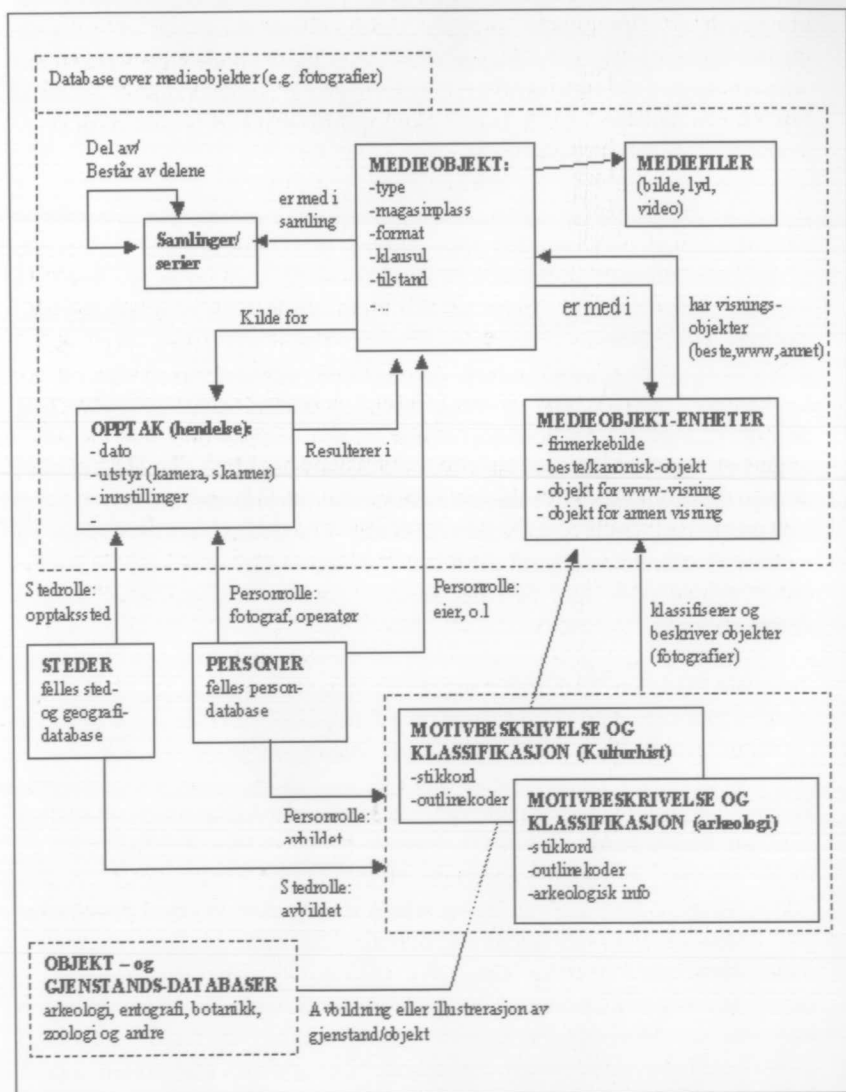
Et medieobjekt skapes gjennom en handling eller hendelse som vi noe teknisk har valgt å kalle et opptak. For et analogt fotografi vil det si prosessen som omfatter eksponering og fremkalling, mens for et digitalt vil det bare være selve eksponeringen eller innskanningen. Det å kopiere, komprimere eller utføre digitalt mørkeromsarbeide på et allerede eksisterende digitalt bilde slik at en får et nytt digitalt bilde vil også betegnes som et opptak denne tekniske betydningen. I Åndsverkloven forstand omfatter opptak det en der kaller 'eksemplarframstilling', men også selve skapelsesprosessen av originalverket.

En medieobjektenhet er en samling av medieobjekter som alle er kopier av samme original. Hva som skal ligge i "kopier av samme original", er det opp til databasebrukeren å bestemme. Vi trenger en slik abstrakt gruppering for å kunne omtale en samling av nesten like versjoner. Bildene som skannes inn i Tromsø eller Bergen, foreligger i minst fire forskjellige versjoner i databasen: original innskanning, komprimerte versjoner med høy og lav oppløsning samt et lite frimerkebilde. I tillegg kommer eventuelle fargeforbedrede bilder. For å forenkle fremvisningen på Internett og i andre brukerapplikasjoner inneholder en medieobjekt-enhet i fotodatabasen et utvalgt (beste) bilde som er en representant for alle bildene i enheten samt et frimerkebilde. Motivbeskrivelse og andre klassifikasjoner er koblet til medieobjekt-enhetene og ikke direkte til medieobjektene. En medieobjekt-enhet er nær men ikke helt det samme som en betegner som 'verk' i Åndsverkloven.

Den siste firkanten "Samlinger/serier" er tatt med for å kunne lagre alle slags måter å gruppere medieobjekter på. For fotografier vil dette svare til serier og samlinger slik fotografiske samlinger ofte er ordnet.

Datamodellen som er vist i nedenstående figur, er utvidelse av den modellen som en kan utlede av NMUs forslag til ny feltkatalog og som er implementert i museumssystemene PRIMUS og WinRegimus. Vi har inkludert selve fotograferingen og arbeid (manipulasjon) med fotografier (og andre medietyper) gjennom opptaks-begrepet. NMUs feltkatalog modellerer dette som en variant av akkvisisjon. Det mener vi er noe søkt og begrepsmessig feil. Datamodellen vår betrakter et medieobjekt som en enhet som ikke kan deles opp. Serie-/samlingsbegrepet er bare brukt for å kunne uttrykke enheter som er større enn et enkeltfotografi. En slik modell er ikke helt tilstrekkelig. Internasjonale standardiseringsforslag som MPEG-7 [REF] for multimedieobjekter (herunder digitale fotografier) opererer med klassifikasjonssystemer som gjør det mulig å dele opp multimedieobjekter

både langs tidsaksen og i deler av bilder. Når brukerbehovene melder seg, vil vi utvide databasen slik at den er kompatibel med eksempelvis MPEG-7.





## Digitalisering av foto - fotobevaring

Tidligere i artikkelen nevnte vi at vårt arbeid med fotomaterialet i Tromsø kom i stand som et ledd i en prosess for å gjøre fotosamlingen lettere å bruke. De digitale bildene ble lagt til for å lette søkingen. Det å bevare og restaurere eksisterende analoge fotografier er noe som har skjedd og vil skje uavhengig av datamaskiner. Riktignok kan en ved hjelp av digital teknologi rekonstruere ødelagte bilder lettere enn ved tradisjonelt mørkeromsarbeid. Men de digitale lagringsmediene er svært ustabile. Det beste lagringsmedium for fotografier er fremdeles analog film av høy kvalitet.

Vi har nå rundt 100 000 digitaliserte fotografier. Mesteparten ble laget som gode katalogkopier. Så lenge originalene finnes, vil det bare være en økonomisk katastrofe om de digitale bildene ble slettet eller uleselige. Men for flere titalls tusen falmene dias finnes etter hvert ikke noen original. Forsvinner den digitale kopien, forsvinner også bildet. Dette og de økonomiske omkostningene med å skanne de andre bildene på nytt gjør det viktig å finne en mest mulig sikker måte å oppbevare bildene på. I utgangspunktet ble de digitaliserte fotografiene lagret på CD. Det viser seg at særlig skrivbare CD-er ikke er spesielt stabile. Allerede etter 5 år har vi sett at en viss prosent av bildene som er lagret på CD er blitt uleselige. Etter vår vurdering er det derfor to gode måter å lagre bildene på: En kan "skrive" bildene tilbake på standard fotografisk film for så å reskanne bildene senere om nødvendig. Alternativt kan en lagre bildene på magnetbånd kombinert med en driftsorganisasjon som sørger for en omspoling og oppfriskning av båndene med noen års mellomrom. Museumsprosjektet har etter lengre diskusjoner med USIT (Universitetets Senter for Informasjonsteknologi) ved Universitetet i Oslo fått i stand et system med mangnetbånd-roboter. De digitale bildene lagres på magnetbånd på to slike lagringsenheter plassert i brannsikre rom i to ulike bygninger på Blindern.

Når fotografier i den nære fremtid kommer som digitale originaler, burde en muligens supplere den digitale lagringen med en utskrift på en analog film av et standardformat. Dette kan gjøres maskinelt og bildene kan lagres på store ruller i egne klimarom.

## **Publisering av bilder på Internett – Nyere kulturhistorie i Tromsø**

Høsten 2000 ble det bestemt at den digitale fotosamlingen til Avdeling for nyere kulturhistorie (NyK) ved Tromsø Museum skulle publiseres på Internett. Museumsprosjektets rolle var å komme med forslag til grafiske og tekniske løsninger, implementere dem og koordinere hele prosessen. NyK skulle være faglig godkjennende instans og komme med egne ønsker og spesifikasjoner i prosessen. Koordinator ved Tromsø Museum var Ketil Zakkariassen og ansvarlig hos Museumsprosjektet var Jon Holmen.

NyK hadde på det tidspunktet ingen erfaring med å lage internettutstillinger, mens Museums-/Dokumentasjonsprosjektet hadde laget flere digitale utstillinger (Norsk mynt i 1000 år og Folkemusikkautomaten), diverse publikumstjenester og hadde den nødvendige tekniske kompetansen. Den grafiske utformingen av en internettutstilling er svært viktig. Her har vi liten kompetanse. Derfor har vi for de aller fleste internettutstillingene våre satt bort den grafiske utformingen til folk som er profesjonelle digitale formgivere. Den grafiske utformingen av fotoutstillingen ble overlatt firmaet Web-veveriet.

### **Elementene i en fotodatabase på Internett**

De fleste bildesamlinger på Internett har muligheter for å lete frem bilder etter ulike kriterier, vise dem som frimerkebilder og i et større format. I de senere årene har det også blitt mulig å lagre bilder en har funnet i "personlige album" og også å bestille kopier av bildene via epost. Det ble derfor besluttet at nettversjonen av NyKs fotosamling skulle ha disse funksjonalitetene: Skjemasøk, visning som frimerke og mellomformat, lagring av brukernes bilder i album og bestilling av bilder over nettet.

### **Søking**

Alle bildene i databasen er, som nevnt over, blitt klassifisert med fotograf, datering, sted, motivbeskrivelse og annet. Et skjema med alle disse søkemulighetene vil bli ganske stort og virke noe avvisende på en utrent bruker. Vi bestemte derfor å lage et enklest mulig standard-søkeskjema med feltene motivbeskrivelse og sted. I tillegg introduserte vi et innstillingsskjema der brukerne selv kan krysse av for hvilke søkefelt de vil ha frem. Dermed kan en avansert bruker lage sitt eget søkeopplegg. For å muliggjøre lagring i album måtte vi også lage et system der en bruker kunne

opprette sin egen brukerkonto. En registrert bruker kan også lagre sine innstillinger fra gang til gang.

## **Visning**

Det er selvsagt at brukerne skal kunne få se resultatet sitt som et slags lysbord med frimerkebilder. Databasen inneholder over 50 000 bilder og det er ingen begrensning på antall bilder en kan få tilslag på. For ikke å kvele brukerens maskin og nettet, vises tolv bilder av gangen. Ved å klikke på et bilde får en mellomformatet og også muligheten til å se all informasjon som er lagret om bildet. Dette er standard for bildedatabaser. Databasesystemet kan levere mer. Brukerinnstillingene gjorde det derfor mulig å la brukerne selv bestemme hvordan fremvisningen skal være. I prinsippet kan brukeren dermed bestemme det meste med hensyn til resultatets utforming. For NyK-databasen åpnet vi for at brukerne selv kunne bestemme om det ved siden av lysbordene med frimerkebilder skulle komme en statistikk som fortalte hvordan bildene i søkeresultatet fordeler seg på ulike kategorier som perioder, årstall, steder, kommuner med mer. Hvilke kategorier som skal vises, bestemmer brukeren selv ved å krysse av i innstillings skjemaet. Ved å klikke på en av kategoriene, for eksempel en bestemt tiårsperiode, blir bildeutvalget automatisk ytterligere begrenset til den kategorien. Denne typen fordelingsstatistikk har vi ikke sett i noen annen norsk bildedatabase på Internett.

## **Album**

En registrert bruker kan også lagre bilder i album som vedkommende kan hente frem siden. En bruker kan opprette så mange album en vil. Det er heller ingen begrensning på antallet bilder i et slikt album.

## **Bildebestilling**

Bildedatabasen fungerer også som en elektronisk butikk. Brukerne kan plukke ut bilder fra søk og legge dem i en "handlekurv". Deretter kan bestillingen sendes til Tromsø Museum som en epost. Personalet ved museet sørger så for å ta kontakt med bestilleren for å ordne med betaling og lignende. Vi har ikke introdusert noen mulighet for bruk av kredittkort eller annen elektronisk betaling. En slik tjeneste er ikke vanskelig å gjøre rent teknisk. Men det forutsetter at museet har rutiner for å håndtere denne.

## Erfaringene med fotodatabasen etter et halvt år

Fotodatabasen ble offisielt åpnet 1. oktober 2001. Museet hadde fått god dekning i lokalprosessen og interessen var svært stor. Dagen etter åpningen var det rundt 120 000 filforespørsler mot databasen. Det skapte visse tekniske problemer på våre datamaskiner og hos USIT. Som mange andre hadde vi undervurdert nyhetens interesse. To uker etter den offisielle åpningen hadde antallet forespørsler sunket til rundt 50 000 i døgnet. To uker etter åpningen ble det vist et innslag om bildedatabasen på NRK2s faste program Norge i dag. Det resulterte igjen i en voldsom vekst i antallet forespørsler helt opp til 400 000 i døgnet. Denne gangen kollapset ikke systemet men det gikk veldig langsomt. Utover høsten 2001 sank antallet forespørsler til rundt 40 000 per døgn. Det var ved juletider 2001 omlag 600 registrerte brukere.

I april 2002 er antallet registrerte brukere steget til 747 og det er i gjennomsnitt 13 000 filnedlastninger i døgnet. Det vil si at det er i rundt 1000 søk mot databasen i døgnet. Norsk Folkemuseums bildedatabase har omtrent 2000 søk i døgnet. Det er vanskelig å gi noe eksakt antall brukere siden mange av forespørslene kommer via såkalte "proxy-server". Disse mellomlagerer sider og anonymiserer dem som sitter på den andre siden. De ti første dagene i april var det 415 forskjellige maskinadresser som spurte mot bildedatabasen. For de fleste enkeltdagene er det 60-70 forskjellige maskiner som spør. Det betyr at det er en rekke gjengangere eller faste brukere av tjenesten.

Tallene viser at en ny tjeneste som blir markedsført godt, får en svært stor oppmerksomhet til å begynne med. Fotografier av ulik kvalitet fra Nord-Norge ser imidlertid ikke ut til å ha samme appell som en dagsavis. Det er imidlertid mange måter å registrere brukere på. For eksempel Digitalarkivet ser ut til å telle en bruker for hver gang en går til forsiden. Dette er en tellemåte som gir et kunstig høyt brukerantall. Dagens bruk av fotodatabasen ved Tromsø Museum er på den annen side svært god i forhold til den manuelle bruken av arkivet. En kan forestille seg hvordan det skulle vært om det kom 70 personer hver dag til Tromsø Museum for å lete i fotosamlingen!

## **Gjør vi ting riktig? Kvalitetssikring av digitaliserte data**

”Gjør vi ting riktig?”, var spørsmålet Arkivarforeningen har bedt meg om å svare på. Jeg skal forsøke å gjøre dette ut fra minst to perspektiver, for det ene min stilling som faglig leder og forsker ved Registreringssentral for historiske data (RHD), og for det andre ut fra mitt engasjement i internasjonale prosjekter som også registrerer og forsker på nominative massedata. Prinsipielt er det vanskelig å svare noe annet enn ”nei” på spørsmålet. Jeg kan umulig tenke meg at de fagfellene, som etter hvert vil ta over etter oss, ikke vil utvikle nye metoder, gjøre mangt annerledes og være kritiske til hvordan vi har forvaltet vårt pund. Dessuten er det lett å finne prosjekter som har anvendt ressursene annerledes og som vektlegger andre målsettinger når de gjør historiske mikrodata tilgjengelige for databehandling. På den annen side er det mulig å hevde at vi har gjort gode ting ut fra våre forutsetninger. Mitt hovedpoeng er at vi har vært flinke til å kvalitetssikre avskriftene av kildematerialet, mens vi har vært svake på å kvalitetssikre brukervennligheten i form av kodede versjoner til bruk i forskning.

Den som overfører et historisk materiale fra original kilde til maskinlesbart medium bør ha tre hovedmålsettinger for virksomheten: Kildetro avskrifter, effektiv registrering og brukervennlige mikrodata - her nevnt i prioritert rekkefølge. Alle er enige om at formålet med å lage digitale kopier av kildene er at brukerne stort sett skal slippe å konsultere originalen. Dessuten vil vi ikke en gang finne fram til de rette personene dersom det er alvorlige feil i avskriften. Skriver vi at Anton Hansen er 55 i stedet for 65 år i 1865-tellinga, øker sjansen for at han ikke kan identifiseres betraktelig. Imidlertid koster det betydelige ressurser å kvalitetssikre registrerte persondata, og vi kan sannsynligvis alltid finne flere feil og tvilstilfeller ved å lese korrekturen ytterligere en omgang. Disse kostnadene må veies opp mot alternativ bruk, for eksempel er det for de fleste formål bedre å registrere to folketellinger med 99% nøyaktighet enn å dataføre en telling med 99,5% nøyaktighet. Vi må altså huske at kvalitetssikring er en aktivitet med sterkt fallende grensenytte. Dessuten er det en rekke feil og uoverensstemmelser mellom ulike kilder som omhandler de samme personene. Vi kan dermed omformulere vårt krav til kvalitetssikring i retning av at det gjelder å ikke øke denne andelen feil i vesentlig grad når vi overfører informasjonen i kildene til maskinlesbart format.

Her griper det tredje hovedhensynet vi må ta ved slik overføring inn. De som skal anvende informasjonen ønsker at den skal være tilrettelagt slik at den forenkler deres spesifikke bruk av kildene. De første som anvendte data-behandling på nominative persondata var statistisk orienterte samfunnsvitere. For dem var det nærliggende å lage en digital versjon av kilda som var tilrettelagt for bruk i statistikkprogramvare, og det innebar at informasjonen ble oversatt til tallkoder. Arkivarer har til vanlig det stikk motsatte behovet: For å tilby brukerne alternative kildeversjoner som reduserer slitasjen på originalmaterialet, bør den digitale versjonen avspeile kildene så bokstavtro som mulig. De siste 35 årene har dette dilemmaet vært løst ved først å lage en kildetro avskrift på maskinlesbart medium, og deretter foreta en maskinell koding for statistiske formål. Dermed har vi kunnet tilgodese begge typer brukere.

Spørsmålet er imidlertid om vi dermed fullstendig har fjernet den interessekonflikten som i utgangspunktet er der mellom ulike brukere. Dette spørsmålet har sitt utgangspunkt i to forhold, et ressursmessig og et informasjonsvitenskapelig. Det sistnevnte har å gjøre med at noen av de standardiseringsoppgavene vi ønsker å løse, gjøres så mye bedre av mennesker enn av dataprogrammer. Der hvor registreringsassistenten ser opplagte feilstavinger og forvirrede stavemåter, må maskinen fortelles tegn for tegn hvilke variabel-verdier som er de samme. For eksempel skrev folketellerne noen ganger "snedker" og andre ganger "snekker". Og enkelte navn, som Margrete, kan staves på 30 til 40 ulike vis. Mens dataprogrammene må fortelles i detalj om hvordan slikt kan standardiseres, vil en registreringsassistent kunne fjerne slike variantstavinger uten kostnader, ja med en effektiv bruk av forkortelser kan arbeid spares.

Dermed er vi inne på det andre momentet: kostnader. Sett i sammenligning med andre land kunne vi kanskje være såre fornøyd med at Norge har fullregistrert tre historiske folketellinger, et par gårdsmatrikler og betydelige deler av annet kildemateriale. Men mengden av annet materiale som det er ønskelig å registrere er enorm, tenk bare på kirkebøkene. Dessuten er historikere flest mer orientert mot tekstlig materiale, jf. den utgaven av Eilert Sundts verker som nå ligger på Internett hos RHD. Det å kvalitetssikre avskrifter er en hel liten vitenskap i seg, det er forsket mye på hva som er avgjørende for at en registrator klarer å lage en korrekt avskrift. I slike studier har det vært lagt vekt på dagsformen; det er klare forskjeller ut fra ukedag, årstid, værforhold med mer. Personlighetstype betyr også mye, noen av oss har evnen til konsentrasjon om detaljer i lengre tid, andre mister evnen til å se feilstavinger etter kort tids korrekturlesing. Mye kan gjøres for å legge forholdene til rette for et godt resultat, det er viktig med godt lys og

annen ergonomi, spesialtilpassede registreringsprogram, samt raske tilbakemeldinger fra korrekturleser til registrator. (At dette skal være to ulike personer er en selvfølge.) Imidlertid har en annen faktor vist seg å ha avgjørende innflytelse ved registrering fra historiske kilder: Kvalitet og stil på håndskriften i originalen, dokumentets utforming og hvor tydelig eventuelle kopier har latt seg framstille. Siden den feilprosenten som er funnet ved korrekturlesing av henholdsvis 1700-talls kirkebøker og folketellinger fra 1900-tallet er svært forskjellig, gir det liten mening å bruke like mye tid på korrektur av sistnevnte kildetype som på den første.

Når det for eksempel finnes 370 000 ulike yrkesbetegnelser, medregnet alle yrkeskombinasjoner og ulike stavemåter i folketellinga 1900 for Norge, blir det kostbart både å sørge for at alle variantene blir overført riktig til digital avskrift, og å foreta koding og standardisering av yrkesbetegnelsene på hoved- og byrker ifølge ulike codesystemer. Slik standardisering konkurrerer om ressursene med ny registrering av annet kildemateriale. Dette har også sammenheng med at kostnadene ved lagring og overføring av data nå er en brøkdel av hva de var da de systemene vi nå anvender i Norge ble utviklet. Det er ikke lenger urealistisk å overføre mye av innholdet i de viktigste kildeseriene til digitale bilder ved hjelp av skanning. Siden mye av materialet allerede er mikrofilmet, kan skanningen langt på vei automatiseres. Dermed kan vi tilby brukerne flere digitale produkter. For det ene et bilde av originalen, for det andre en avskrift hvor mange variantstavinger er fjernet, og for det tredje en kodet versjon for statistisk bruk. Avskriften blir lettere å søke i fordi brukerne slipper å tenke på alle de ulike variantstavinger av navn med mer som kan forekomme, mens originalens detaljrikdom kan inspiseres på dataskjermen slik at brukerne selv kan ta stilling til tolkning av utydelig skrift med mer.

For å illustrere at arbeidet med digitalisering av nominative data kan prioriteres og utføres på ulikt vis, skal vi sammenligne Norge med USA, der the Minnesota Population Centre ved University of Minnesota har verdens største samling historiske individdata i maskinlesbart format.

Norge	USA
Lokalhistorisk tradisjon	Internasjonal tradisjon
100% registrering, til dels hele landet	Landsdekkende 1% utvalg, andre land
Tidsserien 1801-1900	Tidsserien 1850 - 1990
Ufullstendig koding	Standardiserte koder
Fullstendig korrektur	Korrektur på lite utvalg
Uten konstruerte variabler	Konstruerte variabler
Genealoger, historikere, navnegranskere	Samfunnsvitere, økonomer, historikere

Denne skjematisk oversikten forsøker å oppsummere hvordan "The Census Project" ut fra forutsetninger som er forskjellige fra de norske har valgt andre mål og metoder ved dataføring av de amerikanske folketellingene. Forskningsmessig står lokalhistorie sterkere i Norge, og det har vært vanlig at profesjonelle historikere har skrevet bygdebøker og byhistorier, gjerne fra eget bo- eller fødested under mottoet "grav der du står". Selv om rikshistorie, for eksempel USAs sosialhistorie eller demografiske utvikling står i fokus, er det ikke uvanlig at historikere i USA velger å belyse den historiske utviklinga i et lite geografisk område. Men det anses som noe lugubert å skrive om eget nærområde, stedet velges fordi det har en særegen historisk utvikling (hvalfangst, slavedrevne sukkerplantasjer, innvandrere med en bestemt etnisk bakgrunn), og det er de særegne historiske fenomenene som står i fokus. Det betyr at kilder med mer spesifikt innhold enn folketellingene (USA har i liten grad kirkebøker) blir viktigere i en lokalstudie slik at en fullstendig registrering av folketellingsmateriale har blitt prioritert lavere. Fra de nominative tellingene 1850 til 1960 finnes representative utvalg av en prosent av husholdene, fra seinere tellinger har The Census Bureau også levert fem prosent utvalg. Bare én folketelling er registrert i sin helhet, og typisk nok er dette arbeidet utført på frivillig basis av mormonerkirken i Utah med slektsgranskning som siktemål. (Registreringen dekker også Canadas og Storbritannias 1881-tellinger.) Men den digitale utgaven er selvsagt interessant også for forskning, og The Census Project tilrettelegger derfor dette materialet i samarbeid med



forskere i andre land. At utviklingstrekk i hele USAs befolkning kan følges fra 1850 og framover i (utvalg fra) tretten sammenlignbare folketellinger, gjør dette materiale til en gullgrube for den store majoritet av historikere, og samfunnsforskere som ønsker å følge utviklingen i hele landet og er svært opptatt av regionale sammenligninger.

Vekten på spesifikke, historiske emner betyr også at amerikanske historikere ofte velger å skrive om lokalsamfunn eller om den rikshistoriske utvikling i andre land. For eksempel har to tredjedeler av personalet ved Department of History i Minneapolis hovedtyngden av sin forskning på andre lands historie. Dette er bakgrunnen for at The Census Project søkte om og fikk gjennomslag for en stor bevilgning til å starte bearbeiding også av folketellingsmateriale fra andre land. Hovedsiktemålet er i første omgang å redde materiale som ble digitalisert i forbindelse med folketellingene fra 1960 og framover, men som i dag oppbevares uten den nødvendige sikring mot avmagnetisering. Mens perioden dermed blir tidsrommet etter 2. verdenskrig, har prosjektet hittil inngått avtaler med Columbia, Mexico, Vietnam, Ghana, Kenya, Spania, Frankrike og Ungarn. Det forhandles også med blant annet China, Italia og Norge om overføring av representative utvalg av hushold fra de moderne folketellingene. Utvalg av så moderne datasett må selvsagt trekkes og anonymiseres slik at risikoen for at enkeltpersoner kan identifiseres blir ubetydelig. I tillegg til de moderne folketellingsutvalgene kommer et annet internasjonalt prosjekt hvor mormonermaterialet 1880-81 fra USA, Canada og Storbritannia samkjøres med tilsvarende fullstendige folketellinger fra Island og Norge i NAPP (North Atlantic Population Project).

Med visse unntak for det moderne materialet fra andre land, gjøres alle de folketellingene som bearbeides i Minneapolis, tilgjengelig for nedlasting via Internett. Den som laster ned en slik telling vil imidlertid knapt kjenne igjen originalen. Mesteparten av informasjonen i skjemaenes kolonner har nemlig blitt oversatt til tallkoder, det gjelder for eksempel kjønn, sivilstand, yrke, familiestilling og fødested. Person- og stedsnavn er riktignok bevart, men bare for materiale som er eldre enn 72 år, USAs grense for offentliggjøring av denne type persondata. Imidlertid finnes i alle tellingene en variabel som sier hvilke personer innen husholdet som hadde samme etternavn (surname similarity), noe som er nyttig når vi avgjør familierelasjoner. Dette er eksempel på en konstruert variabel, altså opplysninger som ikke kan leses direkte ut av folketellingsskjemaet, men som programvare kan lage på grunnlag av ulike felt i folketellinga. Andre eksempler er variabler som forteller eksplisitt hvilke relasjoner personene innen en familie eller hushold har til hverandre. Altså for gifte personer hvor vi kan finne ektefellen, og for

barn hvor vi kan finne hver av foreldrene; sammen med andre variabler som angir hvilken regel programmet brukte for å bestemme relasjonen. Især for familiehistorisk forskning gjør dette materialet brukervennlig idet forskeren ikke er avhengig av å bruke en klassifikasjon av familier og hushold som er gjort en gang for alle, men selv kan utforme sitt eget klassifikasjonssystem. Et annen viktig prinsipp er at de redigerte utgavene av folketellingene skal ha samme utforming for de ulike tellingsårene. Variabelen for kjønn finnes alltid i samme kolonne, og det fins en yrkesvariabel hvor snekker er gitt samme tallkode i alle tellingene.

Selv om mye av kodingen og konstruksjonen av hjelpevariabler kan gjøres maskinelt, brukes betydelige ressurser på manuelle kontroller og endringer i datasettene for å kvalitetssikre sluttproduktet. Disse ressursene har prosjektet i betydelig grad fått ved til gjengjeld å bruke mindre tid på å kvalitetssikre avskriften fra originalkilda. Etter avskrift leses korrektur bare på et lite, randomisert utvalg av de registrerte postene, og forskere som ønsker å anvende et mest mulig feilfritt materiale, henvises til å avgrense sin analyse til disse postene som er markert i den digitale versjonen. Bakgrunnen for en slik prioritering er at det amerikanske materialet i all hovedsak anvendes til statistisk orientert forskning hvor hele USA er analyseenhet. Da kan selv et lite utvalg gi reliable resultater. Samtidig er det slik at en mindre feilprosent i de registrerte postene med liten sannsynlighet vil innvirke på den feilmarginen man i alle fall må kalkulere med når man anvender et representativt utvalg. For slektsgranskere forholder dette seg ganske annerledes: Her kan en enkelt, feilaktig opplysning om en person lede helt galt av sted når genealogien settes opp. Visse typer forskning har samme problemet: En genetiker som følger en arvelig sykdom, får også motstridende resultater hvis en fremmed grein podes inn i slektstreet. Slike hensyn har veid lite hos The Census Project, de avfeies nærmest med "A warning to genealogists", mens de har vært tungtveiende i de norske prosjektene. Så anvendes da også det norske materialet i hovedsak av slektsgranskere, mens samfunnsvitere, økonomer og historikere er hovedbrukere av de digitale folketellingene i USA.

For å illustrere at kvalitetssikring av dataført kildemateriale i form av koding er nødvendig i forskning, skal vi kort se på tre anvendelser av norsk materiale innen migrasjonshistorie med tilknytning til *Innvandringsprosjektet* under ledelse av Knut Kjeldstadli. Det første er en bearbeidelse av folketellinga 1801, som undertegnede har gjort sammen med Sølvi Sogner. Siden denne folketellinga ikke inneholdt opplysninger om individenes fødesteder, var den i utgangspunktet lite egnet til å studere innvandring til Norge i tidlig nytid. For å bøte på dette hadde imidlertid

Sølvi Sogner, som skriver om innvandring i denne perioden, en idé om å i stedet anvende etternavnene. Materialet ble velvillig stilt til rådighet av Jan Oldervoll ved Historisk institutt i Bergen, og det viste seg snart at en betydelig andel av befolkninga hadde navn som kunne klassifiseres som fremmedlandske. Denne klassifikasjonen ble gjort av en vitenskapelig assistent i samarbeid med forskerne, og det ble anvendt en rekke avhandlinger om navn, oppslagsverk og konsultasjoner med navnegranskere. 8169 ulike familienavn ble klassifisert som verken patronymika eller norske gårdsnavn og gruppert etter nasjonalt opphav. Om lag 25 000 individer kunne dermed defineres med innvandrerbakgrunn ifølge etternavnet. Av disse hadde omtrent 14 000 tysk bakgrunn, omtrent 6000 dansk bakgrunn, omtrent 2000 svensk/finsk bakgrunn og resten annet, vesentlig britisk og nederlandsk opphav. Vår klassifikasjon har klare svakheter, især at vi ikke skiller mellom første generasjons og senere generasjoners innvandring og at danskene er underrepresentert på grunn av de ensartede patronymikonkikkene. At hovedstrømmen av innvandrere til Norge før 1800 gikk fra Tyskland og Danmark kan det imidlertid ikke herske tvil om. Destinasjonene var som ventet byene og kysten langs Sørlandet, mens vi var mer overrasket over den relativt høye konsentrasjonen av ikke-norske navn i det nordafjelske Norge (se figur 1).

Det andre eksemplet er hentet fra Registreringsentralens bearbeiding av 1900-tellinga. Her er variabler som sivilstand, familiestilling og fødested allerede klassifisert og kodet, mens de nærmere 370 000 yrkene først kan ferdigstilles i 2002. De omtrent 70 000 ulike fødestedene, inklusive ulike stavemåter, lot seg i de aller fleste tilfellene oversette til firesifrede tallkoder, hvorav de to første sifrene angir fylke og de to siste kommune innenfor fylket. Dessuten ble alle som var født utenlands gitt en kode for vedkommende land. Dette er til hjelp både ved bruk av materialet til statistiske formål og ved søking etter enkeltpersoner. Det siste er en del av Registreringssentralens tilbud under knappen "Avansert søk" på våre nettsider, og lar brukerne finne fram til folk født i en bestemt kommune selv om stavemåten av kommunenavnet er svært fantasifull, eller et tettsted i kommunen er angitt i stedet for kommunenavnet. Et eksempel på statistisk bruk av materialet er gitt i figur 2, hvor de nær 50 000 svenskfødte er fordelt etter bosted i Norge. Siden så mange svensker oppga "fødesteds-län", ville en avgrensning til folk med "Sverige" som fødested gi et svært mangelfullt bilde av den omfattende svenske innvandringen. Med dette kartet som utgangspunkt kan forskeren så gjøre statistiske dypdykk, for eksempel se på hvordan innvandrerne var sysselsatt i Østfold, eller i hvilken grad de hadde giftet seg norsk eller med hverandre.

Den sistnevnte problemstillingen er også bakgrunnen for mitt eksempel på bruk av folketellinga 1960. Dette materialet er stilt til rådighet av Statistisk sentralbyrå, i form av et ti prosent utvalg av hushold fra den digitale utgaven som ble laget i forbindelse med byråets opprinnelige utarbeidelse av aggregater. Den aktuelle utgaven er standardisert med kodete variabler som er helt sammenlignbare med tilsvarende bearbeidinger av folketellingene 1970 og 1980, men disse har jeg ennå ikke hatt tilgang til. (SSB har også laget en utgave hvor folk kan følges på individnivå i de tre tellingene ved hjelp av personnummer, men dette tillates ikke brukt utenfor byrået.) Når SSB stiller slikt materiale til rådighet for forskere, er det anonymisert ved at direkte identifikatorer som navn, fødselsdato og adresse er fjernet, men andre variabler i kombinasjon (sjeldne yrker, alder) kan gjøre det mulig å identifisere noen personer. Derfor er min bruk av tellinga godkjent gjennom Det datafaglige sekretariat ved Norsk samfunnsvitenskapelige datatjeneste. Liksom i de amerikanske tellingene, er det i "min" utgave av 1960-tellinga lett å lage statistikk som kombinerer data om flere ulike personer som er relatert til hverandre. Mens dette i Minneapolis ble gjort ved hjelp av pekere eller referanser (konstruerte variabler - se ovenfor), har SSB gjort det særlig lett å sammenføre opplysninger om ektefellene fordi hans variabler er lagt til i hennes datapost og vice versa. Når vi leser datafila inn i et statistikkprogram får vi dermed tilgang både til en variabel som angir personens fødested, og en annen variabel som angir ektefellens fødested. På grunnlag av den koding av fødestedsland som er brukt i SSBs versjon, kunne jeg gruppere landene i noen få, grove kategorier og lage en samkjøring av ektefellers fødested (se figur 3). Vi skal huske at innvandrerbefolkningen i 1960 var liten, og at det dermed ikke er uventet at en stor andel av immigrantene hadde norske ektefeller.

Eksemplene ovenfor skulle vise at koding og standardisering av nominative kilder er viktig for å kvalitetssikre materialets brukervennlighet i forskningen. Det er også viktig at materialet er bearbeidet etter ensartede prinsipper over tid, fordi vi for mange forskningsformål ønsker å studere den historiske utviklingen gjennom et forholdsvis langt tidsrom. Videre vil det være ønskelig om historikere og samfunnsvitere i større grad settes i stand til å samarbeide om utnyttelsen av materialet, men det forutsetter at vi lettere får tilgang til helt ut anonymiserte utvalg fra de moderne tellingene. Slike undersøkelser lar seg fortrinnsvis utføre med hele landet som analyseenhet, og med fylkene som primære, komparative enheter. Det er en stor oppgave både å kode opplysningene, konstruere nye variabler og registrere utvalg fra folketellingene 1920 til 1950 for å lage en lang tidsserie

med lett sammenlignbare tellinger i digitalt format. Disse ressursene kan vi blant annet få ved å bruke mindre tid på korrekturlesing av overføring fra originalkilde til maskinlesbar versjon.

## Kilder på Internett

The Census Project: url <http://www.ipums.org>

Registreringsentral for historiske data: url <http://www.rhd.uit.no>

## Litteratur

Goyer, Doreen S., *The Handbook of National Population Censuses. Europe*, New York: Greenwood Press, 1992.

Kelly Hall, Patricia, McCaa, Robert & Thorvaldsen, Gunnar (red.), *Handbook of International Historical Microdata for Population Research*, Minneapolis: Minnesota Population Center, 2000.

Nygaard, Lars m.fl., *Histform - Norsk standard for registrering og utveksling av nominative folketellingsdata for årene 1865-1910*, Tromsø: Registreringsentral for historiske data, Universitetet i Tromsø, 1995.

Ruggles, Steven m.fl. (red.). Historical Methods. The Minnesota Historical Census Project, bd. 28 (theme issue) i *Historical Methods*, 1995.

Ruggles, Steven m.fl. (red.), Historical Methods. IPUMS - International Public Use Microdata Samples, bd. 32 (theme issue) i *Historical Methods*, 1999.

Sogner, Sølvi og Thorvaldsen, Gunnar, "Surnames as Proxies for Place of Origin in the 1801 Census for Norway", i *Scandinavian Population Studies*, bd. 13, 2002.

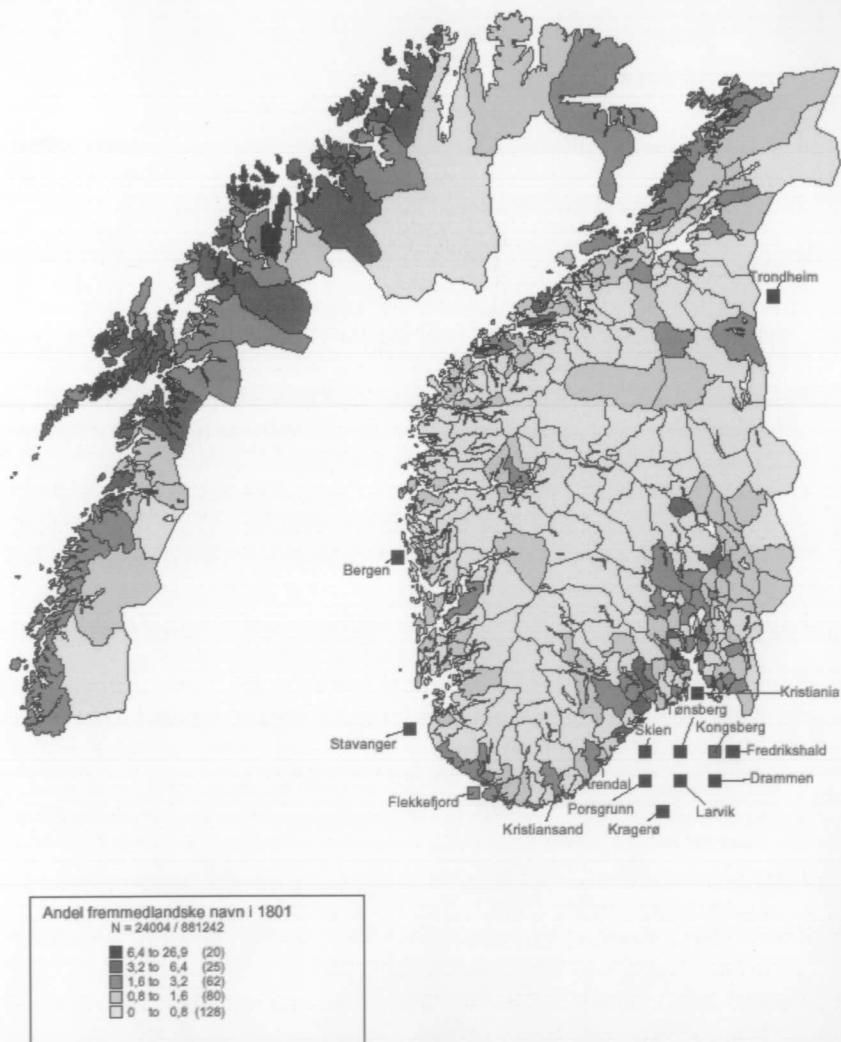
Thorvaldsen, Gunnar, The Encoding of Highly Structured Historical Source, i *Computers and the Humanities*, bd. 28, 1995.

Thorvaldsen, Gunnar, *Håndbok i registrering og bruk av historiske persondata*, Oslo: Tano Aschehoug, 1996.

Thorvaldsen, Gunnar, *Databehandling for historikere*, Oslo: Tano Aschehoug, 1998.

Vassenden, Kåre, *Folke- og boligtellingsene 1960, 1970 og 1980. Dokumentasjon av de sammenlignbare filene*, Oslo: Statistisk Sentralbyrå, 1987.

Figur 1

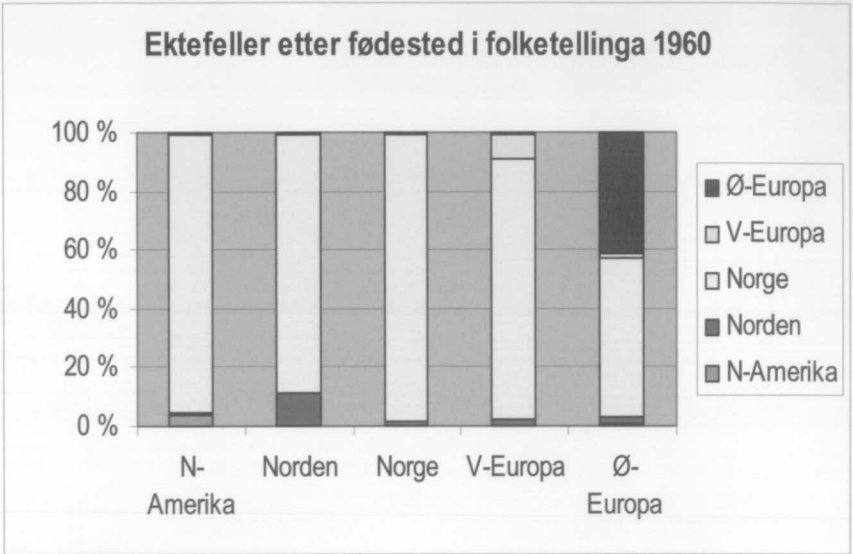


Figur 2:

Swedes in Norway in the 1900 census



Figur 3:





**Del III.**

**Debatt:**

**Har Norge en digitaliseringspolitikk?**



*John Herstad*

## **Riksarkivarens digitaliseringspolitikk**

Takk for invitasjonen. Det er i seg selv et spennende tema som Arkivarforeningen har valgt, og det er ekstra spennende fordi feltet er preget av problemstillinger som springer ut av det forholdet at vi har å gjøre med en relativt sett ny dimensjon i formidlingssammenheng.

Jeg viser til det Yngve Nedrebø sa om forskjellige aktiviteter og prosjekter i Arkivverkets regi, og skal selv innledningsvis konsentrere meg om spørsmålsstillingen: "Har Arkivverket en digitaliseringspolitikk?"

Mitt enkle svar er ja, det har vi. Vi har en politikk, og vi har faktisk hatt det en god stund allerede. Men som med alt nytt – det tar tid før en politikk manifesterer seg og gir synlige eller tydelige resultater. Og det tar tid å endre en politikk.

Jeg skal redegjøre for elementene i denne politikken, og tar utgangspunkt i Strategiplanen for Arkivverket 2001-2005 (i det følgende sitert som Strategiplan). Denne planen viser veiene og veivalgene, strategiene for Arkivverkets politikk over et bredt knippe av problemstillinger der digitalisering av kulturalrvmateriale er det sentrale. I praksis betyr dette at ni arkivinstitusjoner, Riksarkivet og de åtte statsarkivene, er blitt enige om hva som er de viktigste tiltakene sett fra vår side, når det gjelder digitalisering av arkivmateriale.

### **Digitalarkivet**

Et av de viktigste punktene i vår digitaliseringspolitikk er å bestemme Digitalarkivets fremtidige skjebne. Intensjonen vår har lenge vært å gjøre Digitalarkivet til en permanent institusjon i Arkivverket innen utgangen av 2002. Det er nå klart at Digitalarkivet blir permanent i løpet av 2002, og at det vil bli drevet fra Statsarkivet i Bergen slik det også har vært gjort i prosjektperioden. Dette vil være en garanti for at digitaliseringsarbeidet i Arkivverket fortsetter med stor kraft når det gjelder befolkningshistoriske massedata, som jo er Digitalarkivets hovedarbeidsområde.

Når nå registreringen av 1900-folketellingen og emigrantmaterialet er fullført i løpet av inneværende år, er det 1910-folketellingen som står for døren. For Digitalarkivet vil det altså i overskuelig fremtid være befolkningshistorisk massemateriale som vil stå i fokus for registreringen. Digitalarkivet har samarbeidet med mange instanser for å nå dit vi er i dag,

og den modellen vil bli fulgt også i fremtiden. For eksempel vil Digitalarkivet fortsetter samarbeidet med Registreringssentral for historiske data (RHD) om 1910-prosjektet med sikte på at hele tellingen skal være digitalisert og tilgjengelig på nettet i 2010.

I tillegg har vi inngått en avtale med DIS-Norge som sikrer at digital registrering av kirkebokmateriale skjer innenfor en nasjonal standard. Det er viktig for oss på denne måten å trekke inn frivillige krefter. I tillegg til at det medvirker til brukerstyring av prosjektene, tjener det til å øke registreringsvolumet. Derved frigjøres ressurser hos oss som kan brukes på andre viktige prosjekter som vi enten velger å stå for alene eller i samarbeid med andre institusjoner i ABM-sektoren (jf. Strategiplan, punkt 3.1 –3.2). Vi venter oss meget av samarbeidet med DIS-Norge og yter selv en betydelig innsats ved å stå for kvalitetskontrollen.

Jeg forlater nå Digitalarkivet og skal konsentrere meg om forskjellige typer av digitaliseringstiltak som vi har omtalt i vår strategiplan for årene 2001-2005.

### **Digital kataloginformasjon/Digitaliserte hjelpemidler**

Digitalisert kataloginformasjon som veiviser til arkivmaterialet er en viktig del av politikken. I løpet av årene frem til 2005 skal overordnet informasjon om alle arkivene som inngår i Arkivverkets bestand, være tilgjengelig på Internett gjennom Felleskatalogen. Det betyr ikke at de enkelte dokumentene vil være å finne på nettet, men informasjon om det enkelte arkiv som finnes i Riksarkivet og i statsarkivene. Til denne kataloginformasjonen kan det så bli knyttet databaser av ulik slag, for eksempel kartdatabaser, mikrofilmdatabaser med mer. Slike databaser eksisterer allerede og er i bruk som redskap i saksbehandlingen i våre institusjoner. Flere slike databaser er under utarbeidelse og vil etter hvert kunne legges ut som web-tjenester.

### **Elektronisk publisering**

Arkivverkets digitaliseringspolitikk omfatter også elektronisk publisering av kildeutgaver. Strategiplanen (punkt 3.2.3) slår fast at slik publisering skal vektlegges sterkere i perioden frem til 2005. Igjen er det tale om noe som kommer i tillegg til – ikke som erstatning for – tidligere tjenester. Når det gjelder elektronisk publisering av allerede trykte ting, står Regesta Norvegica høyt på prioritetslisten.

Det arbeides dessuten med å lage et felles navneregister for hele Diplomatarium Norvegicum. Dette er et prosjekt som gjennomføres i samarbeid med Universitetet i Oslo, og det er et prosjekt som vil gi oss en betydelig *merverdi* i forhold til den bruk av Diplomatarium Norvegicum som hittil har vært mulig. Prosjektet skal være avsluttet på nyåret 2002. Vi er dessuten kommet langt i arbeidet med et elektronisk navneregister til vår samling av utrykte diplomene etter 1570, der tusenvis av kort skrives inn. Dette prosjektet skal være slutført på vårparten 2002.

### **Digitalisering av kulturarvmateriale**

Arkivmateriale som av kulturformidlingshensyn bør være bredt tilgjengelig i bildeform, det være seg dokumenter, kart, tegninger og fotografier, skal velges ut, digitaliseres og legges ut på Internett i løpet av planperioden, det vil si innen 2005 (Strategiplan, punkt 3.2.4).

Vi skal dels vise smakebiter, og dels hele arkiver og samlinger. Smakebitene vil være enkle å integrere i den nye internettløsningen vår. En del smakebiter ligger klar for visning, men vi er forsinket med nettpresentasjonen fordi omleggingen av hjemmesidene våre har tatt lengre tid enn beregnet. Vi har forankret det faglige ansvaret for å digitalisere denne type arkivdokumenter i en av avdelingene i Riksarkivet, og vi har startet der vi er svakest – nemlig i håndteringen av fotografier og fotosamlinger. Vi har nå skaffet oss en første samlet oversikt over fotografier i Riksarkivet, og vi har fra før en tilsvarende oversikt som dekker statsarkivene. Digitalisering av dette kulturmateriale vil etter hvert få et betydelig omfang, og vi regner med at vi vil være en interessant samarbeidspartner for mange andre institusjoner i den forbindelse.

### **Elektroniske arkiver – arkiver som er skapt i digital form**

Det forholder seg ofte slik med arkiver at det som er et administrativt arkiv hos en arkivskaper i dag, kan være kulturarvmateriale i et arkivdepot i morgen. Dette gjelder også for de elektroniske arkivene som nå skapes, og jeg skal derfor si noe om hvordan vi ser på de elektroniske arkivene i et digitaliseringspolitisk perspektiv. Men la meg for ordens skyld presisere at elektroniske arkiver ikke er arkivmateriale som er digitalisert. Det er elektronisk skapt arkivmateriale. (Strategiplan, punkt 1.3.)

Arkivverket har allerede tatt i mot betydelige mengder elektronisk arkivmateriale. Det må sies at avleveringene har ikke nådd et tilfredsstillende omfang, og avleveringene har heller ikke kommet inn i en fast rytme. Dette jobber vi med, men ressursituasjonen er slik at vi må

regne med at betydelige informasjonsmengder går tapt før vi har situasjonen under kontroll. Det er etablert en depotordning som sikrer oss de elektroniske arkivene i 25 år før vi selv etablerer brukertjenester på dem. Når de elektroniske arkivene formelt overtas av Riksarkivet er det meningen at vi skal yte brukertjenester på dem så snart som mulig. De eldste avleveringene nærmer seg nå denne grensen og vi er derfor i ferd med å utvikle visse brukertjenester på dette materialet.

Enkelte avleveringer vil kunne presenteres raskt som et digitalt tilbud når det ikke knytter seg personvern hensyn eller andre hensyn til det. Andre vil det måtte ta lenger tid med. Dødsfallregisteret som Yngve Nedrebø nevnte i sitt foredrag, er et eksempel på hva som kan representere den fremtidige kulturarven. Og vi har andre store registre som ligger og venter på å bli lagt ut. De vil komme etter hvert de nærmeste årene.

Vi er kommet langt med hensyn til metodene for lagring og bruk av elektroniske arkiver, men det gjenstår et betydelig arbeid med å tilrettelegge tjenester. Det er en ressurskrevende oppgave å skape gode brukertjenester. Vi har derfor under overveielse å inngå samarbeid med andre miljøer for å vinne erfaring med å utvikle brukertjenester på denne type materiale. Hvis vi får i land et slikt samarbeid om brukertjenester på digitalt født materiale, vil mye kunne være vunnet sett både fra vår og brukernes side. Dette tilretteleggings- og utviklingsarbeidet må sees i sammenheng med hva vi gjør på digitaliseringsområdet.

### **Mikrofilming og skanning av mikrofilmet materiale**

Mikrofilming av arkivmateriale er en godt etablert og stor komponent i Arkivverkets politikk for å gjøre arkivene tilgjengelige og samtidig skåne dem for sterk slitasje. Nå kommer dette oss til gode i vårt arbeid med å digitalisere arkiver. Vi bruker mikrofilm som grunnlag for å digitalisere arkivmateriale fordi det både er arbeidsbesparende og en hensiktsmessig måte å sikre og gjøre arkiver tilgjengelig på. Informasjonsfangsten er stor når vi mikrofilmer, og vi har et arbeidsopplegg som på en effektiv måte, det vil si ut fra en samlet vurdering av arbeidstid, kvalitetskontroll og enkel sikring av primærkildene, fører frem til etterspurt digitalt materiale. (Strategiplan, punkt 3.1.)

Vi starter langt fra på bar bakke når det gjelder mikrofilmet materiale. Et samarbeid med mormonerne om mikrofilming av kirkebøker og annet personbasert materiale startet allerede i 1947. Etter at samarbeidet ble avsluttet ved inngangen til 1990-årene som følge av at mormonerne ville prioritere andre land i Europa, har mikrofilmingen fortsatt i vår egen regi, og

vi har i dag en mikrofilmbeholdning som dekker millioner av sider med arkivdokumenter. Vi begynner altså ikke på nytt, men bruker den tidligere plattformen, det vil si mikrofilmen, som drar informasjonen med seg over på en ny digitalisert plattform.

Det vi nå ser begynnelsen til er i prinsippet parallelt til hva vi har gjort tidligere, når det gjelder å bringe 1801-folketellingen fra å være en database som kun var tilgjengelig for de få eller i form av papirkopier, til å bli tilgjengelig for de mange som en web-tjeneste. Dette er løsninger som er meget arbeidsøkonomiske sett både fra et arkivsynspunkt og fra forskersiden, og det er slike løsninger vi vil søke gjennomført i vår digitaliseringspolitikk.

Den tidligere investeringen i mikrofilm er altså ikke bortkastet, og vi investerer i ny mikrofilming. Vi har nylig gjennomført en intern utredning om hvilke arkivserier som skal mikrofilmes i fremtiden, og vi har besluttet at den nye mikrofilmingen skal gjennomføres av en ny produksjonsenhet. Den endelige avgjørelsen om dette prosjektet vil bli tatt i løpet av inneværende år. Vi har videre nettopp gjennomført en utredning av hvilke mikrofilmede kilder som nå skal digitaliseres, og hvilke tekniske og andre krav som skal legges til grunn for denne digitaliseringen. Budsjettene i årene fremover avgjør i hvilket omfang vi kommer i gang med dette prosjektet. Kirkebøker og tinglysingsmateriale står fremst i køen over det materialet som skal overføres fra mikrofilm til digital nettpresentasjon.

En av konklusjonene i den helt ferske rapporten om digitaliseringspolitikken i Arkivverket, slår fast at kvaliteten på mikrofilmet materiale skannet i gråtoner, er god nok for de fleste formål, og at Arkivverket vil være tjent med å ha en samling av meget brukt arkivmateriale digitalisert på denne måten. Denne råskanningen er første steget i en tilrettelegging av arkivmaterialet i digital form. Etter skanningen må/kan materialet gjennomgå ulike former for viderebehandling før det kan tas i bruk til forskjellige formål. I prioriteringen av slike formål/oppgaver vil vi i Arkivverket måtte være ekstremt opptatt av å gjøre valg som er riktige i den forstand at de er sterkt arbeidsbesparende innad i etaten, samtidig som de representerer hyppig etterspurte brukertjenester. Basis for alle slike tjenester vil være en indeksering av det skannede materialet. En av hovedformålene med Arkivverkets mikrofilmmateriale er å gjøre materialet lettere tilgjengelig for det store publikum på Internett, som en web-tjeneste fra Digitalarkivet.

## Avsluttende betraktninger

Uansett hvordan de enkelte produktene fremstår, må politikken vår være basert på at vi velger registreringsløsninger som er robuste, det vil si at de holder seg over tid. Samtidig må det være løsninger som er tidsøkonomiske og som fungerer godt for brukerne, det vil si at de blant annet skal være raske. Og siden dette er å bevege seg i ABM-landskapet, må vi hele tiden skjele til at vi velger løsninger som er samkjørte med aksepterte løsninger/standarder både nasjonalt og internasjonalt slik at sektorene kan bindes sammen.

Arkivverkets digitaliseringspolitikk kan sies å bygge på en solid tradisjon for konvertering av arkivinformasjon fra en teknologiplattform til en annen. Mikrofilmingen vår og konverteringen av 1801-folketellingen er solide uttrykk for denne politikken. Dette føres videre gjennom registreringsaktiviteten i Digitalarkivet samtidig som vi bruker mikrofilm som grunnlag for digitaliseringen. Parallelt med dette bindes vår egen digitaliseringsinnsats nært opp mot de arkivene som nå skapes elektronisk. Den elektronisk skapte arkivmengden vil i stadig sterkere grad komme til å prege den arkivverden som våre brukere etter hvert vil oppleve. Vi har tro på at kombinasjonen av erfaringsmessig rasjonelle løsninger basert på gode tradisjoner og gjennomtenkt anvendelse av ny teknologi, vil sette oss i stand til å betjene dagens og morgendagens brukerne på en god måte med digitaliserte tjenester.



## **Nasjonalbibliotekets digitaliseringsprinsipper og -politikk**

Opprinnelig hadde jeg fått forespørsel om å si noe på Arkivarforeningens seminar om temaet *Har Nasjonalbiblioteket en digitaliseringspolitikk?* I programmet for dagens møte har dette blitt til en presentasjon av Nasjonalbibliotekets digitaliseringsprinsipper og -politikk, noe som er atskillig vanskeligere å besvare entydig.

For Nasjonalbiblioteket kan det her være behov for å forsøke å tegne et samlet bilde av de utfordringer institusjonen står overfor. Hvis vi skulle forsøke å beskrive oppgaver kunne det bli noe i retning av at *"Nasjonalbibliotekets samfunnsmessige forpliktelse er å gi samtidige og fremtidige generasjoner mulighet og anledning til å studere og oppleve seg og sin samtid på bakgrunn av egen fortid. Biblioteket skal videre gi tilgang til kunnskapsressurser for kompetanseutvikling, forskning og innovasjon"*.

I praksis oppleves og oppfattes dette i stor grad til å dreie seg om innsamling, bevaring, tilgjengeliggjøring og formidling av materiale utgitt for allmennheten. Biblioteket består av to avdelinger: Rana og Oslo hvorav Rana blant annet forvalter pliktavleveringsloven som er det viktigste instrumentet for innsamling av mest mulig av vår (i tradisjonell mening) trykte kulturarv. I tillegg inneholder pliktavleveringsloven krav om at materiale utgitt i andre former også skal pliktavleveres, men her har det tatt tid å få etablert gode ordninger. Nasjonalbiblioteket Oslos hovedoppgaver er knyttet til å være bibliotekets publikumsavdeling i tillegg til at det er her man utarbeider Nasjonalbibliografien – som et uttrykk for dokumentasjon av den nasjonale trykksakproduksjonen.

Arbeid med digitalisering av de eldre samlingene har i hovedsak vært gjennom å etablere større eller mindre prøveprosjekter som søkes gjennomført på en slik måte at erfaringer og resultater kan danne basis for det videre arbeid og at de har stor overføringsverdi til liknende materiale. Selve utvalget og prioriteringen har tatt utgangspunkt i litt ulike momenter, men et viktig aspekt har vært å sette i gang prosjekter hvor biblioteket har fått ekstern delfinansiering til gjennomføringen.

I likhet med en rekke andre institusjoner står Nasjonalbiblioteket i en situasjon som preges av overganger fra gammelt til nytt. Nasjonalbibliotekfunksjonen og pliktavleveringsloven er relativt sett gamle ordninger med tradisjon og sedvane knyttet til oppgavens løsninger.

Nasjonalbiblioteket i seg selv er en ung organisasjon: Rana ble opprettet i 1989, Nasjonalbibliotekarembetet i 1994 og Nasjonalbiblioteket Oslo ble skilt ut fra Universitetsbiblioteket i 1999. Inn i dette bildet kommer så utfordringene fra den teknologiske utviklingen – hvor jeg i denne sammenheng først og fremst vil legge vekt på tre aspekter:

Den første utfordringen består i utviklingen og den raske utbredelsen av web-teknologien på siste halvdel av 90-tallet (en tendens som ikke har avtatt etter årtusenskiftet) – som i vår sammenheng nå først og fremst får betydning for tilgangs- og formidlingsdelen av bibliotekets oppgaver.

Den andre utfordringen ligger i veksten i digitale utgivelser – og her menes utgivelser i betydningen gjort tilgjengelig for allmennheten – ikke bundet til en bestemt teknologi som eksempelvis mangfoldiggjøring ved hjelp av trykketeknologi, men også omfattende for eksempel digitalt baserte utgivelser hva enten de distribueres i form av en fysisk gjenstand som diskett/CD-rom eller via webteknologi.

For det tredje ser jeg problemstillinger knyttet til langtidsbevaring av digitalt lagret materiale.

Det første av disse punktene har i stor grad sammenheng med temaet for dagens seminar: digitaliseringspolitikk med vekt på arkivmateriale (som for eksempel Nasjonalbibliotekets håndskriftsamlinger) og billedmateriale – det er altså et spørsmål om å bruke teknologien for tilgjengelighetsformål. De to andre punktene er selvsagt også relevante her, men de har - i forhold til Nasjonalbibliotekets oppgaver - et videre perspektiv. Vi står overfor store utfordringer når det gjelder etablering av varige ordninger for innsamling av digitale dokumenter, slik at pliktavleveringens intensjon om medieuavhengighet kan innfris. Nasjonalbiblioteket har akkurat etablert et internt treårig prosjekt kalt "Pliktavlevering av digitale dokumenter" hvor det blant annet er meningen å etablere seleksjonskriterier for hva som skal samles inn/pliktavleveres, hvordan materialet skal organiseres for å sikre gjenfinning, hvordan det skal dokumenteres med mer. Samtidig har institusjonen allerede etablert et såkalt Digitalt sikringsmagasin (i fjellhall i Rana) hvor det for øyeblikket er lagringskapasitet til 50 TB og som enkelt kan bygges ut til 350 TB.

Det arbeides videre med problemstillinger knyttet til langtidsbevaring av digitalt materiale eller digital bevaring hvor sentrale problemstillinger er knyttet til strategivalg som emulering eller migrering. Hovedpoenget i denne sammenheng er at for Nasjonalbiblioteket – troligvis i likhet med for eksempel Riksarkivet – er bevaring i evighetens perspektiv en sentral

oppgave hvor løsningene naturligvis har sammenheng med og legger føringer for spørsmålene rundt bibliotekets digitaliseringsprinsipper og – politikk.

Ved Nasjonalbiblioteket Oslo har man en rekke spesialsamlinger som Håndskriftsamling, Billedsamling, Norsk-amerikansk samling, Plakatsamling, Småttrykk/Krigstrykk, Kartsamling med mer, i tillegg til den store – mer eller mindre komplette – nasjonale samlingen av det som er utgitt i Norge siden trykkekunsten kom til landet.

Omfanget av samlingene er slik at det i utgangspunktet er vanskelig å se for seg at det noen gang vil være aktuell politikk å digitalisere alt – og det er kanskje også vanskelig å se rasjonalitet i en massedigitalisering av eldre materiale. Av dette følger naturlig at det er et behov for å etablere kriterier for utvelgelse av hva som skal digitaliseres. For å kunne føre den diskusjonen på en noen lunde meningsfull måte er det viktig å stille spørsmål om hvorfor.

I hovedsak har det blitt anført to typer argumenter for digitalisering. Det ene har sammenheng med tilgang mens det andre har sammenheng med bevaring, selv om det finnes en indre sammenheng mellom disse to aspektene.

- Materiale det er stor interesse og etterspørsel for og hvor man ønsker å gi større brukergrupper tilgang til materialet.
- Materiale som er i hyppig bruk hvor det vil ligge effektiviseringsgevinster i selve håndteringen av dette dersom det digitaliseres.
- Materiale som er i dårlig forfatning slik at det bør skånes for bruk.
- I tilfeller hvor det kan være viktig å lage en sikringskopi av originalen dersom denne skulle gå tapt – særlig aktuelt for unika av en viss kulturhistorisk eller faglig betydning.

Et slags hovedformål her kan være at man digitaliserer for å forhindre stor slitasje på materiale ved at brukerne primært får tilgang til den digitale kopien. I dette ligger ingen argumenter for at en digital kopi fullt ut kan erstatte originalen for alle formål – i en rekke sammenhenger vil det være vesentlig å ha en original som utgangspunkt for for eksempel forskningsformål. Jeg vil også kort nevne at det i en rekke sammenhenger har blitt utført digitaliseringsarbeid hvor resultatet mer er av politisk karakter for institusjonen og resultatet av digitaliseringsprosessen mer er å

betrakte som rent illustrerende. I slike tilfeller blir det digitaliserte objektet i hovedsak brukt til å "vise fram hva man har" til omverdenen – noe som kanskje kan ha betydning for muligheten til å få midler til digitaliseringsaktiviteter. Men det er ikke dette aspektet som er sentralt i den videre gjennomgangen.

Som nevnt tidligere er bevaring og tilgjengelighet de sentrale aksene diskusjonen om digitalisering dreier seg rundt. Sett i Nasjonalbibliotekets perspektiv er dette også gyldig, og fordi vårt perspektiv nettopp er å bevare for evigheten, får disse utfordringene en ekstra dimensjon ved seg som tidligere ble berørt når jeg nevnte det digitale masselageret i Rana. Det er vel riktig å si at når det gjelder infrastruktur, står Nasjonalbiblioteket godt rustet til å møte utfordringene, problemene ligger mer i å finne løsninger for langtidsbevaring av digitale objekter samt til utfordringer knyttet til organisering av samlinger av digitale objekter slik at de blir gjenfinnbare og verifiserbare. Til disse siste poengene hører metadataproblematikken – som dreier seg om hvilken informasjon som må knyttes til det enkelte digitale objekt for at det skal inngå i en organisatorisk struktur som gjør det gjenfinnbart. Det dreier seg her dels om metadata knyttet til lagringsformater eller tekniske forhold, det er metadata som beskriver innhold og metadata som beskriver tilgang – det vil si om det finnes klausuler eller opphavsrettslige begrensninger som berører bruken av det (eller tilgangen).

Når det gjelder tilgjengelighetsaspektet kommer man ikke utenom å nevne at gjennom de tradisjonelle måtene å gi tilgang på – gjennom organisering av materialet, katalogisering, registeroppbygging med mer har publikum og ansatte kunnet finne fram i samlingen. For Nasjonalbiblioteket er situasjonen nå slik at på kort sikt er en av de aller mest sentrale digitaliseringsoppgavene knyttet til konvertering av tradisjonelle kortkataloger. I vår situasjon har dette et meget kortsiktig perspektiv: det er meningen at Nasjonalbiblioteket Oslo skal ut av nåværende lokaler i Drammensveien 42 høsten 2003 på grunn av rehabilitering. Før det må vi ha konvertert den gamle hovedkatalogen (norsk og utenlandsk materiale) som dekker om lag 100 kvm fordi vi ikke har plass til å ta den med oss til midlertidige lokaler. Inn i dette ligger ikke konvertering av spesialsamlingenes kataloger – kun hovedkatalogen.

I den grad man kan snakke om prinsipper for digitalisering er det to områder som peker seg ut (og hvor vi kan si noe mer konkret):

- 1) For billedmateriale skal digitalisering skje med slik kvalitet at man i ettertid kan reprodusere trykkkvalitet på grunnlag av det vi kan kalle en digitale "morfil". Denne skal videre være så nær som mulig originalen; den skal ikke arbeides på og den vil oppbevares i DSM. I tillegg vil man generere ulike kvaliteter/komprimerte formater som for eksempel JPG for bruk til formidling, eksempelvis via web. Og endelig: det finnes (i alle fall inntil videre) et analogt sikringseksemplar.
- 2) Det andre som peker seg ut er sammenhengen mellom konservering og digitalisering. Når skjørt/slitt materiale tas ut av samlingen for konservering i tradisjonell form av originalen, vil det samtidig bli laget en digital kopi fordi man ønsker å minimalisere antall ganger slikt materiale må håndteres.

## Samtidslitteraturen

Jeg vil forsøke å bringe trådene sammen. Nasjonalbiblioteket står overfor en rekke utfordringer i forhold til både det eldre og det nyere materialet. Mot slutten av dette året er det planen å arrangere et seminar i samarbeid med, i denne omgang, de skjønnlitterære rettighetshaverne for å reise problemstillinger rundt hva som allerede er "de nye manuskriptene". De aller fleste forfattere (og forlag for den del) arbeider i dag med ulike tekstbehandlingsverktøy som gjør at dagens håndskrifter i betydningen manuskripter – finnes bare i digital form. Nasjonalbiblioteket vil svikte sin forpliktelse dersom man ikke tar mål av seg til også å samle inn, oppbevare, tilgjengeliggjøre og formidle litteraturen som skapes i samtiden – til glede for oss og for fremtidens generasjoner.

For Nasjonalbiblioteket blir det også et tankekors dersom effekten av økt eller lettere tilgang som et resultat av digitalisering, skulle bli at det materiale som digitaliseres også blir det det forskes i fordi det er det som er enklest tilgjengelig. En slik effekt vil neppe framstå som ønsket, men er heller ikke noe vi som institusjon har et godt svar på.

Avslutningsvis kan det kanskje passe å si at Nasjonalbiblioteket neppe kan sies å ha en digitaliseringspolitikk – men vi har definitivt en politikk for digitalisering!



## Norsk Arkivforum - tidligere utgivelser

- 1 (1980) Rapport fra Norsk Arkivseminar 1976 (Vang, Hedmark).
- 2 (1980) Rapport fra Norsk Arkivseminar 1978 (Søgne).
- 3 (1981) Rapport fra Norsk Arkivseminar 1980 (Rollag).
- 4 (1983) Rapport fra Norsk Arkivseminar 1982 (Inderøy).
- 5 (1985) Metodeutvikling i Arkivarbeidet. Festskrift til Carlo Larsen.
- 6 (1985) Rapport fra Norsk Arkivseminar 1984 (Fana).
- 7 (1986) Rapport fra Norsk Arkivseminar 1986 (Sola).
- 8 (1988) Andres A. Svalestuen: Medisinalvesenets sentraladministrasjon 1809-1940.  
Anne Marie Bøhmer: Helsevesenet 1940-1983, en administrasjons-historisk oversikt. Tor Breivik: Landbruksdepartementet 1940-1983.
- 9 (1989) Nils Johan Stoa: Arkivmateriale fra dansk-norske sivile sentral-institusjoner før 1814.
- 10 (1990) Rapport fra Norsk Arkivseminar 1988 (Hamar).
- 11 (1993) Jørgen H. Marthinsen: Ordningsskjemaer, serier og arkivskaper. Liv Mykland: Et forspilt liv? (Foredrag på De norske historikerdagene 1992.) Rolf Petter Amdam: Glassverker. (Foredrag på Norsk Arkivseminar 1990.) Rapport fra Norsk Arkivseminar 1990 (Halvorsbøle).
- 12 (1996) Festskrift til Helge Paulsen.
- 13 (1998) Innlegg fra Norsk Arkivseminar 1993 (Tromsø), 1995 (Hurdalsjøen).
- 14 (2000) Norsk Arkivseminar 1998 (Rosendal).
- 15 (2001) Artikler og seminarinnlegg.
- 16 (2001) Arkivfagets profesjonalisering. Festskrift til Jørgen H. Marthinsen.
- 17 (2002) Norsk Arkivseminar 2001 (Kristiansand).
- 18 (2004) Frank Meyer (red.), Digitalisering av kulturarvmateriale.

Bestilling av tidligere utgivelser kan rettes til kassereren i Arkivarforeningen, for tiden Åse Lange, Riksarkiv-bygningen. E-post: [aase.lange@riksarkivaren.dep.no](mailto:aase.lange@riksarkivaren.dep.no). Tidligere utgivelser koster kr 50,- per hefte. Pakkepris for ett sett med alle utgivelser er kr 250,-. Prisen for siste utgivelse (nr. 18) er kr 75,-.











